Online Appendices, Not for Publication

Appendix A. Network definitions

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

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

At the hamlet level, we consider:

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

Appendix B. Kalman Filter, Estimation and Simulation

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

B.1. **Model.**

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

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

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

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

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

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

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

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

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

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

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

$$\mathbf{s}^{i,t} := \left(s^{i,t}_1, ..., s^{i,t}_{t-d(i,j)}\right).$$

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

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

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

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

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

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

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

- state equation is given by

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

- measurement equation given by

$$s_{\tau}^{i,t} = w_{j,\tau} + v_{\tau}^{i,t},$$

where $v_{\tau}^{i,t} \sim \mathcal{N}\left(0, \sigma_{\tau,t,i}^2\right)$.

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

- Time update equations:

$$\begin{aligned} \hat{w}_{\tau}^{-} &= \rho \hat{w}_{\tau-1} + c \\ P_{\tau}^{-} &= \rho^2 P_{\tau-1} + \sigma_{\epsilon}^2. \end{aligned}$$

- Measurement update equations:

$$\begin{aligned} K_{\tau} &= \frac{P_{\tau}^{-}}{P_{\tau}^{-} + \sigma_{\tau,t,i}^2}. \\ \hat{w}_{\tau} &= \hat{w}_{\tau}^{-} + K_{\tau}\left(s_{\tau}^{i,t} - \hat{w}_{\tau}^{-}\right). \\ P_{\tau} &= \left(1 - K_{\tau}\right) P_{\tau}^{-}. \end{aligned}$$

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

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

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

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

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

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

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

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

(1) Pick lattice $\Xi \subset \Theta$. For $\xi \in \Xi$ on the grid:
   (a) For each network $r \in [R]$, compute
$$d(r,\xi) := \frac{1}{S}\sum_{s\in[S]} m_{sim,r}(s,\theta) - m_{emp,r}.$$
   (b) For each $b \in [B]$, compute
$$D(b,\xi) := \frac{1}{R}\sum_{r\in[R]} \omega_r^b \cdot d(r,\xi)$$
   where $\omega_r^b = e_{br}/\bar{e}_r$, with e_{br} iid exp(1) random variables and $\bar{e}_r = \frac{1}{R}\sum e_{br}$ if we are conducting bootstrap, and $\omega_r^b = 1$ if we are just finding the minimizer.
   (c) Find $\xi^{\star b} = \text{argmin } Q^{\star b}(\xi)$, with $Q^{\star b}(\xi) = D(b,\xi)'D(b,\xi)$.[33]
(2) Obtain $\{\xi^{\star b}\}_{b\in B}$.
(3) For conservative inference on $\widehat{\theta}_j$, the j^{th} component, consider the $1-\alpha/2$ and $\alpha/2$ quantiles of the $\xi_j^{\star b}$ marginal empirical distribution.

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

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

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

Appendix C. Details on Poverty Targeting Procedures

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

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

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

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

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

Table D.2A: The Correlation between Household Network Characteristics and the Error Rate in Ranking Income Status of Households, conditional on offering assessments

	(1)	(2)	(3)	(4)
	Outcome variable: Error rate conditional on reporting			
	Panel A: Consumption Metric			
Degree	-0.00280***			-0.00175
	(0.000713)			(0.00116)
Clustering		-0.0117		-0.00830
		(0.00893)		(0.00992)
Eigenvector Centrality			-0.0857***	-0.0434
			(0.0233)	(0.0383)
R-squared	0.664	0.663	0.665	0.665
	Panel B: Self-Assessment Metric			
Degree	-0.00384***			-0.00266**
	(0.000713)			(0.00122)
Clustering		-0.00248		0.000797
		(0.0100)		(0.0109)
Eigenvector Centrality			-0.104***	-0.0471
			(0.0248)	(0.0410)
R-squared	0.671	0.669	0.671	0.671
Village Fixed Effect	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village. The sample comprises 5,633 households. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table D.2B: The Correlation between Household Network Characteristics and the Error Rate in Ranking Income Status of Households, Controlling for Household Characteristics, conditional on offering assessments

	(1)	(2)	(3)	(4)
	Outcome variable: Error rate conditional on reporting			
	Panel A: Consumption Metric			
Degree	-0.00215***			-0.00122
	(0.000698)			(0.00112)
Clustering		-0.0115		-0.00828
		(0.00879)		(0.00979)
Eigenvector Centrality			-0.0694***	-0.0387
			(0.0231)	(0.0375)
R-squared	0.668	0.667	0.668	0.668
	Panel B: Self-Assessment Metric			
Degree	-0.00301***			-0.00200*
	(0.000697)			(0.00119)
Clustering		-0.00239		0.000599
		(0.00973)		(0.0107)
Eigenvector Centrality			-0.0828***	-0.0406
			(0.0243)	(0.0401)
R-squared	0.676	0.674	0.676	0.676
Village Fixed Effect	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village, controlling for the household's characteristics as in Table 2B. The sample comprises 5,630 households for panel. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table D.3: The Correlation Between Inaccuracy in Ranking a Pair of Households in a Village and the Average Distance to Rankees, conditional on offering assessments

	(1)	(2)	(3)	(4)
	Outcome variable: Error rate conditional on reporting			
	Panel A: Consumption Metric			
Inverse of the Distance	-0.00573	-0.0170***	-0.00830*	-0.0204
	(0.00655)	(0.00654)	(0.00485)	(0.0130)
Average Degree		0.00134	0.00602**	0.00663**
		(0.00139)	(0.00305)	(0.00318)
Average Clustering Coefficient		0.0226	0.0594**	0.0672**
		(0.0226)	(0.0266)	(0.0279)
Average Eigenvector Centrality		-0.0448	-0.185*	-0.164
		(0.0543)	(0.0995)	(0.105)
R-squared	0.000	0.008	0.082	0.138
	Panel B: Self-Assessment Metric			
Inverse of the Distance	-0.00647	-0.0130*	-0.00401	0.00183
	(0.00682)	(0.00686)	(0.00502)	(0.0129)
Average Degree		0.000986	0.00183	0.00126
		(0.00145)	(0.00294)	(0.00309)
Average Clustering Coefficient		-0.0180	0.0215	0.0235
		(0.0232)	(0.0286)	(0.0296)
Average Eigenvector Centrality		0.0293	0.0210	0.0269
		(0.0544)	(0.0977)	(0.102)
R-squared	0.001	0.002	0.166	0.168
Physical Controls	No	Yes	Yes	Yes
Village FE	No	No	Yes	Yes
Ranker FE	No	No	No	Yes

Notes: This table provides an estimate of the correlation between the accuracy in ranking a pair of households in a village and the characteristics of the households that are being ranked. In Panel A, the dependent variable is a dummy variable for whether person i ranks person j versus person k incorrectly based on using consumption as the metric of truth (the sample mean is 0.497). In Panel B, the self-assessment variable is the metric of truth (the sample mean is 0.464). The sample is comprised of 117,492 ranked pairs in Panel A and 116,338 in Panel B. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcome variable: Error rate				*Outcome variable: Share of don't knows*			
				Panel A: Consumption Metric				
Degree	-0.00281***			-0.00182	-0.00333***			-0.00169
	(0.000712)			(0.00116)	(0.000677)			(0.00115)
Clustering		-0.0118		-0.00859		0.00269		0.00577
		(0.00889)		(0.00986)		(0.0109)		(0.0119)
Eigenvector Centrality			-0.0847***	-0.0409			-0.102***	-0.0662
			(0.0232)	(0.0380)			(0.0256)	(0.0411)
R-squared	0.667	0.666	0.667	0.668	0.722	0.720	0.722	0.722
				Panel B: Self-Assessment Metric				
Degree	-0.00386***			-0.00276**	-0.00333***			-0.00169
	(0.000712)			(0.00122)	(0.000677)			(0.00115)
Clustering		-0.00283		0.000172		0.00269		0.00577
		(0.00999)		(0.0108)		(0.0109)		(0.0119)
Eigenvector Centrality			-0.103***	-0.0439			-0.102***	-0.0662
			(0.0247)	(0.0408)			(0.0256)	(0.0411)
R-squared	0.674	0.672	0.673	0.674	0.722	0.720	0.722	0.722
				Panel C: Simulations				
Degree	-0.00835***			-0.00459***	-0.0168***			-0.00943***
	(0.000569)			(0.000770)	(0.00112)			(0.00152)
Clustering		-0.0782***		-0.0702***		-0.157***		-0.142***
		(0.00729)		(0.00705)		(0.0145)		(0.0141)
Eigenvector Centrality			-0.308***	-0.176***			-0.616***	-0.344***
			(0.0182)	(0.0259)			(0.0361)	(0.0512)
R-squared	0.900	0.891	0.910	0.919	0.902	0.893	0.912	0.921
Village Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village. The sample comprises 5,633 households. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. Details of the simulation procedure for Panel C are contained in Appendix B. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcome variable: Error rate				*Outcome variable: Share of don't knows*			
				Panel A: Consumption Metric				
Degree	-0.00215***			-0.00127	-0.00267***			-0.00110
	(0.000697)			(0.00112)	(0.000670)			(0.00113)
Clustering		-0.0116		-0.00860		0.00182		0.00544
		(0.00877)		(0.00974)		(0.0107)		(0.0117)
Eigenvector Centrality			-0.0684***	-0.0363			-0.0854***	-0.0634
			(0.0230)	(0.0372)			(0.0254)	(0.0403)
R-squared	0.671	0.670	0.671	0.671	0.726	0.724	0.726	0.726
				Panel B: Self-Assessment Metric				
Degree	-0.00302***			-0.00209*	-0.00267***			-0.00110
	(0.000697)			(0.00118)	(0.000670)			(0.00113)
Clustering		-0.00279		-5.26e-05		0.00182		0.00544
		(0.00972)		(0.0106)		(0.0107)		(0.0117)
Eigenvector Centrality			-0.0819***	-0.0376			-0.0854***	-0.0634
			(0.0243)	(0.0398)			(0.0254)	(0.0403)
R-squared	0.679	0.677	0.678	0.679	0.726	0.724	0.726	0.726
				Panel C: Simulations				
Degree	-0.00835***			-0.00457***	-0.0168***			-0.00939***
	(0.000573)			(0.000768)	(0.00113)			(0.00151)
Clustering		-0.0784***		-0.0700***		-0.157***		-0.141***
		(0.00725)		(0.00703)		(0.0144)		(0.0140)
Eigenvector Centrality			-0.308***	-0.176***			-0.615***	-0.345***
			(0.0182)	(0.0259)			(0.0362)	(0.0512)
R-squared	0.900	0.891	0.910	0.919	0.726	0.724	0.726	0.726
Village Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: This table provides estimates of the correlation between a household's network characteristics and its ability to accurately rank the poverty status of other members of the village, controlling for the household's characteristics as in Table 2B. The sample comprises 5,630 households for panel. The mean of the dependent variable in Panel A (a household's error rate in ranking others in the village based on consumption) is 0.50, while the mean of the dependent variable in Panel B (a household's error rate in ranking others in the village based on a household's own self-assessment of poverty status) is 0.46. Details of the simulation procedure for Panel C are contained in Appendix B. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

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

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Outcome variable: Error rate				*Outcome variable: Share of don't knows*			
				Panel A: Consumption Metric				
Inverse of the Distance	-0.0594***	-0.0401***	-0.0226***	-0.0128	-0.0771***	-0.0451***	-0.0294***	-0.00434
	(0.00836)	(0.00823)	(0.00569)	(0.0125)	(0.00946)	(0.00985)	(0.00718)	(0.0130)
Average Degree		-0.00508***	0.00267	0.00266		-0.00968***	-0.00258	-0.00295
		(0.00174)	(0.00315)	(0.00321)		(0.00210)	(0.00306)	(0.00306)
Average Clustering Coefficient		-0.00643	0.0282	0.0286		-0.0393	-0.0167	-0.0196
		(0.0252)	(0.0272)	(0.0275)		(0.0302)	(0.0286)	(0.0284)
Average Eigenvector Centrality		0.0668	-0.0735	-0.104		0.153*	0.109	0.0392
		(0.0668)	(0.0912)	(0.0945)		(0.0816)	(0.0969)	(0.103)
R-squared	0.007	0.012	0.136	0.202	0.020	0.064	0.332	0.445
				Panel B: Self-Assessment Metric				
Inverse of the Distance	-0.0680***	-0.0403***	-0.0224***	-0.00351	-0.0771***	-0.0451***	-0.0294***	-0.00434
	(0.00938)	(0.00904)	(0.00607)	(0.0134)	(0.00946)	(0.00985)	(0.00718)	(0.0130)
Average Degree		-0.00620***	0.000102	-0.000553		-0.00968***	-0.00258	-0.00295
		(0.00193)	(0.00336)	(0.00345)		(0.00210)	(0.00306)	(0.00306)
Average Clustering Coefficient		-0.0404	0.00400	0.00403		-0.0393	-0.0167	-0.0196
		(0.0270)	(0.0301)	(0.0301)		(0.0302)	(0.0286)	(0.0284)
Average Eigenvector Centrality		0.126*	0.0532	0.0130		0.153*	0.109	0.0392
		(0.0746)	(0.103)	(0.106)		(0.0816)	(0.0969)	(0.103)
R-squared	0.033	0.035	0.183	0.264	0.011	0.012	0.136	0.202
				Panel C: Simulations				
Inverse of the Distance	-0.261***	-0.224***	-0.202***	-0.238***	-0.522***	-0.451***	-0.405***	-0.480***
	(0.00195)	(0.00254)	(0.00465)	(0.00775)	(0.00367)	(0.00483)	(0.00916)	(0.0152)
Average Degree		-0.00589***	-0.00538***	-0.00406***		-0.0115***	-0.00970***	-0.00700***
		(0.000577)	(0.00119)	(0.00126)		(0.00108)	(0.00240)	(0.00255)
Average Clustering Coefficient		-0.0948***	-0.128***	-0.124***		-0.184***	-0.250***	-0.241***
		(0.0101)	(0.0124)	(0.0135)		(0.0195)	(0.0243)	(0.0267)
Average Eigenvector Centrality		-0.0731***	-0.182***	-0.115***		-0.146***	-0.390***	-0.251***
		(0.0207)	(0.0397)	(0.0408)		(0.0388)	(0.0789)	(0.0803)
R-squared	0.673	0.692	0.757	0.786	0.745	0.762	0.826	0.856
Physical Controls	No	Yes	Yes	Yes	No	Yes	Yes	Yes
Village FE	No	No	Yes	Yes	No	No	Yes	Yes
Ranker FE	No	No	No	Yes	No	No	No	Yes

Notes: This table provides an estimate of the correlation between the accuracy in ranking a pair of households in a village and the characteristics of the households that are being ranked. In Panel A, the dependent variable is a dummy variable for whether person i ranks person j versus person k incorrectly based on using consumption as the metric of truth (the sample mean is 0.497). In Panel B, the self-assessment variable is the metric of truth (the sample mean is 0.464). The sample is comprised of 117,492 ranked pairs in Panel A and 116,338 in Panel B. Details of the simulation procedure for Panel C are contained in Appendix B. Standard errors are clustered by village and are listed in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Appendix F. Tables without Physical Covariates

Table F.6 without Controls: Numerical Predictions on Correlation between Village Network Characteristics and Village-Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Average Degree	-0.0290***						0.0809***
	(0.00231)						(0.00740)
Average Clustering		-0.355***					0.435***
		(0.0350)					(0.0735)
Number of Households			0.00110***				0.000661**
			(0.000253)				(0.000269)
$\lambda_1(A)$				-0.0286***			-0.0558***
				(0.00212)			(0.00453)
Fraction of Nodes in Giant Component					-0.386***		-0.741***
					(0.0230)		(0.0495)
Link Density						-0.560***	-0.573***
						(0.0698)	(0.0878)
R-squared	0.198	0.135	0.030	0.267	0.297	0.118	0.512

Notes: This table provides village network characteristics and the error rate in ranking others in the village. Columns 1-6 show the univariate regressions, while column 7 provides the multvariate regressions. The sample comprises 631 villages. Results for error rates using simulated data, as described in Appendix B. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Table F.8 without Controls: Empirical Results on Correlation between Village Network Characteristics and Village-Level Error Rate

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
			Panel A: Consumption Metric				
Average Degree	-0.0200***						0.0356***
	(0.00274)						(0.0112)
Average Clustering		-0.361***					-0.359***
		(0.0406)					(0.0953)
Number of Households			0.000892***				0.000305
			(0.000301)				(0.000391)
$\lambda_1(A)$				-0.0168***			-0.0211***
				(0.00217)			(0.00578)
Fraction of Nodes in Giant Component					-0.264***		-0.205***
					(0.0300)		(0.0699)
Link Density						-0.349***	0.108
						(0.0780)	(0.138)
R-squared	0.076	0.114	0.016	0.075	0.113	0.037	0.153
			Panel B: Self-Assessment Metric				
Average Degree	-0.0276***						0.0294**
	(0.00294)						(0.0124)
Average Clustering		-0.495***					-0.476***
		(0.0431)					(0.106)
Number of Households			0.00135***				0.000266
			(0.000337)				(0.000418)
$\lambda_1(A)$				-0.0206***			-0.0165**
				(0.00251)			(0.00660)
Fraction of Nodes in Giant Component					-0.355***		-0.219***
					(0.0319)		(0.0779)
Link Density						-0.524***	0.163
						(0.0816)	(0.148)
R-squared	0.115	0.170	0.029	0.090	0.161	0.066	0.198

Notes: This table provides village network characteristics and the error rate in ranking others in the village. Columns 1-6 show the univariate regressions, while column 7 provides the multvariate regressions. The sample comprises 631 villages. Panel A presents results for error rates using the consumption metric. Panel B presents results for error rates using the self-assessment metric. Robust standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1.

Appendix G. Alternative Parameters

Table G.5: Numerical Predictions on Stochastic Dominance with Alternative Parameters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Panel A: (α = 0.04, τ = 3)				*Panel C: (a = 0.36, τ = 3)*			
I fosd J	-0.116***	-0.235***	-0.119***	-0.225***	-0.124***	-0.253***	-0.125***	-0.239***
	(0.0157)	(0.0243)	(0.0157)	(0.0247)	(0.0149)	(0.0230)	(0.0151)	(0.0230)
J fosd I	0.121***		0.123***		0.129***		0.129***	
	(0.0172)		(0.0174)		(0.0165)		(0.0162)	
Observations	199,396	147,460	199,396	147,460	199,396	147,460	199,396	147,460
	Panel B: (α = 0.04, τ = 5)				*Panel D: (α = 0.36, τ = 5)*			
I fosd J	-0.145***	-0.281***	-0.148***	-0.277***	-0.123***	-0.227***	-0.131***	-0.226***
	(0.0163)	(0.0251)	(0.0167)	(0.0260)	(0.0162)	(0.0248)	(0.0163)	(0.0253)
J fosd I	0.134***		0.141***		0.0985***		0.106***	
	(0.0178)		(0.0182)		(0.0180)		(0.0179)	
Observations	199,396	147,460	199,396	147,460	199,396	147,460	199,396	147,460
Non-Comparable	Yes	No	Yes	No	Yes	No	Yes	No
Physical Controls	No	No	Yes	Yes	No	No	Yes	Yes
Stratification Group FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Same as Table 5, with alternative parameters generating the simulations.

Table G.6: Numerical Predictions on Correlation between Village Network Characteristics and Village-Level Error Rate with Alternative Parameters

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
	Panel A: ($\alpha = 0.04$, $\tau = 3$)									*Panel C: ($\alpha = 0.36$, $\tau = 3$)*				
Average Degree	-0.0304***						0.0836***	-0.0312***						0.0782***
	(0.00267)						(0.00810)	(0.00272)						(0.00814)
Average Clustering		-0.362***					0.326***		-0.374***					0.319***
		(0.0433)					(0.0753)		(0.0433)					(0.0761)
Number of Households			0.000942***				0.000520*			0.000963***				0.000510*
			(0.000251)				(0.000272)			(0.000251)				(0.000278)
$\lambda_1(A)$				-0.0294***			-0.0580***				-0.0297***			-0.0558***
				(0.00236)			(0.00503)				(0.00233)			(0.00499)
Fraction of Nodes in Giant Component					-0.422***		-0.689***					-0.429***		-0.680***
					(0.0273)		(0.0528)					(0.0271)		(0.0531)
Link Density						-0.534***	-0.633***						-0.540***	-0.573***
						(0.0730)	(0.0900)						(0.0760)	(0.0923)
R-squared	0.242	0.181	0.106	0.309	0.322	0.173	0.510	0.248	0.184	0.103	0.311	0.327	0.171	0.504
	Panel B: ($\alpha = 0.04$, $\tau = 5$)									*Panel D: ($\alpha = 0.36$, $\tau = 5$)*				
Average Degree	-0.0295***						0.0752***	-0.0290***						0.0780***
	(0.00278)						(0.00782)	(0.00278)						(0.00786)
Average Clustering		-0.359***					0.374***		-0.355***					0.372***
		(0.0431)					(0.0722)		(0.0434)					(0.0730)
Number of Households			0.000879***				0.000253			0.000897***				0.000298
			(0.000252)				(0.000300)			(0.000249)				(0.000282)
$\lambda_1(A)$				-0.0281***			-0.0508***				-0.0278***			-0.0519***
				(0.00241)			(0.00462)				(0.00238)			(0.00455)
Fraction of Nodes in Giant Component					-0.446***		-0.811***					-0.442***		-0.810***
					(0.0269)		(0.0490)					(0.0269)		(0.0493)
Link Density						-0.476***	-0.424***						-0.475***	-0.453***
						(0.0734)	(0.0862)						(0.0757)	(0.0871)
R-squared	0.232	0.176	0.094	0.292	0.356	0.150	0.546	0.224	0.171	0.092	0.285	0.350	0.146	0.543

Notes: Same as in Table 6, with alternative parameters generating simulations.