Microfinance network analysis and optimal seeding prediction - Group 53

Alan Newman (awnewman), Sam McLaughlin (thesam), V. Govind Manian (vgmanian)

Acknowledgements: We thank the TAs for their advice and Arun Chandrasekhar, author of [1] for many helpful comments and suggestions.

I. Introduction

1. Motivation and Objectives

The spread of microfinance throughout the developing world has the potential to affect tremendous positive change by spurring entrepreneurship, economic growth, and development in areas and demographics cut off from traditional forms of financing. One of the largest barriers to the greater use of microfinance is simply lack of information about what microfinance is, how it works, and how an individual might take advantage of it.

Our top-level objective is to predict which individuals within a village would be most likely to participate in microfinance. We plan to proceed as follows: first, we will characterize the structure of the village networks in the dataset gathered by Banerjee et al., extracting various features with a focus on the status of an individual within the network, their similarity with other individuals in the network, and whether or not links within the network are "supported" (see 2.B below) as particularly interesting and potentially useful features. We then use these features in a simulated diffusion model. Finally, we use the outcomes of the diffusion model construct and train a logistic regression classifier to both improve our model's performance and gain understanding into the relative importance of the features used.

2. Prior Work

A. The Diffusion of Microfinance

Banerjee et al [1] explore the diffusion of participation in a microfinance program in South Indian villages, concentrating on three major points. First, the nature (in the context of the node's properties in the network) of the "injection point"—that is, who first learns about the program—provides a robust measure of the overall participation rate of the village. Second, whether a particular person is informed about the program is sufficient to extend the diffusion to his/her contacts—whether he/she participates in the program does not influence whether his/her contacts will also participate (beyond the contacts merely learning about the program). And third, although standard diffusion models utilize a "contagion" based paradigm, adopting a sort of two tiered approach instead (keeping track both of (a) whether a person is informed and (b) whether he/she participates) is substantially preferable.

While this paper provides an analysis that is both novel and insightful, we believe that by utilizing more informative graph structure features such as status, similarity, and support (see below), we can better predict the optimal injection points for microfinance information.

B. Social Capital and Social Quilts: Network Patterns of Favor Exchange

Jackson et al. [3] develop a game theoretic model of a robust network of favor exchanges, and then apply that model to the microfinance dataset gathered by Banerjee et al. The resulting model produces tree-like structures of cliques (generally, 3-cliques), which the authors term "social quilts". The authors

then apply the model to the microfinance dataset and find high levels of "support" (a supported link being one between two agents who each have a favor exchange link with a third agent in common) in the favor exchange networks in the South Indian villages. The level of support exceeds clustering within the network by an order of magnitude, and correlates negatively with participation in microfinance.

It seems natural to wonder whether information about microfinance might be more likely to propagate along a supported link as opposed to an unsupported one. The motivation behind this is straightforward: whether or not a link is supported may be indicative of the quality of that social link, and that quality may in turn influence the likelihood of information propagating across that link. The natural follow-up question is whether there are variations in the effects of support from one favor exchange network to another.

C. Status and Similarity Metrics

a. Positions in Networks

Burt develops a series of metrics for measuring the similarity of two different "actors" (ie, nodes) within a social network. One of these measures is what he calls the structural equivalence of or Euclidean distance between two actors: if the actor's individual distances to every other actor in the network are the same, then the two actors occupy structurally equivalent positions in the network. If the notion of distance that we use between two nodes is simply the presence or absence of an edge in the social network, then two nodes are structurally equivalent if they have exactly the same set of neighbor nodes, not including one another. This version of structural equivalence can be quantified as follows:

$$x_{ij} = \sqrt{\sum_{k \neq i,j} (A_{ik} - A_{ij})^2},$$

where A_{ij} represents the presence or absence of an edge between nodes i and j (alternately, the value of cell ij in an adjacency matrix). Structural equivalence is method commonly used by sociologists [5] to quantify the similarity between two individuals within a network, and may be a useful measure of similarity in our favor exchange networks.

b. Detecting Community Structure in Networks

Newman's work on community structure detection is largely tangential to our own interests, but his discussion of several measures of similarity between individuals is a useful starting point for our own work. The first measure he considers is structural equivalence as discussed above; he also makes use of the Pearson correlation between the columns of an adjacency matrix as well as the Jaccard similarity between two nodes in an adjacency matrix. We use all of these measures as network features.

c. Extracting Reputation in Multi Agent Systems by Means of Social Network Topology

Pujol et al. [6] develop an algorithm for determining the reputation of an individual within a social network that they dub *NodeRanking* which, as the name might imply, is a modified version of the well-known *Pagerank* algorithm. There are some differences from *Pagerank* with regard to how the principal eigenvector is calculated (*NodeRanking* uses only local information for each node, ie information about just its immediate neighbors), and how the probability of randomly teleporting to another node in the graph is determined, but the underlying random walker strategy is the same. Pujol et al. then compared the reputations

generated by their algorithm to actual reputations for a real citation network and found that the reputation values that their network produced correlated well with the actual reputation values for the network. Given the success that Pujol et al. found in using PageRank as a status measure in a social network, we likewise make use of it as a network feature, but discard the authors' modifications on the grounds that we are already considering relatively small villages and that the favor-exchange networks do not necessarily encoded information about which villagers know and interact with one another.

3. Dataset

The dataset was collected by Banerjee et al. from 75 villages in rural, southern India. Included in the dataset is an array of demographic information concerning the members of the households surveyed (age, sex, religious affiliation, occupation, and so on), but also a wide array of information charting the social structure of the village. Survey respondents were asked whom they would go to for various favors such as borrowing money or kerosene, or or advice. Similarly, they were asked who would come to them for the same favors, and additionally the people with whom they attended religious and social events. It is important to note that only about half of all households within the villages were actually surveyed, and so the data cannot be construed as presenting a complete social network graph. All demographic and social information was collected before a microfinance program was initiated in 43 of the 75 villages; once the program began, Banerjee et al. tracked which individuals were first informed about the program and the spread of that information through the village, including information about whether individuals actually participated in the program or simply heard about it.

The favor-exchange networks consist of the individuals listed by survey respondents when asked who they:

- borrow money from
- give advice to
- help with a decision
- borrow kerosene or rice from
- lend kerosene or rice to
- lend money to
- obtain medical advice from
- engage socially with
- are related to
- go to temple with
- invite to one's home
- visit in another's home.

Additionally, networks were constructed from the intersection and union of the above networks. We did not use the former as the resulting network was extremely sparse and also improperly keyed.

II. Methods & Results

Figure 1 presents a high-level view of the procedure we used to predict microfinance participation; the following subsections cover each step in detail. Steps 1-4 cover the feature generation, weighting, diffusion modelling, and logistic regression used to actually predict microfinance participation, while steps 5 and 6 are intended to characterize the structure of the network and verify the results of the logistic regression classifier.

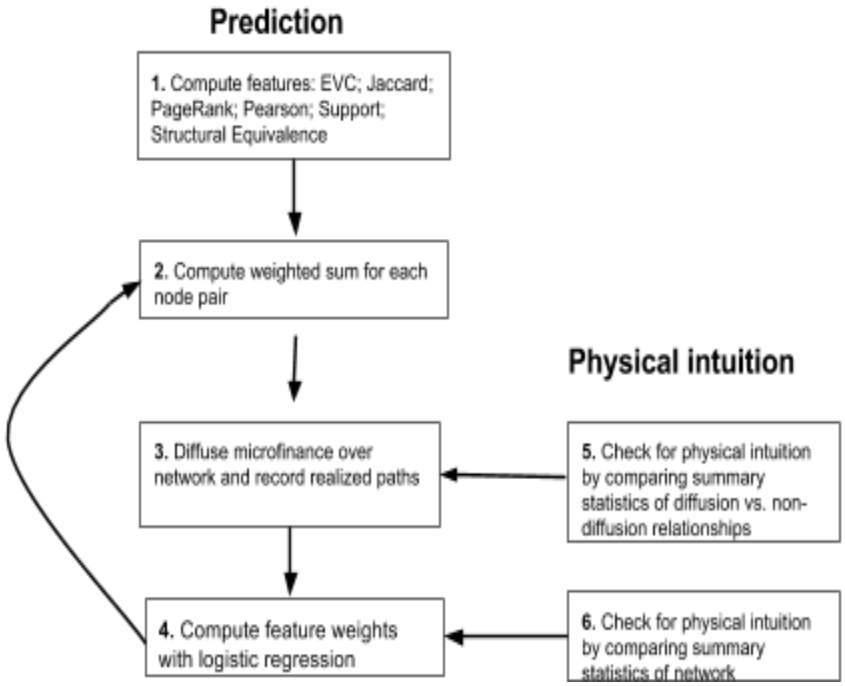


Figure 1: Procedure for microfinance participation prediction.

1. Compute features

We computed a number of metrics for determining the status of a node within a network and the similarity between two nodes in a network. Additionally, we computed which network edges were supported per [3]. Status scores were computed for each favor-exchange network for a given village (ie, villages were considered entirely separate from one another). Similarity measures were likewise computed for each favor-exchange network for a given village and were calculated for every combination of nodes within a given network regardless of whether an edge existed between those two nodes in that particular network.

A. Status

We initially used the SNAP library functions to compute Eigenvector Centrality and PageRank scores for each node in each favor-exchange network and used these as status the status of a given node. The two features received very similar weighting, which should be unsurprising given the similarity of the two measures, so we ultimately dropped EVC as a feature and used PageRank alone as a status measure.

B. Similarity

To measure the similarity of two nodes within a given favor-exchange network, we computed the Structural Equivalence of the two nodes as detailed in [4] and section I.2.C.a above. We additionally computed the Pearson correlation (following [5]) with the means and variances of the columns from the adjacency matrix for our two nodes as:

$$\mu_i = \frac{1}{n}\sum_j A_{ij}, \quad \sigma_i^2 = \frac{1}{n}\sum_j (A_{ij} - \mu_i)^2,$$

And the resulting correlation coefficient as:

$$x_{ij} = \frac{\frac{1}{n}\sum_k (A_{ik} - \mu_i)(A_{jk} - \mu_j)}{\sigma_i \sigma_j}.$$

where A_{ij} represents the presence or absence of an edge between nodes *i* and *j*. We also computed the Jaccard similarity for each pair of nodes; Jaccard similarity is the size of the intersection of the two node's sets of neighbors divided by the size of the union of their sets of neighbors:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}.$$

Finally, we also computed a naive measure of similarity between two nodes using the household demographic data that is part of the dataset. The household demographic data contains information about the religion of the head of the household, whether the household has access to electricity and running water, the type of material that the roof of the house is made of, and so on. Our naive similarity measure is just the number of matching features divided by the total number of features for two household entries in the demographic data.

C. Support

For each favor exchange network for a given village, we computed the set of edges that were "supported" as defined by [3]--that is, the edges that are part of a triangle within the favor-exchange network. This was a simple matter of finding all the pairs of nodes in a network that each have an edge to some third node.

2. and 3. Compute weighted sum of features and diffusion model

As an overview, we use a diffusion model to determine the optimal thresholds to determine whether a household will participate in microfinance, diffuse information about microfinance, or both. Following [1], because the villagers are under-surveyed and therefore under-sampled, we take our agents to be households, which are fully sampled. The network of households is constructed by taking the union of the networks of individuals. We assume that if one member of a household decides to tell others about microfinance, all members of that household will tell others about microfinance. The threshold to participate is greater than the threshold to diffuse information. The input features are the set of status and similarity features, as previously discussed, of each household. Less than 1% of households surveyed had a keying error, i.e. a reference to a node that did not exist. These were removed.

The initial conditions for the diffusion are the 'injection points', which are local leaders approached by BSS about microfinance. Not all local leaders choose to do microfinance. We assume all of them tell others about microfinance.

The pseudocode for the diffusion model is as follows:

```
Initialize TELLMF, the set of local leaders approached to do microfinance
Initialize DOMF, the set of local leaders agreeing to do microfinance
For each threshold combination of (do, tell), such that do > tell:
    For each node n in TELLMF:
        For each neighbor e of n:
            If the status + similarity between n and e > do:
                Add e to list of DOMF
                Add e to list of TELLMF
```

7. If the status + similarity between *n* and *e* > tell:
Add *e* to list of TELLMF
8. If the length of DOMF has not changed in 5 iterations, go to 3. Else, go to 4.

We then determine the performance of each (do, tell) threshold combination by finding the average number of misclassifications, i.e. the sum of |y - y_hat|, for each threshold over all villages.

4. Logistic regression

For the first diffusion run, all features are re-scaled to [0, 1], weighted equally, and summed. They are rescaled because e.g. Jaccard similarity can be 100x larger than e.g. PageRank, and when summed, PageRank would have no impact on the diffusion. They are weighted equally because we do not make any a priori assumptions about the relative importance of the features.

In order to determine which favor exchange networks are relatively important for the diffusion and uptake of microfinance, and which metrics best capture this importance, we train an L1-norm regularized logistic regression classifier on the data. The classifier determines the weights of each feature, relaxing our assumption that all features are equally important. Furthermore, logistic regression is invariant to the scale of the input space, and we can handle data types at multiple scales, e.g. Jaccard similarity and PageRank.

An immediate difficulty in using logistic regression is that many target nodes have more than one neighbor that could potentially persuade them to do microfinance. Coding for this conditionality would be very difficult. However, because our diffusion model is deterministic, we can observe the actual node that persuaded a target node to do microfinance, which we achieve by using the realized paths from the diffusion model with the lowest number of misclassification.

Formally, a logistic regression transforms the Ordinary Least Squares (OLS) regression model **y = XB + e**, where **y** is an stacked vector of outputs, in this case an [n x 1] vector of 0 or 1; **X** is an [n x p] matrix of input features, in this case measures of status and similarity; and **B** is the [p x 1] vector of weights. OLS is transformed by the sigmoid function, **y** = 1 / (1 + exp{-**z**}), where **z** = **XB**. Due to the dimensionality of the input space and that we expect significant overlap between some favor exchange networks, we penalize this regression model in order to shrink the weights of some input features to zero. In particular, we use an L1-norm regularization (LASSO), so that determining the optimal weights **B** requires solving the optimization problem with penalty parameter λ [2]:

$$\min_{\beta} \sum_{i=1}^{N} \left(y_i - \beta_0 - \sum_{j=1}^{p} x_{ij}\beta_j \right)^2 + \lambda \sum_{j=1}^{p} |\beta_j|,$$

In order to choose the best model, i.e. the best choice of the shrinkage parameter, we will use k-fold cross validation, for k = 5. K-fold cross validation works by partitioning the data into k folds, and then training or building the model on k - 1 folds and testing the performance of this model on the kth fold. In order to choose the optimal model, we use a modified loss function. Physically, BSS is constrained by the amount of cash on hand necessary to finance a villagers. This means that they can tolerate a certain number of false positives, but can not tolerate as many false negatives. This choice is influenced by BSS's financials, which we do not have access to. Therefore, we choose the model that minimizes the difference in the number of predicted false positives and false negatives, effectively seeking to find the model that equates the two.

Finally, we feed the computed feature weights back into our diffusion model and re-run the diffusion. The paths realized by the diffusion model are contingent on the feature weights used to compute the importance of the status of each node and the relative weights of similarity between each node pair. Similarly, the feature weights computed by the logistic regression classifier are contingent on the paths realized by the diffusion model. Therefore, as indicated in the flowchart, we iterate the diffusion model and classifier until the performance of the classifier converges.

Logistic regression computes the following non-zero feature weights with a classification rate of 83% and intercept of -1.82.

Similarity metrics	**Favor-exchange network**	**Feature weight**
Support	Whom do you borrow money from?	-0.044
Support	Who borrows kerosene or rice from you?	-0.017
Support	Whom do you lend money to?	-0.038
Pearson	Whom do you lend money to?	0.24
Pearson	Whom do you borrow kerosene or rice from?	0.42
Pearson	Whom do you engage with socially?	0.91
Status metric	**Favor-exchange network**	**Feature weight**
PageRank	Whom do you give advice to?	-0.40
PageRank	Whom do you go to visit?	0.21

Figure 2: Network features of non-zero weight.

Starting from 65 features, computed from 5 metrics and 13 favor-exchange networks, we find substantial shrinkage down to 8 features.

It may be surprising at first to find that some of these weights are negative. Observe, however, that the feature weights for Support, while negative, are very small, and unlikely to have an impact. This is further supported by investigating the histogram of Support, which shows that the frequency of non-zero support values is approximately 1%. Therefore, in practice, the features worth considering are those computed for Pearson and PageRank.

The final negative weight is -0.4, computed for PageRank over the 'Who do you give advice to?' network. This information, however, is self-reported, and therefore may be confounded by a respondent thinking highly of their advice-giving abilities and the value of that advice, perhaps to the extent that they are perceived as arrogant, or givers of too much or not very useful advice, which would negatively impact their ability to persuade others to participate in microfinance. In short, this merits further investigation.

The remaining weights are very plausible. Relationships surrounding money, borrowing food, and engage with socially are very important relationships, particularly in rural India, and signify a high trust between individuals. We therefore expect that similarity, in this case computed by Pearson, over this network has a substantial positive effect on diffusion. Furthermore, the status feature computed by PageRank over the 'who do you go to visit?' network indicates households that are of high status and visited often. The importance of PageRank is also consistent with [1], which found support for eigenvalue centrality as an import feature.

The final two sections further investigate the physical interpretation behind these metrics.

5. Diffusion vs. Non-diffusion relationships, Summary Statistics

To more carefully evaluate the feature weights generated by the logistic regression classifier, we compared the average PageRank scores, Pearson correlation values, and probability of an edge being supported for the respective favor-exchange networks picked out by the classifier. For the PageRank scores, we separated all nodes in all the villages in the relevant networks into those with a probability of participating in microfinance that was greater than 0.7 ("diffusing"), and those with a probability lower than that ("non-diffusing"). We then compared the average PageRank scores for those two sets of nodes. Similarly, we partitioned all edges in all the village networks into those with a probability of at least 0.7 of diffusing microfinance participation, and those with a probability lower than 0.7. We did the same for support, but note that the "average support" for a set of edges is effectively the probability of any one of those edges being supported.

A full table of these values appears in Figures 5-7 in Appendix A. The overall picture is clear: the average PageRank scores, Pearson correlation values, and probability of an edge being supported where all substantially greater for the nodes and edges that participation in microfinance diffused to. This indicates that there is some property of these relationships that is amenable to the spread of microfinance information and participation.

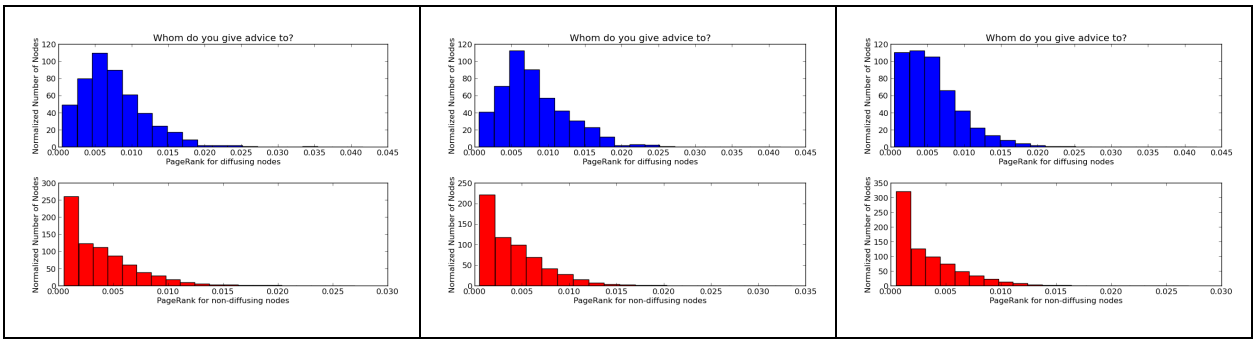


Figure 3: PageRank histograms for Do >.7, Tell>.7 and Local Leaders. Diffusing nodes in blue, non-diffusing in red. See Appendix A for additional histograms.

Additionally, we repeated the above procedure, but instead of requiring that nodes and edges have a probability of at least 0.7 of participating in microfinance, we required that they have at least a 0.7 chance of telling another person about microfinance--in effect, we examined the diffusion of information about microfinance rather than participation in it. Again, we found substantially greater PageRank scores, Pearson correlation values, and probability of being supported for the relevant nodes and edges. Interestingly, these average values were actually higher than for participating in microfinance. Most

importantly, this again reinforces that there is some external quality to the relationships in the network that exists independently of anything the logistic regression classifier is doing and therefore that the weighting it gave to the features in Figure 2 are not spurious. Why the values were actually higher for the diffusion of information about microfinance versus participation in it is a more difficult question to answer, but we would suggest that higher threshold required by the diffusion model to participate in microfinance coupled with the myriad other features involved in the logistic regression may have excluded enough nodes and edges from meeting our 0.7 probability threshold and driven down the averages. In any case, the differences are slight and we consider the overall comparison between both the participation and telling nodes and edges and the non-participating, non-telling nodes and edges to be the more significant result.

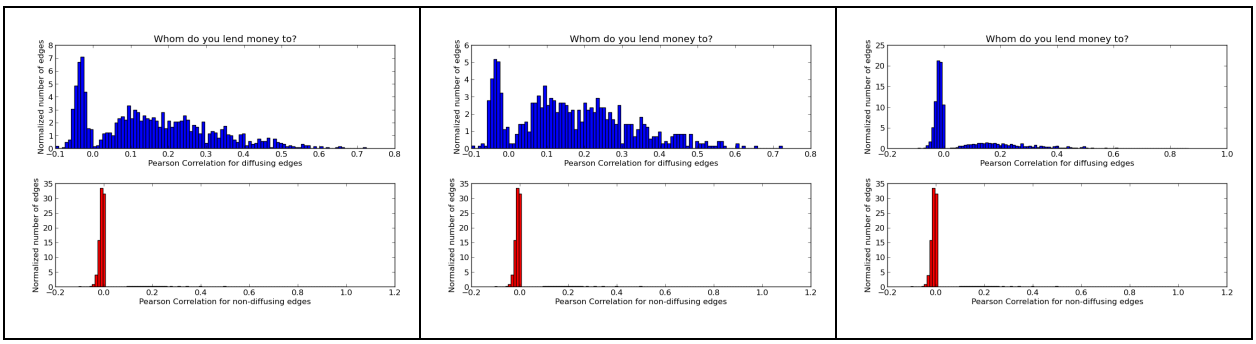


Figure 4: Pearson correlation histograms for Do>.7, Tell>.7, and Local Leaders.
Diffusing edges in blue, non-diffusing in red.

Finally, we ran the same analysis as before, but instead of partitioning nodes based on participation in microfinance, we compared nodes marked as local leaders in the demographic dataset with all other nodes. Again, we found that the PageRank scores for the local leader nodes and the Pearson correlation values and probability of being supported for edges involving a local leader node where higher than non-local leader nodes and edges, however, these values were not nearly as high the averages for the nodes and edges that participated in microfinance in our diffusion model. This serves yet again to reinforce our confidence that our features are in fact useful, and further suggests that our logistic regression classifier picked out a series of features more relevant to the diffusion of microfinance information and participation than the local leaders identified by Banerjee et al.

6. Diffusion vs. Non-diffusion relationships, Structural Features

In order to investigate the differences between the favor networks, we ran a number of metrics including: diameter, expected degree, clustering coefficient, and expected excess degree (See Figure 10 in Appendix B for the full data table and Figure 11 for a graphical representation of the networks' averaged degree distributions). These metrics were applied to each network for each village and then averaged across villages--to end up with one value for each metric for each network type. From there, to establish a mechanism by which we could cluster the networks according to similar graph properties, we ran K-means using the relevant properties as features.

However, none of these metrics are sufficient to account for the distinction between networks that were chosen as useful by our classifier and those that were not. Specifically, the mean diameter, expected degree, excess expected degree, and clustering coefficient for the non-selected group of networks were: 13.77, 5.54, 13.54, and 0.79; those of the selected group of networks were: 14.08, 5.7, 13.862, and 0.75.

Particularly as the sample size was small, the intra-group variance was high enough to explain the inter-group differences, and we can consequently conclude that at least in the context of the presently computed metrics, we were not able to adequately predict which networks were useful for our diffusion model purely through structural features.

In addition, to further examine the relationships between the networks (to establish that there are robust differences), we counted the number of edges shared across different favor-exchange networks within each village (between 1 and 14), and plotted the frequencies of each of these shared edges (Figure 12, Appendix B). That our chart is skewed right indicates that there is not much overlap between the networks. However, this distribution doesn't distinguish between different networks--that is, it considers them all as a set. In order to further investigate the similarity between any two networks, we ran a version of full-graph Jaccard similarity on every combination of two networks for each village. This pairwise Jaccard similarity distribution can be found in Figure 13 in Appendix B.

The two highest scores were from the combination of the Visit Come and Visit Go networks (0.73) and that of the Kerosene Come and Kerosene Go networks (0.72), and the remaining comparisons' scores are below 0.6. Two points become clear from this analysis: first, no two networks are identical (as this would produce a Jaccard similarity score of 1.0). More importantly, the vast majority of scores are below 0.5, and the mean is 0.267, which indicates that most of the networks are substantially dissimilar.

Lastly, we constructed visualizations of our networks using Fruchterman Reingold layout in order to clearly show the stark contrast observed in the similarity scores. Figure 14, Appendix B, shows three graphs: 'Whom do you go to temple with?' (Temple Company), 'Whom do you visit?' (Visit Go), and 'Whom do you give advice to?' (Give Advice). There are stark differences in the resulting visualizations. Temple Company is highly clustered with many disconnected components, while Visit Go is far more interconnected, and Give Advice is somewhere in between. These robust differences are confirmed both by the metrics in Appendix A and intuition--that is, we contend that the networks essentially measure something amounting to "strength of relationship". For example, it makes sense that religious groups will be highly concentrated for the most part, with few, if any, connections to outsiders--that is, only the strongest relationships appear. The Visit Go network, on the other hand, is far more open and interconnected, as people tend to visit others with whom they have less in common. It also follows that the Give Advice network models a balance of the two--insofar as individuals tend to be more picky about those to whom they give advice versus those that they merely visit. Nevertheless, the strength of a relationship required to give advice, while stronger than that of mere visiting, is substantially weaker than that of the temple.

III. Conclusion

Using the social, demographic, and microfinance data gathered by Banerjee et al. [1], we have developed a set of network features that combine some traditional measures of status (eg, PageRank) and similarity (eg, Jaccard similarity) with the notion of support developed by Jackson et al. [3] and then used a diffusion model coupled with a logistic regression linear classifier to predict microfinance participation. As noted in section II.4, we achieved a classification rate of 83%.

In addition to predicting with a respectable degree of accuracy the individuals most likely to participate in microfinance, we determined that the Pearson correlation coefficient and PageRank score of edges and nodes, respectively, had the most impact on accuration microfinance participation prediction given the high

weighting given to them by the logistic regression classifier.

As a check against the logistic regression classifier's feature weighting, Wwe examined various network-level metrics such as clustering coefficient and average diameter, but found no significant differences between the networks given non-zero weight by the logistic regression classifier and those given zero weight.

Finally, we partitioned our network edges and nodes between those with a high likelihood of participating in or sharing information about microfinance and found that the Pearson correlation coefficient, PageRank score, and probability of being supported were much higher for the group with a high likelihood of participation than for the others. This suggests that there are in fact substantial differences in the quality or nature of relationships that affect the diffusion of information about and participation in microfinance.

References

[1] Banerjee, Abhijit, et al. *The diffusion of microfinance*. No. w17743. National Bureau of Economic Research, 2012.
[2] Trevor. Hastie, Robert. Tibshirani, and J. Jerome H. Friedman. *The elements of statistical learning*. Vol. 1. New York: Springer, 2001.
[3] Jackson, Matthew O., Tomas Rodriguez-Barraquer, and Xu Tan. "Social capital and social quilts: Network patterns of favor exchange." *The American Economic Review* 102.5 (2012): 1857-1897.
[4] Ronald S. Burt. *Positions in networks*. Social Forces, Vol. 55, No. 1 (Sep., 1976), pp. 93-122
[5] Newman, Mark EJ. "Detecting community structure in networks." The European Physical Journal B 38, 321–330 (2004)
[6] Josep M. Pujol, Ramon Sangüesa, and Jordi Delgado. 2002. Extracting reputation in multi agent systems by means of social network topology. In *Proceedings of the first international joint conference on Autonomous agents and multiagent systems: part 1* (AAMAS '02).

Appendix A

Network Type	Diffusing, Do >= 0.7	Non-diffusing, Do >= 0.7	Diffusing, Tell >= 0.7	Non-diffusing, Tell >= 0.7	Local Leader	Not Local Leader
From whom do you borrow money?	0.5732700	0.0058958	0.6628308	0.0060130	0.1768903	0.0057143
Who borrows kerosene or rice from you?	0.5312075	0.0065027	0.5949367	0.0066185	0.1531494	0.0063720
Whom do you lend money to?	0.4871099	0.0050154	0.5627157	0.0051152	0.1328393	0.0049176

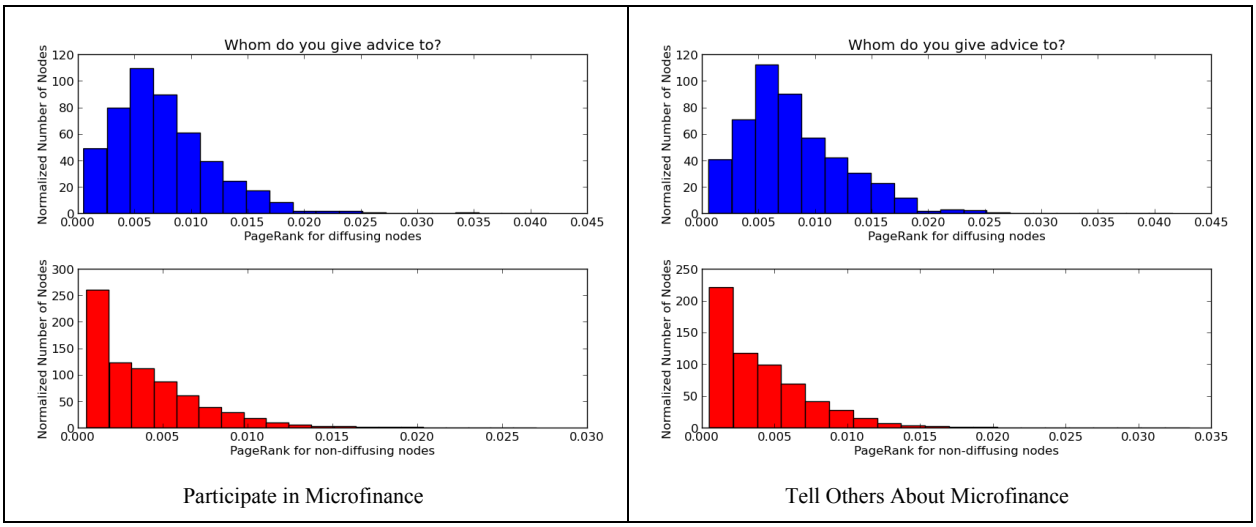
Figure 5: Probability of an edge being supported for various networks. "Diffusing, Do >= 0.7" indicates the probability of support for edges with a probability greater than 0.7 of diffusing microfinance participation. "Non-diffusing" indicates a probability of diffusing participation less than 0.7. "Tell" corresponds to diffusing information about microfinance.

Network Type	Diffusing, Do >= 0.7	Non-diffusing, Do >= 0.7	Diffusing, Tell >= 0.7	Non-diffusing, Tell >= 0.7	Local Leader	Not Local Leader
Whom do you give advice to?	0.0075894	0.0038927	0.0081143	0.0040635	0.0057291	0.0033584
Whom do you visit?	0.0074835	0.0039127	0.0079288	0.0040848	0.0057448	0.0033443

Figure 6: The average PageRank score for diffusing and non-diffusing nodes in various networks.

Network Type	Diffusing, Do >= 0.7	Non-diffusing, Do >= 0.7	Diffusing, Tell >= 0.7	Non-diffusing, Tell >= 0.7	Local Leader	Not Local Leader
Whom do you lend money to?	0.1549396	0.0048592	0.1676831	0.0048944	0.0579300	0.0047859
Whom do you borrow kerosene or rice from?	0.1830286	.00408020	0.2080044	.00411843	0.0710393	0.0039809
Whom do you engage with socially?	0.1571081	0.0082188	0.1720724	0.0082528	0.0696339	0.0081178

Figure 7: The average Pearson correlation coefficient for diffusing and non-diffusing edges in various networks.



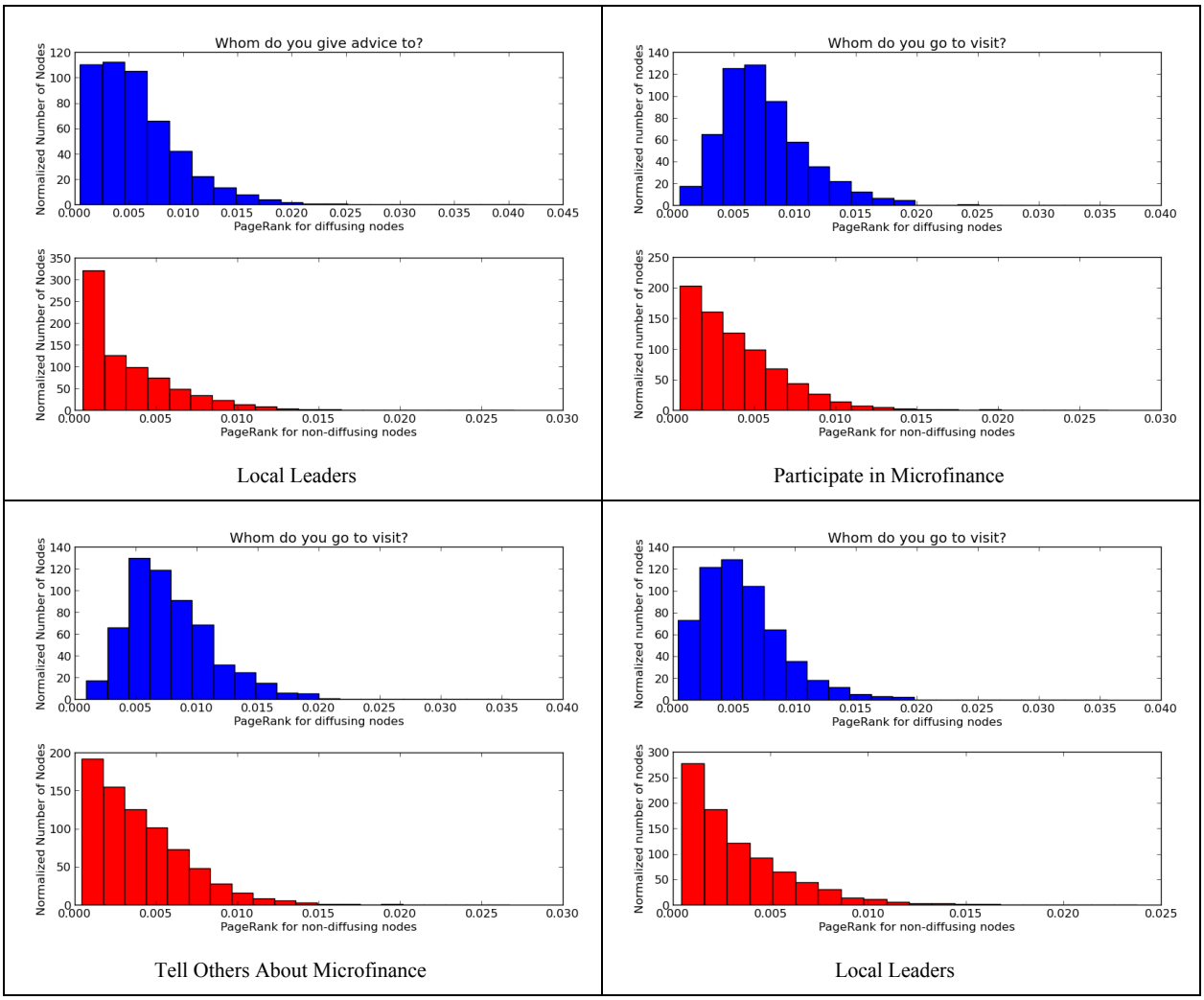
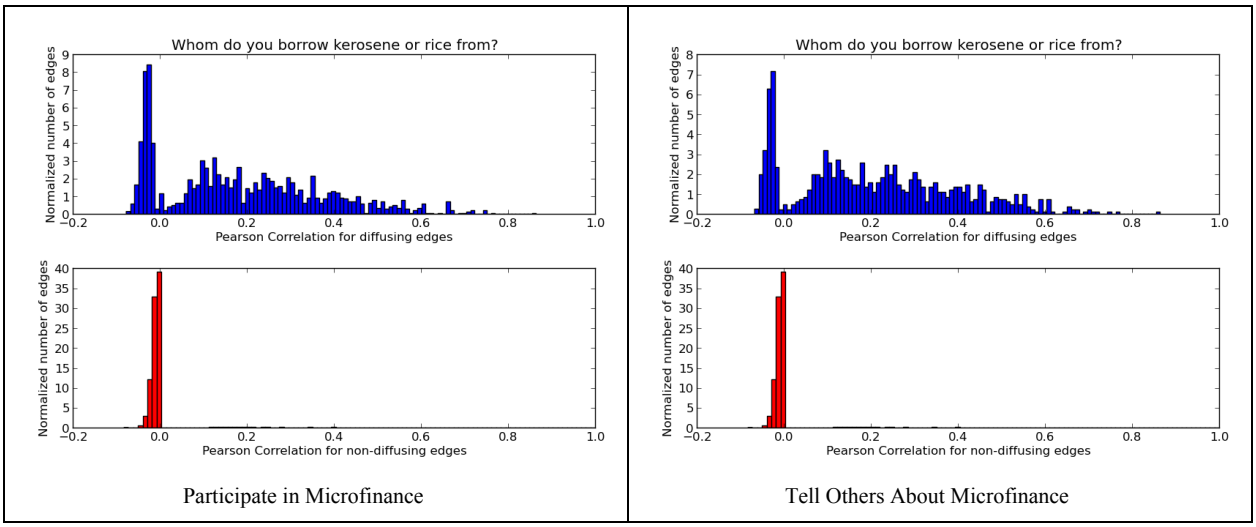
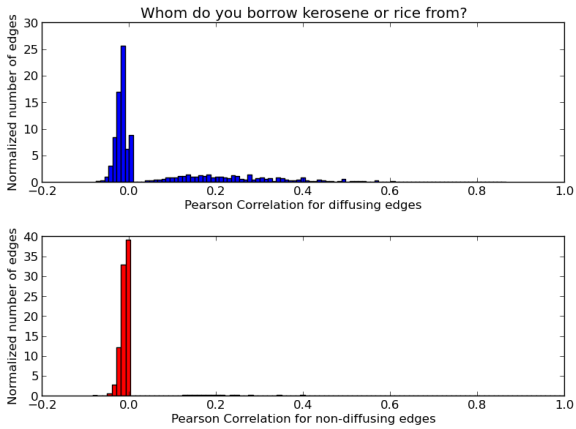
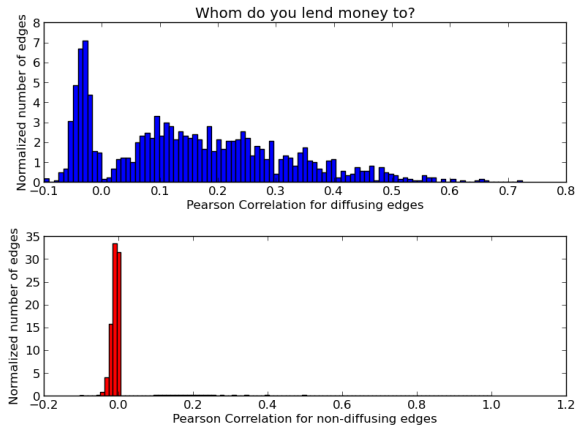


Figure 8: Normalized histograms of PageRank scores. "Participate in Microfinance" indicates partitioning between nodes with at least a 0.7 probability of microfinance participation. "Tell Others About Microfinance" indicates a similar partitioning, but for diffusion of information about microfinance. "Local Leaders" indicates partitioning between nodes marked as local leaders and all others.

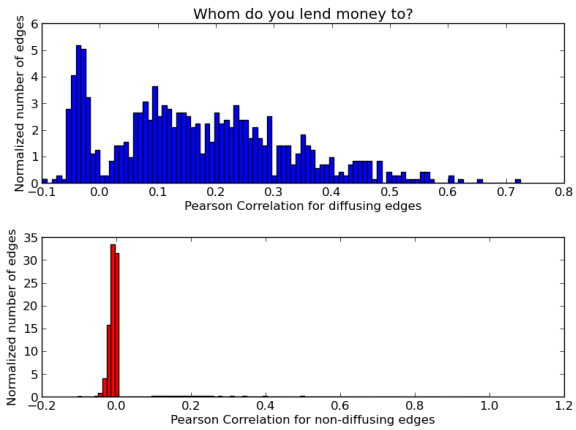




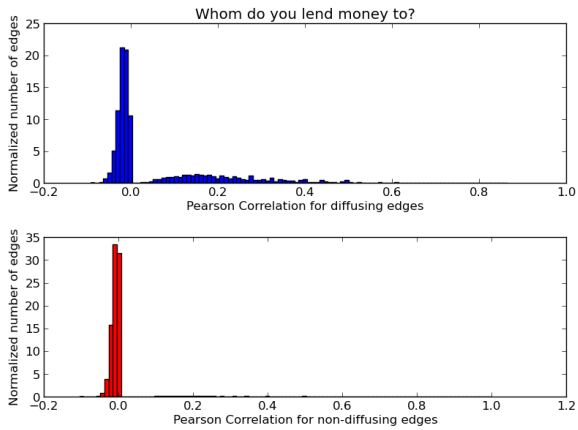
Local Leaders



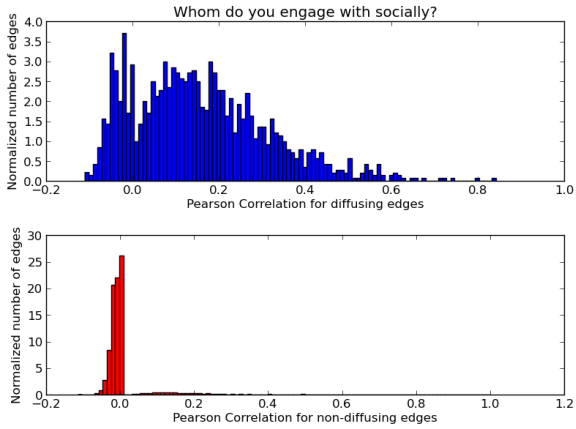
Participate in Microfinance



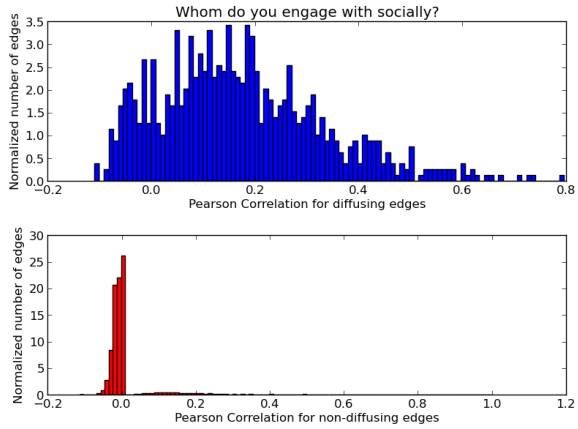
Tell Others About Microfinance



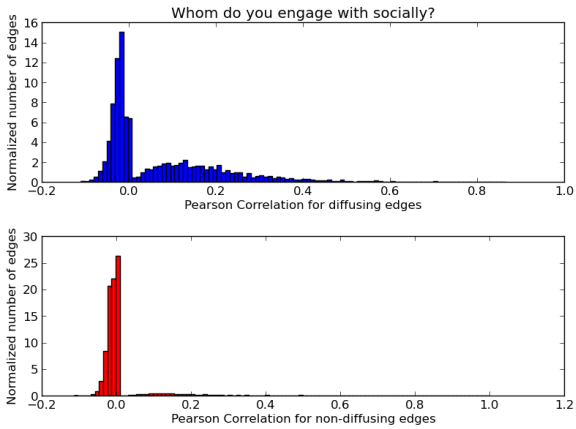
Local Leaders



Participate in Microfinance



Tell Others About Microfinance



Local Leaders

Figure 9: Normalized histograms of Pearson correlation coefficients. “Participate in Microfinance” indicates partitioning between edges with at least a 0.7 probability of diffusing microfinance participation. “Tell Others About Microfinance” indicates a similar partitioning, but for diffusion of information about microfinance. “Local Leaders” indicates partitioning between edges involving a local leader node and all other edges.

Appendix B

	Cluster # (kmeans)	Diameter (Avg)	Expected Degree (Avg)	Expected Excess Degree (Avg)	Clustering Coefficient (Avg)
Lend Money	0	14.37	5.57	13.44	0.75
Visit Come	5	11.65	6.02	14.68	0.71
Give Advice	4	18.12	5.24	12.68	0.81
Rel	0	18.79	5.60	13.73	0.81
All Vill Rels	1	7.87	8.43	24.19	0.65
Non Rel	3	11.63	5.82	14.25	0.73
Borrow Money	6	12.84	5.67	13.87	0.75
Medic	0	14.84	5.60	13.57	0.76
Help Decision	7	15.49	5.33	13.11	0.80
Kerorice Go	3	14.71	5.81	14.14	0.75
And Relationships	8	1.03	4.67	11.55	0.95
Visit Go	5	11.43	6.08	14.88	0.71
Kerorice Come	3	14.85	5.79	14.06	0.75
Temple Company	1	8.05	4.76	11.71	0.92

Figure 10: Assorted network characteristics.

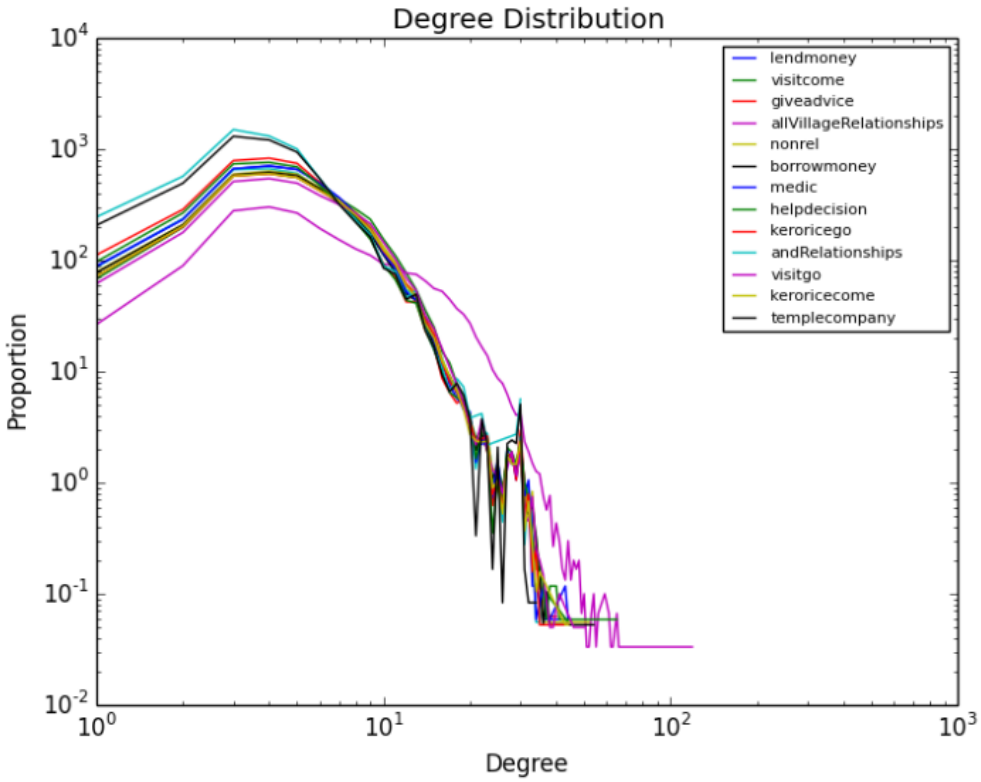


Figure 11: Degree distributions averaged over all villages for all favor-exchange networks.

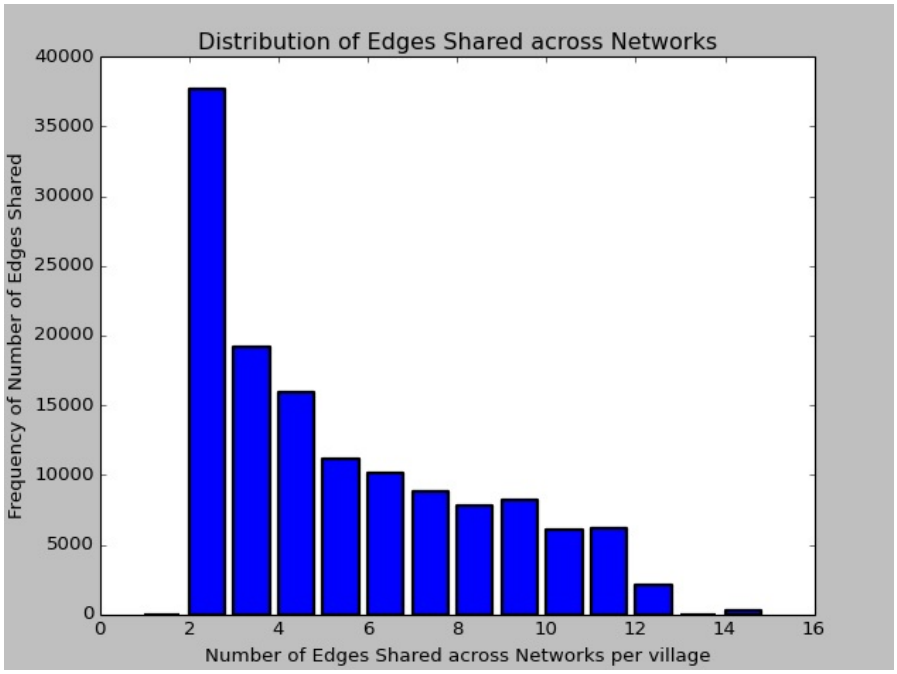


Figure 12: Frequencies of shared edges across networks

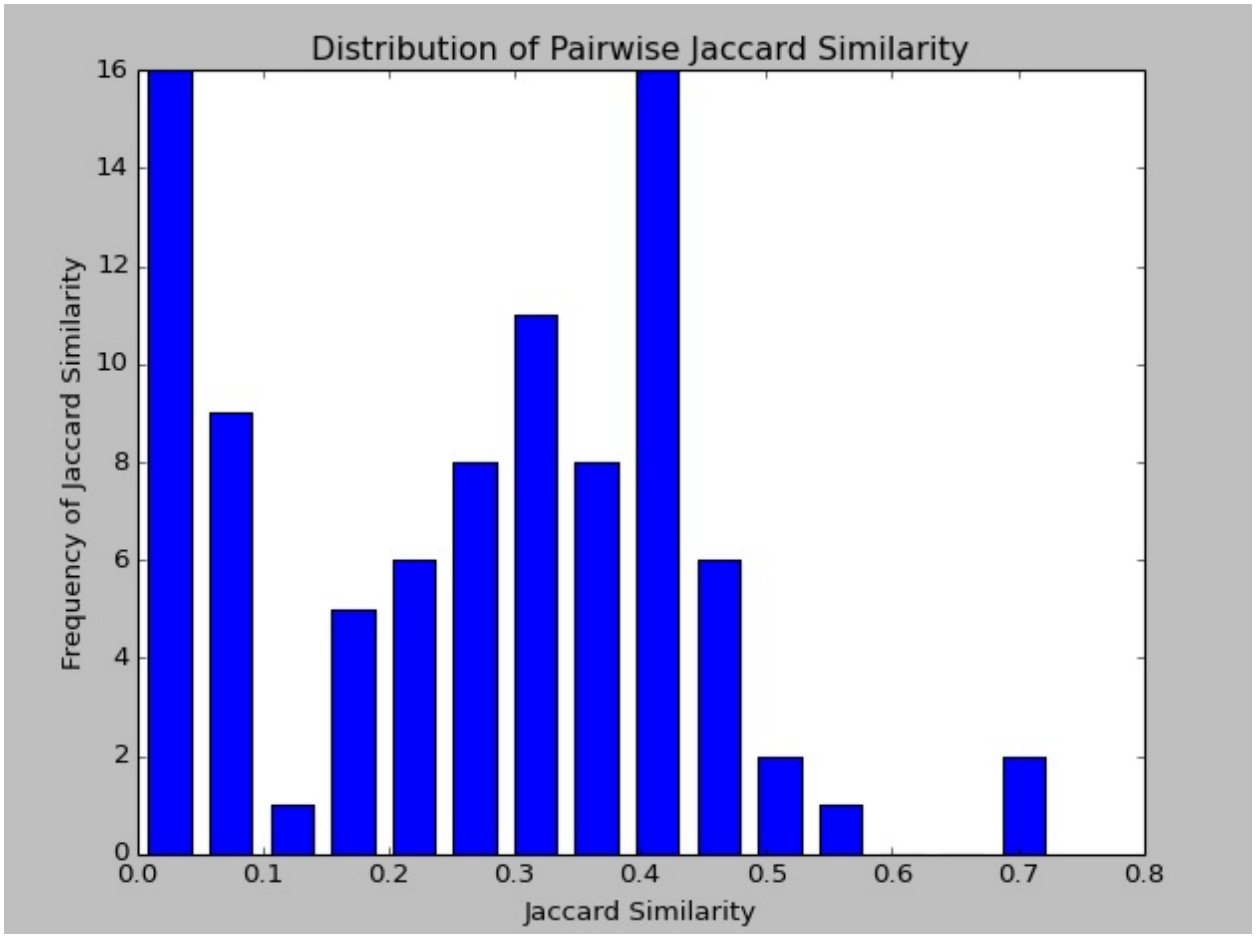
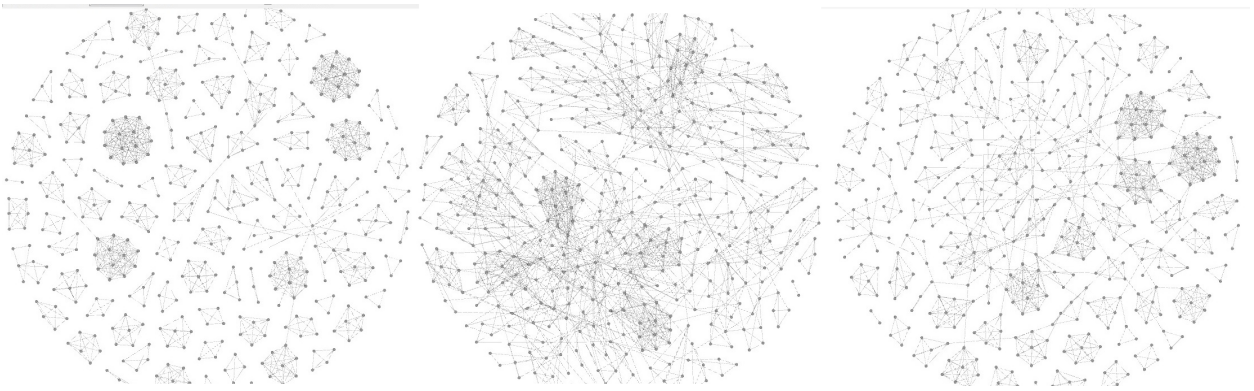


Figure 13: Frequency of pairwise Jaccard Similarity across all favor-exchange network types.



Temple Company Visit Go Give Advice

Figure 14: Fruchterman Reingold network layouts