

Network Structure Matters

A Network Analysis of Micro-loan
Borrowing Groups

SI 508 – Adamic
December 18, 2007

David Hutchful
Pieter Kleymeer

Background

Microfinance borrowing groups are groups of individuals—typically economically disadvantaged—who as a group receive joint-liability loans that would otherwise not be available to them as individuals. Usually these groups are socially connected, either by culture or geographic location, but often not formally acquainted. The rationale behind group lending is that borrowing peers, who have an individual interest to financially prosper and maintain social standing in the community, will be incentivized to monitor and ensure that other members repay their part of the loan. Since lenders within a developing market often do not have the resources to directly monitor borrower behavior, peer monitoring provides an efficient and low cost alternative. Our literature review indicated a relationship between social connections in group banking and repayment rates and social standing. Defining and experimenting with this social connection and its effect on financial outcomes is the basis for our project.

Purpose/Research Question

Our work is inspired by the empirical field work¹ of Professor Dean Karlan² at Yale University, which compiled observations of and lending data from FINCA Perú.³ His work arrived at conclusions regarding the Karlan's data shows that groups with stronger social connections have higher rates of loan repayment and individual savings. In addition, failure on a member's part to repay affects his or her existing social standing with the borrowing group. Given the constraints of field work, Karlan does not explore all possible variables (e.g. different microfinance lending policies, group size, or other various externalities) that could affect the performance of a group. Network analysis will allow us to do this. Specifically, we hope to answer the question: What boundaries of social connectedness exhibit a positive optimality of both social well-being and financial stability?

Model Setup

Data Sources

We begin our exploration by constructing a network based on Karlan's statistical results for cultural similarity and distance proximities of borrowing group members. Without access to individual member data we were forced to recreate these characteristics for individual borrowers (agents) based on averages and standard deviations reported by Karlan. We used a random number generator through NetLogo, an agent-based modeling program, to create a Gaussian distribution of agent characteristics. The agents interacted in a designed financial market over a period of time equivalent to a 4-month loan cycle. The financial and social outcomes of the agents and the borrowing group are dependent upon the social correlations analyzed by Karlan. We model these interactions to create measurable data and then apply network metrics to gain further insights into group and individual behavior.

¹ "Measuring Trust in Peruvian Shanty Towns" by Dean Karlan and Markus Mobius - <http://www.nber.org/~rosenbla/peru/perustrust.pdf>.

² <http://research.yale.edu/karlan/deankarlan/index.html>

³ FINCA is a microfinance institution that operates globally in a number of developing regions.

Assumptions

In designing this model, we made several assumptions. First, because we generated our own data, we were not able to directly map real-world geographic data to the NetLogo model. Instead we chose to organize agents according to the z-score of their distance proximity (created by the random number generator). Second, we assume that the cultural and distance effects on agent performance are equally weighted. Third, we assume zero externalities such as probabilities of natural disaster, effects of family relations, or external financial circumstances. However, we created a dummy variable that assumed a social mechanism for decreasing relational ties between agents. In other words, our dummy variable accounts for the dissolution of social connections.

Model Design

Overview

The model creates a network of n of agents. Each agent n_i carries a vector of attributes v_i describing its cultural background and distance from other group members. Each agent or node also carries a dummy variable that can be used to represent a multitude of potential outcome-affecting factors such as social status, monitoring or punishment costs, or group size preferences.

The model sets up the network by randomly generating values for culture and distance proximities based on averages and standard deviations reported by Karlan. Once each agent has been initialized, it is represented on the graph connected to each other group member by an edge. Each node is symbolized by its cultural leaning—graduate student (representing the majority culture) or lumberjack (representing the minority culture)—and its savings or debt (represented by node size). Each node is placed on the graph according to its generated distance value. This provides a visualization of each agent’s physical proximity to the rest of the group through a concentric circular layout. Agents on the inner circle are, on average, closer to the rest of the group members. The edges that connect nodes are based on a cultural proximity score. The cultural proximity score denotes how culturally similar node n_i is to node n_j . This is represented on a color scale where a lighter shade of green represents greater distance in cultural similarity. Each edge also carries a social influence weight. The amount of influence an agent may have on another agent is dependent upon the distance tier(s) of the agents. Agents closer to the inner tier have greater influence than agents on

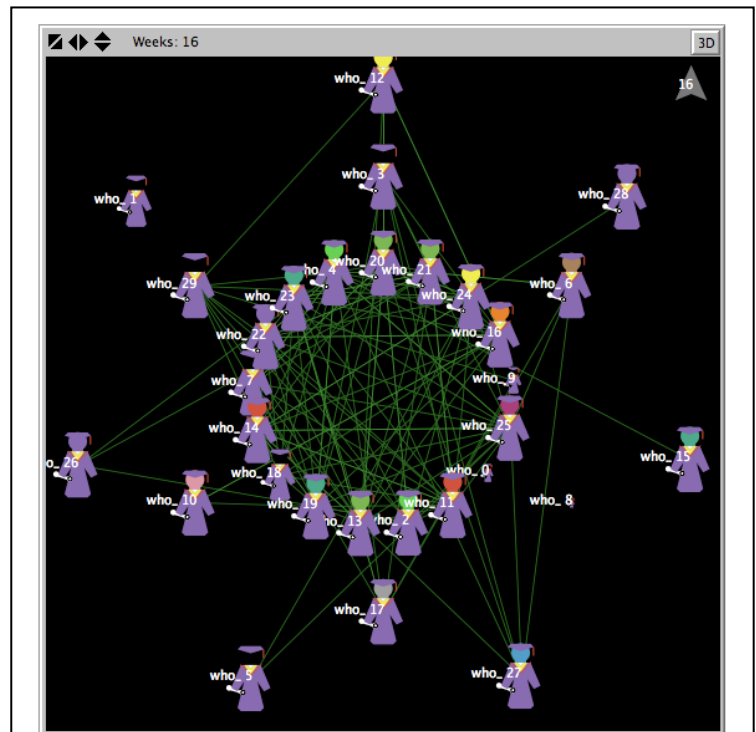


Figure 1: 3-Tier Layout

the outer tiers. Influence is also determined by cultural similarity. Those linked agents of a similar culture have greater influence on one another than on a linked agent of a different culture.

Once the network is arranged, the model then simulates the financial and social outcomes of an x-month borrowing cycle. The user of the model can dictate a number of inputs including the initial loan amount, the length of the borrowing cycle, the number of members in the borrowing group, the standard deviation in cultural proximity and the average distance proximity. The user will also be able to vary the dummy variable to accentuate seemingly hidden externalities. During the simulation, the user can visualize the changes to the network by viewing savings increases (agent size), internal loan default (black face), group member punishment (loss of links), and group member exit (smiley face).

As the model simulates, it calculates typical financial processes for lending. Each group member is required to pay a positive amount on the principal plus interest of the loan. The loan carries an interest rate of 3% that is compounded each month of the borrowing cycle. Each member is also required to hold a savings balance of 20% over the initial loan amount by the end of the borrowing cycle. For example, if the loan amount was \$100, in order to not “externally” default⁴ by the end of a cycle, the borrower would be required to make a payment each week on the principal plus interest, carry a principal plus interest balance of zero at the end of the cycle, and save \$20 by the end of the cycle. The probability that each agent will make a payment each week and will save 20% of the loan amount at the end of the cycle is determined by its cultural and distance proximity scores. The model designates this probability by calculating the z-score for a borrower’s cultural score and computing a normalized distance score from the distance proximity. The dummy variable is also included in determining this probability and is normalized for consistency. If a borrower is not able to make a payment in a given week, she is then required to draw money from her savings to cover the payment. If her savings are depleted she must then borrow money from the group—an internal loan as opposed to the initial “external” loan. If a borrower is required to request an internal loan to make payment on the initial loan that borrower has “internally” defaulted.

Punishment

When a borrower externally defaults, that is, cannot make a payment during the week due to a lack of both personal and community funds, the borrower is removed from the borrowing group (all the links to it’s link neighbors are removed) and thus it has no influence on the other agents. The lending institution assumes any debt incurred by the expelled borrower.

In the case of an internal default, the social relationships of the borrower (B1) are taxed: first, a link to a randomly chosen friend (F1) is broken; second, for each of B1 and F1’s mutual friends, the influence level between F1 and the mutual friend is decreased; and third, if the influence level between B1 and F1 falls below the zero mark, the link to that friend is also removed.

⁴ External default in this scenario is a combination of three events. 1) an agent is not able to make payment on her loan for a given week; 2) the agent does not hold enough money in savings to cover the payment amount; 3) the combined group contributions (“the pot”) does not have enough money to cover the agent’s payment.

Each borrower has a dummy variable (described in “Model Design” section) that is factored into the calculations of their weekly payments. When a borrower internally defaults, some fraction x of the dummy is subtracted from the dummy value until the dummy value is less than or equal to 0, at which point it is set to 0.1 (this is to avoid multiplying by 0 in the payment calculations). If the defaulting borrower has friends, then an influence score is calculated as a fraction of the borrower’s initial influence level (sum of all weights on incident links) and the borrower’s current influence level. Based on the influence score only a fraction of x is deducted from the borrower’s dummy variable.

The intuition here is as a borrower internally defaults, they greatly upset an immediate friend and to a lesser extent mutual friends with the first friend. However, based on the strength (influence level) of ties to the borrower’s remaining friends, the borrower is “encouraged” to meet their payments (this is done by punishing them less as described above). So, although friends “get mad,” the social ties continue to act as influence to get the defaulting borrower to pay back. However, if that borrower continually defaults and upsets the same friend(s), those links will disappear and that influence will no longer encourage repayment.

Outputs

The simulation produces weekly reports as it runs: the first two are Pajek and GUESS format files that represent the network at the end of each week and the third is a CSV file, which contains information about the average degree centralization, average clustering coefficient, the percentage defaulters and total contributions. The CSV file is generated at the end of each loan cycle.

Model Analysis

Measures

Using this model, with a small measure of accuracy, we have replicated Karlan’s data results. Specifically, we have generated a network of 30 individuals borrowing \$150 each, with an average cultural similarity of ~80%, an average distance proximity of ~13 minutes, and an average internal default rate of 21.5%—corresponding to Karlan’s data of 80%, 13.3 minutes, and 20%. Any network data created using our model can be analyzed by applying network metrics. While we completed some analysis using NetLogo, we have also created an export of our network data to Pajek, GUESS, SPSS, and Excel to complete any further analysis. Emphasizing relationships in a graph using Pajek or GUESS may improve the effectiveness of the visualization. Also, using statistical measures allows us to determine if our data is correlated and if our results are significant.

We begin our analysis by recording the degree centrality and clustering coefficient of the network at each week of the loan cycle. The degree centrality is a good measure of how social influence is distributed among members in the borrowing group. A low degree centrality means that social influence is more widely distributed around the group while a high degree centrality indicates a concentration of social influence in the network. The clustering coefficient measures how quickly an agent’s connections will dissolve the social ties between them. A low clustering coefficient means that an agent’s social connections are not interconnected and will not negatively influence each other. Conversely, a high clustering coefficient means that an agent’s social connections are interconnected and have a greater effect on one another.

We ran 1000 iterations of the model for each of five group parameter settings.

1. The first setting reflected a group of borrowers who were very close in average distance (~3 minutes away from one another), and all of one culture (named "Zero").
2. The second setting used Karlan's averages for FINCA-Peru (~13 minutes average distance and 80% similarity, named "Karlan").
3. The third setting modeled a high average distance (~60 minutes) and a high diversity (50%) of borrower culture (named "Max all").
4. The fourth set of parameters maintained the high average distance, but set all borrowers to the same culture (named "Max proximity").
5. The fifth and final setting was the reciprocal of the fourth setting, where agents had a low average distance (~ 3 minutes) and a high diversity (50%) (named "Max culture").

Using NetLogo we exported a csv file of the degree centrality and clustering coefficient measures of each setting for each week of the loan cycle. Dumping the CSV file into Excel and SPSS, we ran statistical measures to find the mean values and correlation between the network measures and the financial outcomes (default rate and total contributions). In performing these analyses, we intend to determine which and to what degree network metrics can predict agent behaviors and financial outcomes. Below are images showing the visual outputs of the data that were analyzed for one sample run:

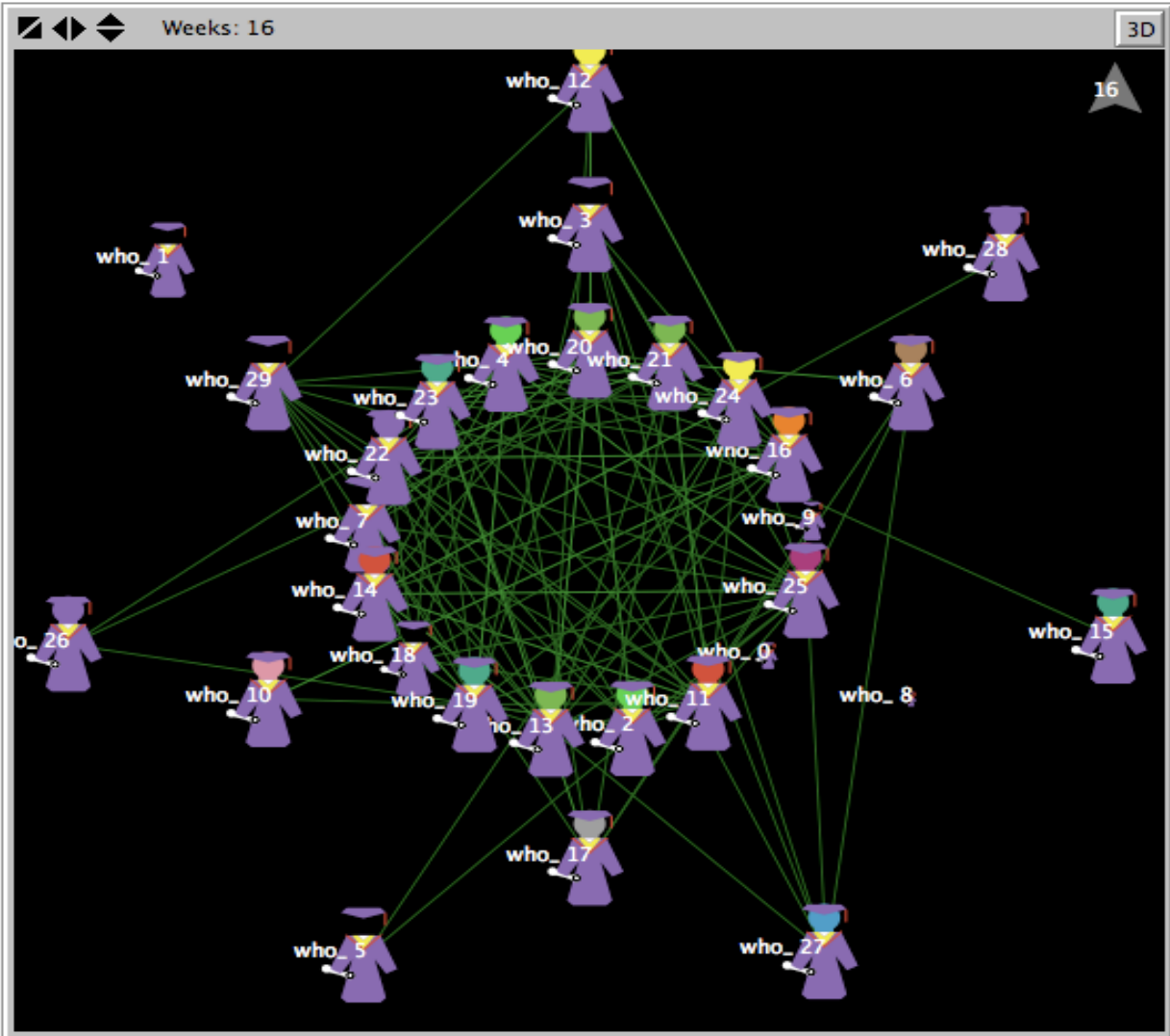


Figure 2: End of loan cycle in N-Logo: defaulters are marked with black faces.

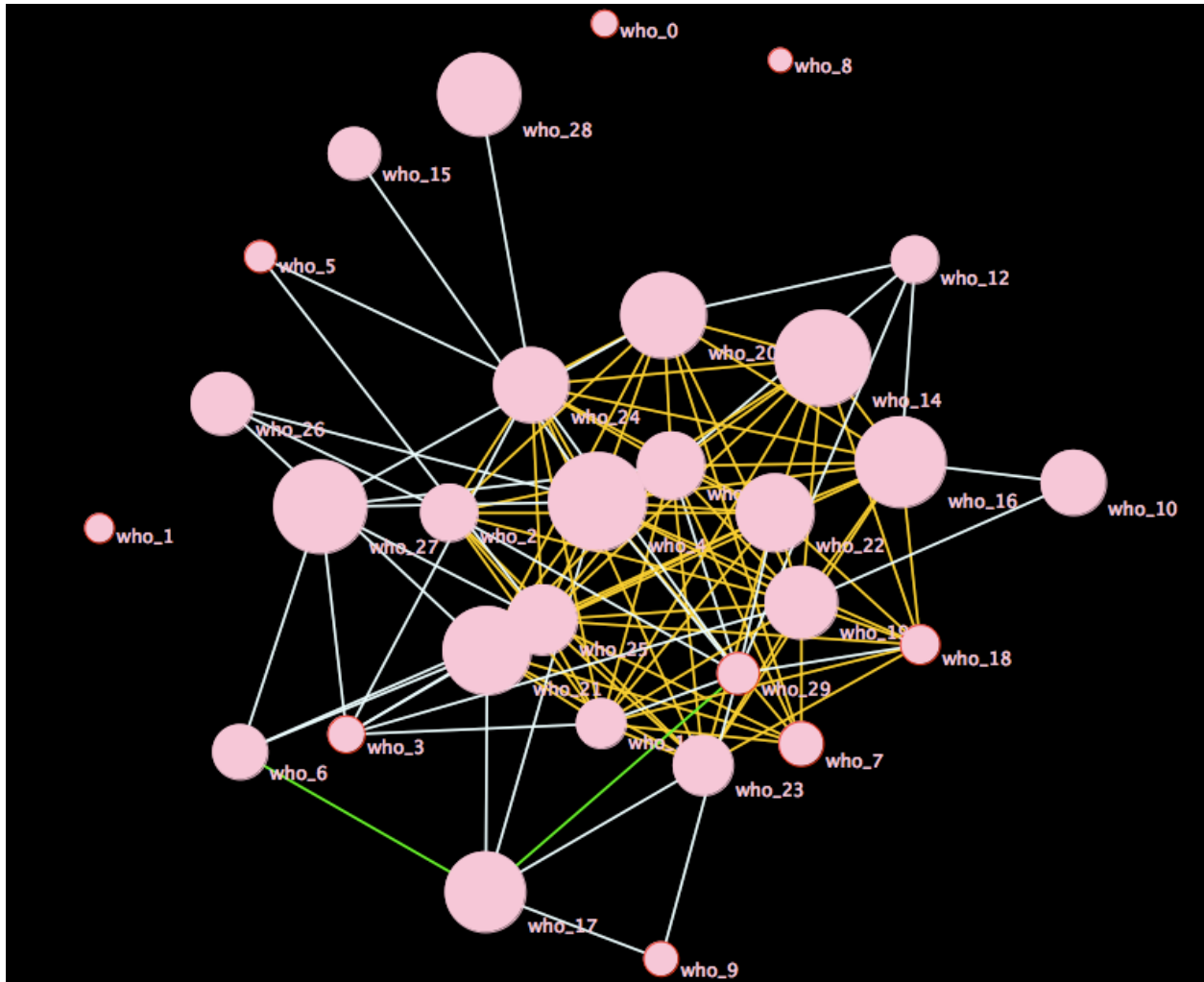


Figure 3: GUESS output - red border stroke indicates defaulters. Node size represents savings and the link colors represent the 3-tier structure: orange is tier 1, green is tier 2 and white is tier 3 links.

The following two images present the network in Pajek at the beginning of the loan cycle and at the end. For both images, the nodes are sized according to clustering coefficient; the colors of the links represent the 3-tier system and a red border stroke marks a defaulter. Note the loss of links between the two images.

For example, in Figure 4, node “who_28” (tier 3) which is in the top right hand corner is connected to nodes “who_3” (tier 2) and “who_24” (tier 1) who are also connected to each other. In Figure 5, the only link left is that between “who_24” and “who_28” (now in bottom left hand corner) and both “who_28” and “who_3” have defaulted. This suggests that “who_28” defaulted at some point, lost a link to “who_3” and some influence from “who_24.” Furthermore, “who_3” defaulted and lost his link to “who_24.”

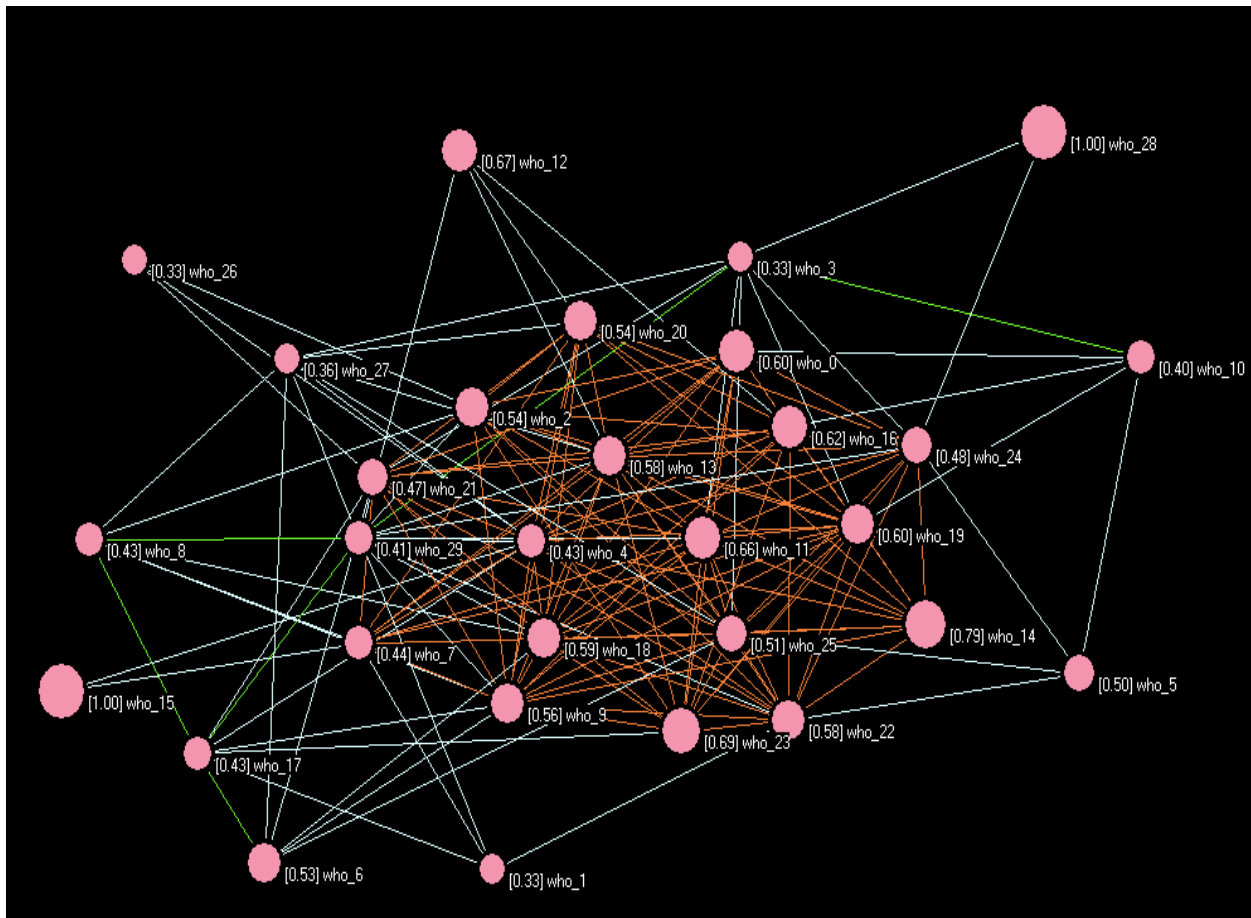


Figure 4: Pajek output at beginning of loan cycle.

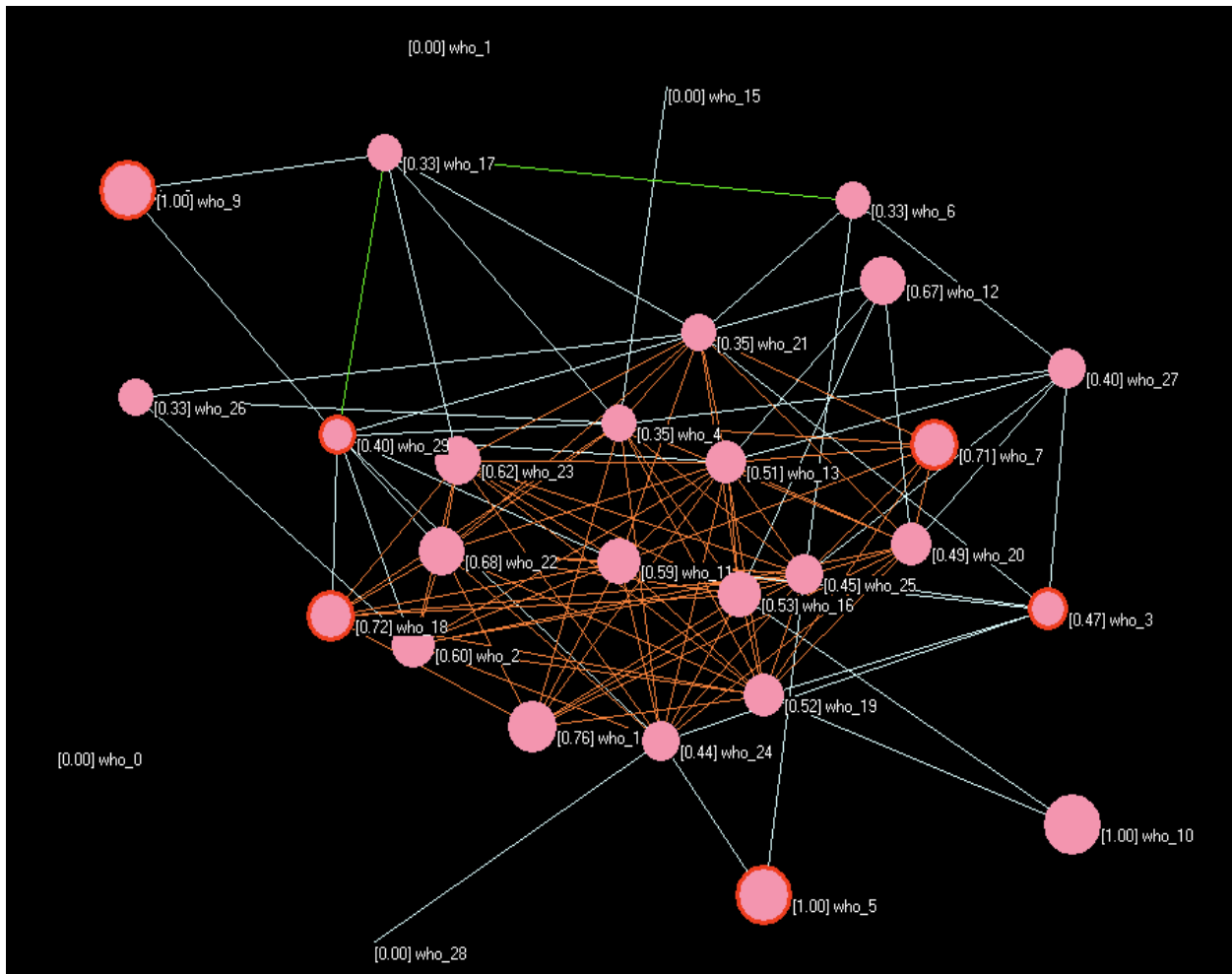


Figure 5: Pajek output at end of loan cycle.

Results

This section presents the results of our analysis. Unfortunately, the correlations we found were low (~0.05 to ~0.43). However, there were significant differences between the numbers to warrant an explanation.

Table 1: Results and Correlations between Network Metrics and Parameters

	Zero	Karlan	Max all	Max proximity	Max culture
Degree (Deg)	High (0.27)	Average (0.25)	Low (0.16)	Low (0.20)	High (0.27)
Clustering Coefficient (CC)	High (0.55)	Average (0.46)	Low (0.13)	Low (0.16)	High (0.55)
Repayment Rate	High (0.85)	Average (0.78)	Low (0.65)	Low (0.58)	High (0.86)
Top Correlation to Repayment Rate	CC	CC	Deg/CC	Deg/CC	CC

Explanation

Our results provided a few interesting insights into the world of microfinance. We see three major findings: 1) culture and distance are independent of one another, and distance is the most critical factor for loan repayment; 2) degree centrality of the network does not directly indicate a predictable financial outcome for a borrowing group; 3) clustering coefficient of the network is weakly correlated with the financial outcomes of a borrowing group.

From the above table, the following can be inferred:

- When degree centrality and clustering are low, clustering is weakly correlated to a very low repayment rate.
- When degree centrality and clustering are high, clustering is weakly correlated with a high repayment rate.

This result indicates that, on the network level, financial outcomes are determined not entirely by how many people a borrower is connected to, but whether those people are connected with each other as well.

Conclusion

For this project, we carried out an exploratory data analysis of the social structure of micro finance borrowing groups to uncover determinants of positive financial and social outcomes.

Our analysis indicates that a high clustering coefficient of the network is necessary for positive outcomes. However, the lack of individual data points and the general difficulty in defining a model that perfectly fits the real world scenario weakened the correlations on which we base our findings and thus, the significance and generalizability of the findings. For future work, a more comprehensive data source and further analysis looking at individual node network metrics and edge weights (influence values) could strengthen our findings and illuminate other results we were unable to uncover with our current model.