L109 Report 1: Signed Networks in Social Media

Robert Hoff

February 10, 2012

1 Analysis

Summary

Most research on social media networks has been related to the analysis of graphs that exhibit positive (indicating friendship) edges only. It has been the case that many online applications don't allow negative edges to be formed, Facebook for example only allows the creation of friendships (not foes), and one can only 'like' (not dislike) various items. Just considering positive interactions may limit our understanding of online behaviour, because compared to real-world networks we have, to varying degrees, both amiable and antagonistic relationships.

The paper considered [2] from 2010 is one of the first to consider signed large-scale datasets, these are evaluated against Heider's classic theories of balance [1], and a new proposed organising principle of *status*, suggested to be more applicable to the different nature of online social graphs. Data from three different sites are used that allow users to express both negative and positive sentiment between each other. In these cases the edges between nodes (individuals) are either positive or negative (binary), but the interpretatino of the edges of the three sites are slightly different in nature. Edges on Epinions[1] indicate user's trust or distrust related to each other's ability of providing product reviews. Slashdot[2] enables the creation of 'friendships' or 'foes' (in a style similar and opposite to Facebook), and data from Wikipedia[3] are votes between users in favour or disfavour of becoming administrators.

Balance theory can be applied between nodes where each of two people simultaneously like or dislike each other. In three individual relations, or triad, balance is observed if the multiplication of signs in the triad is positive. This behaviour stems from real-world observations that people tend to ally or disjoin based on the exposure to similar events or backgrounds. For example, if we have a population of people that support a soccer team A, and another population of people that support a team B. If any relational edge drawn from a node A to B is antagonistic and any edge drawn from within either A and B is friendly, it implies that any triad drawn from the entire population will be in balance.

In the real-world situations are not as black-and-white because people have the opportunity to have opinions and personalities for a wide range of reasons. The same individuals that support a soccer team may for example have fractured political opinions. Such fragmentations may cause negative edges within similar populations, resulting in certain triads to be in imbalance. In places where there are only a few leading factions, such as between countries at was, observed balance is high, in cases where there are more and diverse causes, and allegiance is not critical to fundamental needs, network balance is less pronounced.

In the case of the online populations discussed, an observed majority of triad type T_3 and under-representation of T_2 supports balance theory. The other types of triads are inconclusive, these findings suggests that the relationships that are being formed between users are conceptually different from how we understand the terms friend and enemy in the real-world.

[1] http://www.epinions.com/
[2] http://slashdot.org/
[3] http://www.wikipedia.org/

The authors propose the concept of *status* that relates to a notion of skill or authority of a user in the given context of the website. In this sense a negative edge between users does not indicate hostility, but rather indicates perceived and relative skill-level. In this case imbalance between two nodes makes sense if there is an agreement between the two users of their difference in status. Observing the data, many such imbalances do exist indicating that the interpretation of the edges does not correspond well with balance theory alone.

We see that when considering the time-evolution of nodes, status theory provides better predictions than balance theory. In most of the triad configurations, there is matching correlations between status theory and observation. Consider for example triad t_1 from Figure 1; A has rated X positively, and X has rated B positively. Both balance theory and status theory predict that if an edge later forms between A and B, it should also be positive in the majority of cases, which is overwhelmingly the case in this instance (97% of cases). In the case of t_{11} where the edges are still positive, but are now reversed compared to t_1 the links formed are cyclic. Here status theory seems to get it right because the population count for these triads are lower than average, indicating that A is reluctant to form a link to B, and in the cases where a link does occur it more often confirms relative status than that of balance. For many of the other triads, such as t_2, often the observation of the status or balance principle are not that easily pronounced. In t_2 the relative status between A and B cannot be easily inferred because accumulating the values along a path of mixed signs effectively cancels out. In this case balance theory conclusively predicts that a relative abundance of negative edges will be formed.

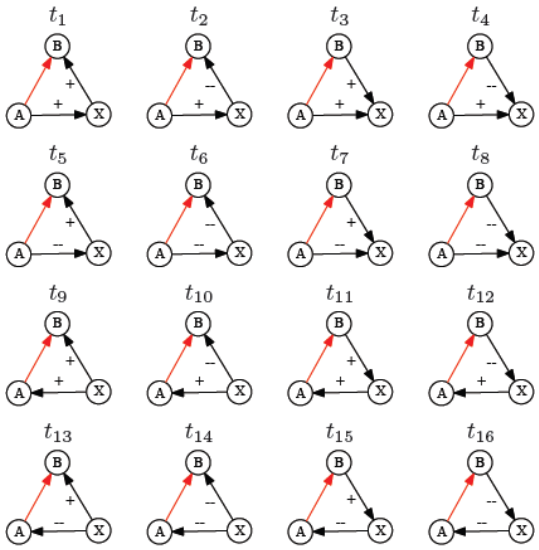

Figure 1: Courtesy Lekovec *et al* [2]; all possible triad configurations.

Evaluation

The authors clearly demonstrate that for the online social graphs considered, while subsets of the data exhibits properties from classical balance theory. There are cases, specially where asymmetric edges are being formed, that balance theory does not account for. The attempt to formulate a new theoretical model of status seems good, and confirmed by looking at the data, such as the triad t_{11} resulting in a relative abundance of negative edges. This line of thinking also seems conceptually valid, by recognising that the social graphs being formed on these online sites are contextually and environmentally dependent, and very different from real-world interactions. The social landscape in an online environment is reduced to a limited domain, and in all three websites considered are largely oriented around task-based and performance related activities. Real-world type interactions necessarily still occur because interactions are between humans, and communication between them (the nodes) is rich and complex. Because of this individuals may fall out, or gain friendships, for other reasons than the service-oriented goals of the website. We are therefore probably observing that both behavioural traits are taking place (balance and status). The variation in results may depends on the specific individuals and their interpretation of what the edge formation actually means to them personally. But in the general case the proposed theory of status provides a new dimension of exploration in the domain of signed networks.

2 Ideas for Future Research

In order to forward the general idea of *status*, the paper is convincing. But the data-analysis is quite shallow for a few reasons. First, only triads are considered, so the analysis excludes much of the network topology in the local vicinity of these nodes. There are two approaches, albeit limited, to mediate this problem. The factors of generative and receptive baseline, a characteristic distinction between node, are used in the estimation of standard deviations for the receptive and generated edges. The connectedness of nodes are also taken into account in a separate analysis to show that the positive edges is some function of, and increases, as a number of common neighbours. This makes sense because a person who is well connected has been active for longer than average and will be more influential. The second reason, I believe, where the analysis may be subject to considerable noise is additional properties that are not sufficiently considered in the underlying population. For example, results between different websites vary. Slashdot exhibits stronger properties from balance theory, while Epinions and Wikipedia show tendencies towards status. This corresponds with the advertised interpretation of what the edges should mean. Slashdot emphasises that edges formed should be personal relationships, while the other two advocate a more skill-based and professional approach towards edge creation. This observation means that the user's interpretation of what the edges are supposed to mean is very important. We can note that no matter how well a website explains to the user what the edge creation should be actually mean, there will be variation between users on the exact interpretation and methodology. Therefore a future line of research would be to perform analysis on individuals behavioural traits to determine what their conceptual model of the system is. If different user classifications are defined it would allowed more fine-grained analysis on various groups. This type of analysis would also be beneficial for the site provider to verify that the users are behaving according to the intentions of the system. Another line of research to mitigate the first issue is to explore the influence of other network variables such as clustering and centrality.

A clear understanding of the mechanics of the online environment, and the factors that are shaping people's perception is important. Another line of inquiry is the behaviour of different users depending on their conceptual model. We could forward the hypothesis that some users have an affinity for balance, while other's have an affinity for status. Separating out these groups, and determining the comparative proportions would say something about the success the online community has in conditioning its users, and would possibly produce better analytical results between groups, that may explain features of the data unaccounted for in the paper.

Finally, note that the edges from the obtained data are binary, it could possibly be interesting to consider other environments that have rank-based edges to provide the basis for richer analysis. In this case, perhaps the stronger ties would have more pronounced characteristics of the principles considered. In terms of applications, propagation of information and transactions, or the time-evolution of edges beyond triads could also be worth considering.

References

[1] Heider F. Attitudes and cognitive organization. *The Journal of Psychology*, 21:107–112, 1943.

[2] Lekovec J., Huttenlocher D., and Kleinberg J. Signed networks in social media. *Proceedings of the 28th international conference on Human factors in computing systems*, 2010.