Dynamic centrality in complex networks

Final project for Introduction to Complex Systems

Hira Ali

December 5, 2021

1 Introduction

The analysis of dynamical networks is usually based on aggregate networks over the entire period of observation. Braha *et.al.* [1], suggested that a good strategy will be to monitor a dynamical network over a small period of time in order to understand the topology of the network more efficiently. This short report is a review of the above article, where they explored a large email network in the context of centrality and how it differs for dynamical networks when represented with small time scale networks. In the coming sections, I will briefly describe the data used by the authors and the one I used to reproduce their results. Then, I will review their analysis and results.

2 Setup

2.1 Data set used by Braha et.al.

The data for this study has been collected from 57,158 email users on campus over a period of 113 days. Every day, the network starts with zero edges, and the edge is established between any pair of vertex (email users) whenever they exchange an email. The data was then cleaned from spam emails to consider emails only that represent the flow of valuable information. This process was repeated for 113 days, during which a total of 447,543 emails were exchanged between the nodes.

2.2 Data used by the author

The data [2] I used consists of 10 source computers which over time connect to a set of 2005 computers. Data include info like, on the day D computer i was connected to n computers from the set, where i goes from 0 to 9, D goes from 0 to 90 and n goes from 0 to 2005.

For the analysis of the above described dynamical networks, the paper considered centrality and correlation to be the center of attention. Correlation is basically the measure of how two random variables are similar to each other. While the concept of centrality is central to network theory. Given a graph $G(V, E)$ with V vertices/nodes and E edges, centrality assigns ranking to nodes according to their degrees. This ranking then in turn characterizes the importance of a node within the graph. Knowing this information is helpful for further analysis in a network, for example, finding the efficient paths for information flow.

3 Analysis and results

The analysis starts with finding the correlation between the corresponding edges of the daily networks (Fig. 1a). From routine communication one can assume that the daily networks will be highly correlated, but surprisingly, the networks obtained on each day are substantially different from each other. The much smaller correlation (0.12±0.02) represents the fact that if a link has appeared between any two nodes on a specific day, does not make it more likely to be there for some other day.

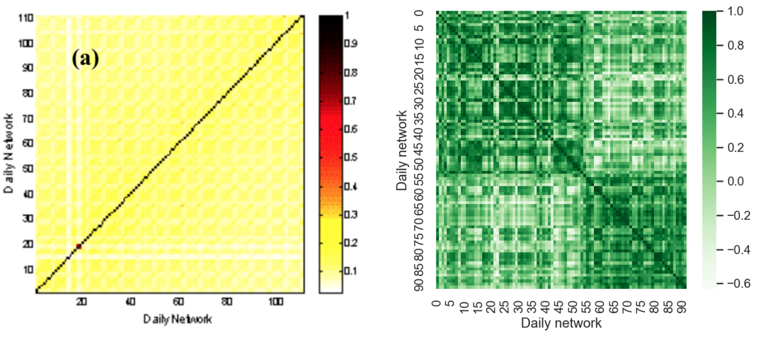

Figure 1: (a) Correlation matrix for corresponding edges of daily networks. (b) Correlation matrix for corresponding edges for computer network data.

One might think to increase the correlation by dividing the sampled data into groups. To test this hypothesis, the data is divided into two subgroups of weekends and weekdays. This did not improve the correlation, but it did show that the two subgroups are more correlated among themselves (weekends: 0.17 ± 0.03 and weekdays: 0.16 ± 0.05), than to each other. Second approach is to see if correlation increases when aggregate networks are considered with time scales from 2 to 40 days. The increase in correlation occurs at a very slow pace in aggregate networks, and never reaches a larger value even if the aggregation is done over a month or more.

The nodal degree in daily networks is well described by a power law distribution, indicating that these local hubs predominantly influence the connectivity of the network. But are these local hubs consistent over the entire period of observation, or do they fluctuate in time? This can be answered by identifying the top 1000 nodes, which makes about 1.7% of the network, from each daily network according to there degree centrality. The average centrality overlap is calculated to be (0.27 ± 0.06), which is a measure of how many nodes that appeared on the top ranking list on a certain day will appear once again on the days following.

The low centrality overlap represents the fact that very few of the nodes frequently show up on the set of top ranking 1000 nodes, the majority of them are replaced by other nodes. This can also be verified by looking at the time series of a particular hub (Fig. 2) This phenomenon is called as Temporary Fame by the paper.

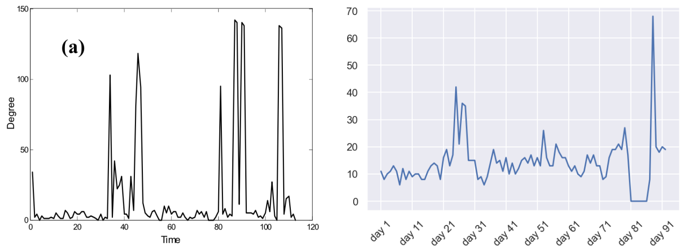

Figure 2: (a) Time series of degrees associated with nodes 724 (hub in day 34). (b) Time series of degrees associated with nodes 4 (hub in day 88)for computers data.

If we vary the percentage of nodes in the top-ranking list, the mean centrality overlap decreases to 0.2 at around 4%, before increasing slowly to 1 when the list includes all the nodes.

Finally, a significant difference was found between the daily network and the aggregate network. Nodes that appeared with high frequency in the top ranked 1000 nodes, are also ranked high in the aggregate network, but a significant number of nodes are not. Surprisingly, the nodes that are highly ranked in the aggregate network are not even on-average important in daily networks (Fig. 3). Here, the white represent those nodes which appear in both aggregate networks and daily networks while black represent only those nodes which appear in daily 1000 high centrality nodes set.

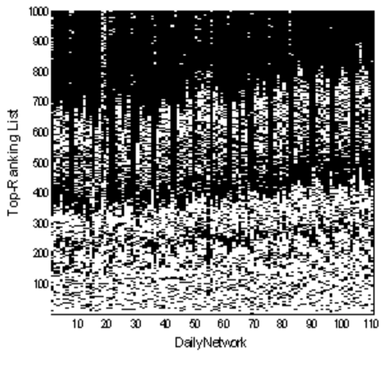

Figure 3: Comparison of the aggregate network with daily networks

4 Conclusion

The results that we just saw represent the fact that dynamic node centrality contradicts with the existing analysis in network theory, that is to disrupt a network, we need to attack high degree nodes. In the case of dynamical networks, we can see that targeting high degree nodes in aggregate networks will not be a good idea all the time. The paper suggests that a good strategy would be to monitor the nodes centrality over time.

References

[1] Dan Braha and Yaneer Bar-Yam, From centrality to temporary fame: Dynamic centrality in complex networks, *Complexity* 12: 59-36 (2006).

[2] http://statweb.stanford.edu/ sabatti/data.html