DATA5441: Networks and High-dimensional Inference S1 2021 Student Project

Topic 7: Weak Ties

Jimena Espinoza SID: 500052750

Maria Lim SID: 500386873

1 Background

1.1 Existing work

Onnela et al. (2007) examined a social network of calls between millions of people. They explored the consequences of removing links from higher to lower weight and from lower to higher weight and the same for the overlapping of nodes that share a link. The results are shown in Figure 1-1. A relationship between the strength in the interactions and the local structure of the network was observed. A phase transition was found to be signalled by the $\tilde{S}$ value, shown in plots C and D, when the removal links were done by both lower to higher weight and overlapping. A peak in plots C and D is seen right before the network collapsed and fell apart; this is when R_{gc}, representing the relative size of the largest connected component in the network, approaches 0. R_{gc} is defined in Equation 1. The $\tilde{S}$ value calculation is shown in Equation 2, where n_s is the number clusters of size s, and N in the total number of nodes. $\tilde{S}$ is an approximation of the susceptibility that neglects the largest connected component from the sum.

Equation 1 The relative size of the largest connected component in the network after removing a fraction f of links.

$$R_{gc}(f) = \frac{N_{gc}(f)}{N_{gc}(f=0)}$$

Equation 2 Approximated susceptibility of the network.

$$\tilde{S} = \frac{\sum_{s<s_{max}} n_s s^2}{N}$$

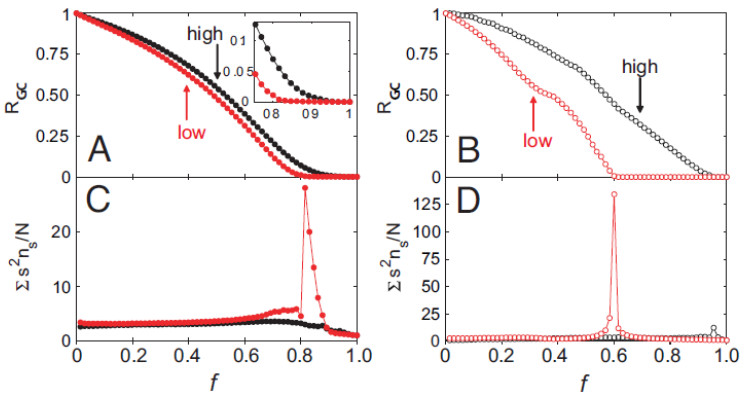

Figure 1-1 Stability of the network to link deletion. A and C show removal by weight, and B and D exhibit removal by overlapping. Removal from the highest to the lowest value is seen in black, while removal from the lowest to the highest is shown in red. A and B shows R_{gc}, while C and D are showing $\tilde{S}$.

1.2 Theory

1.2.1 Concept of 'Bridge' and Weak Ties

In the pursuit of an explanation of how removing a fraction of links leads to the shrinking of the largest component in a network, Figure 1-2 illustrates the concept of the bridge. Figure 1-2 shows a simple social network example, where A, C, D, E and F are a group of closely-knit friends, and A's acquaintance B is the other different part of the network. Edge <A, B> is a bridge as its deletion will segregate the network into two components. According to Granovetter (1973), a bridge is always a weak tie in a social network. Thus, this study ought to test this claim by removing links in different sequences based on their strength and overlap, and comparing the results.

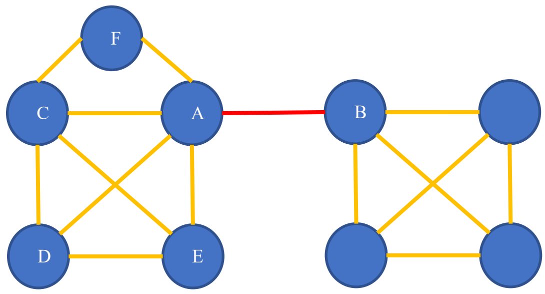

Figure 1-2 Edge <A, B> (in red) is a bridge for this simple network when its removal will separate the network into two components.

1.2.2 Neighbourhood overlap and tie strength

Considering Figure 1-2, common friends of A and F can be calculated. There are four friends of A or F, but two of them only has one friend in mutual, who is C. Thus, the overlap between A and B is ¼. On the other hand, the overlap of a bridge <A, B> is 0. Hence, based on the weak ties theory by Granovetter (1973), neighbourhood overlap increases as the tie strength increases since bridges are mostly weak edges.

2 Experiments

Two experiments were carried out using Python and various Python libraries to investigate the behaviour described in Onnela et al. (2007). The stability of networks in both experiments was analysed by using the same method proposed by the authors.

2.1 Real-world dataset

As the dataset used in the paper was mega (N = 4.6 million, L = 7 million) and not available on the internet, it was decided to find a similar nature network, aka mobile call graph from the internet. However, a weighted call network was found (Sapiezynski, Stopczynski, Lassen, & Jørgensen, 2019). Due to a small dataset after pre-processing of data applied as described in work done byOnnela et al., the initially sparse network became extremely sparse. The network was not comparable with the one in the paper, either in terms of the properties or the behaviour. We were forced to choose another type of network as there were no other decent-sized weight call graphs available.

After all, many networks were tried out, including co-citation networks, transport network, animal network, social networks, co-authorship network, scientific collaboration network, insect-plant relationship, and the food web. A social network that describes friendships among students living in a residence hall at Australian National University (ANU) (Freeman, Webster, & Kirke, 1998), which will be referred to as "residence hall network" in the remaining sections, is presented in this study as it shows similar behaviour as observed in the paper. In this network, the tie direction indicates that the resident *i* named resident *j* as a friend, and its weight represents their friendship level: 5 (best friend), 4 (close friend), 3 (friend), 2, 1.

2.2 Network model

A 'Power Law Clustering' graph was used to explore the robustness of a random graph to link deletion by ascending and descending overlapping and weight. The algorithm to create this type of random graph available in Networkx library ("Networkx generators of random graphs: PowerLaw Cluster graph," 2020) uses the Holme & Kim algorithm (Holme & Kim, 2002). The algorithm mentioned extends the scale-free network model introducing a probability P of adding a triangle after adding a random link. As a result, it generates networks with power-law degree distribution and small average shortest path length but with a high clustering coefficient. As social networks tend to have high clustering, and overlapping needs to be part of the structure of the network to observe it collapsing when deleting edges according to their overlapping from lowest to highest, the highest probability of triangle formation was used. The Powe-Law Cluster model was generated using the parameters shown in TABLE 1.

TABLE 1 Power Law Cluster graph parameters.

Number of nodes (N)	1000
Number of random edges to add (M)	2
Probability of adding a triangle (P)	1

As there is no information about weight in a generated network, and according to the Strength of Weak Ties Theory, there is a correlation between overlapping and the strength of ties, where nodes that have more common neighbours tends to have stronger ties for each link the overlapping value was assigned as its weight.

Different proportion of the link weights are shuffled to test the consistency of the results. The larger the proportion changed, the more the correlation of weights and overlapping values will be lost.

3 Results

3.1 Real-world dataset

3.1.1 Descriptive measures of network

TABLE 3-1 Summary of descriptive measures

Measure	No. of nodes, N	No. of edges, L	Network transitivity	Clustering coefficient	Average geodesic distance	Diameter	Density	Average degree of nodes	Average weight of nodes
Result	214	833	0.293	0.344	3.392	8	0.037	7.785	7.206

Only reciprocal ties in the original directed residence hall network are kept as an edge between two nodes when it was transformed to an undirected network, with the weight of the edge as a sum of weights of two directed ties. As shown in TABLE 3-1, the undirected residence hall network consists of 214 nodes and 833 edges. It is known that social networks typically have higher transitivity (as compared to a random network). Indeed, this is what is observed here, where the transitivity coefficient of the network is 0.29, which means, on average, the probability of two students that share a common friend is almost one-third. From the probability distribution of node degree and edge weight (see Figure 3-1A and B), it is observed that most of the nodes have degree less than 10, and a strong tie consists of a higher proportion. This is different from the reference network in Onnela et al. (2007), in which both distributions are exponential. Figure 3-1C shows the residence hall network generated using Gephi.

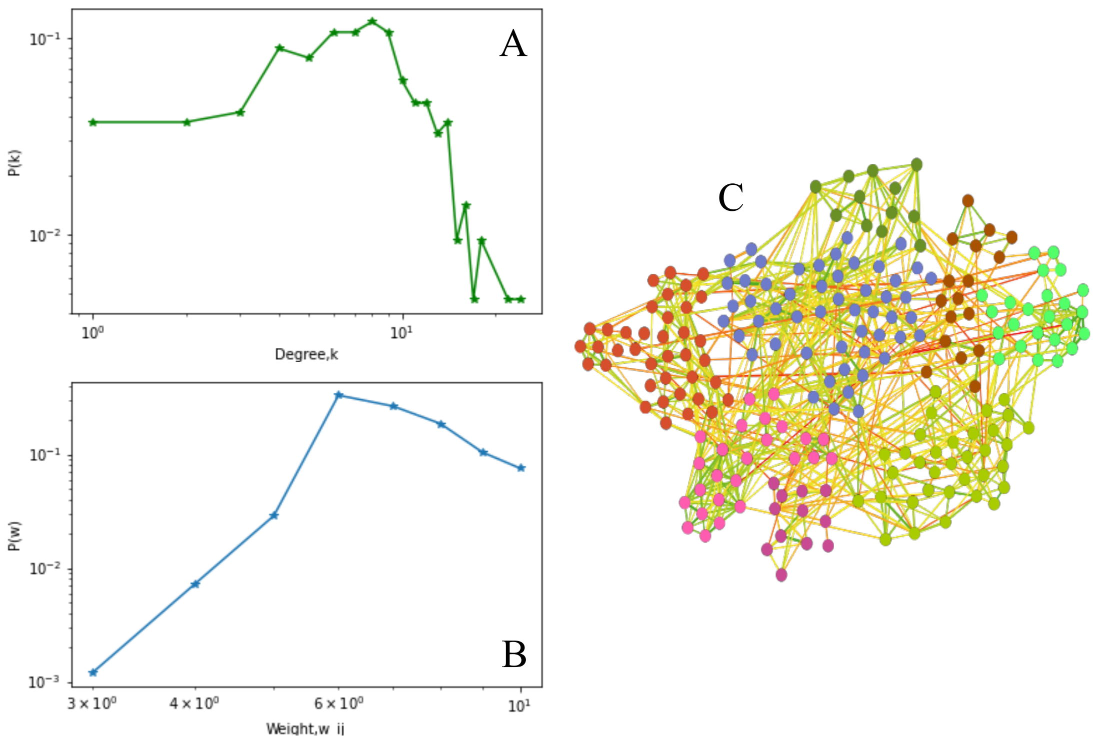

Figure 3-1 (A and B) Node degree (A) and tie strength (B) distribution. (C)Residence Hall network. Clustering method: Louvain modularity. The Colour of edges denotes the edge betweenness of centrality, continuous mapping from red(strongest) to green (weakest).

3.1.2 Stability of residence hall network

The stability of the residence hall network is tested using the methods proposed in Onnela et al. (2007). When the deletion of links is done either based on link strength (weight removal) or overlap (overlap removal), it is observed that link removal from the lowest to highest value results in a cascade of the network before all links are removed (see Figure 3-2, red trends). This behaviour is not seen in the deletion of links in the other way round (black trends), neither in the weight removal nor in the overlap removal. Critical fraction of removed link, f_c is the proportion of link being removed when the network collapsed, or when approximate susceptibility $\tilde{S}$ is peaked. It is noted that f_c is approximately 0.6 in the case of overlap removal, which is lower than that in the case of weight removal ($f_c \approx 0.8$).

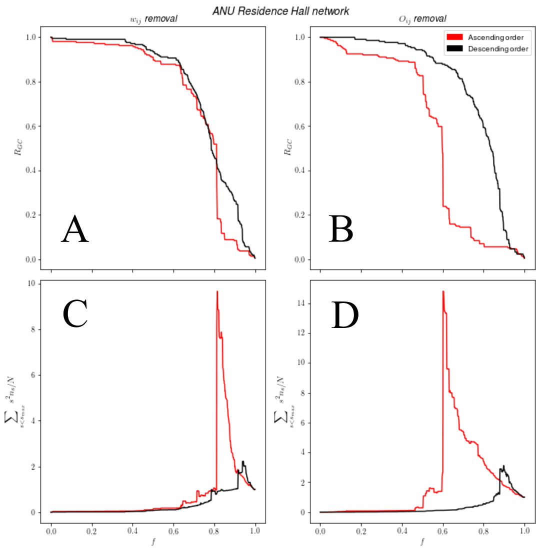

Figure 3-2 The stability of the residence hall network to link removal. Graphs A and C correspond to the case where deletion of link based on its weight, whereas graphs B and D correspond to the deletion of link based on its overlap. Red lines indicate the removal of link in the ascending order (weakest to strongest strength, or lowest to highest overlap), while black lines denote the opposite.

3.2 Network model

Starting with the generated Power Law Cluster graph, the weights and the overlapping values are initially the same. It can be noticed that when the proportion of link weights shuffled is zero, the R_{gc} and $\tilde{S}$ values are the same. Figure 3-3 in plots A1, B1, C1 and D1, is showing p=0 with the strongest red and black, confirming the behaviour is the same as weight is assigned according to the overlapping.

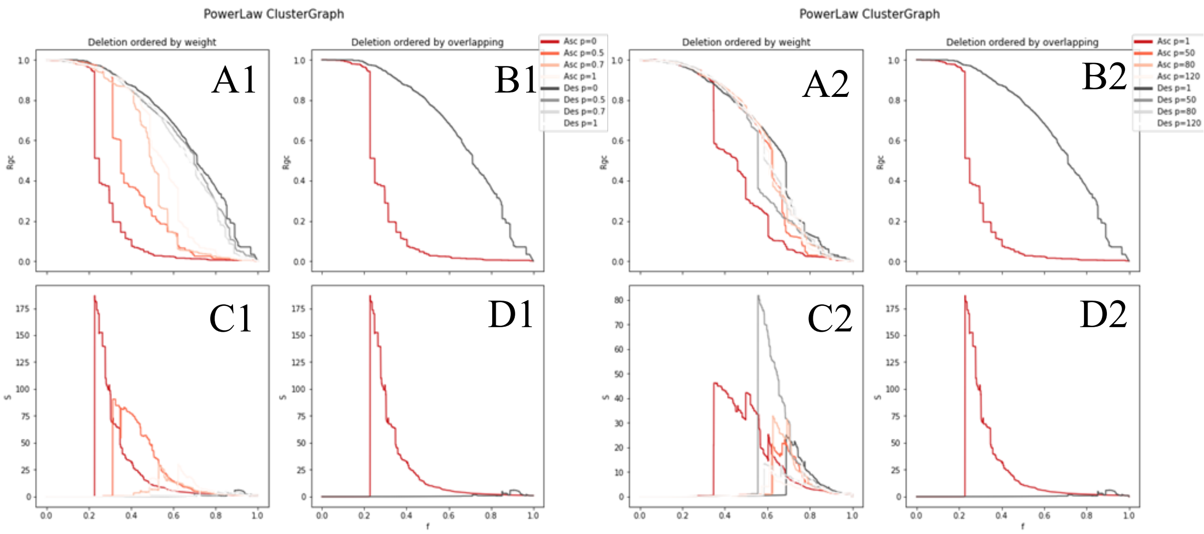

Figure 3-3 Stability of the Power-Law Cluster network to link removal, where f denotes the fraction of deleted links, R_{gc} is the relative size of the largest component in the network, defined in Equation 1 and s is the approximation $\tilde{S}$ defined in Equation 2.

The overlapping is not being modified in the experiment. Consequently, B1 and B2, as well as D1 and D2, are showing the same.

As the proportion of shuffled weight grows, the link weight starts to lose its correlation with the overlapping, exhibiting different patterns of R_{gc} and $\tilde{S}$ over *f*. It is noticed that when a higher proportion of weights are shuffled, the beginning of the collapse of the network signalled by the peak in $\tilde{S}$ is delayed towards a higher *f*, meaning the critical point f_c is higher when links are deleted by ascending weight.

Moreover, the R_{gc} trends for removal by weight in ascending and descending order get closer to each other, as shown in A1 and A2. Furthermore, descending order trend overcomes ascending order in C2 when p = 80, showing a higher peak for removal by descending weight order first. However, f_c for ascending weight order is always lower than for descending order, even with high values of p.

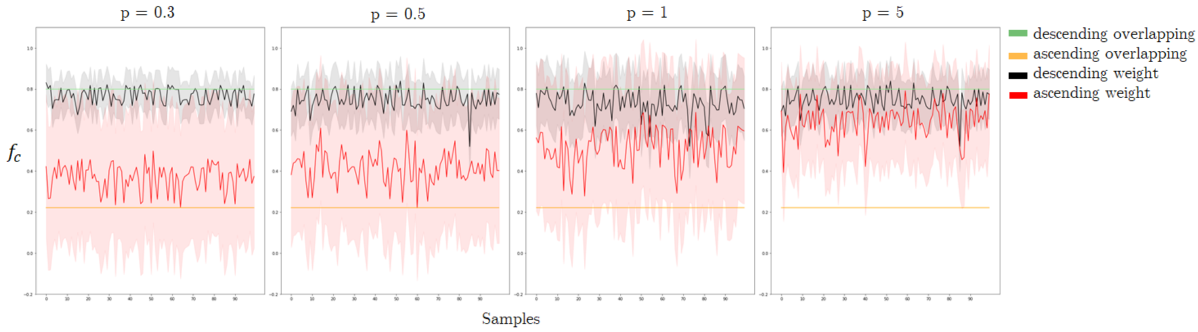

Figure 3-4 Critical fraction deletion measured in a Power-Law Cluster Network when links are removed by overlapping and by weight in descending and ascending order, with different proportion of weights shuffled 'p'.

Figure 3-4 shows the result for 100 samples using the same initial network and shuffling a different proportion of p. Scipy.signal library ("Signal processing (scipy.signal)," 2021) was used to detect peaks higher than 50% of the maximum $\tilde{S}$ value. It is noticed that the f_c value for removal by descending weight order is fluctuating close to f_c=0.8 for removal by overlapping order. When links are removed by ascending order of weights, the f_c is higher than f_c for removal by ascending overlapping with p = 0.3. With higher values of p, f_c for removal by ascending weight starts to get closer to f_c for removal by descending weight. This tendency is seen for all 100 samples, with some fluctuation. As the weights resemble less the overlapping distribution over the links, the collapse that was seen in the beginning (with weights equal to overlapping) is diluted.

4 Discussions and conclusion

There are a few points to outline the similarity in phenomena observed between the first experiment and Onnela et al. (2007). First, they are both social networks. Although the previous is a friendship network while the latter a communication network, the strength of ties in both networks represents the closeness of the relationship between two nodes. The way the weights of directed reciprocal ties aggregated to form an undirected weighted edge increase their similarity. Then, although the residence hall network is far smaller than the mobile call graph (MCG) in the paper, both cover a 'small world' – the previous consists of a relationship network enclosed in a residence hall, and the latter encompasses an extensive call network of individuals. This is also why the smaller MCG found could not replicate the results seen in the paper, as the coverage is limited, and the network is far away from "whole". Lastly, it is observed that not all real-world social networks exhibit the same behaviour.

Taking about limitations of this study, in the second experiment, creating weights initially equals overlapping starts with the assumption that they are correlated. Testing further this assumption is needed to obtain more generalisable conclusions.

As for future work, sensitivity analysis to the random graph model used will be needed to extend the conclusions. The parameters used (number of nodes, number of random edges to add for each new anode and the probability of adding a triangle after adding a random edge) to generate a network needs to be tested to see if there is any boundary set of conditions in which we will not see this phenomenon being reproduced. For example, performing a sensibility analysis on the probability of adding a triangle after creating an edge will give us information about the interval of clustering that Power-Law networks need to reproduce the network's fragility after a fraction of weak ties is removed.

References

Freeman, L. C., Webster, C. M., & Kirke, D. M. (1998). Exploring social structure using dynamic three-dimensional color images. *Social Networks, 20*(2), 109-118. doi:https://doi.org/10.1016/S0378-8733(97)00016-6

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology, 78*(6), 1360-1380. Retrieved from http://www.jstor.org/stable/2776392

Holme, P., & Kim, B. J. (2002). Growing scale-free networks with tunable clustering. *Physical Review E, 65*(2), 026107. doi:10.1103/PhysRevE.65.026107

Networkx generators of random graphs: PowerLaw Cluster graph. (2020). Retrieved from https://networkx.org/documentation/stable/reference/generated/networkx.generators.random_graphs.powerlaw_cluster_graph.html

Onnela, J.-P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., . . . Barabási, A.-L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences, 104*(18), 7332-7336. doi:10.1073/pnas.0610245104

Signal processing (scipy.signal). (2021). Retrieved from https://docs.scipy.org/doc/scipy/reference/signal.html