# Strength of Weak Ties
# Evidence from Multiple Villages
# A simulation study

*Casey Breen*

*09 May, 2019*

## Introduction

The theory of weak ties puts forth that are aquaintances (weak ties) are more likely to be the the crucial ties (bridges) that connect two individuals from distinct closely-knitted social groups. In this capactity, the bridges serve as an important way for simple contagions to spread within a network. This paper investigates the strength of local bridges within a series of 74 networks of social connection in Indian villages, using two distinct definitions of a tie strength. Empirical evidence is found in support of the theory that bridges are disporportianately weak ties.

Central to Granovetter (1973) theory of weak ties is the idea that bridges are important channels for the flow of communication within networks. He claims that local bridges, unlike the strong ties within tightly-knit clusters, tend to be disproportiantely weak ties. This paper assess the claim that weak ties are bridges, finding strong evidence that bridges are disproportianately weak ties.

Past work has empirially shown that bridge are weak ties. Friedkin (1980) collected analyzed a social network of biologists, finding support for Granovetters theory. The importance of bridges is debated, but there has been considerable evidence that information diffuses through weak ties.

## Data

I analyze a set of 75 social networks collected in Southern India first collected by Banerjee et al. (2013) to assess the diffusion of microfinance products. The 75 networks, collected across 5 districts in Karnataka, are a median distance of 46 kilometers apart from their closest neighboring village (Gee et al. 2017). The data was collected within each village under the assumption that each village was a distinct system, and network data was collected within villages and not between them.

First, a household-level census was administered collecting data on characteristics of the household (roof type, access to electricity, etc.) and information on household head. The household census did not collect information on social networks. After the household-level census was completed, an individual questionnaire was administered in each village. Individual questionnaires were administered to households with a woman between ages 18-50. The individual level questionnaire was administered in all Christian and Muslim households, and Hindu househoulds were clustered by geography and then 50% of households were randomly sampled. Once elegible households were selected, the individual questionnaire was administered to the household head, the spouse of the household head, other women ages 18-50, and their spouses.

The individual questionnaire contained a module asking respondents about 12 different dimensions of social relationships:

1. Borrow money from
2. Give advice to
3. Help with a decision
4. Borrow kerosene or rice from
5. Lend kerosene or rice to
6. Lend money to
7. Obtain medical advice from
8. Engage socially with
9. Are related to
10. Go to temple with
11. Invite to one's home
12. Visit in another's home

An additional module was administered to a sample of those individuals asking about age, religion, caste, etc.

(**???**) created used the information collected on the social relationship module to create a set of 74 undirected networks, one for each village. The decision to create an undirected network was made as the authors were primarily interested in communication, TODO. So, a certain tie exists between two individuals even if only one individual reports that social relationship. Reciprocity is an important component of any relationship, and defintiely an important component of measuring tie strength, so obtaining a directed version of this dataset could be valuable.

## Methods

Two distinct definitions of tie strength were used in this analysis. The first definition of tie stength is was a measure of the count of the number of distinct social relationships between two persons, as per (**???**). If person A gave medical advice, borrowed money from, and engaged socially with person B, the strength of the tie between them $T_s = 3$. This definition of tie strength ranging from 0, where people have no shared connections, to 12, where people are connected along every social dimension. Tie strength is always reciprocal, as the network is undirected. A high proportion of the population is connected along every social dimension

Definition 1 of tie strength, while incorporating all information, is not without drawbacks. Their has been much discussion on how best to define tie strength (**???**, @Marsden, @Granovetter). While (**???**) seminal work measured tie strength as frequency of contact between two individuals, introducing his theory that tie strength is a combination of amount of time, the emotional intensity, intimacy, and reciprocal services (**???**). Additionally, he included a pertinent discussion of whether tie strength should be operate as a binary weak vs. strong or whether a continous measure of tie strength would be more appropriate. Several additional asepcts of of tie strength have been introduced, namely social distance and structure within network topology (**???**).

A second measure of tie strength was calculated to reduce redundancy within a network. The first definition, while making use of all social relationships, has two main shortcomings. The first is that all social relationships are weighted the same. It is hard to believe that "lending rice or keroscene" and "being related" make equal contributions to the strength of a tie between two individuals.

The second qualm of the first tie definition double-counting: certain highly-correlated variables are asked about twice (lend and borrow money, give advice or receieve advice), while other questions are only asked once. This leads to a measure of reciprocated borrowing/lending count twice, while being related only counting once. Further, there may be redundancy between these certain

To address these two concerns, I use k-means clustering to find the social dimensions with the most redundancy. I collected all dyads in the 75 villages, meaning there were 12 potential ties between two individuals. Then I removed all null-dyads (no connection for any social relationship between two individuals) and conducted the principal component analysis (PCA) to reduce the number of dimensions of the data. PCA was perfomed by computing the correlation matrix of the 12 social relationships within the dyad matrix and taking the eigenvectors of the correlation matrix. No normalization was perfomed as all data was already binary for either having a tie or not. The PCA dimension reduction found a very high propotion of the variance could be explained by the first principal component, hinting that having any social relationship was a strong predictor of having any other social relationship. PCA was performed prior to k-means clustering to ensure

PCA was first conducted to reduce the dimensionality, and then a sillhouette analysis was conducted to determine the optimum number of clusters for k-means clustering. Sillhouette analysis provided a graphical interpretation for which social dimensions lie well within a cluster and which are bordeline cases. Further, it allows for the selection of am "appropriate" number of clusters. In this case, the sillhouette analysis recommended 4 clusters. A k-means clustering was performed upon the principle components to lump the 12 different social relationships into following clusters were found:

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|
| Related to | Borrow money from | Visit in another's home | Borrow kerosene or rice to |
| | Lend money to | Invite to one's home | Lend kerosene or rice to |
| | Help with a decision | Engage socially with | |

| Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 |
|---|---|---|---|
| | Obtain medical advice from | | |
| | Go to temple with | | |
| | Give advice to | | |

Once the clusters were found, a measure of tie strength was created. Cluster 1 represents to family / kin, Cluster 2 represents to those an individual would discuss important matters with, Cluster 3 represents a traditional friendship, and Cluster 4 represents a casual acquaintance. I measure tie strenght as the number of distinct categories a social ties enters. For example, if an individual has a tie in Cluster 2 and a tie in Cluster 4, they would have a tie strength of 2. This definition of a tie strength is also far from perfect, as it seemingly weights being related (Cluster 1) and having serious converstations (Cluster 4), but it represents a very distinct way of measuring tie strength that promotes robustness.

To identify bridges and local bridges within the the 75 village networks, every edge was was removed and the distance between its two endpoints was calculated. A local bridge was defined as any edge, which if deleted resulted in distance of more than 2 (the endpoints had no shared common neighbor). The span of a local bridge was defined as the total number of steps between the two endpoints if the edge was deleted. A bridge occured when removing the tie resulted in the creation of a new component.

### Results

## References

Brashears, Matthew E., and Eric Quintane. 2018. "The Weakness of Tie Strength." *Social Networks* 55 (October): 104–15. https://doi.org/10.1016/j.socnet.2018.05.010.

Gee, Laura K., Jason J. Jones, Christopher J. Fariss, Moira Burke, and James H. Fowler. 2017. "The Paradox of Weak Ties in 55 Countries." *Journal of Economic Behavior & Organization* 133 (January): 362–72. https://doi.org/10.1016/j.jebo.2016.12.004.

Granovetter, Mark S. n.d. "The Strength of Weak Ties," 22.