The Diffusion of Microfinance
Abhijit Banerjee *et al.*
Science **341**, (2013);
DOI: 10.1126/science.1236498

The Diffusion of Microfinance

Abhijit Banerjee,* Arun G. Chandrasekhar,* Esther Duflo,* Matthew O. Jackson*

Introduction: How do the network positions of the first individuals in a society to receive information about a new product affect its eventual diffusion? To answer this question, we develop a model of information diffusion through a social network that discriminates between information passing (individuals must be aware of the product before they can adopt it, and they can learn from their friends) and endorsement (the decisions of informed individuals to adopt the product might be influenced by their friends' decisions). We apply it to the diffusion of microfinance loans, in a setting where the set of potentially first-informed individuals is known. We then propose two new measures of how "central" individuals are in their social network with regard to spreading information; the centrality of the first-informed individuals in a village helps significantly in predicting eventual adoption.

Methods: Six months before a microfinance institution entered 43 villages in India and began offering microfinance loans to villagers, we collected detailed network data by surveying households about a wide range of interactions. The microfinance institution began by inviting "leaders" (e.g., teachers, shopkeepers, savings group leaders) to an informational meeting and then asked them to spread information about the loans. Using the network data, the locations in the network of these first-informed villagers (or injection points), and data regarding the villagers' subsequent participation, we estimate the parameters of our diffusion model using the method of simulated moments. The parameters of the model are validated by showing that the model correctly predicts the evolution of participation in each village over time. The model yields a new measure of the effectiveness of any given node as an injection point, which we call communication centrality. Finally, we develop an easily computed proxy for communication centrality, which we call diffusion centrality.

Results: We find that a microfinance participant is seven times as likely to inform another household as a nonparticipant; nonetheless, information transmitted by nonparticipants is important and accounts for about one-third of the eventual informedness and participation in the village because nonparticipants are much more numerous. Once information passing is accounted for, an informed household's decision to participate is not significantly dependent on how many of its neighbors have participated. Communication centrality, when applied to the set of first-informed individuals in a village, substantially outperforms other standard network measures of centrality in predicting microfinance participation in this context. Finally, the simpler proxy measure—diffusion centrality—is strongly correlated with communication centrality and inherits its predictive properties.

Discussion: Our results suggest that a model of diffusion can distinguish information passing from endorsement effects, and that understanding the nature of transmission may be important in identifying the ideal places to inject information.

FIGURES AND TABLES IN THE FULL ARTICLE

Fig. 1. Diffusion of information and participation.

Fig. 2. Microfinance participation versus measures of leader centrality.

Table 1. Parameter estimates of the structural model.

Table 2. Time series validation.

Table 3. Microfinance participation versus centralities of leaders.

SUPPLEMENTARY MATERIALS

Materials and Methods
Supplementary Text
Tables S1 to S7
References

Diffusion of information and participation. (**Left**) First-informed households have decided whether to participate and stochastically pass on information to their neighbors. (**Right**) Participation may affect the probability of passing information. Newly informed nodes make their decisions, possibly being influenced by the decisions of their neighbors. After newly informed nodes make their participation decisions, all informed nodes engage in another round of stochastic communication.

Information is passed on by leaders; leadership participation affects probability of information sharing.
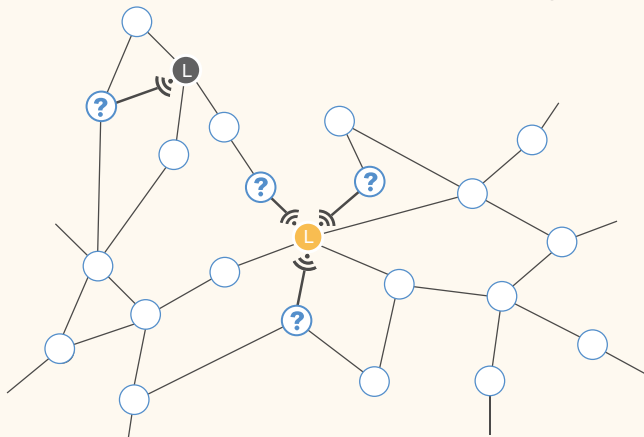
All informed nodes pass on information further; the probability of information sharing is, again, based on participation.
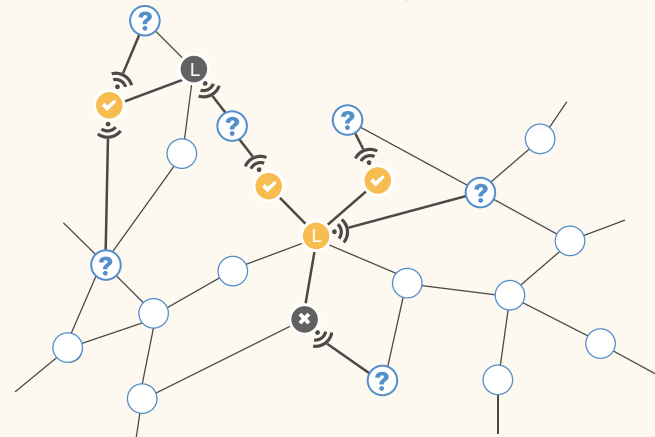
The list of author affiliations is available in the full article online.
*Corresponding author. E-mail: banerjee@mit.edu (A.B.); arungc@stanford.edu (A.G.C.); eduflo@mit.edu (E.D.); jacksonm@stanford.edu (M.O.J.)

The Diffusion of Microfinance

Abhijit Banerjee,[1]* Arun G. Chandrasekhar,[2,3]* Esther Duflo,[1]* Matthew O. Jackson[3,4,5]*

To study the impact of the choice of injection points in the diffusion of a new product in a society, we developed a model of word-of-mouth diffusion and then applied it to data on social networks and participation in a newly available microfinance loan program in 43 Indian villages. Our model allows us to distinguish information passing among neighbors from direct influence of neighbors' participation decisions, as well as information passing by participants versus nonparticipants. The model estimates suggest that participants are seven times as likely to pass information compared to informed nonparticipants, but information passed by nonparticipants still accounts for roughly one-third of eventual participation. An informed household is not more likely to participate if its informed friends participate. We then propose two new measures of how effective a given household would be as an injection point. We show that the centrality of the injection points according to these measures constitutes a strong and significant predictor of eventual village-level participation.

How to implant useful information into social networks so that it benefits the maximum number of people is a question of great importance for policy-makers. Although simulations (*1*) and analytic results (*2*) suggest that the choice of initial injection points in a diffusion process (the first persons to be informed) affects the ultimate reach of the process, there is little empirical evidence concerning whether this is actually the case in real-life examples, and if so, in what ways (*3*). Moreover, the answer to this question crucially depends on the model of information transmission. As information about a new product diffuses through a social network, what are the factors that influence whether an individual chooses to adopt or purchase that product?

We consider two main factors. First, individuals have to be aware of the product before they can adopt, which is more likely when more of their friends can tell them about it. Second, the adoption decisions of informed individuals might be influenced by the decisions of their friends. To account for these factors, we developed a simple model of information diffusion that allows us to (i) distinguish information passing among neighbors from direct influence of neighbors' participation decisions, and (ii) distinguish information passing by participants versus nonparticipants.

We then proceeded in four steps. First, we derived model parameter estimates from a uniquely rich data set on network structure and participation in microfinance in 43 rural villages in Karnataka, a state in southern India. The network data were collected in anticipation of the

introduction of services in these villages by a microfinance institution, Bharatha Swamukti Samsthe (BSS), and were drawn from detailed surveys of households covering a wide range of interactions. BSS then entered these villages and provided us with data on participation in microfinance at regular intervals.

An important feature of this context is that the set of potential injection points is known. BSS relies on word-of-mouth communication to reach potential borrowers. When its representatives start working in a village, they begin by inviting a set of "leaders" (e.g., teachers, shopkeepers, savings group leaders) to an information meeting, and then asking those leaders to spread the information. Network distance to these leaders therefore offers a proxy for access to information about microfinance, and our estimation of the structural model is based on the correlation between access to information about microfinance and participation in microfinance.

The results from the structural exercise are of independent interest. We show that to explain the observed patterns, we need to allow for information about microfinance to be transmitted both by those who choose to participate in microfinance and those who do not. On the other hand, once the household is informed that microfinance is available, it does not seem to matter whether the information came from a participant or a nonparticipant (we find no "endorsement" effects).

The second step was to validate these parameter estimates. Validation is important because the variation that identifies these estimates is non-experimental and may be partly driven by homophily (the tendency of individuals to be linked to others with similar characteristics) or correlated unobserved stimuli or shocks (*4*). We show that the spreading pattern generated by simulating the model, given the observed injection points and estimated parameters, is similar to what we observe in the time series data, which were not used in the estimation of the model.

In the third step, we used the model as a basis for a measure of what we call communication centrality, which identifies how effective any given node would be as an injection point according to the parameters of the model. We show that the communication centrality of the set of original injection points in a village is strongly correlated with participation in that village. Communication centrality of the injection points outperforms the other standard measures of centrality (of the same injection points) in terms of predicting microfinance participation in the cross section of villages.

Although these findings show that BSS would benefit from using communication centrality to maximize participation, this measure has the disadvantage of being informationally demanding: To compute it, we need to know not only the network structure, but also the patterns of diffusion in a sample of villages required for model estimation. Thus, the fourth and final step is to develop a measure in the spirit of communication centrality that does not require estimation of the model and yet can still serve as a proxy for communication centrality. We propose a simple measure along these lines that we call diffusion centrality, and we show that it is strongly correlated with communication centrality and inherits much of its predictive properties (at least in our context).

Context and Data

Our initial sample was a list of 75 villages where BSS was planning to start operating within the following year. These villages are spread across five districts in Karnataka, with a median distance of 46 km from other villages in the sample; typically the villages are far enough from each other that we can regard them as independent systems for the questions that we are asking. These villages are, by and large, linguistically homogeneous but heterogeneous in terms of caste. The most common primary occupations are agricultural work, sericulture, and dairy production. Until BSS's entry, these villages had almost no exposure to microfinance institutions and had limited access to any type of formal credit.

In 2006, 6 months before BSS's entry into any village, we conducted a baseline survey in all 75 villages. This survey consisted of a village questionnaire, a full census that collected data on all households in the villages, and a detailed follow-up survey fielded to a subsample of individuals. In the village questionnaire, we collected data on the village leadership, the presence of preexisting nongovernmental organizations (NGOs) and savings self-help groups (SHGs), and various geographical features of the area (such as rivers, mountains, and roads). In the household census, we gathered demographic information, GPS coordinates, and data on a variety of amenities (e.g., roofing material, type of latrine, quality of access to electric power) for every household in each village.

[1]Department of Economics, Massachusetts Institute of Technology, Cambridge, MA 02142, USA. [2]Microsoft Research New England, Cambridge, MA 02142, USA. [3]Department of Economics, Stanford University, Stanford, CA 94305, USA. [4]Santa Fe Institute, Santa Fe, NM 87501, USA. [5]CIFAR, Toronto, Ontario M5G 1Z8, Canada.

*Corresponding author. E-mail: banerjee@mit.edu (A.B.); arungc@stanford.edu (A.G.C.); eduflo@mit.edu (E.D.); jacksonm@stanford.edu (M.O.J.)
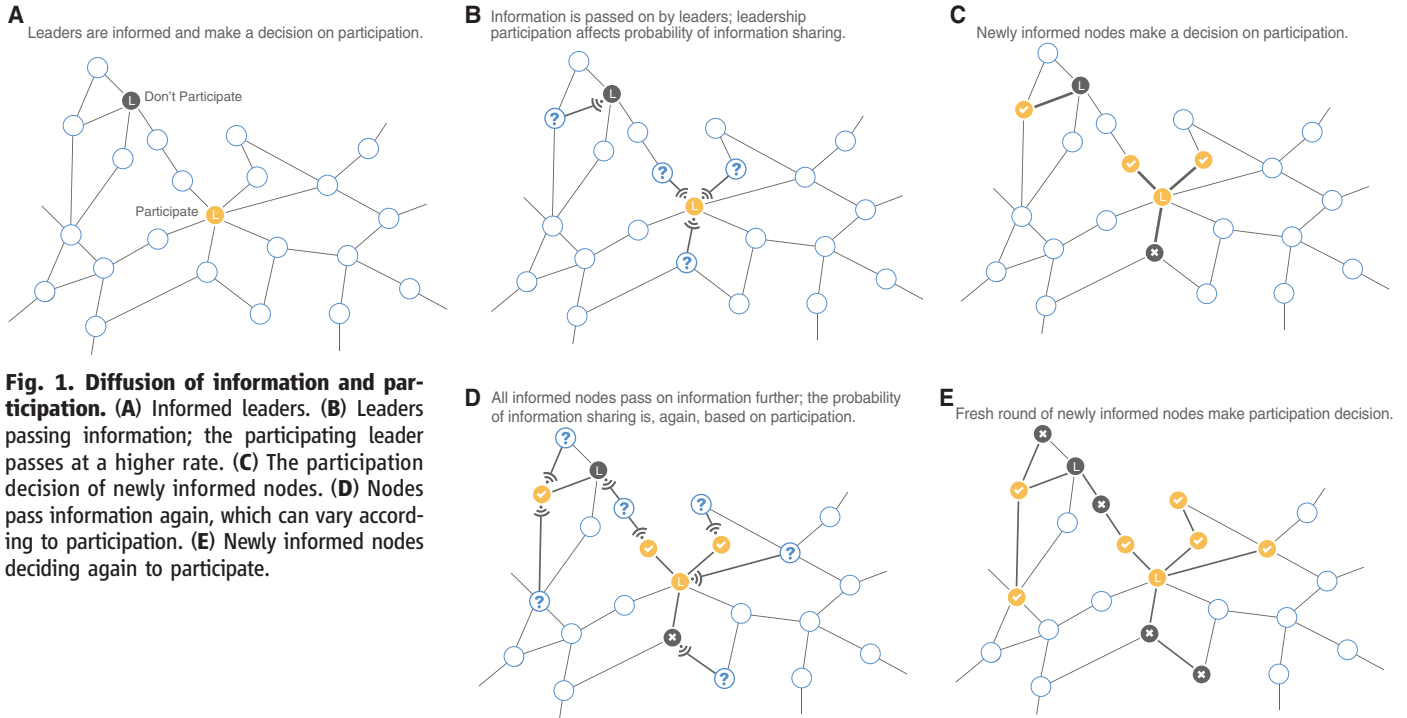
Fig. 1. Diffusion of information and participation. (**A**) Informed leaders. (**B**) Leaders passing information; the participating leader passes at a higher rate. (**C**) The participation decision of newly informed nodes. (**D**) Nodes pass information again, which can vary according to participation. (**E**) Newly informed nodes deciding again to participate.

After the village and household modules were completed, a detailed individual survey was administered to a subsample of villagers. Respondents were randomly selected, and we stratified sampling by religion and geographic sublocation. More than half of the BSS-eligible households (i.e., those with females between the ages of 18 and 57) in each stratification cell were randomly sampled. Individual surveys were administered to eligible members and their spouses, yielding a sample of about 46% of all households per village, and we corrected some of our measures for missing data. The individual questionnaire asked for information including age, caste, education, language, native home, and occupation. So as to not prime the villagers to join BSS or suggest any possible connection with BSS (which would enter the villages later), we did not ask for explicit financial information.

These individual surveys also included a module that collected social network data along 12 dimensions: names of those who visit the respondent's home, those whose homes the respondent visits, kin in the village, nonrelatives with whom the respondent socializes, those from whom the respondent receives medical advice, those from whom the respondent would borrow money, those to whom the respondent would lend money, those from whom the respondent would borrow material goods (kerosene, rice, etc.), those to whom the respondent would lend material goods, those from whom the respondent gets advice, those to whom the respondent gives advice, and those with whom the respondent goes to pray (at a temple, church, or mosque) (*5*).

In 2007, after we finished data collection, BSS began operations in some of these villages. By the time we finished collecting data for this study in early 2011, BSS had entered 43 of the villages. Across a number of demographic and network characteristics, the villages they entered look similar to the ones they did not (*6*). Our analyses focus on the 43 villages in which BSS introduced its program.

In these villages, BSS provided us with regular administrative data on who joined the program, which we matched with our demographic and social network data. When BSS started to work in a village, it sought out a number of predefined leaders whom they expected to be well-connected within the village (e.g., teachers, leaders of self-help groups, and shopkeepers). BSS first held a private meeting with leaders that were amenable to it, and credit officers explained the program and asked the leaders to help organize a meeting to present information about microfinance to the village. These leaders play an important part in our identification strategy, as they function as injection points for microfinance in the village. We used the full set of predesignated leaders, as opposed to the subset of leaders who actually worked with BSS in each village (both because this is endogenous—whether or not a leader worked with BSS could correlate with omitted variables that may bias estimation—and because we did not always have this information).

Model and Structural Estimation

Our simple model of diffusion on a network is depicted in Fig. 1. We model the diffusion of participation as a process on the household-level

network, with participation decisions being made at the household level. As such, a node represents a household (which is the appropriate unit for microfinance). The model can be summarized as follows:

1) An initial set of households is informed (injection points).

2) The initial households decide whether to participate.

3) In each subsequent period, households that have been informed in previous periods pass information to each of their neighbors, independently, with probability q^P if they are participants and with probability q^N if they are not.

4) Newly informed households then decide whether to participate. This decision may depend on a newly informed household's characteristics and potentially on the previous participation choices of their neighbors who told that household about microfinance (*7*). Previously informed households do not have a second chance to decide.

5) The process stops after T periods of information passing.

If $q^N = 0$, so that only participating households pass information, and $T = \infty$, then this is a variant of the standard Susceptible, Infectious, Recovered (SIR) model (*8*, *9*). By allowing it to operate only for T periods, we study what happens in finite time (because after enough rounds, everyone would be informed). Both the finite horizon and the fact that nonparticipants can pass information are important realistic features in most applications.

To capture endorsement effects, let p_{it} denote the probability that an individual who was just

Table 1. Parameter estimates of the structural model. q^N represents the probability that a household that is informed about microfinance but has decided not to participate transmits information to a neighbor in a given period, and q^P represents the probability that a household that is informed and has decided to participate transmits information to a neighbor in a given period. λ is the coefficient in Eq. 1 on the fraction of neighbors that informed a household about microfinance who themselves decided to participate. We use village-level Bayesian bootstrap estimates of the model parameters with 1000 draws to compute the standard errors of the parameter estimates (in parentheses).

	q^N	q^P	λ	$q^N - q^P$
Information model	0.050 (0.01)	0.350 (0.124)	—	−0.300 (0.126)
Information model with endorsement	0.050 (0.007)	0.500 (0.234)	−0.200 (0.146)	−0.450 (0.234)

informed about microfinance decides to participate, where p_{it} is a function of the individual's characteristics X_i [which can account for homophily based on observables (10, 11)] and peer decisions. We model it as a logistic function:

$$\log\left(\frac{p_{it}}{1 - p_{it}}\right) = X_i'\beta + \lambda F_{it} \qquad (1)$$

where F_{it} is a fraction whose denominator is the number of i's neighbors who informed i about the program and whose numerator is the number of these individuals who participate in microfinance, λ represents the change in the log-odds ratio of household i participating because of a change in the fraction of neighbors who informed i that chose to participate, and β represents a vector of coefficients that describe how the log-odds ratio of participation changes as characteristics X_i change. (In table S4, we also experiment with different weights on different neighbors based on their centrality in the network.)

In what follows, the information model constrains λ to be equal to zero, whereas the information model with endorsement effects estimates λ.

Estimation

The model was estimated using the method of simulated moments (MSM) (12, 13). Specifically, we chose parameters to minimize an objective function that is a quadratic form of the distance between moments observed in the data and the same moments as predicted by the model for a specific combination of parameters. We selected parameters so that the following moments predicted by the model best matched the actual moments in the villages:

1) The share of leaders who participate in microfinance.

2) The share of households with no participating neighbors that participate.

3) The share of households in the neighborhood of a participating leader that participate.

4) The share of households in the neighborhood of a nonparticipating leader that participate.

5) The covariance of household participation with the share of its neighbors that participate.

6) The covariance of household participation with the share of its second-degree neighbors that participate.

Model estimation proceeds in three steps. First, we estimate β via logistic regression using the participation decisions among the set of leaders (who are assumed to be informed of the program). X consists of a rich set of covariates including quality of access to electricity, quality of latrines, number of beds, number of rooms, the number of beds per capita, and the number of rooms per capita. These covariates vary substantially across the nearly 1140 leaders throughout the 43 villages, and so the parameters are tightly estimated.

Second, to estimate q^N, q^P, and λ (or any subset of these in the restricted models), we proceed as follows. The parameter space Θ is discretized (henceforth we use Θ to denote the discretized parameter space) and we search over the entire set of possible parameters. For each possible choice of $\theta \in \Theta$, we simulate the model 75 times, each time allowing as many rounds of communications in the diffusion process as the number of trimesters that a given village was exposed to microfinance (typically 5 to 8). For each simulation, moments 2 to 6 are calculated. Next, we take the average over the 75 runs. This gives us the vector of average simulated moments, which we denote $m_{\text{sim},r}$ for village r. We let $m_{\text{emp},r}$ denote the vector of empirical moments for village r. Finally, we choose the set of parameters that minimizes the criterion function, namely

$$\hat{\theta} = \arg\min_{\theta \in \Theta}\left[\frac{1}{R}\sum_{r=1}^{R}m_{\text{sim},r}(\theta) - m_{\text{emp},r}\right]'$$
$$\times \hat{W}\left[\frac{1}{R}\sum_{r=1}^{R}m_{\text{sim},r}(\theta) - m_{\text{emp},r}\right] \qquad (2)$$

where $\hat{W}$ is a weighting matrix obtained by first using this estimation with an identity matrix weight, obtaining a first-stage estimate $\tilde{\theta}$, and then estimating the optimal weighting matrix making use of $\tilde{\theta}$.

Third, to estimate the distribution of $\hat{\theta}$, we use a simple Bayesian block-bootstrap algorithm. The bootstrap takes into account the fact that decisions are interrelated across household within villages but assumes independence of networks across villages (because villages are far apart). This method allows us to estimate standard errors in a computationally simple manner for an MSM model that requires numerous runs of a compli-

cated diffusion process (14). The estimation procedure and bootstrap are explained in detail in the supplementary material.

Identification of the Diffusion Model

The first set of moments combined with the injection points allow us to identify the parameters of the model. The intuition behind the identification of endorsement effects and differential information effects in our application can be clarified by a simple two-by-two example. Imagine, for example, that $q^N = 0.05$ and $q^P = 0.35$ (these are the estimated parameters that we report in Table 1). Also assume for sake of discussion that an informed individual with no participating friend joins with probability 0.22, whereas an informed individual with all participating friends joins with probability 0.32 (corresponding to $\lambda = 0.5$).

Consider four individuals: One of them has one friend who is a leader, and this leader takes up microfinance; the second one has one friend who is a leader but does not participate in microfinance; the third has four friends who are leaders, and all participate in microfinance; the fourth has four friends who are leaders, and none of them participate in microfinance. On average, if the model runs for seven periods (the average number of periods we observe in a village), the probability that the first person is informed is $0.95 = 1 - (0.65)^7$. Similarly, the probability that the second person is informed is 0.3. The probability that the third person is informed is essentially 1, and the probability that the fourth person is informed is 0.76. Although the difference in the fraction of informing friends who take up microfinance, which is the source of the endorsement effect, is exactly the same for person 1 versus person 2 as it is for person 3 versus person 4 (it is 1 in both cases), the difference in participation between persons 1 and 2 (0.24) is much larger than the difference in participation between persons 3 and 4 (0.15). This difference captures the pure information effect.

To see the pure endorsement effect, let one of the four leaders who are friends with person 4 participate. The probability that person 3 and person 4 are informed is now more or less the same (about 1) and therefore, in a pure–information effect world, they would behave identically. However, if there is an endorsement effect, these two will behave quite differently. Under the parameter assumptions above, person 3, who has a

higher fraction of informing friends who participate, is more likely to participate (0.32 versus 0.24 for person 4).

The fact that these parameters are formally identified in the context of our model does not mean that they could not be spurious. As often in network-based studies, causal interpretation of the correlation between the decisions of connected people as the result of information transmission or endorsement (which is partly what we use to identify the model) is potentially questionable. This is, of course, the standard identification problem with observational data on networks.

First, imagine that the data are generated by a model wherein all households know about microfinance but differ on preferences. The probability of participating is modeled as a logistic probability model,

$$\log\left(\frac{p_i}{1-p_i}\right) = X_i'\beta + \nu_i \qquad (3)$$

where ν_i is a preference shock that can be correlated with ν_j, where, for instance, i and j are neighbors in the graph. In such a model, one may worry that the correlations in preference generate cross-sectional participation patterns that look like diffusion. It is possible that, for example, households with neighbors who participate are themselves more likely to need microfinance (in ways that we cannot pick up with our demographic information) because, for example, neighbors may share a common activity or may have common access to finance. In that case, the interpretation of our parameters as capturing the effect of j's decision on i's choice would be inappropriate.

In this alternative model, however, the diffusion of microfinance would not follow specific time patterns of participation based on network distances from first-informed individuals, whereas our model makes specific predictions about the pattern of diffusion over time. A test of the model is thus whether the empirical time series of adoption matches a corresponding simulation of the model, given the set of parameters we estimate.

Second, even if our model were correct, the presence of unobserved correlated effects influencing participation could bias our estimate of endorsement effects. Consider what would happen if we modify Eq. 1 by including a shock term, ν_i:

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = X_i'\beta + \lambda F_{it} + \nu_i \qquad (4)$$

Again ν_i is a time-invariant term, unobserved to the econometrician, that may be correlated across network neighbors. In this case, diffusion over time in the data may look similar to what the model would predict, but our cross-sectional identification in the presence of positively correlated

shocks could lead us to estimate a spurious positive λ.

Parameter Estimates

Table 1 presents the result of the estimation. The first row presents the parameters of the information model without any endorsement effects: $q^N = 0.050$ and $q^P = 0.350$. Both of these values are significantly different from zero ($P < 0.01$, t test). In addition, we are able to reject equality of the two parameters ($P < 0.01$, t test).

These results highlight the role of nonparticipants in the diffusion process. Even though they pass information at a much lower rate than participants, there are many more nonparticipants in a village than participants, and thus they end up playing an important role. In fact, our estimates indicate that information passing by nonparticipants is responsible for nearly one-third of overall informedness and participation. We calculate this figure by simulating information spread in the model, constraining q^N to be equal to 0 (and setting q^P at what we estimate). The eventual participation would then drop from 20.0% to 13.97%.

The second row of Table 1 presents estimates of the model in which endorsement is included and the villagers potentially pay attention to the participation decisions of their informed neighbors. There is no significant evidence for a positive endorsement effect: Once a household is informed, its decision to participate in microfinance is not significantly affected by whether its neighbors chose to participate themselves. If anything, the point estimate of the endorsement effect is generally negative, and in the supplementary materials we show that depending on the weighting of the friends, it is sometimes (marginally) significantly negative (table S4).

The finding that we cannot reject that $\lambda = 0$ in most specifications, and can reject that it is positive, provides some reassurance that the estimates are not driven by unobserved correlated heterogeneity between neighbors; in that case, we would find a positive λ.

Robustness and Model Validation

There are two main identification concerns: First, we have treated leaders symmetrically with everyone else, but they may be different in ways that are not entirely captured by the impact of observed characteristics on participation. For example, they may be more likely than anyone else to inform their friends about microfinance (e.g., they may have brought some of them with them to the initial meeting) but no more likely to participate themselves. In this case, we would observe a high participation among friends of people who are informed but do not participate themselves, which would drive q^N up, even if no one except the leaders does that.

To address this issue, we estimated three nested variations of the model (table S6). First, we removed moments 3 and 4, which explicitly

rely on participation decisions within a leader's neighborhood for identification (table S6, case 1). Second, in addition to removing moments 3 and 4, we excluded observations for which the node in question is a leader when constructing moments 2, 5, and 6 (table S6, case 2). Third, we modified case 2 by also excluding leaders from the neighborhoods of other nodes in the computation of moments 2, 5, and 6 (table S6, case 3). In each case, the parameter estimates were not statistically different from those in our baseline specification and the point estimates were stable.

The second concern is the possibility of correlated shocks we discussed above. As already mentioned, one way to address this issue is to use the estimated parameters to predict the time series of how microfinance participation spreads by village. The time structure of diffusion was not used in the estimation of the parameters; we used correlations between neighbors' ultimate decisions. Thus, checking the time series of participation under the model and seeing whether it matches the data provides a way to validate the model. In particular, if the appearance of diffusion is caused by neighbors having correlated needs but these needs do not exhibit a diffusion pattern over time, then there is no reason why the model would match the data.

In this validation exercise, we focus on what happens after the first trimester has elapsed. This is because in our model a period is simply a round of communication and may not correspond to a fixed period of time. In the villages, more adoption and more rounds of communication occur in the first trimester, when microfinance is new. In other words, there are more "periods" within the first trimester. Moreover, some people may be informed in the first trimester for reasons that have nothing to do with the model—for example, because they happen to have found themselves at the first BSS meeting.

What is the equivalent of the end of the first trimester within our model? We estimate it as the number of model periods required for simulated participation in the model to reach the actual level of participation achieved in that village at the end of the first trimester. For example, in village 1, the actual participation at the end of the first trimester was 13%. In the simulation for that same village, period 1 participation is 9% < 13%, and period 2 participation is 15% > 13%. In this case, we assume that the first trimester is equivalent to two model periods for that village. (In 11 villages, the simulated participation never reached the observed participation at the beginning of period 1, so we could not perform the exercise and thus dropped them from the analysis).

We then ran a regression of the observed microfinance participation rate on the simulated participation rate, using fixed effects for both village and time period (Table 2). Fixed effects for time period are important here to avoid a spurious correlation generated by regressing two monotonically increasing processes (and ones

Table 2. Time series validation. Values shown are coefficients from ordinary least-squares regressions. Each column represents a different regression. The dependent variable is the empirical participation rate in a village in a given period, from trimester 1 to the final trimester corresponding to December 2010. The main independent variable, simulated participation rate, is the mean participation rate in a village in a period (where period 1 corresponds to the empirical period 1 participation rate) across 1000 simulations of the diffusion model. All regressions include fixed effects for village and time period. The demographic controls include number of households, savings, self-help group participation, caste composition, and fraction of households that are leaders, each interacted with time. Standard errors (in parentheses) are clustered at the village level.

	Without demographic controls	With demographic controls
Simulated participation rate	0.921 (0.612)	0.652 (0.261)
Observations	125	125

Table 3. Microfinance participation versus centralities of leaders. Values shown are coefficients from ordinary least-squares regressions. Each column represents a different regression. The dependent variable is the microfinance participation rate of nonleader households in a village. The covariates are various measures of centrality—averaged over the set of leaders—as well as control variables. The controls include number of households, savings, self-help group participation, caste composition, and fraction of households that are leaders. Standard errors (in parentheses) are robust to heteroskedasticity.

Centrality measure	Regression			
	1	2	3	4
Communication	0.766 (0.335)		0.713 (0.428)	
Diffusion		0.022 (0.007)		0.018 (0.009)
Degree			−0.005 (0.006)	−0.003 (0.006)
Eigenvector			3.572 (2.330)	3.692 (2.265)
Betweenness			1.709 (1.776)	1.710 (1.687)
Katz-Bonacich			−0.072 (0.070)	−0.106 (0.063)
Decay			0.032 (0.047)	0.034 (0.045)
Closeness			−1.077 (1.551)	−0.891 (1.496)
R^2	0.406	0.442	0.530	0.515

that both taper off after a few periods) on each other. Similarly, unobserved village-level heterogeneity may generate spurious correlations, which we want to eliminate as well.

Table 2 presents this regression without and with demographic controls. The point estimates suggest strong correlation between incremental changes in the simulated microfinance participation rate and empirical participation rate, even when accounting for the time effects and village fixed effects. Without demographic controls, the coefficient does not have statistical significance at conventional levels under a two-sided test ($P = 0.142$, t test), whereas with demographic controls the estimate is statistically significant ($P = 0.018$, t test). To ensure conservative inference given the likelihood of temporally correlated errors in the regression model, we cluster our standard errors at the village level. This close correspondence between the time series as predicted by the model (after period 1) and the time series we observe in the data suggests that time-invariant omitted variables and homophily are not driving the results.

Application: The Impact of Injection Points

Suppose a microfinance organization like BSS relies on word-of-mouth diffusion to spread information about the availability of microfinance. How will the eventual participation in microfinance depend on whom they approach first?

To analyze this question, we compute for each leader (our injection points) a score. This score is the fraction of households who would eventually participate if this household were the only one initially informed. To compute this fraction, we simulate the model with information passing and participation decisions being governed by the estimated values of q^N, q^P, and β. We call this score the communication centrality of a node.

In our data, because BSS relies on a fixed rule for choosing leaders, there is considerable variation across villages in the average communication centrality of the set of leaders in a village (which is 0.001 at the 10th percentile and 0.13 at the 90th percentile). This is true for other measures of centrality as well. Moreover, this variation does not come from any information BSS has about the village, and hence is likely to be independent of village characteristics. The identification assumption is that the centrality of the leaders is not correlated with the demand for microcredit, including control variables, and it does not seem to be problematic. In table S2, we regress the various measures of leader centralities on the village characteristics used in Table 3, and we find little relationship between the network positions of the leaders and the characteristics of the villages.

Figure 2 exhibits plots of village-level participation in microfinance as a function of communication centrality and degree centrality (for comparison). The communication centrality of the leaders is strongly and significantly correlated with eventual village-level participation, and the correlation is stronger than for degree.

Table 3 shows the results of a series of four regressions of village-level adoption on various measures of average centrality of leaders, after including village-level controls for savings behavior, self-help group participation, caste composition, and fraction of households with BSS-designated leaders (*15*). The first regression shows that communication centrality is strongly correlated with eventual take-up. The third regression shows that this remains true after controlling for a hosts of other measures of centrality, none of which are significant in this specification.

An Approximation of Communication Centrality: Diffusion Centrality

Communication centrality of the injection points is strong predictor of eventual participation in microfinance and should therefore provide guidance to anyone trying to spread the news about microfinance in similar villages. However, it cannot be computed without those estimates, which could be very different if we were interested in the diffusion of other products or even microfinance in a very different context (say, a city). Thus, we propose an approximation of communication centrality—diffusion centrality—that is highly correlated with communication centrality, at least in this setting, but requires considerably less data. In particular, it does not rely on estimating the diffusion model.

We start from our model with $q^N = q^P = q$. Although we have shown that q^N and q^P differ in the data, this can be useful as an approximation in settings where the full model may be difficult to
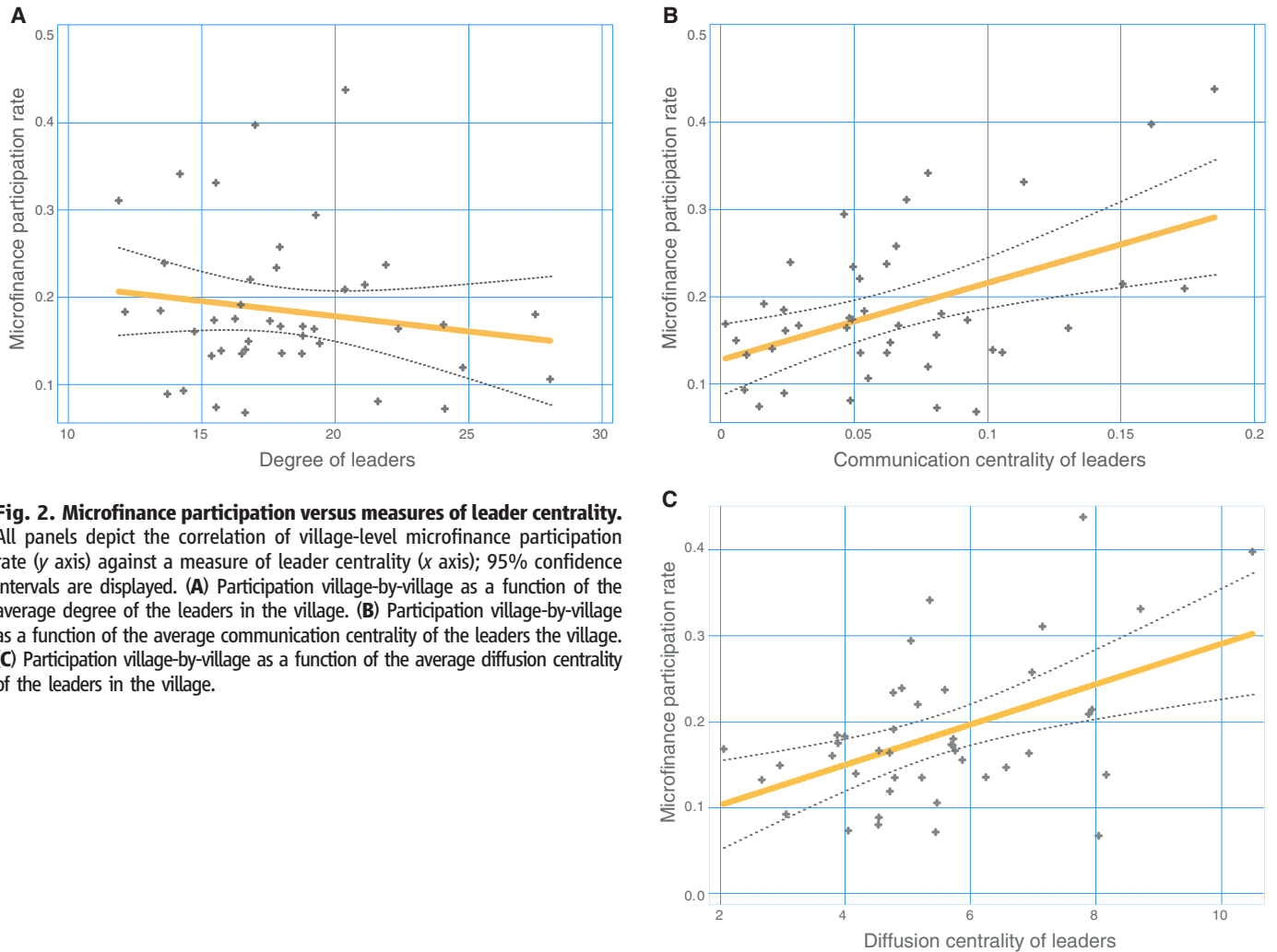
Fig. 2. Microfinance participation versus measures of leader centrality. All panels depict the correlation of village-level microfinance participation rate (*y* axis) against a measure of leader centrality (*x* axis); 95% confidence intervals are displayed. (**A**) Participation village-by-village as a function of the average degree of the leaders in the village. (**B**) Participation village-by-village as a function of the average communication centrality of the leaders the village. (**C**) Participation village-by-village as a function of the average diffusion centrality of the leaders in the village.

estimate. This suggests a simple measure of the centrality or potential influence of each node. We define the diffusion centrality of a node *i* in a network with an adjacency matrix **g**, passing probability *q*, and iterations *T*, as the *i*th entry of the vector

$$DC(\mathbf{g}; q, T) := \left[\sum_{t=1}^{T} (q\mathbf{g})^t\right] \cdot 1 \qquad (5)$$

Consider *T* iterations of information passing from a single initially informed node *i* where at each iteration every informed node tells each neighbor with probability *q*. The diffusion centrality of node *i* then corresponds to the expected total number of times that all nodes taken together hear about the opportunity. If *T* = 1, diffusion centrality is proportional to degree centrality. As *T* → ∞ it becomes proportional to either Katz-Bonacich centrality or eigenvector centrality, depending on whether *q* is smaller than the inverse of the first eigenvalue of the adjacency matrix or exceeds it, respectively (*16*). In the intermediate region of *T*, the measure differs from existing measures.

Any method of computing a measure of diffusion centrality that does not rely on the estimation of the model requires the choice of an appropriate value for *q*. Extreme values of *q* lead either to no diffusion or to complete diffusion, and so do not distinguish nodes. We choose a prominent intermediate value of *q*: the inverse of the first eigenvalue of the adjacency matrix, $\lambda_1(\mathbf{g})$. This is the critical value of *q* for which the entries of $(q\mathbf{g})^T$ tend to 0 as *T* grows if $q < 1/\lambda_1$ and some entries diverge if $q > 1/\lambda_1$.

In essence, diffusion centrality uses our model as a starting point but assumes that everyone spreads information with the same probability *q*, which is selected such that information spreads at a rate that neither saturates too quickly nor dies out. For each village, we set *T* to the number of trimesters during which the village was exposed to BSS (6.6 on average). The choice of *q* and *T* can be important. However, in our data, diffusion centrality is not very sensitive to the choice of *q* and *T* within a reasonable range. In table S7 we compute it for other values of *q* in the neighborhood of $1/\lambda_1$ and a range of values of *T*. The identity of the most diffusive leader is robust to

these changes. Diffusion centrality is strongly correlated with communication centrality; the correlation is 0.86. Consequently, the average diffusion centrality of the leaders performs equally well in predicting eventual participation (Fig. 2C). This is true even when accounting for demographic control variables (Table 3, column 2) and other standard centrality measures (degree centrality, eigenvector centrality, Katz-Bonacich centrality, betweenness centrality, decay centrality, or closeness centrality). This is robust to choosing values of *q* and *T* within a range of 25% around their assumed value here.

These findings highlight the importance of injection points: The correlation with eventual village-level participation (Fig. 2C) implies that an increase in the diffusion centrality of leaders from the 10th percentile to the 90th percentile would lead to an increase in eventual participation in microfinance by 10.7 percentage points.

Conclusion

We estimate a model of diffusion that allows both for information and endorsement effects. In this context, we find no evidence of strong endorsement effects: The role of neighbors in the diffusion

process is to pass information, and even those who are not taking up the program themselves play a role in this process, although their probability to pass information is lower than that of those who adopt the product. This model of pure information diffusion motivates a new centrality measure for measuring the effectiveness of alternative injection points that differs from standard centrality measures and, in our sample, performs better than them. This has important implications for policy makers and firms that are trying to pick the right people to inform in order to ensure that a new idea or product or piece of information reaches the maximum number of people in the network.

The results suggest a number of directions for future work: first, to test how well this new centrality measure does in predicting diffusion patterns when we purposefully vary the centrality of the injection point; second, to understand to what extent it is possible to improve on the theory of information passing and learning used here by introducing more sophisticated approaches to learning and strategic motives for information sharing, and what that implies for choosing the right centrality measures; and third, to use the theoretical insights that come out of that work to understand where, and in what ways, endorsement effects may be important and how best to model them.

References and Notes

1. M. Kitsak *et al.*, Identification of influential spreaders in complex networks. *Nat. Phys.* **6**, 888–893 (2010). doi: 10.1038/nphys1746
2. C. Ballester, A. Calvó-Armengol, Y. Zenou, Who's who in networks. Wanted: The key player. *Econometrica* **74**, 1403–1417 (2006). doi: 10.1111/j.1468-0262.2006.00709.x
3. There is an old and large literature that studies information diffusion (*17–25*). However, empirical evidence on the role of injection points is still sparse.
4. For experimental approaches to controlling for and analyzing the role of homophily in diffusion, see (*26, 27*).
5. Individuals were allowed to name as many as five to eight network neighbors, depending on the category. The data exhibit almost no top-coding; fewer than 10% of the respondents named the maximum number of individuals in any single category.
6. The main difference is the number of households: Villages that BSS entered had 223.2 households on average (SD = 56.17); those it did not enter had 165.8 households on average (SD = 48.95).
7. One could imagine that households need some time to think, or to process information from their neighbors, before adoption (*21, 28*). However, in our setting, one period (4 months) encompasses much of this decision time. It may be that in the long run (over a period of years) they would reconsider their decisions, which our model does not address, and so we do not capture long-run reactions to neighbors' experiences. For instance, it could be that there are coordination problems as people wait for others to participate before participating themselves. This seems unrealistic in our context: It takes about 1 to 2 years to learn how neighbors did with microfinance (whether they managed to repay, whether they benefited, etc.). Our data show that most of the adoption of microfinance takes place within the first year in a village (before the first cycle of loans is complete).
8. W. Kermack, A. McKendrick, A contribution to the mathematical theory of epidemics. *Proc. R. Soc. London Ser. A* **115**, 700–721 (1927). doi: 10.1098/rspa.1927.0118
9. N. Bailey, *The Mathematical Theory of Infectious Diseases and Its Applications* (Griffin, London, 1975).
10. S. Aral, Social science: Poked to vote. *Nature* **489**, 212–214 (2012). doi: 10.1038/489212a; pmid: 22972291
11. S. Aral, L. Muchnik, A. Sundararajan, Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21544–21549 (2009). doi: 10.1073/pnas.0908800106; pmid: 20007780
12. D. McFadden, A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57**, 995–1026 (1989). doi: 10.2307/1913621
13. For an example of a method of simulated moments approach used in a spatial setting, see (*29*).
14. For example, 1000 runs of a simulated annealing search would be prohibitively slow.
15. The results are essentially identical without controls.
16. For background on centrality definitions, see (*30*).
17. B. Ryan, N. C. Gross, The diffusion of hybrid seed corn in two Iowa communities. *Rural Sociol.* **8**, 15 (1943).
18. P. Lazarsfeld, B. Berelson, G. Gaudet, *The People's Choice: How the Voter Makes Up His Mind in a Presidential Campaign* (Duell, Sloan and Pearce, New York, 1944).
19. E. Katz, P. Lazarsfeld, *Personal Influence: The Part Played by People in the Flow of Mass Communication* (Free Press, New York, 1955).
20. J. Coleman, E. Katz, H. Menzel, *Medical Innovation: A Diffusion Study* (Bobbs-Merrill, Indianapolis, IN, 1966).
21. A. Foster, M. Rosenzweig, Learning by doing and learning from others: Human capital and technical change in agriculture. *J. Polit. Econ.* **103**, 1176–1209 (1995). doi: 10.1086/601447
22. T. Conley, C. Udry, Learning about a new technology: Pineapple in Ghana. *Am. Econ. Rev.* **100**, 35–69 (2010). doi: 10.1257/aer.100.1.35
23. M. Kremer, E. Miguel, The illusion of sustainability. *Q. J. Econ.* **122**, 1007–1065 (2007). doi: 10.1162/qjec.122.3.1007
24. M. O. Jackson, L. Yariv, in *Handbook of Social Economics*, J. Benhabib, A. Bisin, M. Jackson, Eds. (North Holland, Amsterdam, 2010).
25. R. M. Bond *et al.*, A 61-million-person experiment in social influence and political mobilization. *Nature* **489**, 295–298 (2012). doi: 10.1038/nature11421; pmid: 22972300
26. D. Centola, The spread of behavior in an online social network experiment. *Science* **329**, 1194–1197 (2010). doi: 10.1126/science.1185231; pmid: 20813952
27. S. Aral, D. Walker, Identifying influential and susceptible members of social networks. *Science* **337**, 337–341 (2012). doi: 10.1126/science.1215842; pmid: 22722253
28. B. Jovanovic, Y. Nyarko, Learning by doing and the choice of technology. *Econometrica* **64**, 1299 (1996). doi: 10.2307/2171832
29. G. Topa, Social interactions, local spillovers and unemployment. *Rev. Econ. Stud.* **68**, 261–295 (2001). doi: 10.1111/1467-937X.00169
30. M. O. Jackson, *Social and Economic Networks* (Princeton Univ. Press, Princeton, NJ, 2008).