Microfinance in Karnataka Villages

1032067

In the interest of increasing participation in a microfinance program in villages of Karnataka, we look at communities within the villages. We find many of the communities to be grouped by caste, and households of the same caste interact with each other at a much higher rate than they do with other castes. Furthermore, communities with well-connected leaders (the households first informed of the program) that are representative of the community size and caste breakdown tend to have higher participation. Therefore, we recommend choosing leaders such that the number of leaders in each community is proportional to the community size, the leaders are diverse across castes in the communities and the leaders are central and well known within their communities. This strategy of allocating leaders by targeting communities consistently increases participation, particularly in low participating communities.

Communities and microfinance participation

As the villages are rather large with 200 – 350 households, it is unrealistic to assume each village acts together as one entity. Rather, the villages are separated into several communities, typically by caste. Hence, the main issue we would like to address is the participation within these communities of each village, and how this can affect overall participation. We first use a measure of “caste preference” to see how divided the village is by caste. A higher value means a greater amount of households mostly interact with other households of the same caste. This leads to more divided communities being formed around castes in the village. **Figure 1** shows this measure against village participation rate. We see an overall positive relationship, suggesting caste preference can be useful in increasing participation.

Importance of Community Leaders

Households chosen to be leaders within their communities are the first to hear about the program and therefore have an important influence on the participation rate of the community and the village. Hence, we develop a three part strategy to optimally allocate leaders to increase participation. We compare communities with high participation to those with low participation. First, we find those with high participation have, on average, at least as many leaders proportional to the size of the community. If a high participating community consists of 20% of the village population, then at least 20% of the village leaders are in that community. Second, we find high participating communities have leaders who are more central to their community. This means leaders are both centrally located and have many connections to other households in the community. The occupation of a leader is not nearly as important as how well connected and central that leader is. Third, we consider caste preference to assure representation in choosing the leaders. Villages with high caste preference will have mostly caste homogenous communities, so the leaders will be of the same caste. This is more conducive to interactions and participation. Villages with lower caste preference will have mixed communities, so the leaders’ castes should represent the community. In other words, if the community is 50% OBC and 50% Scheduled Tribe, the caste of the leaders should be similarly split. We implement this strategy on low participating communities across villages and consistently produce equal or higher community participation, especially in communities with originally low participation.

- **Goal: Increase microfinance participation**
- **Villages partitioned into numerous communities by caste**
- **Target communities with leader selection strategy**
  - Number of leaders proportional to community size
  - Leaders are central and well known in community
  - Leaders representative of community

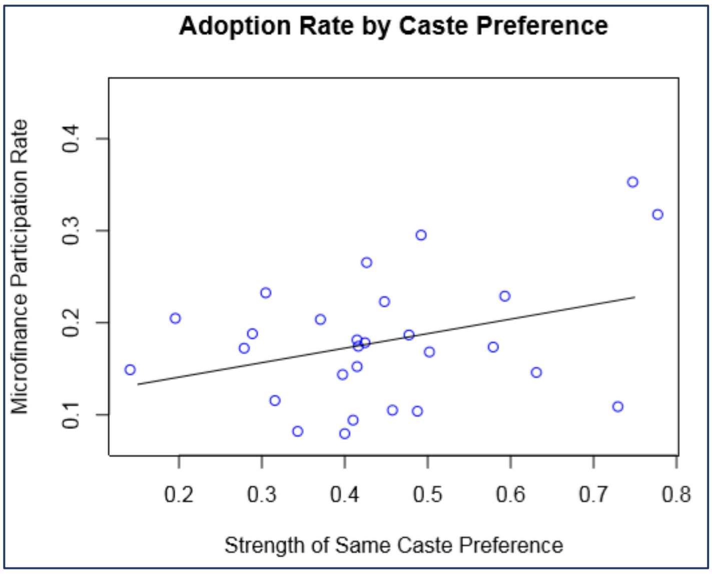

Figure 1. *Source: Investigating Social Network Community Structure in the Adoption of Microfinance. (2019)*

Investigating Social Network Community Structure in The Adoption of Microfinance

1032067[a]

[a]University of Oxford Candidate Number

This manuscript was compiled on March 27, 2019

Identifying community structure is important to understanding how information spreads through a social network. Previous work on the diffusion of microfinance (1) found the centrality of leaders in the network to be an important predictor of adoption, but did not explicitly consider social stratification. We address this by assessing homophily within the networks via assortativity on multiple household attributes and find caste assortativity in particular to be consistently high. We then use this to inform community detection and look for connections between some of the communities and adoption status. A pattern emerges where some networks with low adoption rates but high assortativity have a majority of adoption stored in one to three communities, and almost nothing in all others. From this, we propose a way to select leaders and develop a diffusion model using household attributes and centrality measures at a community level. We observe that while there is no clear increase or decrease in overall adoption rates, adoption within communities is more evenly spread, suggesting the leader selection strategy compared to the original allocation of leaders benefits some communities at the expense of others.

social networks | communities | clustering | homophily

Community structure and social stratification is often an important component to analyzing how information spreads through social networks. Understanding the attributes that connect communities in a network may help to understand how the network functions as a whole.

This analysis focuses on ideas from *The Diffusion of Microfinance*, first presented by Banerjee et al. (1). They analyze how information and adoption of a microfinance program spreads through household networks in 43 villages in the Indian state of Karnataka. The first main finding of the study is the adoption rate predictability of "leaders" (e.g., teachers, shopkeepers) communication centrality. Intuitively, the more central an individual is by this measure, the more influence they have in a network. The second finding is the relative ineffectiveness of endorsement i.e. the decision to participate in the program is not significantly influenced by neighbor's decisions. This approach implicitly addresses homophily, which is the tendency for nodes to connect to socially similar nodes in a network, by using a logistic function with observable household characteristics to distinguish influence driven and homophily driven behaviors as described in (2). Findings from (2) show that peer influence in dynamic networks is drastically overestimated and that homophily can actually explain greater than 50 percent of contagion. While (1) takes this into account, it does not explicitly consider the social stratification within each of the villages. This is important because we can reasonably hypothesize there to be some social attributes by which households tend to be grouped. We aim to use ideas from (1) and (2) to directly consider this stratification. More tightly connected communities within networks may behave differently than other parts of the network, and networks with a strong community structure may need a modified approach to predicting adoption rates. For this analysis, we will mainly focus on networks with high assortativity, which may imply social stratification on the assortative attribute. Additionally, (1) builds their model given the leaders by BSS, who chose leaders assuming they would be well connected based on their occupation. This leads to a less than ideal set of leaders, and even some leaders who are isolated from the rest of the village. We aim to develop a way to choose leaders based on their network properties and attributes.

Data

The network data for interactions and participation as well as household characteristics were collected by the microfinance institution Bharatha Swamukti Samsthe (BSS) over the course of their microfinance implementation scheme. It includes 43 villages and their social connections, and characteristics such as religion, caste, house specifications, whether or not the household has a "leader" and whether or not they adopted the program. For this study, we do not have access to the temporal data of adoption rates over time, only the eventual adoption rate. The data for each village is rather varied, and we consider each one far enough apart to be independent of each other. An interesting observation when first exploring the data is every village has some households that are not connected to any other household, yet have the status of leadership or adoption. A notable example is village 10, the least connected village, which has 24 isolated households, 16 of which adopted microfinance. There is no way to know for sure how some of these isolated households who are not leaders heard of and adopted the program, but it is reasonable to assume that the social interactions are not being completely captured, and that no village is truly disconnected, despite what the data shows. It is also possible some of these households attended a meeting with BSS before the beginning of the program. In any case, we calculate adoption rates then remove households that are disconnected and have no status of interest. We leave the households that are either leaders or participants in the program for modeling and analysis purposes.

Assortativity and Mixing Matrices

We first want to see if there is a tendency for households who adopted the program to connect to each other on the basis of their adoption status, and likewise for non-adopters. We calculate the mixing matrices and assortativity as a quick and straightforward way to assess this homophily. The mixing matrix gives the proportion of edges between and within the

Table 1. Mixing Matrix on Adoption for Village 30

Status	Non-Adopt	Adopt
Non-Adopt	0.791	0.076
Adopt	0.076	0.056

Assortativity: 0.336

Table 2. Mixing Matrix on Caste for Village 30

Caste	General	OBC	Sched. Caste	Sched. Tribe
General	0.349	0.066	0.005	0.002
OBC	0.066	0.285	0.011	0.004
Sched. Caste	0.005	0.011	0.192	0.000
Sched. Tribe	0.002	0.004	0.000	0.000

Assortativity: 0.729

groups in the network. Assortativity, first defined by (3) is calculated from this as:

$$r = \frac{\sum_s (e_{ss} - a_s b_s)}{1 - \sum_s a_s b_s} \qquad [1]$$

where a_s and b_s are row and column sums of group s to represent outgoing and incoming edges to s, and e_s is the proportion of edges within s. The range is from -1 to 1, meaning complete disassortativity and complete assortativity respectively. A value of 0 means the network is essentially random in regard to that attribute. **Table 1** gives the mixing matrix for the most assortative village by adoption. Note since the networks are undirected, the mixing matrices are symmetrical.

Since adoption rates are low across all villages, assortativity by adoption can be misleading. Even looking at **Table 1** we see assortativity is only high due to non-adopters. If we regress adoption rate on adoption assortativity, we get a p-value of nearly 1. It is likely that higher adoption assortativity is being driven by homophily of other household characteristics. To explore this, we calculate mixing matrices and assortativity for religion, caste, home ownership, latrine type, and electricity type. The latter three behave like adoption in that there are at best some villages that have small values of assortativity ($\sim$ 0.2) that is likely driven by some other characteristic or by having an overwhelming amount of households falling into a single category. We decide not to consider these values on account of the inconsistency. Caste and religion, however, consistently yield high assortativity. This is sensible as both attributes can be important factors in Indian society, particularly in social circles. **Table 2** shows the mixing matrix again for village 30, but by caste. The homophily here is much clearer as we see a relatively high proportion of edges within the castes, but very little between them. We plot assortativity on caste and religion against adoption rate in **Figure 1** to visualize the relationship. Note assortativity is not available for all villages either due to lack of data or all households having the same characteristic. We see there is a positive relationship. Regressing adoption rate on both variables yields an insignificant p-value for religion (0.296) and an almost significant value for caste (0.069). While there is not a direct predictive significance, we hypothesize these attributes will be important in

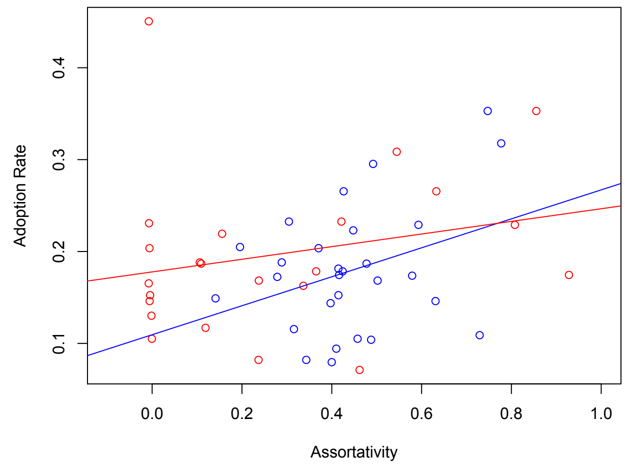

Fig. 1. Scatter plot with fitted lines of assortativity on religion (red) and caste (blue) plotted against adoption rates.

understanding structure from community detection, especially caste.

Community Detection

Looking at assortativity clearly indicates some social structure based on caste and religion. With this in mind, we run a community detection algorithm to first identify these communities and confirm they make sense given the network's assortativity. Then we try to identify villages where the communities align with adoption and non-adoption, as well as community level characteristics that might help explain overall adoption rates.

We use the Louvain Method (4) for community detection. In short, the method begins by treating each node as a community. For a single node, it groups together with each of its neighbors and the modularity is calculated. Whichever grouping produces the highest modularity is kept. Otherwise the node is left unmoved. This is iterated over every node in the network until the modularity stops increasing. The process is then repeated for the resulting clusters or "super nodes" until the modularity stops increasing. Modularity is also a value between -1 and 1 and is defined in (4). Essentially it is the fraction of edges that are within groups minus the expected fraction if the graph was random with an identical degree distribution. A higher modularity indicates a higher level of compartmentalization among the communities. Another thing to consider is the resolution parameter, which controls how large or small the communities are. A parameter greater than 1 will search for a larger number of smaller communities and the opposite for less than 1. As we assume no prior knowledge of the structure, we use the default of 1.

We run the Louvain algorithm for the villages and calculate the modularities. To check for robustness of the communities, we first create a configuration model for each network to act as the null model. We then find community clusters for this model with Louvain. We perturb the original model and the null model by rewiring some number of edges while keeping the degree sequence. The Louvain method is run and Variation of Information, a measure of similarity between two

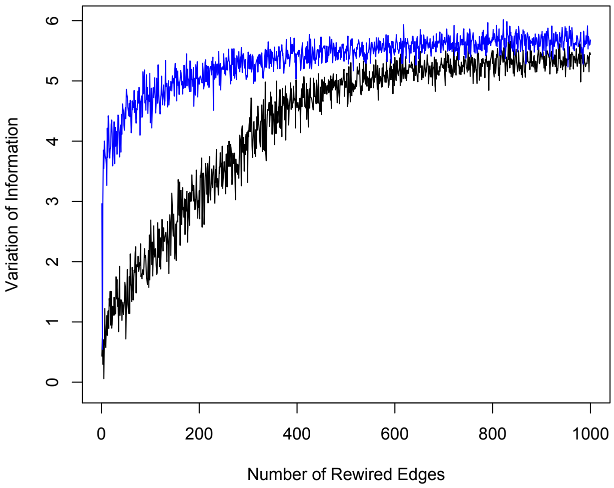

Fig. 2. VI curves for village 30 (black) and its null model (blue). If the community structure were by chance, the black and blue curves would be more similar

sets of clusters, for both models are calculated between the original and new clusters at each rewiring. We do this for up to 1000 rewired edges and plot the curves. If the communities are robust, the $VI_{original}$ curve should deviate from the VI_{null} until the point where both graphs are essentially random. Calculating p-values is complicated so here we rely on visualizations, but the full method is detailed in (5). The curves for village 30 are given in **Figure 2**. The modularity for village 30 is among the highest at 0.596. There could be a connection between robustness and modularity, but we do not explore this.

Adoption Alignment. We now explore the villages to see any alignment or connections with adoption/non-adoption rates. We first look to compare villages that have similarly high assortativity, but significantly different adoption rates. A good example is villages 21 and 30. Both are completely homogeneous on religion and have high assortativity on caste (> 0.7), but the participation rate for village 21 is three times higher than village 30 (0.32 and 0.11). **Table 3** looks at the caste and adoption across communities for village 30 in more detail. We clearly see a large majority of the total adoption in the first community, with very little adoption in the rest. Communities 4 and 6 have no adoption at all despite comprising a fourth of the village. Village 21 (**Table 4**) by contrast does not have this adoption polarization, and is distributed more evenly among communities relative to the size of the community. We see a similar patter when comparing villages 4 and 7. Both are homogeneous on caste with relatively high modularity and religious assortativity , but very different adoption (0.07 and 0.31). The tables are not included for brevity, but the distribution of adoption again shows most adoption for village 4 to be concentrated on a couple communities, whereas village 7 adoption is more evenly spread. This pattern, of varying extremes, emerges in many villages with high assortativity but low adoption. In most of these cases, like village 30 with Scheduled Caste, the adoption

Table 3. Communities and Adoption Village 30

Community	% of Village	% of Adoption	Majority Caste (%)
1	18.2	82.1	Sched. Caste (100)
2	12.9	3.6	OBC (59.4)
3	13.4	0.0	OBC (94)
4	16.6	3.6	OBC/Sched. Tribe (48.8)
5	12.6	0.0	OBC (83.8)
6	16.2	7.1	General (100)
7	9.7	3.6	General (95.8)

Table 4. Communities and Adoption Village 21

Community	% of Village	% of Adoption	Majority Caste (%)
1	24.7	32.8	OBC (97.7)
2	3.4	1.6	OBC (83.3)
3	8.6	9.8	Sched. Caste (46.7)
4	11.5	4.9	Sched. Caste (100)
5	18.9	11.5	Sched. Caste (100)
6	10.3	18	Sched. Caste (100)
7	21.8	19.7	Sched. Caste (97.4)

tends to be concentrated within one caste or religion, though the exact caste or religion varies from village to village. The goal then is to look at these communities to devise ways to potentially improve adoption within them and the network as a whole.

Choosing Leaders. Increasing adoption rates is likely if we ensure that there are enough leaders relative to the community size, that leaders have desirable network traits and that leaders are representative of the household attributes that comprise the community they are in.

To see the effect of number of leaders being proportional to community size, we first calculate the percentage of total households in the village per community. We calculate the percentage of total village leaders per community and take the ratio. A ratio greater than one means there is a higher proportion of village leaders in the community than proportion of households. We then calculate the adoption rate per community and regress this on the leader to household ratio to get a p-value of 0.003. The adjusted R^2 value is low at only 0.03, but we consider using this variable in a leader choosing strategy. The intuition is simple - the more leaders in a community relative to the size, the more opportunity for information to be injected and spread through the community. Of course, this does not mean simply telling everyone and making everyone a "leader" is a practical or sensible option.

One of the main findings from Banerjee et al. is the adoption rate predictive ability of diffusion centrality. This centrality measure was defined by them and acts as a proxy for the computationally heavy communication centrality, which has the same predictive power (1). It is defined as:

$$C_{\text{diff}}(\mathbf{A}; q, T) = \left[\sum_{t=1}^{T} (q\mathbf{A})^t\right] * 1 \qquad [2]$$

where $\mathbf{A}$ is an adjacency matrix and q is the probability at each iteration of an informed node telling each neighbor. T

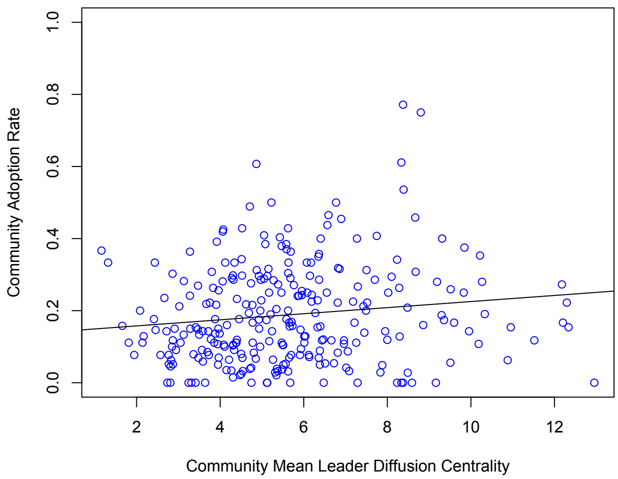

Fig. 3. Community level mean leader diffusion centrality vs community adoption rate. The p-value is 0.026

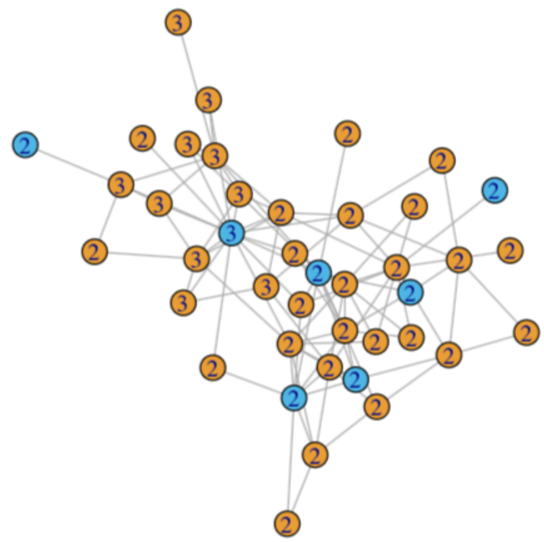

Fig. 4. Village 4 community 1. Blue denotes a leader and 2/3 denotes Hindu/Islamic households. This community had no adoption.

is set from the data as the number of trimesters a village was exposed to the BSS program (average 6.6). Like the paper, we calculate diffusion centrality independent of any model estimation. Hence, we calculate q as $1/\lambda$ where λ is the leading eigenvalue of matrix $\mathbf{A}$. Full details are given in (1). The paper specifically uses the mean diffusion centrality of the leaders, so we calculate this value at the community level and look at the relationship with community level adoption in **Figure 3**. Similar to the network level, we get a significant p-value of 0.026 suggesting mean leader diffusion centrality is important on a community level as well.

The third step in the leader choosing strategy is to appropriately represent household attributes in their respective communities. For example, if a community comprises of 75% Hinduism and 25% Islam, we would want the religious breakdown of the leaders to follow this. To illustrate how this could be an issue, **Figure 4** depicts community 1 in village 4, where adoption was zero. We see a mismatch in the religious breakdown and leader allocation, with only one Islamic leader, even though the village is 27% Islam. We attempt to quantify this across villages in two simple ways: taking community level assortativity and calculating the percentage of the leading caste/religion per community. In the case of village 4, the latter value would be 73%. The intuition is assortativity will have a negative relationship with community adoption rates since it would mean a greater partition and less interaction between the different groups. We hypothesize a positive relationship for the second quantity since a more homogeneous community would likely have more overall interactions. We regress community adoption rates on both these quantities. While the relationships hold, we do not get statistical significance. Using a more sophisticated way to quantify this community level mixing may reveal some significance, but it is also possible the household attributes of the leaders relative to the communities are not important. Ultimately, we decide to test this empirically in a diffusion model.

Network Diffusion Model

The diffusion model we use to test the leader choosing strategy is partially adapted from (1) and (2). First, we implement the first two parts of the leader choosing strategy without considering household attributes. We calculate the community level diffusion centrality for each household and choose the top k with the highest value, with k varying by community. To determine k, the number of leaders we would like for each community, we calculate the percentage of total population for each community, multiply this by the total number of leaders in the village and round to the nearest integer. To implement the third part, we split the leaders for each community proportional to the split on religion/caste, choosing the top leaders in each attribute category. For villages that vary by both religion and caste, we choose to split according to the attribute with the higher assortativity. In all but four villages, this value is caste. The total number of leaders for each set is 1155 and 1130, compared to the 1157 original set. Keeping this number roughly the same is important so we do not observe higher adoption simply due to having more total injection points.

The first part of the model is to fit a logistic regression on the chosen leaders' characteristics. Following from (1) and not considering endorsement effects we have:

$$\log\left(\frac{p_{it}}{1-p_{it}}\right) = X_i^T\beta \qquad [3]$$

where p_{it} is the probability of household i adopting at time t. To estimate β, we fit the logistic model on all the household characteristics and keep those that are significant. We do not want to include too many variables and potentially overfit the model. With the chosen sets of leaders, we end up only keeping the religion and caste characteristics as the others all have very high p-values.

We initialize the diffusion model with the "contagious" and "infected" leaders i.e. households who are informed and can pass information and households who can do the same but have also chosen to adopt. At every time step, the households independently pass information to each of their neighbors with one of four probabilities: q_{ys} if they adopted and share caste/religion, q_{yd} if they adopted but are in a different caste/religion, q_{ns} if they did not adopt and share caste/religion and q_{nd} if they did not adopt and are in a different caste/religion. Each newly informed household i then decides to adopt with probability p_{it} from the logistic model. Note the newly informed households only have that one chance to adopt. This completes one time-step, and the newly informed households become contagious/infected and can pass information. Every household only interacts with each of their neighbors once, either by being informed or informing them. They stop passing information once they have interacted with all their neighbors, or once time runs out.

For simplicity, we use the parameter estimates from (1) and simulate four scenarios. We first set $q_{ys} = q_{yd} = 0.35$ and $q_{ns} = q_{nd} = 0.05$ and run the model for both sets of leaders. This model assumes caste/religion do not affect the probabilities. Then from assortativity and analysis of the community structure, we assume the probability to pass on information is higher/lower given the households share/differ on an attribute. We assume this change in probability to be a constant 20% across villages so that $q_{ys} = 0.42$, $q_{yd} = 0.28$, $q_{ns} = 0.06$ and $q_{nd} = 0.04$, and run the model again on both leader sets. The amount of time-steps we run the simulations for is the same amount we used to calculate diffusion centrality.

Results

We run simulations for villages with high assortativity on either religion or caste (31 villages) and compute the adoption rate for each community at every time step. While it is ideal to run the simulations multiple times and take averages for robustness, it is too computationally intensive with the model. **Figures 5** shows the simulated adoption rates against the actual adoption rates for the four scenarios. There is a clear linear relationship for all scenarios meaning the model overall produces adoption rates that roughly correspond to the observed data. However, there is no clear indication that any set of leaders and assumed probabilities consistently produce higher adoption. Though it may be the case that leader selection and modified passing probabilities did not have much impact on eventual total adoption, we are interested in seeing adoption at a community level. We show the community adoption with modified probabilities for the exemplar village 30 at each time step in **Figures 6** and **7**. The overall adoption in both cases is slightly higher that the actual adoption, though running simulations on this particular village multiple times averages to a negligible difference. What is important is we at least see adoption in all communities, unlike before with adoption only being concentrated in a few communities. Those communities that had a majority of the adoption still have higher than expected adoption, though not as extreme. This is expected as the probability to adopt is dependent only on the household characteristics and not the actual community the household is in. It is also possible that altering the leaders had a negative impact on adoption in these communities, but it is more likely that the model is simply not able to simulate unusually high adoption. Looking at community level adoption for other villages yields the same general results as village 30,

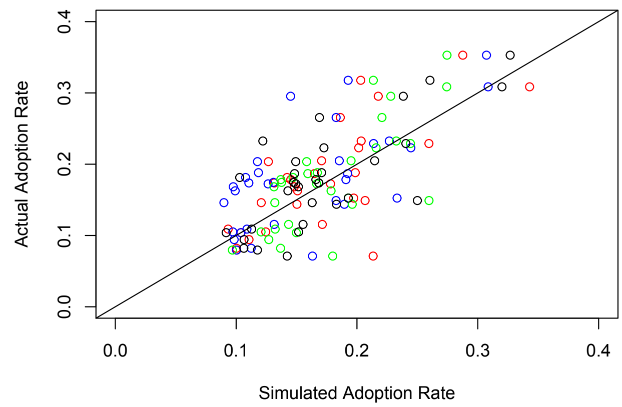

Fig. 5. Simulated vs actual adoption rates. Leaders without attribute consideration and passing probabilities equal (red), with attribute consideration and passing probabilities equal (blue), without attribute consideration and passing probabilities modified (green) and with attribute consideration and passing probabilities modified (black). While all show a clear linear relationship, none indicate better overall adoption

where adoption is more evenly spread among the communities, benefiting those with little to no adoption while hindering those with very high adoption.

Discussion

The leader choosing strategy was devised from a few findings. We first identified high assortativity on caste and religion, suggesting some social stratification and community structure on these attributes. Running Louvain detection gives us communities that are mostly divided on either caste, religion or in some cases both. We then attempt to exploit the pattern of highly assortative, low adoption villages with adoption concentrated in few communities to inform leader selection. Note that in all villages, we find more communities than there are religious or caste affiliations, but a majority of the communities are homogeneous or almost homogeneous. Considering this, it is plausible that some of the communities with the same majority caste or religion can be combined and yield a stronger more connected community. We assumed no prior knowledge of the villages, hence why we used the default resolution. However, suppose we knew roughly how many communities to expect in a village or that a village exhibits some caste/religious dynamic not captured in a simple measure like assortativity. We could then vary the resolution parameter to potentially find a better community partition that more clearly aligns with adoption and non-adoption. Without this knowledge, however, we use what we have to see what influences community level adoption and find leader proportionality and centrality to be significant. Ultimately, the strategy is very intuitive. It is sensible to say having an adequate number of leaders per community who are central and representative of the community is optimal for network diffusion and program participation.

In practice, however, this strategy does not provide much, if any, benefit to overall adoption. Furthermore, running the model for leaders chosen such that they are representative of the community attributes does not show any significant improvement over leaders chosen only on community diffusion

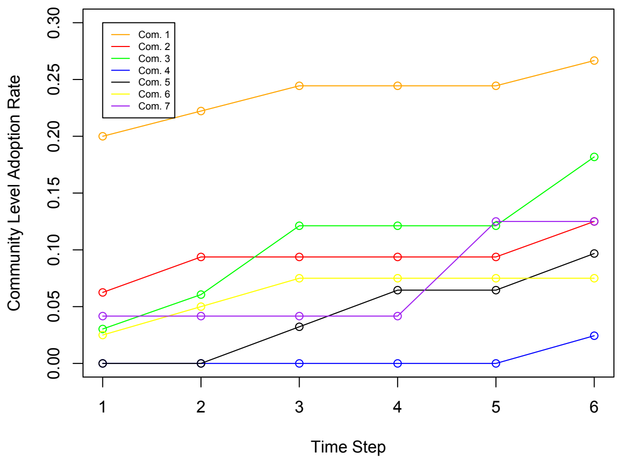

Fig. 6. Community level adoption rates for village 30 at each time step. Leaders chosen with regard to caste.

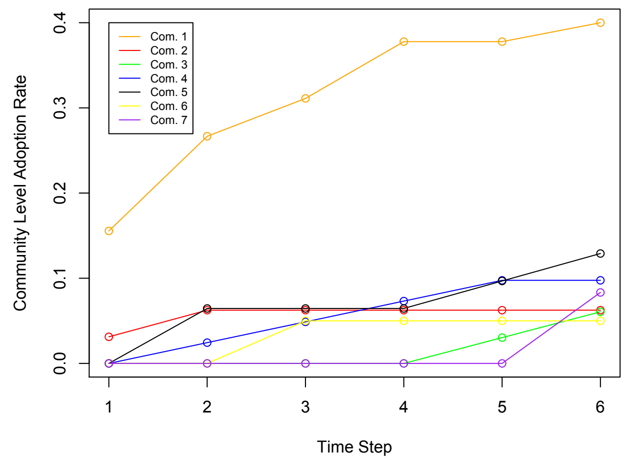

Fig. 7. Community level adoption rates for village 30 at each time step. Leaders chosen only with regard to community diffusion centrality

centrality. This is something we tested empirically without any clear evidence in the analysis of the communities, so it is not unexpected. It is also possible representative leaders actually do help diffusion and adoption, but because a majority of communities are homogeneous, this improvement is tough to reflect in the results as it is very minor even at the community level. This strategy does, however, provide some improvement across communities that previously had very little or no adoption. In this sense, the leader choosing strategy does help on a community level. Essentially we are redistributing the leaders so that adoption and non-adoption is less extreme per community, but overall the same for the village. Something to consider, however, is even though the model shows participation across all communities, there is always a possibility that a particular community with a theoretically optimal selection of leaders still ends up with little or no adoption due to leaders not being the likely household type to adopt. The information passing would end with them and most of the community would be uninformed.

We developed an intuitive and relatively straightforward leader strategy. Going forward, there are many things to consider that this analysis does not which could yield more meaningful results. First is the aforementioned domain knowledge. Gathering more information about social dynamics within villages could lead to better and more informative community detection. It could also help explain and incorporate into the model households which are isolated yet participants. We also only look at one attribute assortativity at a time. When we observed this pattern of adoption being concentrated in only a few communities and the general homogeneity of those communities to make the connection of adoption alignment with a certain caste or religion, we are only considering one attribute. The class of this particular attribute that formed a community alignment with adoption/non-adoption, however, only held on a village level and we did not consider why. Devising a way to consider multiple attributes, such as households of a certain religion, caste and ownership status could reveal more intricate and consistent patterns of adoption and non-adoption across villages. This could also help better explain villages where some communities have very high or little to no adoption. These improvements would naturally help develop a better leader choosing strategy at the community level.

The diffusion model also has its limitations. Simulations using the modified passing probabilities do not yield significantly different results from using the probabilities in (1). The reasoning that households which share an attribute are likely to pass information at a higher rate is sensible, but the model makes a big assumption in using the parameters from (1) and adjusting them by 20%. One improvement could be to adjust the probabilities based on the assortativity. A better model could estimate the probabilities by running simulations over a parameter space and minimizing some objective function. Doing so would allow for a more reasonable comparison of models. Lastly, it should be noted that we take the same approach to every village. In reality, though perhaps not feasibly, the villages are independent and should be approached individually.

This paper provides an initial analysis of the community structure and stratification in Indian villages and how it affects microfinance participation. Though the results show only a slight improvement when targeting communities in this context, the ideas presented can be modified and adapted to further social network analysis.

ACKNOWLEDGMENTS. We thank the C5.10 Mathematics and Data Science for Development team for their guidance throughout the course.

1. Banerjee A. ea (2013) The diffusion of microfinance. *Science* 341:363–370.
2. Aral, S. ea (2009) Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences of the United States of America* 106(51):21544–21549.
3. Newman MEJ (2002) Assortative mixing in networks. *Phys. Rev. Lett.* 89(20):208701–20875.
4. Blondel, V. ea (2008) Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment* 2008.
5. Carissimo, A. ea (2016) Validation of community robustness. *Computing Research Repository* abs/1610.05045.

(1) (2) (3) (4) (5)