Strength of Weak Ties
Evidence from Multiple Villages

Casey Breen

10 May, 2019

Granovetter's seminal Strength of Weak Ties asserts that acquaintances (weak ties) are more likely to be the crucial ties (bridges) that connect two individuals from distinct, closely knit social groups. Central to Granovetter's claim is that bridges are important channels for the flow of communication within networks. Bridges and local bridges, unlike the stronger ties within tightly knit clusters, tend to be disproportionately weak ties. Granovetter outlines a process wherein a stronger tie between two people indicates a higher proportion of friends in common, and subsequently results in a lower likelihood of being a (local) bridge. Past work has empirically shown that bridges are weak ties. Friedkin (1980) collected and analyzed a social network of biologists, finding support that bridges are indeed weaker ties. This paper adds to the existing body of literature by confirming that local bridges are disproportionately weaker ties in a set 75 Indian village networks. The analysis employs two different definitions of a tie strength. Empirical evidence is found in support of the theory that bridges are disproportionately weak ties. Additionally, evidence is found that the local bridges are significantly weaker ties than bridges themselves, a seemingly paradoxical finding that the paper seeks to reconcile. Given the consistency of results across villages, I conclude that while local bridges are disproportionately significantly weaker ties within all network, bridges themselves are only slightly weaker ties than non-bridges.

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 across Indian villages. The 75 networks, spread 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 geographically far enough from other villages to be socially isolated, leading the researchers to collect network data within and not between villages.

Individuals were not sampled randomly, although it seems unlikely that the methods for data collection would bear any significant effect on the analysis. 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 only to households with a woman between the ages of 18-50. Notably, while the individual level questionnaire was administered in all Christian and Muslim households, Hindu households were clustered by geography, and then 50% of households were randomly sampled TODO cite appendix. Once eligible 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 random sample of individuals asking about age, religion, caste, etc.

(**???**) used the information collected on the social relationship module to create a set of 74 undirected networks, one for each village. Authors made the decision to create an undirected network as they were primarily interested in communication, and direction of the communication unnecessary. Of importance to this analysis is that certain ties exist between two individuals even if only one individual reports that social relationship. Reciprocity is an important component of measuring tie strength, so obtaining a directed version of this dataset could be valuable for gaining a better measure of tie strength.

Methods

The best means by which to measure tie strength (**???**, @Marsden, @Granovetter) has long been a topic of debate. While (**???**)'s seminal work measured tie strength as the frequency of contact between two individuals, he also provided a theoretical framework for best understanding tie strength: a combination of amount of time, the emotional intensity, intimacy, and services reciprocated. He also included a pertinent discussion of whether tie strength should be operationalized as a dichotomous weak vs. strong or whether a continuous measure of tie strength would be more appropriate. More recently, several additional aspects of tie strength have been introduced, namely social distance and structure within network topology (**???**).

Two distinct definitions of tie strength were used in this analysis. The first (definition 1) is was a straightforward 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 ranges from 0 (two individuals have no shared connections) to 12 (two individuals 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, as per figure 1. Definition 1 of tie strength, while incorporating all information, is not without drawbacks. While the first definition is tidy and makes use of all social relationships, it has two main shortcomings. The first is that all social relationships are weighted the same. It is hard to believe that "lending rice or kerosene" and "being related" make equal contributions to the strength of a tie between two individuals. The second issue of the definition 1 is double-counting: certain highly-correlated social relationships are both included (lend and borrow money, give advice or receive advice), while other questions are only asked once (related, engage socially with). This leads to certain measures being overrepresented in terms of their contribution to tie strength.

In attempts to address these concerns, I use k-means clustering to find the social dimensions with the most redundancy. I collected all dyads in the 75 villages, keeping track of all 12 potential ties between two individuals. Then I removed all null-dyads and conducted the principal component analysis (PCA) to reduce the dimensionality of the data. PCA was performed by computing the correlation matrix of the 12 social relationships for the 294,626 dyads and taking the eigenvectors of the correlation matrix. No normalization was performed as all data was already binary; a certain social tie existed or it did not. The PCA dimension reduction found a very high proportion 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 first conducted to reduce the dimensionality, a silhouette analysis was conducted to determine the optimum number of clusters for k-means clustering. Silhouette analysis provided a graphical interpretation for which social dimensions lie well within a cluster and which are borderline cases. Further, it allows for the selection of appropriate number of clusters. In this case, the silhouette analysis recommended 4 clusters. A k-means clustering was performed upon the principle components to lump the 12 different social relationships. The following clusters were found:

Table 1: Four clusters selected from a K-Means clustering.

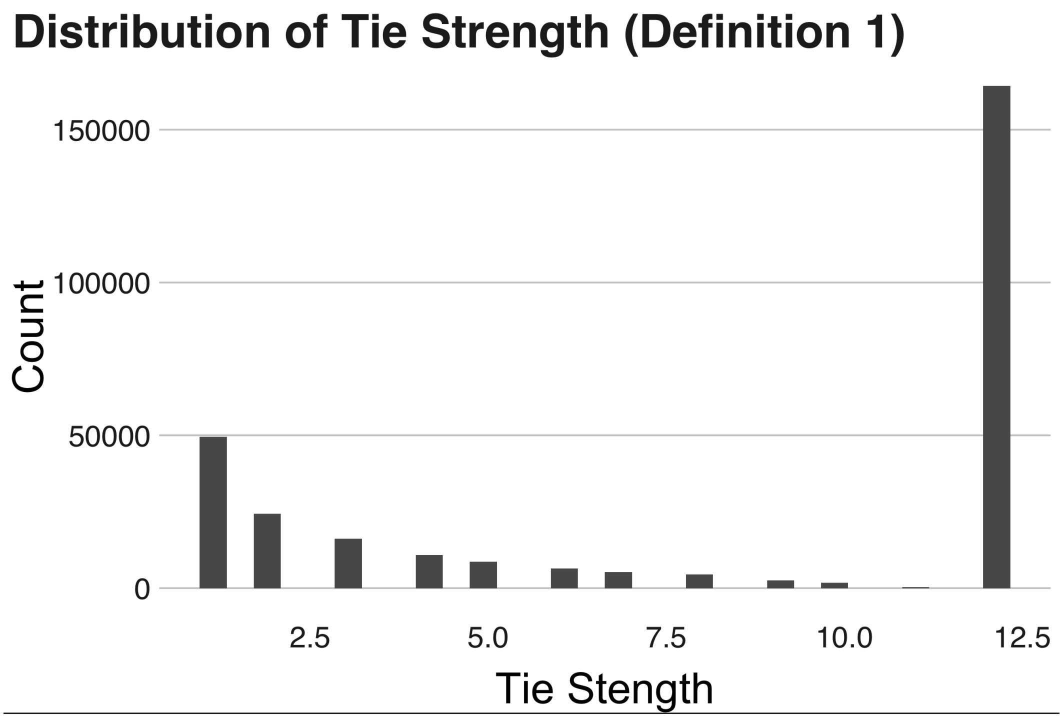
Cluster 1	Cluster 2	Cluster 3	Cluster 4
Related to	Borrow money from	Visit in another's home	Borrow kerosene or rice to
	Lend money to	Invite to one's home	Lend kerosene or rice to
	Help with a decision	Engage socially with	
	Obtain medical advice from		
	Go to temple with		
	Give advice to		

Within this graph, 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 measured tie strength as the number of distinct categories a social tie 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 far from perfect, as both being related (Cluster 1) and being casually acquainted (Cluster 4) contribute equally towards this definition of tie strength. While a fair critique, I argue that this definition actually is more robust – if someone is related to someone else yet won't do a simple favor for them, perhaps that tie is indeed weak.

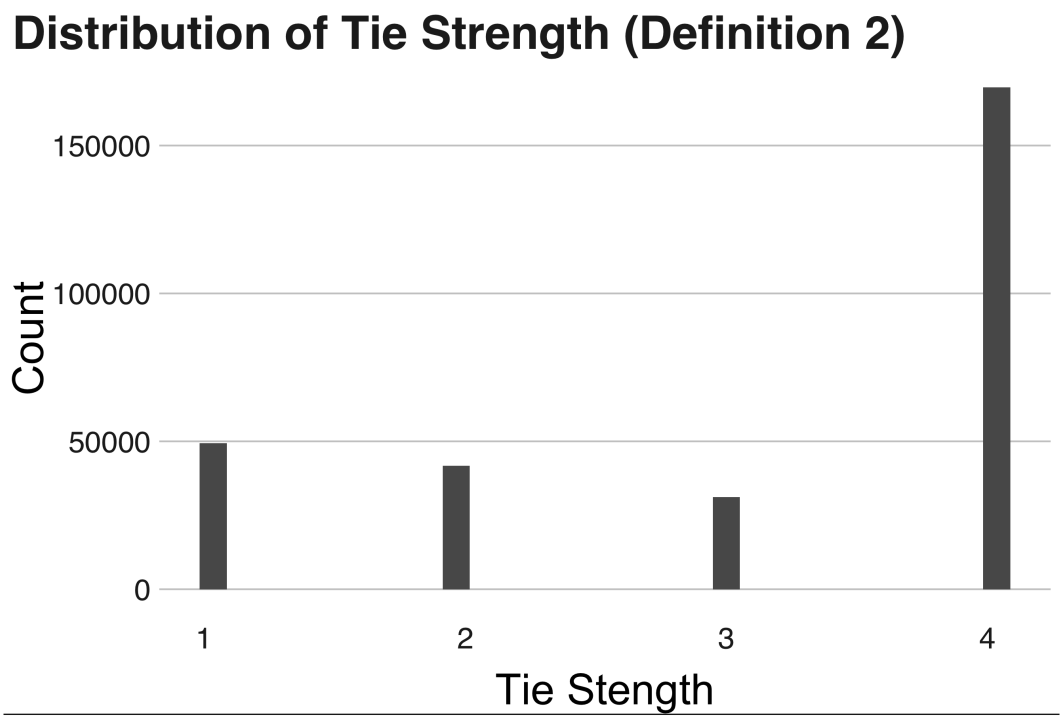
To identify bridges and local bridges within the the 75 village networks, every edge was removed and the distance between its two endpoints was calculated. A local bridge was defined as any edge, which if deleted, resulted in a distance of more than 2 (the endpoints had no shared common neighbor). The span of a local bridge was defined as the total number of steps between the two endpoints if the edge was deleted. A bridge occurred when removing a tie resulted in the creation of a new component. Out of the 294,626 total edges, I found 24,222 local bridges and 1,143 bridges.

Results

In the entire network, the mean of the tie strength (definition 1) of the 294,626 nodes was 8.01 (95% confidence interval 7.99-8.03), with a median tie strength of 12, indicating a high proportion of individuals connected on every social dimension. As expected, for ties that were neither bridges nor local bridges were the strongest, with a mean tie strength of 8.48 with a standard error of 0.0089. Tie strength greatly decreased for local bridges, with a mean tie strength of 2.805 (95% CI: 2.74-2.87). Tie strength varies slightly within local bridges of span 3, 4, and 5. It appears that the tie stength slighly increases for local bridges of span 6 and 7, but the sample size are too small to say definitely whether these increases are significant.

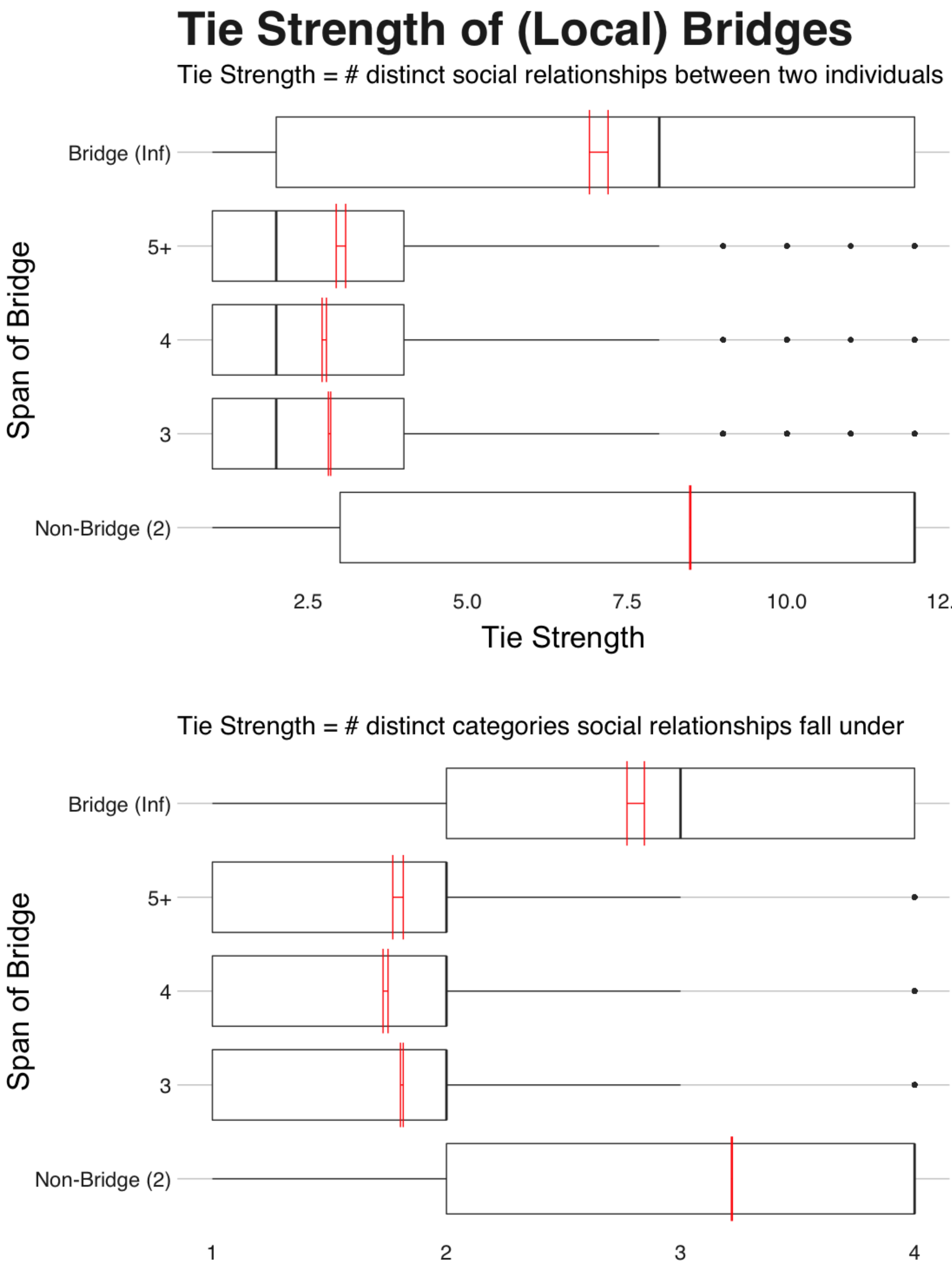


(a)



(b)

Figure 1: Panel (a) shows the distribution of tie strength using the definition 1, a 1-12 scale. Panel (b) shows the distribution of tie strength using definition 2, a 1-4 scale. In both definitions of tie strength, a majority of ties are strong ties.



(a)

Figure 2: Average tie strength of (local) bridges by span, using definition of tie strength 1 (top) and definition of tie strength 2 (below). Tie span is lowest for local bridges and then increases dramatically for bridges. Box plot shows 25 quantile, median, and 75th quantile. Red bars indicate standard errors centered around a mean.

Table 2: Count, mean tie strength, and standard errors for definition 1 and definition 2

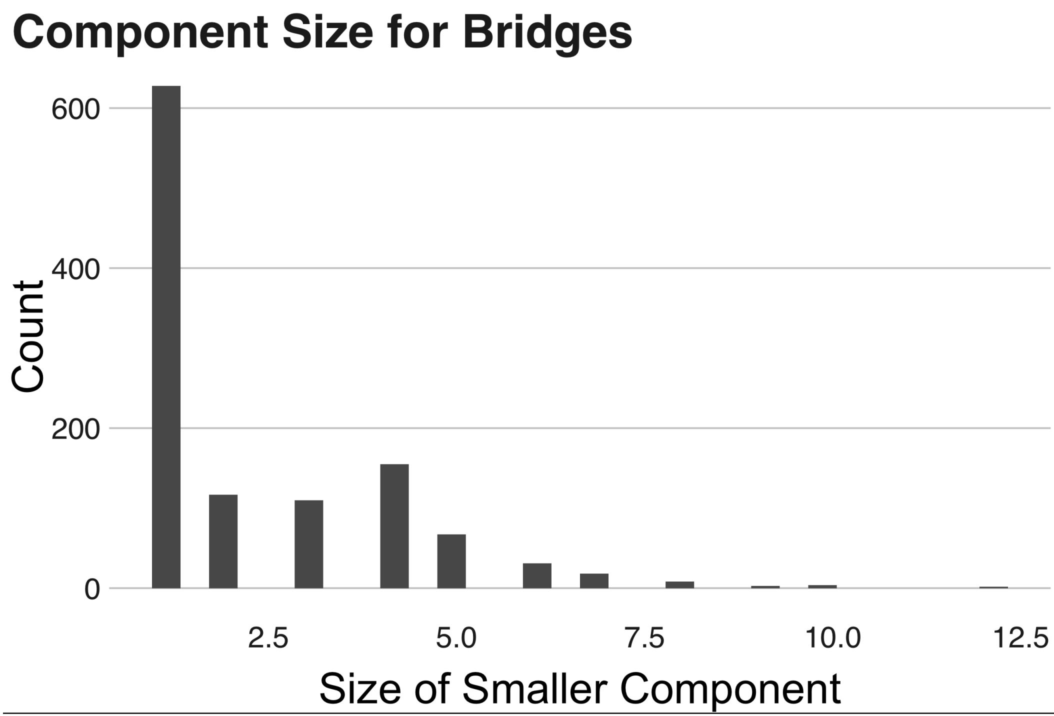
Bridge Type	Count	Mean Tie Strength Def. 1 (max 12)	Mean Tie Strength Def. 2 (max 4)	SE Def. 1	SE Def. 2
Non-Bridge, Span 2	269261	8.48	3.219	0.0089	0.0022
Local Bridge, Span 3	18056	2.83	1.809	0.0177	0.0059
Local Bridge, Span 4	4921	2.75	1.739	0.0338	0.0106
Local Bridge, Span 5	1111	2.94	1.772	0.0768	0.0234
Local Bridge, Span 6	117	3.52	1.945	0.2691	0.0769
Local Bridge, Span 7	17	3.70	2.050	0.8169	0.2233
Bridge	1143	7.05	2.808	0.1464	0.0371

The mean tie strength of bridges is significantly higher than the tie strength of local bridge ($p < 0.01$). This runs counter to Granovetter's strength of weak ties theory. Intuitively, bridges should be even weaker ties, on average, than local bridges, as both endpoints share no friends, or friends of friends.

I hypothesize three reasons for the counter intuitive finding: 1) Bridges are disproportiantely stronger ties are they are predominantly ways for members socially isolated from the group to be connected to the giant component and 2) Bridges over time become stronger ties as they control the flow of information between these two groups 3) Migration creates bridges temporarily before they are more broadly integrated into society. Hypothesis 2 and hypothesis 3 cannot be answered with current cross-sectional data. While migration data was collected for a subsample of idnviduals (~20%), there is no way to asses whether migration

Hypothesis 1: *Bridges are often formed between a smaller, social isolated component and the giant component of a given social network. These ties must be stronger ties as they are important channels for the flow of information, norms, and behavior between a smaller, otherwise isolated community and the larger giant component of the network.* '

For each bridge, by definition the endpoints of the edge belong to two seperate components if that edge is removed. I hypothesized that on of these endpoints would fall within a giant component of a network, and one of these endpoints would fall within a much smaller component. In other words, if the tie between two individuals is a bridge, one individual will be connected to the giant component



(a)

Figure 3: The distribution of the smaller components for the endpoints of a bridge.

and the other will belong to a small socially isolated cluster.

Figure 3 shows the distribution of the smaller of the two components formed when a bridge is removed, and the maximum size of the smaller of the two components is 12, confirms that all bridges have a smaller socially isolated component. To assess whether bridges generally between a giant component and a socially isolate group, I categorized each bridge as having an endpoint in the giant component if the bridge was deleted, one component had a component of size 341 or great (the smallest number of nodes in one village's giant component was 342). Table 3 shows the count of bridges between an individual in the giant component and an individual within a social cluster. The majority of bridges (85.6%) occur between an individual in an otherwise socially isolated component and an individual in the giant component. Bridges between individuals in socially isolated groups occur generally exclusively between those related to one another, and are always strong ties. The bridges between the giant component and socially isolated component are still significantly stronger ties on average than local bridges, connected on an average of 6.21 social relationships compared to only 2.81 social relationships for local bridges

Table 3: Bridge Tie Strength of the Giant Component

One Endpoint in Giant Component	Tie Strength	n
Yes	12	167

One Endpoint in Giant Component	Tie Strength	n
No	6.21	976

Discussion

The results displayed table 2 are broadly consistent with Granovetter's Strength of Weak ties theory, with local bridges exhibiting disporportiantely weaker tie strength across both tie strengths. While both measures of tie strength were constructed quite differently, results are consistent. Evidence was also found that, while not necessarily in direct contrast with Granovetter's theory, did not support it. Slight variation was found within tie strength of span, but sample sizes were too small to find statisitically significant evidence. Bridges themselves were significantly stronger ties than local bridges, which ran counter to Granovetter's theory.

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.