Social Networks 1

Final Paper

A Social Network Analysis of Rural Indian Villages

Salma Taoufiq
Gyorgy Attila Ruzicska
Anastasia Kvasha

I. Introduction

The description of the strength of any type of relationship linking two individuals is not an easy task as this cannot be simply taken as being binary. Rather, we may consider it on a spectrum. Usually, people nurture close relationships with a select few: with family members and/or close friends. A weaker connection links them to a larger group of individuals with whom the interactions are less frequent. Additionally, people generally have an even weaker tie with a bigger number of casual acquaintances. There is a tradeoff between the number of people an individual is linked to through their ties and the strength of these ties. This tradeoff actually stems from important biological factors. These have been studied and described by Dunbar in his famous article "Neocortex Size as a Constraint on Group Size in Primates" [1].

Tie strength is currently one of the topics at the forefront of network science. It is particularly interesting in the current digital age with the myriad of means of communication it has brought such as social media. Numerous studies have shed light on the irrevocable importance of capturing tie strength and the social factors that contribute to it. Knowledge of the strength of a tie has been proven to improve the visualization and modeling of the spread of disease and information [2]. Moreover, it increases the accuracy of link prediction and contributes to better market targeting [3, 4].

As seen in class, one cannot dive into studying the strength of ties without coming across the work of Mark Granovetter, specifically his article "The Strength of Weak Ties". In his paper, Granovetter explains the differences between a strong tie and a weak one. He also presented an interesting hypothesis regarding these ties which can be expressed as: the stronger the tie

between two individuals, the bigger the fraction of common friends they have. This puts forward the importance of one's weak ties as they might bring crucial information. Your relationship with your weak ties should be cultivated as they are the elements that knit together your network and improve the free flow of information between various parts of your network.

We have decided to dive into the study of tie strength using a collection of 75 social networks, each corresponding to a village in Karnataka, India. The interactions covered by this dataset are mostly face-to-face.

II. Dataset Description

a. Context of the dataset:

The dataset we have chosen is about 75 Indian villages that were identified by a microfinance institution called Bharatha Swamukti Samsthe in their diffusion endeavor. These villages are located in the southern rural part of Karnataka. The data was collected by Professor Esther Duflo and her team. Duflo is a distinguished French-American economist who is teaching Poverty Alleviation and Development Economics at the Massachusetts Institute of Technology. She has received multiple academic prizes and honors, notably the 2019 Sveriges Riksbank Prize in Economic Sciences in Memory of Alfred Nobel (with co-Laureates Abhijit Banerjee and Michael Kremer).

The study of these 75 villages entailed conducting a census of households, as well as asking a subset of individuals some detailed questions regarding their relationships with other villagers. This has led to making a rich network dataset which has consequently fueled a lot of research. The data has been used in a variety of work ranging from econometric to theoretical, including field experiments.

The data is two-fold: on the one hand, there is information about households, and on the other about a subset of individual villagers. The former includes facts such as the type of roof the household has, whether it has electricity, number of rooms and beds, whether the inhabitants own or rent the place, etc. As for the latter, it gives a myriad of information about the interviewed individuals, such as their gender, status within the household, age, religion, caste, occupation, education, languages they speak, etc. Additionally, these people were also asked

about the relationships that link them to fellow villagers, for instance: who they borrowed money from, gave advice to, are related to, invited to their homes, as well as with whom they go to the temple. In fact there are 12 types of relationships that were defined in total. Villagers who did not participate in the individual level survey were assumed to be connected to their household members on all the 12 layers. In our analysis, however, we only consider the random sample of villagers who did participate in the individual level analysis and so our results are not distorted by selection issues.

The data is quite intricate, providing us with a lot of information. In our paper, we will make sense of, and use these details as tools to guide us in answering the following research questions:

- ➢ What characteristics determine whether an individual in a village is "strong tied" vs. "weak tied"?
- ➢ How does a villager's centrality relate to his/her level of "tiedness" ?
- ➢ To what extent does Granovetter's hypothesis hold in these villages? Why?

b. <u>Data Exploration:</u>

During our exploration of the data, we have discovered some limitations. In fact, while the available multilayer dataset has rich information on both individual and household levels for 75 villages, there were some differences in the type and availability of information on these two levels, which we had to address in order to perform the analysis. The survey conducted among the households was thorough and thus every node of the developed network had detailed attributes, which include information on:

- religion (hinduism; islam; christianity; other; refuse to say; do not know);
- caste / subcaste (OBC; Schedule caste; General; Schedule tribe; other);
- roof type (thatch; tile; stone; sheet; RCC; other);
- number of rooms;
- number of beds;
- electricity (yes, private; yes, government; no);
- latrine (owned; common; none);

- house ownership / rent (owned; owned nut shared; rented; leased; given by government; other; refuse to say; do not know);
- whether one of the village leaders is living in the household.

However, as mentioned, data collection on an individual level covered only a subset of all households (the proportion of households where an individual was surveyed varied across villages, but was around 50% on average). Households for this detailed assessment were randomly sampled and stratified by religion and geographic sub-region. Main recorded characteristics on an individual level with their respective measurements included:

- gender (male; female);
- status (head of household; spouse of head of household; other);
- age;
- religion (hinduism; islam; christianity; other);
- caste / subcaste (OBC; Schedule caste; General; Schedule tribe; other; refuse to say; do not know);
- mother tongue and knowledge of other languages (kannada; tamil; telugu; hindi; urdu; english; other);
- education (1-9th standard; S.S.L.C.; 1st P.U.C.; 2nd P.U.C.; uncompleted degree; degree or above; other diploma; none);
- whether a person is a village native;
- how long a person lived in a village;
- occupation (did you work last week; how much time spend working last week; occupation; for whom do you work);
- savings (participation in an SHG or other savings group; loans; bank account);
- election card (yes; missing; no; refuse to say; do not know);
- ration card (yes; missing; no; refuse to say; do not know).

Given the information to which household any person belongs to, it was possible to assign household attributes to each individual. An example of data availability on an individual level is shown on Figure 1. The dataset allowed us to develop an undirected weighted network based on the different types of relations between individuals. The network below is an

aggregated version of all 12 layers of possible interactions. Village 52 was selected as an illustration since its network is one of the most dense in the whole dataset. Figure 1 shows the core part of the village's network - most dense part of strongly connected nodes (which also includes individuals with the highest degree). The network reveals the social structure of this village, consisting of complete subgraphs (separate households) connected with each other through much weaker links. In around half of the households, at least one person participated in a detailed survey. Yet, as can be noticed, most of the nodes are grey - meaning that their individual's characteristics are unknown. In case of village 52, the proportion of people for whom full information is available is a bit less than 26% (395 villagers out of the 1525 total population).

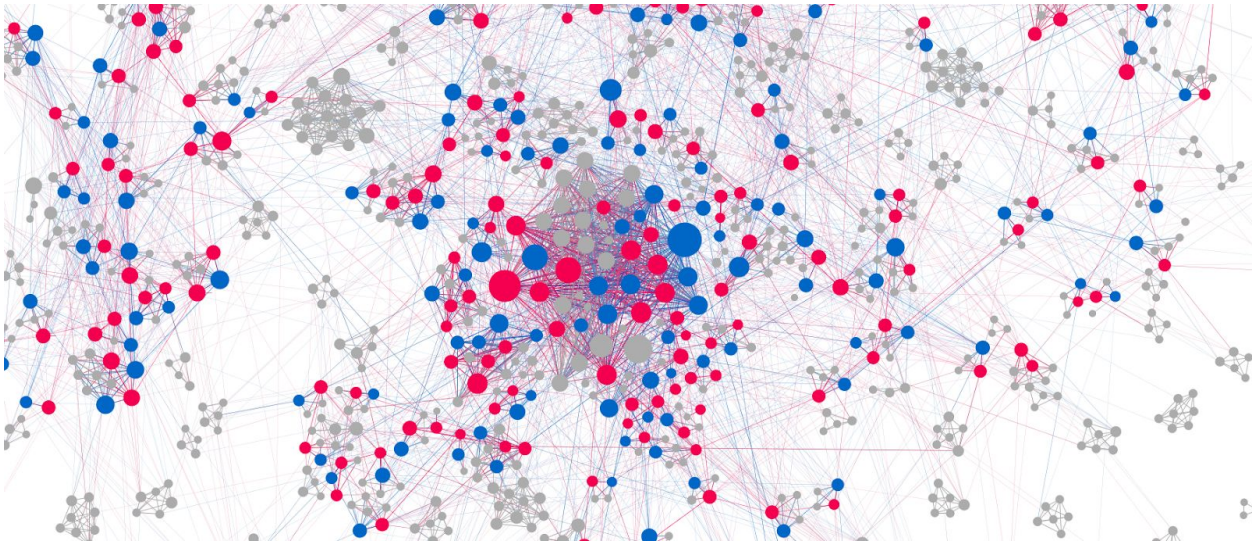

Figure 1. Core section of the aggregated network for village 52. Each node represents an individual: blue - males, red - females, grey - no data. Size of nodes shows their degree (visualized in Gephi).

Taking a closer look at the network's subgraphs (households), we can verify the presence of strong ties between people living within the same household (thicker links - more types of interactions, as will be discussed later) (Figure 2). These complete graphs are connected with other households through fewer much weaker ties.

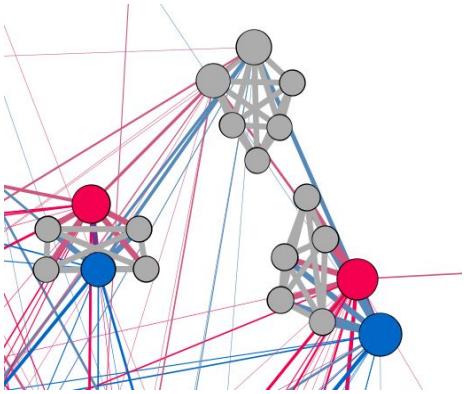

Figure 2. Fragment of aggregated network for village 52 showing three households.

A bit different structure can be noticed when examining the network based on interactions between household, not individuals (Figure 3). We won't notice a large number of small strongly connected clusters which on the previous visualizations represented individuals grouped into households/families. At the household level nodes are forming fewer but much larger groups, however, strongly connected, complete subgraphs still can be found.

Situation in different villages can vary significantly, but just as an example, Figure 3 shows how belonging to a certain caste might be affecting clustering in a village (other characteristics of households did not show grouping patterns as clearly). Caste system in India still can have a strong effect on social hierarchy (social condition and respect), especially in rural areas. With regards to castes, social respect is declining in this order: General Category (GC), Other Backward Caste (OBC), Scheduled Caste (SC) and Scheduled Tribe (ST) [5]. On the figure below, there are two clusters of households quite prominently separated from the larger section of the network: on the right side - group formed exclusively by households from the OBC caste (blue); and on the left - group formed mainly by representatives of the SC caste (yellow). The dynamics within and the structure of the large central component can be of particular interest for analysis - while households still often tend to have stronger connections with people from the same caste (or with those who have similar social status), there is some diversity in distribution and position of nodes in the network.

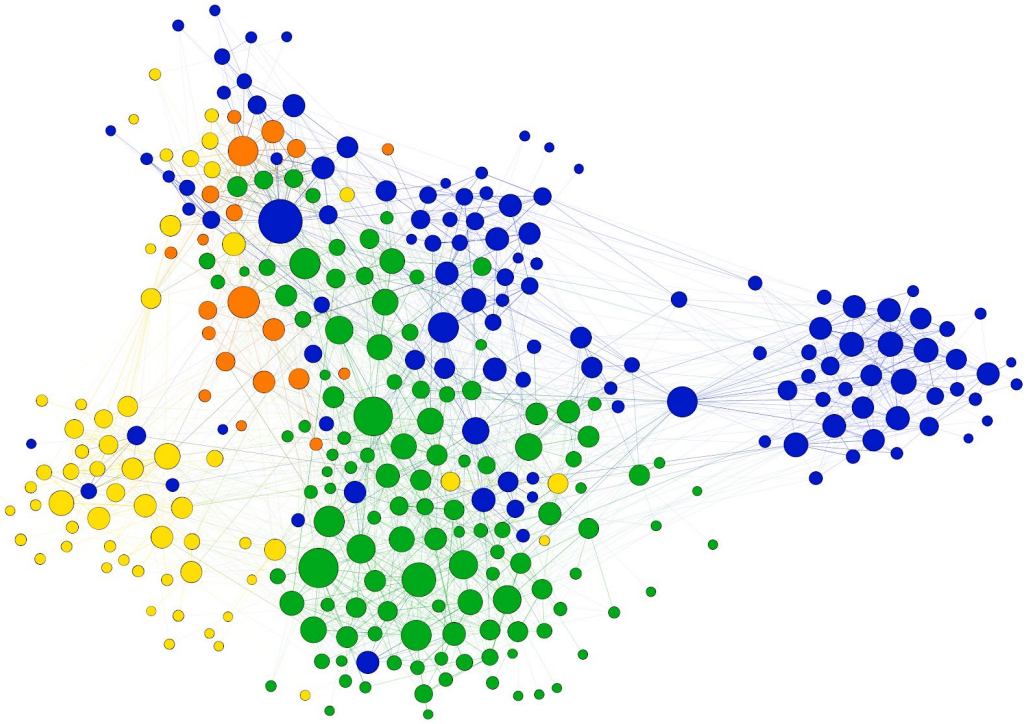

Figure 3. Aggregated network for village 52. Each node represents a household. Colors show different castes/subcastes: green - General Category (GC), blue - Other Backward Caste (OBC), yellow - Scheduled Caste (SC), orange - Scheduled Tribe (ST). Size of nodes shows their degree (visualized in Gephi).

c. <u>Assortativity:</u>

Assortativity (or assortative mixing) is a concept that describes how strongly nodes in the network tend to connect with other nodes which are similar to them. In case nodes in a network are more likely to connect to their opposites (e.g., nodes with low degree try to connect to those with high degree), it is called dissortativity. It is most often defined through the values of node's degree and average degree of its neighbours [6]. Assortativity is common for social networks, however so is the presence of both assortativity and dissortativity elements at the same time. Figure 4 shows the relationship between a node's degree and its neighbor average degree for individuals living in village 52. We can notice more prominent positive slope - there are some nodes with low degree which tend to connect to nodes with much higher degree, however, most nodes are neighbouring the nodes which are quite similar to them (low degree with low degree, high degree with high degree). There are no high degree nodes which would be surrounded by many neighbors with low degree.

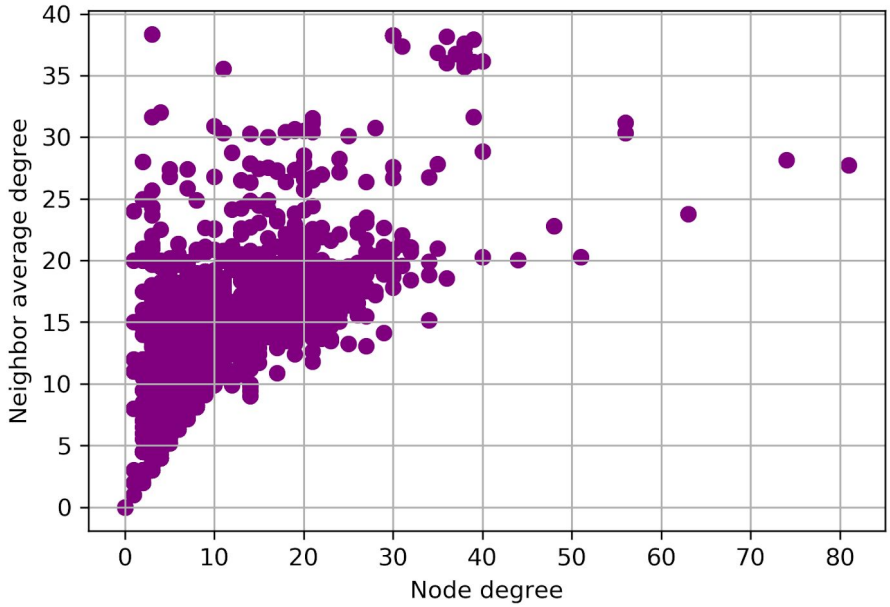

Figure 4. Relationship between node degree and neighbor average degree for individual villagers (data for village 52).

However, the picture is quite different when we look at the network on a household level. Figure 4 shows the same relationship between node degree and neighbor average degree for village 52 but on the household level. This time the distribution is quite different, it combines the features of both assortativity and dissortativity more prominently than it was on an individual level. In this case we can probably call this relation rather neutral. There are some nodes with low degree connected to nodes with very high degree (negative slope in the upper part of the Figure 5 - characteristic of dissortativity). However, there is no clear relationship between node degree and average neighbor degree that we can observe.

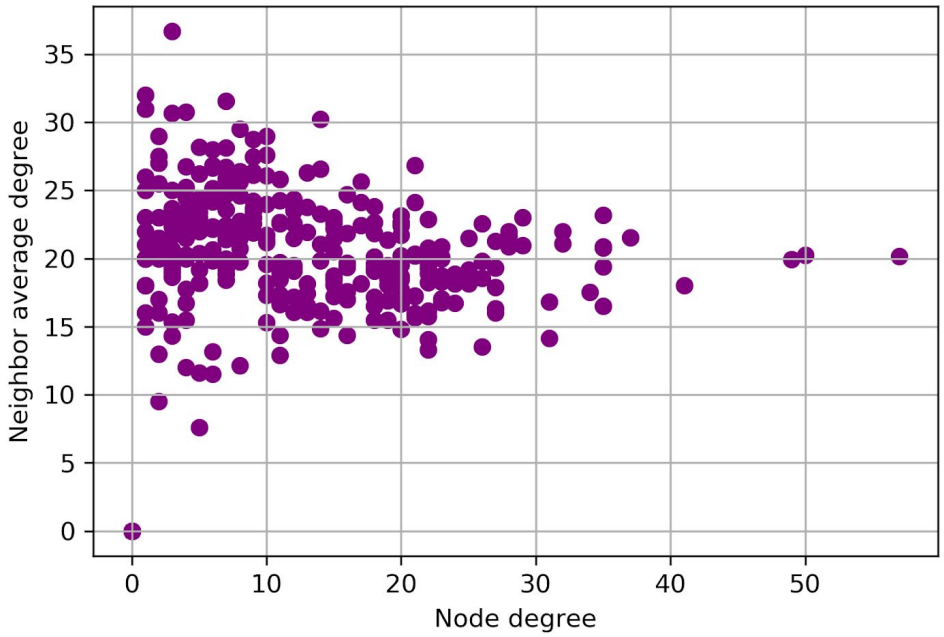

Figure 5. Relationship between node degree and neighbor average degree for households (data for village 52).

III. The Ties Between Villagers:

For the analysis of this dataset, we define the strength of a tie as the number of distinct types of social relationships that exist between two individuals. We have 12 different types of interactions which are:

- whose home he or she visits (VisitGo),
- who (s)he invites to one's house (VisitCome),
- his or her relatives in the village (Related),
- with whom (s)he engages socially (NonRelated),
- who gives him or her medical help (Medic),
- from whom he or she borrows money (BorrowMoney),
- to whom he or she lends money (LendMoney),
- from whom he or she borrows material goods (e.g., kerosene, rice) (KeroRiceCome),
- to whom he or she lends material goods (KeroRiceGo),
- whom he or she asks for help with a decision (HelpDecision),
- to whom he or she gives advice (GiveAdvice), and

- with whom he or she goes to pray (e.g., at a temple, church or mosque) (TempleCompany).

We define "strong tie" as one linking two people who interact with each other across more than 6 distinct layers. It is worth noting that the networks are undirected and so the adjacency matrices symmetric since directionality is not taken into consideration within this dataset. For instance, if a villager A lends money to villager B, and A also gets important advice from B, then the weight of the tie between A and B is 2. If in addition to that, A and B also pray together in the same temple, then their tie strength is 3, etc. Hence, the minimum strength a tie can have is 1 and the maximum is 12 (highest values are very clearly indicating relatives living in the same household).

Three measures have been decided upon to study the strength of an individual's interactions:

1. A villager is considered to be "strongly tied" if (s)he has a higher number of strong ties than that of weak ones;
2. The ratio of strong to weak ties;
3. The ratio of the actual number of layers in an individual's network to the maximum number of interactions with the same neighbours.

On the left panel of Figure 6, for example, we can observe that the majority of the villagers have on average twice as many weak ties than strong ties. The right panel shows that around half of the total layers are present in an individual's relationship structure.

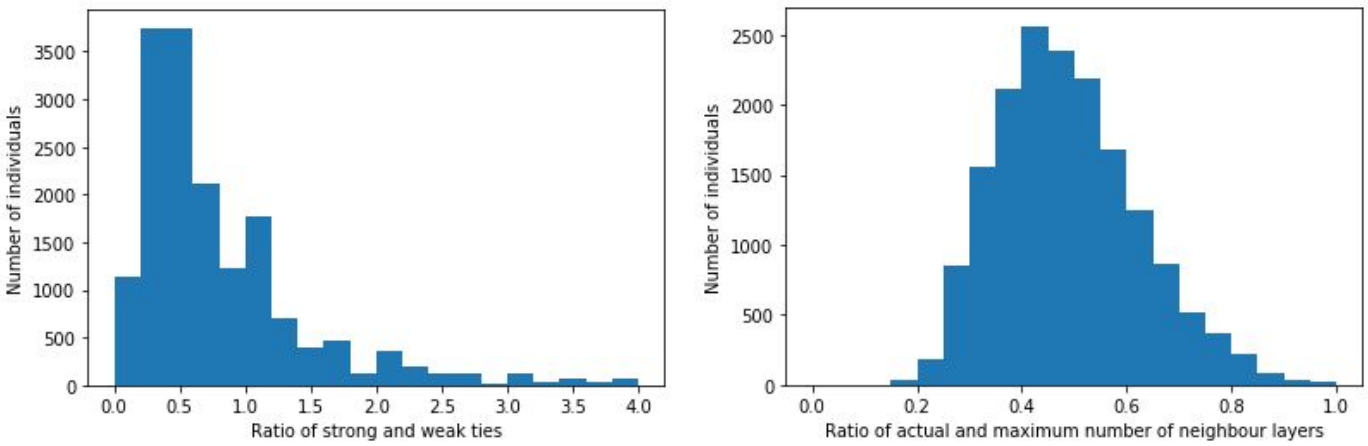

Figure 6. Left - Density of strong/weak ties; Right - Density of actual/maximum number of layers

In an endeavor to answer our first research question, we investigated how household and individual characteristics affect villagers' relationships across the three above-mentioned measures. An Ordinary Least Squares regression (OLS) was performed in which discrete choice variables were included as separate dummies and village level characteristics (e.g. network size) were controlled for using village fixed effects. As such regression specification results in various dummy variables to be perfectly collinear, we have decided on particular base categories to which we compare and interpret the included variables. In particular, base categories include: Caste - did not respond, Mother Tongue - Hindi Language, Latrine - None, Electricity - No, House - Rented. We have performed two regression specifications: in the first one we included degree centrality measures of 8 different layers, in the second we replace those with an aggregated degree centrality measure.

We have summarized our findings in Table 1 and 2 (Appendix) where we have included the independent variables we studied (not showing the village fixed effects) and their impact on our measures as well as the corresponding standard errors. Note that some independent variables, religion and some language dummies specifically, were excluded from the specification because they caused large multicollinearity as measured by the Variance Inflation Factor (VIF). A VIF captures and quantifies the amount of inflation of the variance. In the regression model used in our work, the standard errors — and hence the variances — of the estimated coefficients get

inflated when multicollinearity exists. Thus, an investigation of the VIFs in the model is required. Table 3 summarizes the obtained results.

The cutoff point is 10, and we can readily see that some values exceed it, mostly among the different degree centrality measures. All other estimators are quite low in comparison. The variance inflation factors are fairly large for these measures due to the strong relationship between the different types of interactions that can be established between villagers. A certain redundancy comes with including these measures, entangling their effects on the dependent variables at hand. Relationships that require a certain amount of trust and intimacy are all highly collinear. One cannot borrow money from just anyone, similarly for asking/getting advice. Thus, it is highly probable that these relationships go together, explaining why there are VIF values that exceed the cutoff. Nonetheless, including the centrality measures on these 8 layers results in having all the VIF indicators below 12 which reflects the absence of serious multicollinearity issues. Table 1, then, allows us to interpret more detailed ceteris paribus effects of degree centrality on different layers.

In order to analyze the robustness of the estimates, a regression analysis was reconducted taking the overall degree centrality across all layers as one variable. The results of this regression are recorded in Table 2. Such specification results in having all VIF indicators below 10, therefore, fully avoiding multicollinearity issues.

In the analyzed subsample, our definition results in that approximately 22% of villagers are strongly tied. In contrast, if we defined "strong" ties as interaction across more than 4 layers, this proportion becomes nearly 40%. Clearly, our results are sensitive to these differences in annotations. However, the sign and significance of the estimates are mostly robust to alternative definitions.

Centrality Measures:

In network science, centrality captures the importance of a node in a graph. It can be measured in three different ways, with each one focusing on a certain aspect of that "importance":

1. Degree centrality is simply the total number of edges that are incident upon a given node. This value can be standardized by dividing by the possible number of

edges a node may have, i.e. N-1 with N being the total number of nodes in the given network, and the (-1) accounts for the node for which we compute the degree centrality.

2. Closeness centrality measures how far a node is from all the other nodes in the network.
3. Betweenness centrality measures how many times a node happens to be part of a shortest path between all other pairs of nodes.

We have chosen to dive into the details of degree centrality by studying it individually across 8 of the 12 layers of interactions - as highlighted in color in Table 1 in the Appendix. We have included only 8 in order to avoid high collinearity across variables. The layers included in the regression analysis are: money borrowing, related, non-related, attend the same temple, visit the other's house, help with decision, give advice and material good borrowing. From this table, we can notice that all three types of centralities have significant effects on all three measures, namely: a villager's tiedness, his/her strong-to-weak ties ratio, along with the actual number of interactions for this individual across layers to the maximum number of interactions (s)he can have.

First of all, for the closeness centrality across all layers taken as an independent variable, it affects all three measures negatively. This makes sense as nodes that occupy highly central places in the network are usually the ones with high closeness centrality scores, indicating the low number of hops that must be taken in order to connect to the other distant nodes in the network [7]. The more central villagers are consequently the ones expected to have more connections that can be deemed as weak, thus making their "tiedness" low.

Second, nodes with high betweenness centrality are the ones with the potential of having considerable influence within their village since they are the ones with the most control over the flow of information between others. They are also the ones who once removed from the network will most disrupt communications between remaining village inhabitants due to their strategic position on the largest number of paths taken by messages. Therefore, such villagers are expected to have a higher number of weak ties compared to strong ones. The network of the 75 Indian villages of Karnataka confirms this as can be seen from the results of the regression

performed on the dataset. High betweenness centrality mostly translates into being weakly tied meaning that these villagers have more weak ties than strong ones.

Third, the analytical approach taken for degree centrality is different than the previous two measures. Indeed, a detailed exploration across the 8 layers was taken in this case. Different types of interactions have different effects on the dependent variables, with a general trend that can be expressed as: the higher the number of people a villager interacts with, the higher the chances for him/her to be strongly tied and have larger proportion of layers to be present with his/her neighbours, ceteris paribus. Taking a deeper look at the different types of edges, we observe that the higher the number of people a villager borrows money from, gives/takes advice from, helps with decision, and attends the same religious temple with, the more strongly tied they are. This was expected, as these types of relationships tend to be established with the people closest to an individual. They require a certain amount of existing trust before they can be conducted. Hence, the more such relationships one nurtures, the more strongly tied they may be.

The number of relationships based on family and other social connections have a negative effect on one's tiedness. This has to do with the social organization of the Indian rural scene and the data collection procedure that assigned the maximum number of within-household layers to individuals who did not participate in the individual level survey. An exploration of the data for these 75 villages has shown that approximately 46% of all ties have a maximum strength of 12, as can be seen on Figure 7. This is due to the fact that 96% of all ties within the same households reached a weight of 12. Families and people under the same roof usually know and trust each other, and thus rely on each other first. In contrast, ties across households are less strong. Therefore, the higher the degree centrality across the "Nonrelated" layer, the weaker one's tiedness is. In fact, in this context, "Nonrelated" is equivalent to "engaging socially", this majoritarily involves acquaintances and weak ties, like village shop workers with whom a villager interacts at each transaction for instance, or a neighbor with whom they are not necessarily close but exchange some niceties.

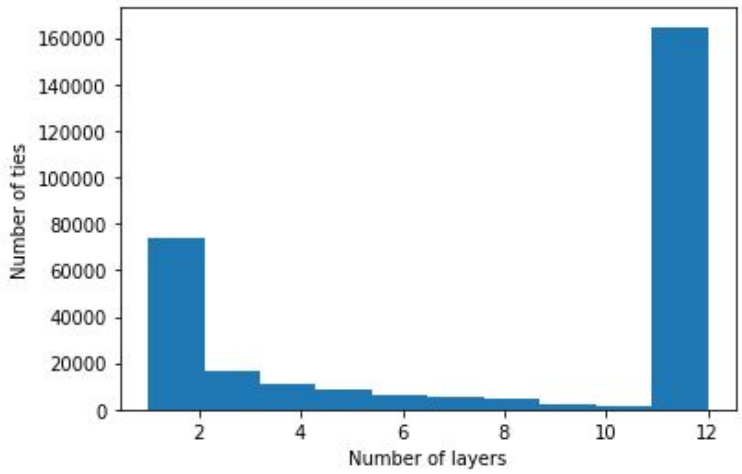

Figure 7. - Histogram of tie strength

As for the "Related" layer, data shows that the larger the families (ie. the higher the "Related" edges within this cluster), the weaker the ties leading to a negative coefficient for this layer. As mentioned earlier, about 4% of same-household interactions do not reach a strength of 12, this is indicative of the fact that sometimes people who are from the same family do not necessarily turn to each other for borrowing rice, kerosene, money, or to ask for medical advice. The households where some of these relationships are not present are the larger ones, increasing the degree centrality of their inhabitants, and resulting in a larger number of weak ties. However, we also must keep in mind that many of the within-household links were artificially created by the data collectors, as many people were not asked individual level questionnaire, and so our estimate on the "Related" variable should be interpreted cautiously.

So far, we have separated the degree centrality measures and studied them for 8 layers. In contrast, when the overall degree centrality is included as a regressor instead, the regression analysis shows that it is not significant which is indicative of the different impacts that the degree centrality has when taken for each layer separately. These results can be observed in Table 2. They also show that for the majority of the other independent variables, the signs of the estimates remain the same (positive/negative). In fact, they only present minor differences in terms of both magnitude and significance.

Individual Characteristics:

Now taking a look at individual people's characteristics, we observe that as people age, their tiedness increases. This is logical, since as a person advances through their life, they tend to form strong bonds. Inversely, with education, the more educated a person is, the smaller the number of strong ties they have. According to India's assessment, evaluation, and research unit ASER Centre's 2018 report called "Annual Status of Education Report (Rural) 2018", an astonishing transformation has occurred in education in rural India within the course of one generation. Young men and women are surpassing their parents' education levels by far [8]. Back in 2001, only roughly 25% of all 18-year-olds in rural India were attending school, as the rest have dropped out by that age. By 2016, the number of 18-year-olds attending schools and colleges had risen by up to 70% [8]. With this surge in education, the youth ends up having to leave the village to continue their studies, reaching better and better positions, and these are the elements that can explain why the more-educated bunch end up having a higher number of weak ties within the village; most probably the only strong ties they have are with family members and close childhood friends, their newer strong ties could be outside of the village (in the city they study in for instance).

Moreover, it is noticeable that all three tiedness-related dependent variables are affected negatively by the "VillageNative" variable. This is not surprising, as a person who is born and raised within the same village know more people. These villages tend to be small communities, so, to a certain extent, most people are able to recognize each other. Village natives and locals tend to experience life in that village throughout their entire lives, they get to "socialize" and somehow interact with more people within the village that people who were born and lived elsewhere before moving to said village. As such, they have much more acquaintances, potentially old friends from different walks of life.

Furthermore, signs of power and social leadership that a person has affects the strength of their ties. There is a "leader" characteristic that is given to a household as a whole. Households that were labeled as "leaders" are those ones that contain shopkeepers, teachers, self-help group leaders - approximately 12% of all households in all villages fall into this category. So, even if only one person in the household is a teacher, and the other members of his family are

unemployed, they all get flagged as “leaders”. The regression results show that if a person is marked as a leader, then they have a stronger tiedness. Unfortunately, as this is a household-characteristic, it does not reflect the effect an individual’s leadership position has on their tiedness. Furthermore, the coefficient on this variable cannot be rejected to be zero in the first regression specification, however, it has a significant positive effect in the second.

Gender as well as individual’s status within the household affects how strongly tied the person is. Women seem to be more weakly tied than men. Similarly, if the person is the head of their household or the spouse of the head, the trend shows that they might actually have more weak ties than strong ones. We could explain this by the fact that this status has specific duties, particularly taking care of the other members of the household. Since they are responsible for the well-being of their household and the people living in it, they tend to work hard to provide for them, and thus have many reasons to engage with other individuals in the village. They end up establishing a high number of weak ties with the people they interact with daily to provide for their families: at work, in shops, etc. Hence, they tend to have a larger number of weak ties as opposed to the rest of the household inhabitants who do not have as many responsibilities.

All in all, a villager’s level of tiedness, the ratio of strong-to-weak ties they maintain, and the amount of interactions they have established across the 12 different types of relationships are strongly linked to a variety of variables. Some of them, like a person’s mother tongue do not have a clear, significant impact on these measures compared to the base category - as can be seen from the high standard errors of the coefficients associated with the mother tongue dummies.

Therefore, we could infer that components like the network structure of the village in which they reside, and the place that individual holds within their entourage in the village have an important impact on tiedness. Furthermore, an individual's personal features such as age and education level have also significant impact on his/her tiedness.

IV. Granovetter’s Hypothesis:

In the 4th column of Table 1 in the Appendix, we have included “Granovetter correlation”. This is a measure we have chosen to analyze the extent to which Granovetter’s hypothesis holds for the individuals in the 75 Indian villages, as well as the features that can

explain the result. The Granovetter-based variable was computed based on the correlation between an individual's tie strength and its neighbours overlap.

Calculating such measure was not possible for all villagers as for some the tie strength was constant across its neighbours. Therefore, we excluded these individuals from the sample.

We may summarize Granovetter's hypothesis as follows:

the stronger the tie between two individuals, the bigger the fraction of common friends they have.

Thus, we began by examining the overlap of the neighboring nodes of the inhabitants of each village. After that, we obtained the "Granovetter_Correlation" which we included in the regression as the dependent variable. We explored the effects of the same regressors as we did for the tiedness measures in order to get a complete view of the strength of ties within the villages of Karnataka. This Granovetter-based score, if high for a certain individual, expresses the fact that (s)he forms strong bonds with the people closest to them -i.e. those who they have multiple common friends with- and weak ones with those they are socially distant from. In contrast, if this score is low, it means that the person forms strong ties even with those who are distant from them socially. From Figure 7 below, we can see that for most villagers (about 6,700 of them), Granovetter's hypothesis holds strongly.

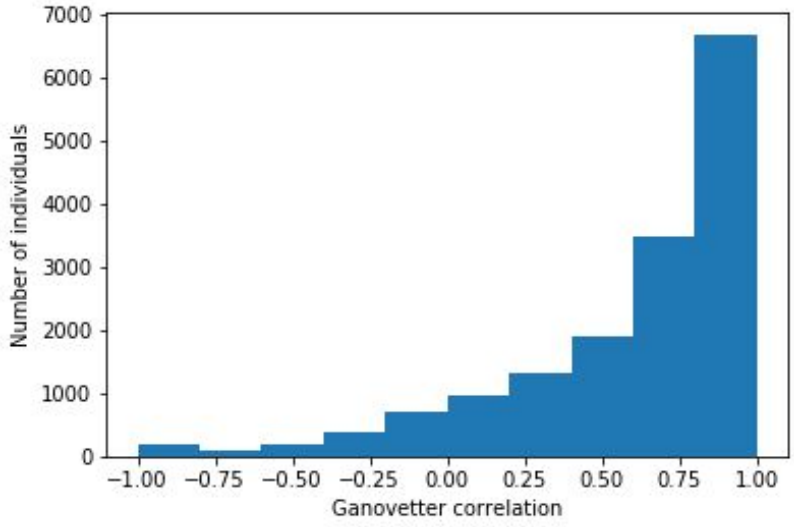

Figure 8. Histogram of the Granovetter correlation scores

Concurrent to our findings regarding closeness centrality discussed above, the effect of this centrality measure is negative. As earlier, the lower the closeness centrality of a vertex, the more peripheral the position it occupies within the network, thus the higher their tiedness, and thus the greater the overlap of common neighbors they have as they usually stay within the same social circle. As for betweenness centrality, it actually has a positive effect on the correlation measure. Individuals showcasing high betweenness centrality are the ones that are on the majority of paths linking other villagers. Hence, the more one holds this key position in their network the higher the Granovetter correlation gets.

Regarding degree centrality, in the analysis we mostly focused on three layers: TempleCompany, Related and NonRelated. The former has a positive, high-magnitude impact on the measure. The more neighboring nodes to a villager attend the same temple as (s)he does, the higher the chances that they would be connecting to the same circles and thus having a lot of common friends. Religion dominates all aspects of Indian rural life. According to A. R. Desai, it is really hard to discern any aspect of rural life that is not determined and permeated with religious rituals [9]. Hence, it makes sense that the individuals who attend the same temples form a tight-knit community.

The two other layers: “Related” and “NonRelated” affect the measure negatively. For the “NonRelated” layer, this means that those who socialize more actively not only form their strong ties with whom their neighbourhood overlap is large, but also with people outside their own circles. This is readily understandable as people with high degree centrality in the “NonRelated” layer are more outgoing and thus not only form their strong ties with people in their close neighbourhood. Furthermore, the coefficient on the “Related” degree centrality measure is also negative, meaning that those who belong to a larger family form bonds more actively with socially distant individuals. This result further confirms our earlier result on tiedness. Together, they imply that ties within larger families are less strong and therefore people from such families seek to form bonds with individuals outside their social circle. In contrast, small families are closely tied together and people within are more introverted. Again, our result on this variable may be biased due to imperfect within household relationship measurement.

Other than the villager's gender, some other regressors which represent a node's individual characteristics have a significant impact on the Granovetter measure. Age, for instance, has a positive coefficient. As a matter of fact, ageing leads people to form closer bonds with individuals in the same social circle. Also, humans tend to associate and befriend people within their age group. In the previous section, we indeed observed that as one ages, they build a larger number of strong ties with multiple other people.

Unlike with strong-weak ratio, whether an individual is educated or illiterate, the length of their studies are not so noteworthy in terms of their effect in checking Granovetter's hypothesis. This might be due to the fact that in rural India, older generations have, for the most part, not attended school or dropped out very early, so education - or lack thereof - does not hinder communications and interactions. For the younger generations, as they tend to interact with people from other villages, besides their families, this variable does not play a significant role here. Although, the more educated a person is, the more weaker ties they have, their strong ties do not necessarily show overlapping social circles.

All in all, the hypothesis presented by the American sociologist Mark Granovetter seems to hold for these 75 Indian villages. However, it could have been very interesting if the data set also included information about the amount of time invested in the relationship between two villagers, as well as the intimacy, and the emotional intensity in order to really be able to capture the strength of ties. Unfortunately, such elements are hard to quantify and also hard to express by the people involved. This is why we had to define our own context by assuming that the number of interactions across the 12 layers approximates the strength of relationships. With this assumption in mind, the hypothesis of the strength of weak ties holds as shown by the histogram of correlations and regression performed on the social network data.

Appendix

Table 1: Results of the OLS regression across all measures (Degree Centrality taken per layer)

	(1)	(2)	(3)	(4)
Variables	**Tiedness**	**Strong:Weak ratio**	**NeighbourLayers ratio**	**Granovetter_Correlation**
Closeness Centrality	-9.015***	-25.97***	-3.035***	-1.258***
	(0.207)	(0.623)	(0.0420)	(0.207)
Betweenness Centrality	-6.649***	-15.78***	-2.048***	3.033***
	(0.448)	(1.338)	(0.0909)	(0.438)
BorrowMoney	38.69***	88.17***	6.936***	-3.432
	(2.205)	(6.575)	(0.447)	(2.143)
KeroRiceCome	12.96***	29.36***	1.433***	-15.60***
	(2.020)	(6.023)	(0.410)	(1.966)
GiveAdvice	5.941**	11.75	3.532***	3.009
	(2.474)	(7.380)	(0.501)	(2.414)
HelpDecision	14.96***	43.07***	5.034***	-6.637***
	(2.080)	(6.205)	(0.422)	(2.016)
NonRelated	-12.95***	-20.67***	-4.115***	-6.252***
	(1.936)	(5.776)	(0.392)	(1.889)
Related	-27.30***	-66.57***	-7.761***	-9.481***
	(1.525)	(4.552)	(0.309)	(1.489)
TempleCompany	16.76***	55.37***	15.19***	75.79***
	(2.710)	(8.099)	(0.549)	(2.642)
VisitCome	7.331***	15.86**	2.239***	-0.947
	(2.095)	(6.249)	(0.425)	(2.044)
Respondent's Gender	-0.0602***	-0.142***	-0.0273***	-0.133***
	(0.0118)	(0.0353)	(0.00239)	(0.0118)
Age	0.00107***	0.00307***	0.000475***	0.00167***
	(0.000322)	(0.000963)	(6.53e-05)	(0.000319)
Education	-0.000476	-0.00654***	-0.000278*	-0.000253
	(0.000819)	(0.00245)	(0.000166)	(0.000808)
VillageNative	-0.0223***	-0.0638***	-0.00813***	-0.0288***
	(0.00827)	(0.0247)	(0.00168)	(0.00821)
WorkFlag	-0.00751	-0.0526**	-0.00381***	-0.00912
	(0.00725)	(0.0217)	(0.00147)	(0.00716)
ShgParticipate	-0.00431	-0.0670**	-0.00549***	-0.0227**
	(0.00983)	(0.0294)	(0.00199)	(0.00969)
Savings	-0.0173**	-0.0316	-0.00678***	-0.0182**
	(0.00751)	(0.0224)	(0.00152)	(0.00739)
ElectionCard	0.0269***	0.0662**	0.00489**	0.000320
	(0.0101)	(0.0301)	(0.00204)	(0.00998)
RationCard	-0.0153*	-0.123***	-0.00727***	0.0139
	(0.00924)	(0.0277)	(0.00187)	(0.00920)
Roof Type 1	0.0545**	0.145*	0.00282	-0.00373
	(0.0266)	(0.0793)	(0.00538)	(0.0261)

Roof Type 2	0.0131	-0.0100	-0.00198	0.00910
	(0.0163)	(0.0486)	(0.00330)	(0.0160)
Roof Type 3	0.0208	0.00767	-0.000967	0.0140
	(0.0156)	(0.0466)	(0.00316)	(0.0153)
Roof Type 4	-0.00505	-0.0273	-0.00707**	-0.00382
	(0.0162)	(0.0485)	(0.00329)	(0.0159)
Roof Type 5	0.0249	0.0192	-0.000843	0.0162
	(0.0173)	(0.0518)	(0.00351)	(0.0171)
Room num	-1.46e-05	0.0129*	-7.14e-05	0.00148
	(0.00258)	(0.00772)	(0.000523)	(0.00256)
Bed num	0.00304	0.00670	0.00134***	-0.00487**
	(0.00237)	(0.00715)	(0.000481)	(0.00233)
Leader	0.0124	0.0334	0.00443**	-0.00438
	(0.00863)	(0.0258)	(0.00175)	(0.00850)
Resp_Status_HeadOfHousehold	-0.0308***	-0.0913***	-0.0246***	-0.0326***
	(0.0112)	(0.0336)	(0.00228)	(0.0112)
Resp_Status_SpouseOfHeadOfHouseh	-0.0534***	-0.214***	-0.0228***	0.0153
	(0.0102)	(0.0306)	(0.00208)	(0.0102)
Caste_OBC	-0.0162	0.0133	-0.00244	-0.00285
	(0.0108)	(0.0323)	(0.00218)	(0.0107)
Caste_ScheduledCaste	-0.0260**	-0.0729*	-0.00963***	0.0112
	(0.0124)	(0.0372)	(0.00252)	(0.0123)
Caste_ScheduledTribe	-0.0217	-0.0321	-0.00427	0.00967
	(0.0166)	(0.0496)	(0.00336)	(0.0163)
MotherTongue_Malayalam	0.271	0.0577	0.0548	
	(0.272)	(1.144)	(0.0552)	
MotherTongue_Marati	-0.484*	-1.830**	-0.107*	
	(0.274)	(0.817)	(0.0556)	
MotherTongue_Tamil	0.00711	-0.0578	-0.00406	-0.0162
	(0.0186)	(0.0556)	(0.00378)	(0.0185)
MotherTongue_Telugu	0.00616	0.0130	0.00348*	-0.00596
	(0.00967)	(0.0289)	(0.00196)	(0.00956)
Latrine_Common	-0.0145	-0.121	-0.0167*	0.0264
	(0.0478)	(0.143)	(0.00969)	(0.0478)
Latrine_Owned	0.0216***	0.111***	0.00796***	0.00115
	(0.00786)	(0.0235)	(0.00159)	(0.00777)
Electricity_YesGovernment	-0.00507	0.0720*	0.00655**	0.0364***
	(0.0141)	(0.0423)	(0.00286)	(0.0139)
Electricity_YesPrivate	0.0146	0.102**	0.0109***	0.0534***
	(0.0140)	(0.0418)	(0.00283)	(0.0138)
OwnRent_GivenByGovernment	0.0442**	0.0339	0.0117***	0.0276
	(0.0218)	(0.0651)	(0.00441)	(0.0215)
OwnRent_Leased	-0.0341	-0.272	0.00854	0.0819
	(0.0701)	(0.209)	(0.0142)	(0.0728)
OwnRent_Owned	0.0203	0.0525	0.0112***	0.0158
	(0.0158)	(0.0473)	(0.00320)	(0.0157)
OwnRent_OwnedButShared	0.0229	-0.0176	0.00455	-0.0580
	(0.0376)	(0.112)	(0.00763)	(0.0373)

Constant	2.596***	7.696***	1.246***	0.913***
	(0.0689)	(0.207)	(0.0140)	(0.0686)
Observations	16,912	16,803	16,912	15,871
R-squared	0.390	0.350	0.671	0.181

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Results of the OLS regression across all measures (Overall Degree Centrality taken across all layers)

	(1)	(2)	(3)	(4)
Variables	**Tiedness**	**Strong:Weak ratio**	**NeighbourLayers ratio**	**Granovetter_Correlation**
Closeness Centrality	-8.024***	-23.87***	-2.806***	-0.588**
	(0.243)	(0.717)	(0.0578)	(0.231)
Betweenness Centrality	-2.516***	-6.020***	-1.078***	3.550***
	(0.487)	(1.428)	(0.116)	(0.453)
Degree Centrality	-1.014	2.946	0.0106	-7.273***
	(0.814)	(2.388)	(0.194)	(0.765)
Respondent's Gender	-0.113***	-0.283***	-0.0463***	-0.160***
	(0.0129)	(0.0378)	(0.00307)	(0.0122)
Age	0.00210***	0.00599***	0.000961***	0.00266***
	(0.000351)	(0.00103)	(8.36e-05)	(0.000331)
Education	-0.00126	-0.00868***	-0.000600***	-0.000542
	(0.000899)	(0.00264)	(0.000214)	(0.000845)
VillageNative	-0.0618***	-0.167***	-0.0213***	-0.0455***
	(0.00903)	(0.0265)	(0.00215)	(0.00855)
WorkFlag	-0.00786	-0.0549**	-0.00573***	-0.0184**
	(0.00796)	(0.0234)	(0.00189)	(0.00749)
ShgParticipate	0.00752	-0.0396	-0.00326	-0.0252**
	(0.0108)	(0.0317)	(0.00257)	(0.0101)
Savings	-0.0213***	-0.0473*	-0.0100***	-0.0256***
	(0.00824)	(0.0242)	(0.00196)	(0.00773)
Election Card	0.0632***	0.160***	0.0164***	0.0101
	(0.0110)	(0.0324)	(0.00262)	(0.0104)
Ration Card	-0.0329***	-0.173***	-0.0152***	-0.00244
	(0.0101)	(0.0298)	(0.00242)	(0.00962)
Roof Type 1	0.0506*	0.141*	0.00169	-0.0106
	(0.0292)	(0.0856)	(0.00694)	(0.0273)
Roof Type 2	0.00839	-0.0206	-0.00422	0.00145
	(0.0179)	(0.0525)	(0.00426)	(0.0167)
Roof Type 3	-0.00401	-0.0577	-0.0117***	-0.0122
	(0.0171)	(0.0503)	(0.00407)	(0.0160)
Roof Type 4	-0.0283	-0.0884*	-0.0165***	-0.0244
	(0.0178)	(0.0523)	(0.00424)	(0.0166)
Roof Type 5	-0.0145	-0.0844	-0.0164***	-0.0154
	(0.0190)	(0.0559)	(0.00453)	(0.0178)

Room num.	0.0287***	0.0899***	0.0114***	0.0243***
	(0.00278)	(0.00818)	(0.000662)	(0.00263)
Bed num.	0.0230***	0.0602***	0.00935***	0.0108***
	(0.00258)	(0.00764)	(0.000615)	(0.00242)
Leader	0.0686***	0.180***	0.0243***	0.0250***
	(0.00944)	(0.0277)	(0.00225)	(0.00885)
Resp_Status_HeadOfHousehold	-0.211***	-0.576***	-0.0978***	-0.191***
	(0.0118)	(0.0347)	(0.00282)	(0.0112)
Resp_Status_SpouseOfHousehold	-0.147***	-0.477***	-0.0665***	-0.0932***
	(0.0109)	(0.0321)	(0.00260)	(0.0104)
Caste_OBC	0.0236**	0.119***	0.0121***	0.0199*
	(0.0118)	(0.0348)	(0.00282)	(0.0111)
Caste_ScheduledCaste	0.0402***	0.101**	0.0144***	0.0507***
	(0.0137)	(0.0401)	(0.00325)	(0.0128)
Caste_ScheduledTribe	0.0378**	0.126**	0.0177***	0.0468***
	(0.0182)	(0.0535)	(0.00433)	(0.0171)
MotherTongue_Malayalam	0.259	0.130	0.0495	
	(0.299)	(1.236)	(0.0713)	
MotherTongue_Marati	-0.574*	-2.058**	-0.140*	
	(0.301)	(0.882)	(0.0718)	
MotherTongue_Tamil	0.00297	-0.0672	-0.00465	-0.00575
	(0.0205)	(0.0600)	(0.00487)	(0.0194)
MotherTongue_Telugu	-0.0127	-0.0353	-0.00251	-0.0131
	(0.0106)	(0.0312)	(0.00253)	(0.0100)
Latrine_Common	0.00116	-0.0906	-0.0125	0.0263
	(0.0525)	(0.155)	(0.0125)	(0.0501)
Latrine_Owned	0.0239***	0.118***	0.00852***	-0.00122
	(0.00864)	(0.0254)	(0.00206)	(0.00814)
Electricity_YesGovernment	0.000225	0.0839*	0.00725**	0.0302**
	(0.0155)	(0.0457)	(0.00370)	(0.0146)
Electricity_YesPrivate	0.0233	0.126***	0.0128***	0.0500***
	(0.0154)	(0.0452)	(0.00366)	(0.0144)
OwnRent_GivenByGovernment	0.0476**	0.0425	0.0140**	0.0447**
	(0.0239)	(0.0702)	(0.00569)	(0.0225)
OwnRent_Leased	-0.0198	-0.224	0.0157	0.100
	(0.0770)	(0.226)	(0.0183)	(0.0762)
OwnRent_Owned	0.0254	0.0671	0.0143***	0.0279*
	(0.0173)	(0.0510)	(0.00412)	(0.0164)
OwnRent_OwnedButShared	-0.00481	-0.100	-0.0109	-0.0949**
	(0.0413)	(0.121)	(0.00983)	(0.0390)
Constant	2.618***	7.873***	1.290***	0.946***
	(0.0772)	(0.228)	(0.0184)	(0.0734)
Observations	16,912	16,803	16,912	15,871
R-squared	0.263	0.241	0.451	0.101

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 3: Variance Inflation Factors and their Inverse Computed for Variables to Check Multicollinearity

Variable	Variance Inflation Factor (VIF)	1/VIF
Closeness Centrality	4.34	0.230638
Betweenness Centrality	2.34	0.427591
BorrowMoney	11.27	0.088710
GiveAdvice	11.59	0.086246
HelpDecision	9.46	0.105671
KeroRiceCome	9.40	0.106425
NonRelated	9.24	0.108211
Related	5.49	0.182029
TempleCompany	10.99	0.090952
VisitCome	10.97	0.091164
Resp_Gender	4.09	0.244720
Age	1.92	0.519995
Education	1.67	0.597585
VillageNative	1.97	0.507530
WorkFlag	1.41	0.710028
ShgParticipate	1.86	0.537432
Savings	1.55	0.644546
ElectionCard	1.37	0.730770
RationCard	1.13	0.882571
RoofType1	1.54	0.647714
RoofType2	6.52	0.153272
RoofType3	5.93	0.168588
RoofType4	4.48	0.223454
RoofType5	4.07	0.245886

Room_no	1.84	0.543226
Bed_no	1.71	0.584002
Leader	1.11	0.902395
Resp_Status_HeadOfHousehold	3.55	0.282050
Resp_Status_SpouseOfHeadOfHousehold	2.85	0.350538
Caste_Obc	3.32	0.301317
Caste_ScheduledCaste	3.39	0.295163
Caste_ScheduledTribe	1.81	0.552260
MotherTongue_Tamil	1.54	0.650449
MotherTongue_Telugu	1.57	0.637164
Latrine_Common	1.04	0.962640
Latrine_Owned	1.51	0.664204
Electricity_YesGovernment	4.70	0.212582
Electricity_YesPrivate	5.09	0.196295
OwnRent_GivenByGovernment	2.23	0.448679
OwnRent_Leased	1.07	0.935288
OwnRent_Owned	2.44	0.409960
OwnRent_OwnedButShared	1.25	0.800730
MEAN VIF	2.88	

References

1. Dunbar, R. I. M. "Neocortex Size as a Constraint on Group Size in Primates". *Journal of Human Evolution* 22, 469–493 (1992).
2. Li N., Feng X., Ji S., Xu K. "Modeling Relationship Strength for Link Prediction". *Lecture Notes in Computer Science*, vol 8039, 62–74 (2013).
3. Sá, H. R. d. & Prudêncio, R. B. C. "Supervised Link Prediction in Weighted Networks". In paper presented at the 2011 International Joint Conference on Neural Networks, San Jose, CA.
4. Linyuan, L. & Tao, Z. "Link Prediction in Weighted Networks: The Role of Weak Ties". *EPL Europhysics Letters* 89, 18001 (2010).
5. Azam, M. "A distributional analysis of social group inequality in rural India". *Journal of International Development*, 24(4), pp.415-432 (2012).
6. Noldus, R. and Van Mieghem, P. "Assortativity in complex networks". *Journal of Complex Networks*, 3(4), pp.507-542 (2015).
7. Derek L. Hansen, Scheinderman Ben, et al. *Analyzing Social Media Networks with NodeXL*. 2nd Edition. (2011).
8. "Annual Status of Education Report (Rural) 2018". Published on the 15th of January 2019. ASER Centre, India. http://www.asercentre.org/Keywords/p/346.html.
9. A. R. Desai. *Rural Sociology in India*. Popular Parkashan. 1994.
10. Prabir C. Bhattacharya. "Economic Development, Gender Inequality, and Demographic Outcomes: Evidence from India". *Population and Development Review*. Volume 32, Issue 2. 26th of June 2006.