An Empirical Test of the "Strength of Weak Ties"

Casey Breen

10 May, 2019

Introduction

Granovetter's (1973) Strength of Weak Ties asserts that ties between 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 (local) 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 weak ties. Granovetter's logic is that a strong tie between two individuals indicates a high proportion of friends in common, and a subsequently lower likelihood of that tie being a (local) bridge. Past work has shown empirically that bridges are disproportionately weak ties. Friedkin (1980) collected and analyzed a social network of biologists, finding support that (local) bridges are indeed weaker ties. I assessing whether bridges are weak ties in using detailed network data collected in 75 villages in India.

Using data from the social networks of 75 Indian villages, I construct two different measures of tie strength using individual-level reports of 12 different social relationships. I find that bridges and local bridges are disproportionately weak ties. Furthermore, local bridges are significantly weaker ties than bridges themselves, a seemingly paradoxical finding. Given the consistency of results across villages, I conclude local bridges are significantly weaker ties than non-bridges; bridges themselves are only slightly weaker ties than non-bridges. To reconcile this seemingly contradictory finding, I theorize that bridges often connect a giant component of a village to a smaller, socially-isolated component, and the importance of this bridge in facilitating the spread of information leads to the tie becoming stronger over time.

Data

I analyze a set of 75 social networks in southern India originally collected by Banerjee et al. (2013) to assess the diffusion of microfinance across Indian villages. The 75 villages, 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 under the assumption that each village was geographically far enough from other villages to be socially independent, leading the researchers to collect network data within and not between villages. A local bridge was defined as any edge, which if deleted, resulted in a distance of more than 2 between the endpoints of that edge (the endpoints had no shared common neighbor). A bridge was defined as any edge, which if removed, resulted in the creation of a new component.

Banerjee et al. (2013) used a multistage sampling process. 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 (Banerjee et al. 2013). 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.

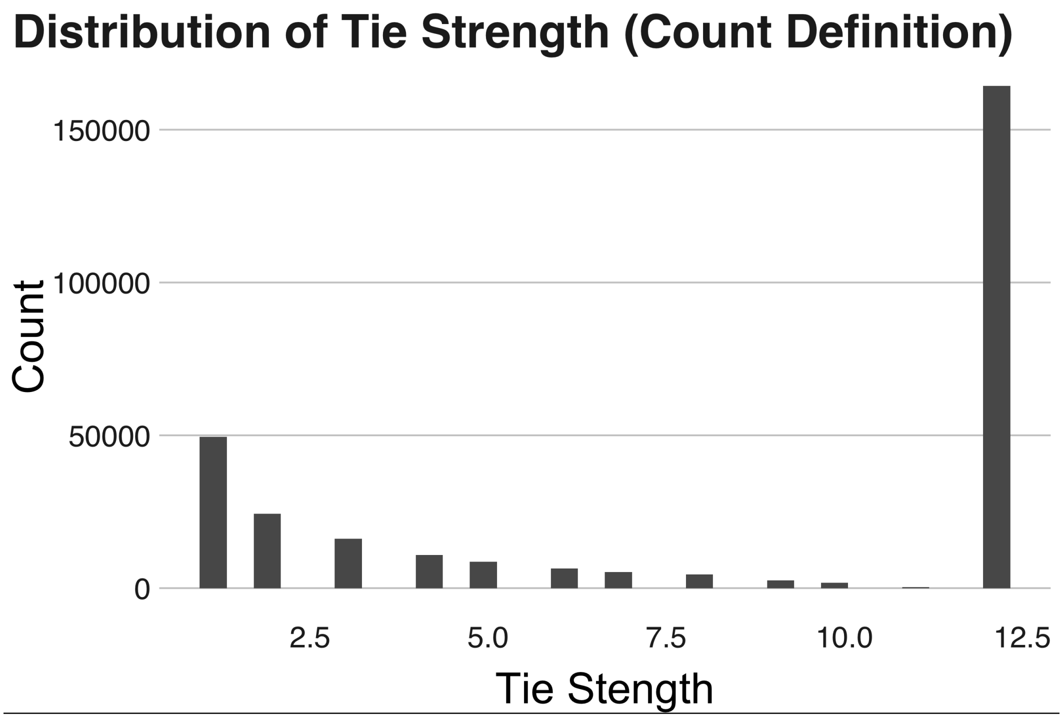
Banerjee et al. (2013) used the information collected on the social relationship module to create a set of 75 undirected networks, one for each village. While the original data collected contained information about direction, the authors made the decision to remove directionality from the networks. Reciprocity is an important component of measuring tie strength, so obtaining a directed version of this dataset in the future could be valuable for constructing a more accurate measure of tie strength.

Methods

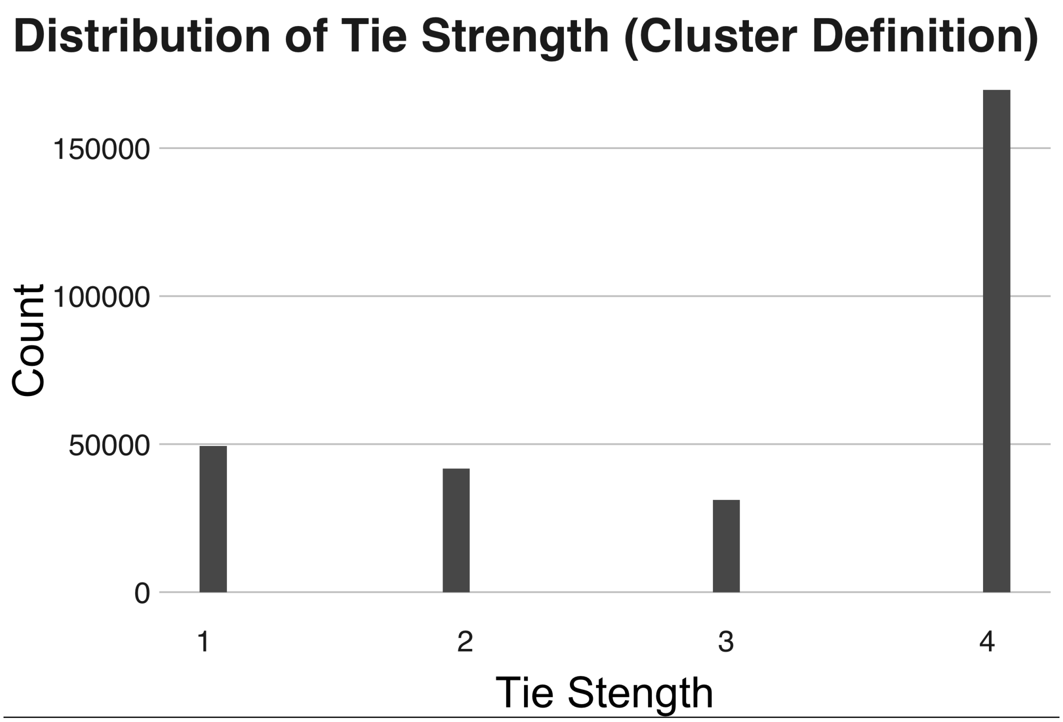
The best way to measure tie strength is widely debated (Gilbert and Karahalios 2009; Marsden and Campbell 1984; Granovetter 1973). While Granovetter (1973) measured tie strength as the frequency of contact between two individuals, he also provided a theoretical framework for best understanding tie strength: a combination of amount of time, the emotional intensity, intimacy, and services reciprocated. He also included discussion of whether tie strength should be measured as dichotomous (weak vs. strong), or as a continuous measure, concluding that dichotomous measures ignored several important issues involved with the creation tie strength (Granovetter 1973). More recently, several additional aspects of tie strength have been introduced, namely social distance and structure within network topology (Lin, Vaughn, and Ensel 1981; DiPrete et al. 2011).

Two distinct definitions of tie strength are used in this analysis. The first, the count definition, is a straightforward count of the number of distinct social relationships between two persons within a given village network. This definition was used in the previous work of Mattie et al. (2018). If an individual gave medical advice to, borrowed money from, and engaged socially with another individual, the strength of the tie between them is 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. Figure 1 shows a high proportion of the population is connected along every social dimension. The count definition of tie strength is not without drawbacks. While simple 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 imagine 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 count definition is double-counting: certain pairs of 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.

To address these concerns, I use k-means clustering to group the social dimensions into related clusters in hopes of reducing the redundancy of the social relationships. I collected the 294,626 dyads across all 75 villages, keeping track of all 12 potential ties between two individuals. 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 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, which all social relationships made roughly (but not exactly) the same contribution to. I interpreted this as having any one social relationship with another individual is a strong predictor of having a second social relationship.



(a)



(b)

Figure 1: Panel (a) shows the distribution of tie strength using the count definition, a 0-12 scale. Panel (b) shows the distribution of tie strength using cluster definition, a 0-4 scale. In both definitions of tie strength, a majority of ties are strong ties.

After I performed PCA to reduce the dimensionality, I performed a silhouette analysis to determine the optimum number of clusters for k-means clustering. The silhouette analysis provided a graphical interpretation for which social dimensions lie well within a cluster and which are borderline cases (Rousseeuw 1987). Further, it allows for the selection of appropriate number of clusters. In this case, the silhouette analysis recommends 4 clusters. I performed k-means clustering upon the principle components to group the 12 different social relationships into following clusters:

Table 1: Four Clusters Selected from a K-Means Clustering.

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

My qualitative interpretation is that Cluster 1 represents to family or 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 cluster a certain dyad had a tie within. For example, if a dyad has any number of ties in Cluster 2 and any number of ties in Cluster 4, a tie strength of 2 would be assigned. I chose not to weight each each social cluster to avoid introducing bias. This definition of a tie strength is far from perfect. For example, both being related (Cluster 1) and being casually acquainted (Cluster 4) contribute equally towards this definition of tie strength. The cluster definition has a minimum of 0 and a maximum of 12.

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 that 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 resulting in the creation of a new component. Out of the 294,626 total edges, I found 24,222 local bridges and 1,143 bridges.

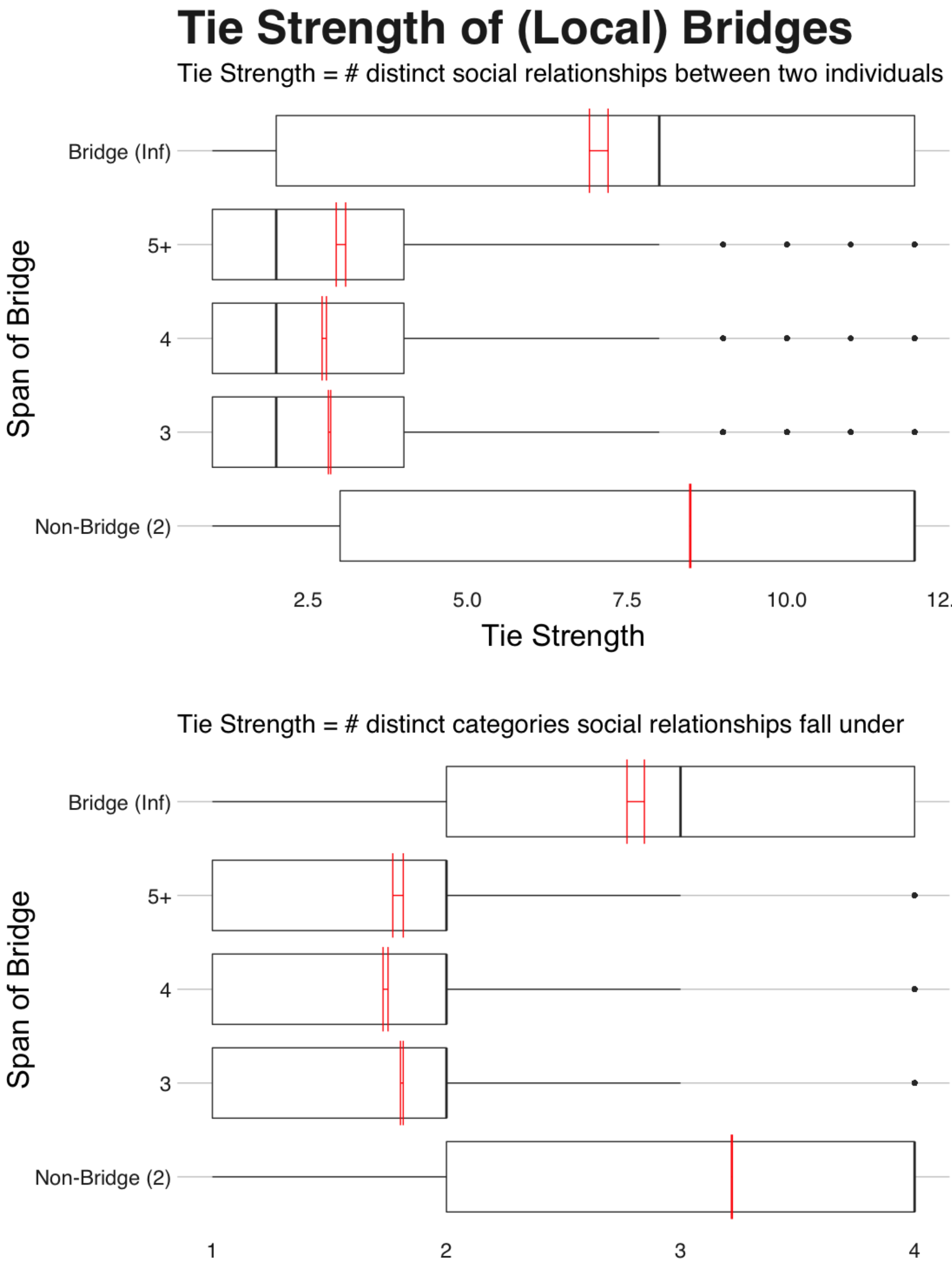
Results

Across all networks, the mean of the tie strength (count definition) of the 294,626 nodes was 8.01 (95% CI: 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 non-bridging ties were the strongest, with a mean tie strength of 8.48 (95% CI: 8.46, 8.50). Tie strength greatly decreased for local bridges, with a mean tie strength of 2.805 (95% CI: 2.74, 2.87). Tie strength is fairly consistent within local bridges of span 3, 4, and 5. Tie strength slightly increases for local bridges of span 6 and 7, but the sample size are too small to say definitively whether these increases are significant. Figure 2 shows that bridges have a significantly higher tie strength (7.05, 95% CI: 6.7572, 7.3428) than non-bridges (95% CI: 8.46, 8.50). This finding is seemingly contradictory, as intuitively, bridges should be even weaker ties on average than local bridges, as both endpoints are more socially disconnected.

Table 2: Count, Mean Tie Strength, and Standard Errors for Count Definition and Cluster Definition

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

To control for village-level effects, I ran an OLS regression with fixed effects for all 75 villages. Tie strength has been standardized to make the regression results more comparable across tie strength definitions. The dependent variable, tie strength, represents the proportion of total social relationships (count definition) or the proportion of the four clusters in which a



(a)

Figure 2: Average tie strength of (local) bridges by span, using count definition of tie strength (top) and cluster definition of tie strength (below). Tie strength is lowest for local bridges and then increases dramatically for bridges. The boxplot shows 25 quantile, median, and 75th quantile, with red bars indicating standard errors centered around a mean.

Table 3: Regression Results

	Dependent variable:			
	Tie Strength			
	Count Def.		Cluster Def.	
	(1)	(2)	(3)	(4)
Local Bridge, Span 3	-0.471^{***}	-0.468^{***}	-0.352^{***}	-0.350^{***}
	(0.003)	(0.003)	(0.002)	(0.002)
Local Bridge, Span 4	-0.478^{***}	-0.486^{***}	-0.370^{***}	-0.374^{***}
	(0.005)	(0.005)	(0.004)	(0.004)
Local Bridge, Span 5+	-0.456^{***}	-0.472^{***}	-0.357^{***}	-0.366^{***}
	(0.011)	(0.011)	(0.008)	(0.008)
Bridge	-0.119^{***}	-0.119^{***}	-0.103^{***}	-0.102^{***}
	(0.011)	(0.011)	(0.008)	(0.008)
Village Fixed Effects		***		***
Intercept	0.707^{***}	0.683^{***}	0.805^{***}	0.774^{***}
	(0.001)	(0.006)	(0.001)	(0.005)
Observations	294,626			
Note:				*p<0.1; **p<0.05; ***p<0.01

given dyad has a tie (cluster definition). Model 1 shows the effect of bridges and local bridge by span on tie strength (count definition). Model 2 includes village fixed-effects to control for village size. Model 3 shows the effect of local bridge by span and bridges on tie strength (cluster definition), and model 4 includes village fixed-effects. Village fixed-effects have a slight effect on coefficients for the dummy variables for (local) bridges, but the direction of the change is inconsistent. The same general patterns and trends are found under count and cluster definitions of tie strength.

I hypothesize three reasons for why bridges are stronger ties than local bridges: (1) Tie strength of bridges become stronger ties over time as they are the only channel for the flow of information between a smaller, socially-isolated community and the broader community (2) Bridges over time become stronger ties as they control the flow of information between these two groups, and (3) Migration creates temporary bridges when individuals first move 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 20% of individuals, there is no way to assess whether whether these bridges are temporary or permanent. I examine hypothesis 1:

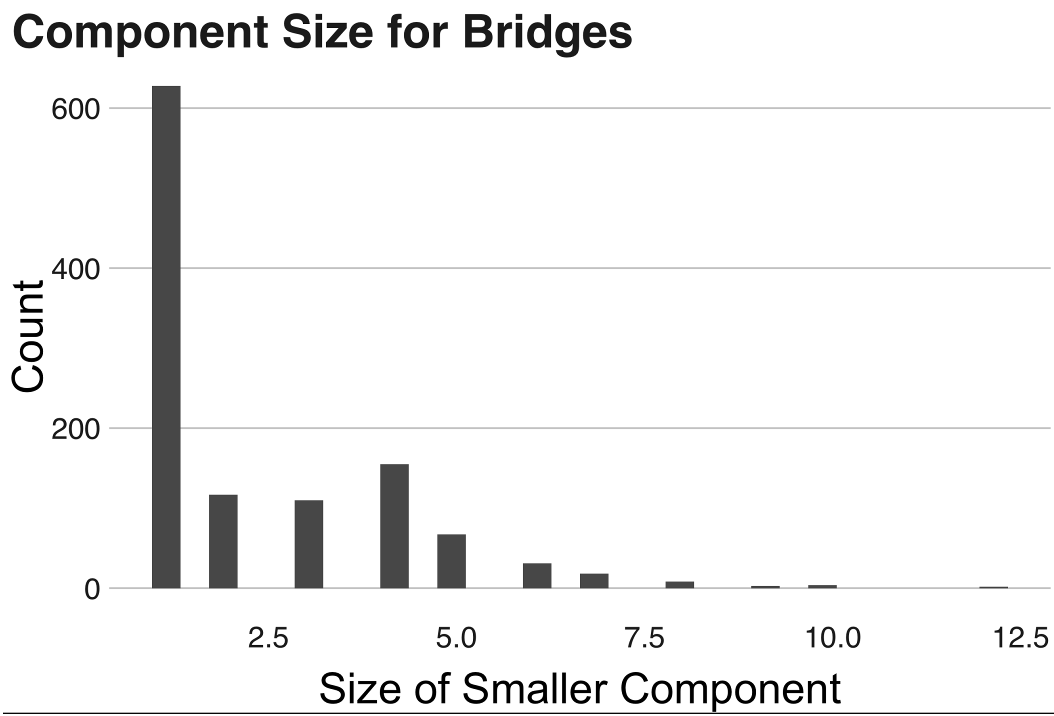
Hypothesis 1: *Bridges are often formed between a smaller, social isolated component and the giant component of a given social network. These ties must be stronger ties as they are important channels for the flow of information, norms, and behavior between a smaller, otherwise isolated community and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

For each bridge, the endpoints of the edge by definition fall into two separate components if that edge is removed. I hypothesized that one 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 belongs to the giant component and the other to a small, socially-isolated cluster.

Table 4: Tie Strength of Bridges by Endpoint of Components

One Endpoint in Giant Component	Tie Strength	n
Yes	12	167
No	6.21	976

Figure 3 shows the distribution of the smaller of the two components formed when a bridge is removed. The maximum size of the smaller of the two components is 12, which confirms that all bridges have at least one smaller, socially-isolated component. To assess whether bridges generally occur between a giant component and a socially isolate group,



(a)

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

I categorized each bridge as having an endpoint in the giant component if the bridge was deleted and one component had a component of size 342 or greater (the smallest size of any 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 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

Discussion

The results displayed in Figure 2 are consistent with Granovetter's Strength of Weak Ties theory; bridges and local bridges are disproportionately weaker ties. Two separate measures of tie strength were constructed, and the results are consistent across both count and cluster definitions. Slight variation in tie strength was found across local bridges of different spans, but bridges themselves were significantly stronger ties than local bridges. Bridges generally occurred between one individual in the giant component and another indivual in a smaller, otherwise socially-isolated component. Bridges were significantly stronger ties than local bridges. I conjecture that isolated individuals with only one tie connecting their smaller, isolated component to the giant component rely on that tie to facilitate the flow of informations, norms, and behavior from the giant component. The value of the tie itself leads to tie strength increasing, while the same is not true for local bridges, as both individuals are likely to be in the giant component and will have other pathways for receiving information and norms.

I found evidence that bridges and local bridges are weaker ties, and further evidence suggesting that bridges involving an otherwise socially-isolated component may have different characteristics than those that do not. Future work could explore whether bridges between otherwise socially-isolated components and giant components have different properties than bridges between components of equal size. The external validity of this study could be explored by looking to see whether similar results could be replicated in different social networks. Additionally, replicating this analysis on another dataset could shed light on whether the paradoxical finding that bridges are stronger ties than local bridges is universal or unique to this particular set of 75 Indian villages.

References

Banerjee, A., A. G. Chandrasekhar, E. Duflo, and M. O. Jackson. 2013. "The Diffusion of Microfinance." *Science* 341 (6144): 1236498–8. https://doi.org/10.1126/science.1236498.

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.

Coleman, James S. 1988. "Social Capital in the Creation of Human Capital." *American Journal of Sociology* 94: S95–S120. https://www.jstor.org/stable/2780243.

DiPrete, Thomas A., Andrew Gelman, Tyler McCormick, Julien Teitler, and Tian Zheng. 2011. "Segregation in Social Networks Based on Acquaintanceship and Trust." *American Journal of Sociology* 116 (4): 1234–83. https://doi.org/10.1086/659100.

Friedkin, Noah. 1980. "A Test of Structural Features of Granovetter's Strength of Weak Ties Theory." *Social Networks*, April.

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.

Gilbert, Eric, and Karrie Karahalios. 2009. "Predicting Tie Strength with Social Media." In *Proceedings of the 27th International Conference on Human Factors in Computing Systems - CHI 09*, 211. Boston, MA, USA: ACM Press. https://doi.org/10.1145/1518701.1518736.

Granovetter, M.S. 1973. "The Strength of Weak Ties." *The American Journal of Sociology* 78 (6). JSTOR: 1360–80.

Lin, Nan, John C. Vaughn, and Walter M. Ensel. 1981. "Social Resources and Occupational Status Attainment." *Social Forces* 59 (4): 1163–81. https://doi.org/10.2307/2577987.

Marsden, Peter V., and Karen E. Campbell. 1984. "Measuring Tie Strength." *Social Forces* 63 (2): 482–501. https://doi.org/10.2307/2579058.

Mattie, Heather, Kenth Engø-Monsen, Rich Ling, and Jukka-Pekka Onnela. 2018. "Understanding Tie Strength in Social Networks Using a Local 'Bow Tie' Framework." *Scientific Reports* 8 (1). https://doi.org/10.1038/s41598-018-27290-8.

Mcpherson, Miller, Lynn Smith-lovin, and Matthew E. Brashears. 2006. "Social Isolation in America: Changes in Core Discussion Networks over Two Decades." *American Sociological Review*, 353–75.

Onnela, J.-P., J. Saramaki, J. Hyvonen, G. Szabo, D. Lazer, K. Kaski, J. Kertesz, and A.-L. Barabasi. 2007. "Structure and Tie Strengths in Mobile Communication Networks." *Proceedings of the National Academy of Sciences* 104 (18): 7332–6. https://doi.org/10.1073/pnas.0610245104.

Rousseeuw, Peter J. 1987. "Silhouettes: A Graphical Aid to the Interpretation and Validation of Cluster Analysis." *Journal of Computational and Applied Mathematics* 20 (November): 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.