Exploring Evolution of Signed Network Over Time

Project Final Report, CS224W

Christina Lee, Nadav Lidor, Danyang Wang

1. Introduction

In our project, we create a theoretical signed network model loosely inspired by real-life social networks with varying parameters that are dependent on the configuration of the network at a given moment. We aim to determine and analyze how different factors affect the network over time by running simulations. The question we want to answer using this project is the following: in a network where the nodes' interactions are represented by signed edges and that develops over time, what are the determining factors of the course of its evolution?

Given an underlying signed network to start with and different parameters, such as the probability of previously disconnected nodes forming a positive or negative link and the probability of an existing edge changing signs, we observe how the network evolves over time and how changing different parameters affects the evolution of the network over time. We also analyze the resulting networks' structure and compare our results against predominant theories of signed network analysis, in particular Balance Theory.

We believe that this project gives us an interesting insight to how signed networks develop over time and what are the distinguishing features of a signed network. As seen in the papers we reviewed, signed networks have many application in real life; a good theoretical understanding of how signed networks work would help us in further exploring these applications.

2. Prior Work

We reviewed three central papers in the field of signed network analysis. We examined the papers to identify the key features we may build upon for our signed network model, and so gain a sound theoretical basis for our proposal. Furthermore, as these papers examine real-world data, rather than propose a sign network model, we extracted the underlying features these scholars identified as the most prominent features of the networks. Thus, in deciding which features to incorporate in our model, we examined their finding to obtain a broader, interdisciplinary basis for our work. Following is a short overview of the research and our main intake.

"Signed Network in Social Media" [1] and "Predicting Positive and Negative Links in Online Social Networks" [2] by Leskovec, Huttenlocher, Kleinberg

In these papers, Leskovec et al. examine relations between users on social media sites as signed network. Using data from Epinions, Slashdot, and Wikipedia, they examine the signed networks focusing on triad distribution across the networks. Furthermore, they propose a linear regression model to predict "hidden" edges and edge signs, basing their model on two central data classes: (1) the (signed) degree of a node and

(2) the nodes' relationship based on a two-step paths, i.e. triangular relationships. Additionally, they incorporate psychology theories - status theory and balance theory (see Section 3.1 for details) - and find that when looking at the networks as undirected, the networks are very consistent with weak structural balance theory, and when looking at the graphs as directed, the networks are mostly consistent with status theory. Building upon this, we will compare our model's triad distribution to that of real-world networks, and examine its consistency with status and balance theories. These papers provides a strong motivation for our model - they find that examining a signed graph, rather than an all-positive (e.g. "friends" graph) enables a significantly better prediction model. Intuitively, this strengthens our hypothesis that an all-positive "friends" graph tells an incomplete story, and our "enemies" are pivotal for a more complete understanding many networks, specifically those which represent human interaction.

"Friends and Foes: Ideological Social Networking" by Brzozowski, Hogg, Szabo [3]

The paper explores people's attitude towards controversial opinions given the response of their friends, allies and nemeses. Looking at data from Essembly, a non-partisan social network that lets users post their opinions on controversial issues, the authors analyze the network present in this data by exploring characteristics of the network such as network homophily, social influences, and decision tree. This paper strengthens the motivation of signed networks, providing evidence for significantly different voting patterns and behavior given a specific friend-foe configuration.

3. Theoretical Background and Model

3.1. Status Theory and Balance Theory

Developing comprehensive models and theories to explain human relations has been a major concern of psychologists and cognitive scientists for decades. Two major theories, Status and Balance, attempt to analyze positive and negative relations, and identify the more likely structures we may expect to observe in a human-relations network. This is done in the context of a third object, analyzing what is commonly referred to as triad distribution.

Heider's Balance Theory emerged in the 1960's as a prominent cognitive model of human behavior. In an attempt to explain human interactions in an agent-based social context, Heider introduced sentiment as a central component, with two distinct classes - *like* or *positive*, and *dislike* or *negative*. In his theory, Heider hypothesizes that relations are formed and altered in an attempt to reach a balanced state of sentiment, defined to be a positive product of a given triad. In other words, unbalanced triads, i.e. with negative products, should be much less common than balanced triads. Additionally, overtime a small but significant alteration of Balance Theory emerged as the Weak Balance Theory, in which the all-enemies triad is also defined to be balanced. Below is a summary of Balance Theory and Weak Balance Theory predictions. As we are primarily concerned with undirected graphs, we only include an overview of undirected balance theory.

Edge Sign			Balance Theory	Weak Balance Theory
+	+	+	balanced	balanced
+	−	−	balanced	balanced
+	+	−	unbalanced	unbalanced
−	−	−	unbalanced	balanced

For clarity, according to Balance Theory, three friends (+ + +), or friends with a mutual enemy (− − +) are balanced triads whereas all enemies (− − −) or having sharing two friends that are enemies (+ + −) are unbalanced.

Another interesting theory of signed networks is Status Theory. In Status Theory, edges are considered to be directed, and a positive edge from A to B suggests that A considers B to be of a higher status than A. This leads to a different prediction regarding the distribution of triads. For example, Balance Theory posits that triads with two positive edges and one negative edge would be underrepresented. On the other hand, Status Theory predicts that such a triad would be very plausible: if A considers B to be of higher status, and B considers C to be a higher status, than it is likely that C considers A to be of lower status.

3.3. Our Model

3.3.1. Choosing Features

In developing the model, we had to make a number of choices regarding which features to include, based on what features we found interesting and realistic constraints.

The first choice we had to make was the initial configuration for our model. Any kind of signed graph - complete positive, complete negative, no edges, etc. - could serve as an initial configuration. In the end, we chose a signed variant of Erdos-Renyi and a signed variant of preferential attachment model, because we believe these models would give us a good distribution of positive and negative edges and various types of triads in the beginning, making for a more interesting analysis. Also, with these models, we can easily control for approximately how many edges will be there in the beginning.

Another important choice we had to make was when a new edge would be created and how edge signs would be determined. As mentioned above, our model is loosely based on social interaction in real life, and as a result, we decided to take cues from real life. We determined that two nodes with a lot of mutual positive neighbors or mutual negative neighbors would be more likely to form a positive edge, and two edges with mutual neighbors with disagreeing connections would be more likely to form a negative edge. Also, we determined that the probability of a new edge forming would depend on the number of mutual neighbors, regardless of the sign of the edge. In addition, we decided that some edges would randomly change signs with a given probability.

With this feature and no way to control edge number growth, our graph would quickly become a complete graph with exponential growth in edge numbers. We decided that this is not particularly interesting, and implemented a way to control for the number of edges. One of the ways we implemented control for edge number growth was to implement random edge dying.

A second method we used to limit edge growth was controlling how often new edges are created. For this, we considered two options. The first option we considered was implementing a ceiling on nodes' degrees. This would be very simple to implement and be an effective way to control for the maximum number of edges. However, the edge growth behavior would be a little strange in that edge number would grow exponentially up to certain point and then suddenly plateau.

Therefore, we considered different feature to achieve edge number control; in addition to edge dying, we decided that the probability of a new edge being created will be negatively correlated with the degree of the two nodes. While this feature would not guarantee a set maximum number of edges like a degree ceiling would, it provides a nice tapering off of edge number growth as opposed to the sudden plateauing of edge number growth demonstrated by a degree ceiling. We decided that this provides for a more interesting analysis, and decided to forego degree ceiling in favor of this feature.

After we have settled on these features, we had to decide when edge updates (creation/sign change/deletion) would happen. There were two obvious option we faced; update as we go, or update only at the end of an iteration. Both are interesting in their own right. For the sake of simplicity, we decided that we will update as we go.

3.3.2. Model Description

In our model, our initial network is either an undirected, signed variant of Erdos-Renyi with 1000 nodes - with probability pEdge * pPos, two nodes will have a positive edge, and with probability pEdge * (1 - pPos), two nodes will have a negative edge - or an undirected, signed variant of preferential attachment model. In this model, we also start with 1000 nodes. Then, for every pair of nodes (j, k) where j < k, with probability (1 - pEdge), we do not create an edge. Otherwise, we choose a neighbor of j, call it j', at random and create an edge (j', k).

Then, at each iteration, for each pair of nodes (say node i, node j), we look at how many mutual neighbors they have, and what the relationships to those mutual neighbors are. We update the existing edge between the pair or create a new edge based on these values:
Let n be the number of mutual neighbors, and sign_agree be the number of mutual neighbor node for which the edge_sign(node i, neighbor node) == edge_sign(node j, neighbor node).

If an edge already exists:

- With probability n / numNodes:
    - with probability pEdgeDying, the edge will be deleted
    - otherwise:

- with probability sign_agree/numNodes, turn the edge positive.
- with probability (n - sign_agree)/numNodes , turn the edge negative.
- with probabilty pEdgeSignRandomChange, the edge will switch signs

If the edge does not exists:

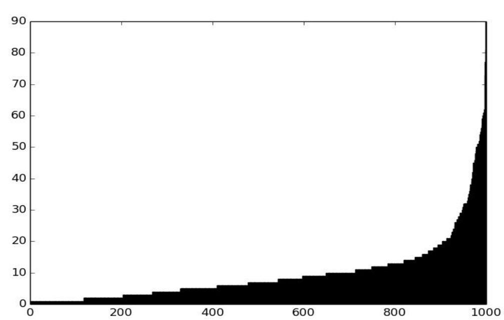
- With probability n / numNodes* (1-degree(i)/numNodes) *(1-degree(j)/numNodes):
  - with probability sign_agree/numNodes, a positive edge will be created.
  - with probability (n - sign_agree)/numNodes, a negative edge will be created.

4. Results and Findings

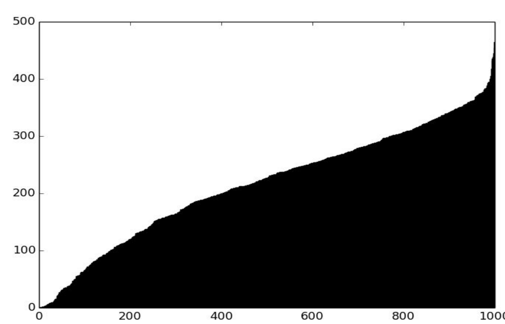
We ran our simulation over different initial graphs - particularly Erdős–Rényi (ER) and Preferential Attachment (PA), as well as different model parameters (edge death, random edge sign change, etc). Below are some of our results which we have found to be particularly interesting. Note that the graphs attached are for particular parameters, though we tested and analyzed the model over a range of values for each parameter.

4.1 Degree Distribution

Preferential Attachment (PA) graph degree distribution after initialization (left) and simulation (right)

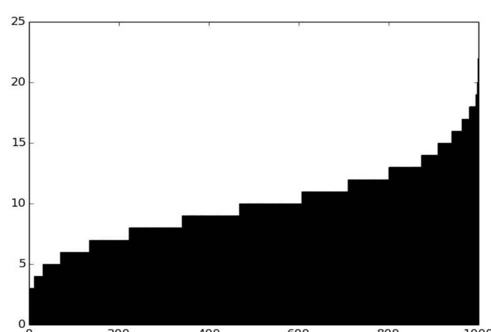


1.1

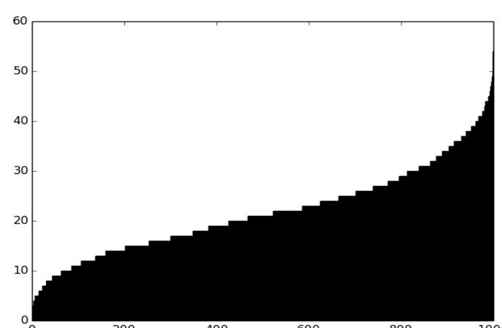


1.2

Erdős–Rényi (ER) graph degree distribution after initialization (left) and simulation (right)



1.3



1.4

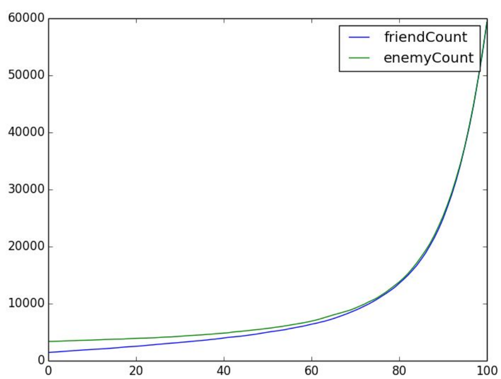
Comparing Erdős–Rényi (ER) and Preferential Attachment (PA)

While PA initially produces high degree clusters (plot 1.1) as opposed to ER which produces a more uniform degree distribution (plot 2.1), we see that over time our model produces a fairly similar degree distribution for both initial conditions (plots 1.2 and 2.2). This can be explained as our graph is becoming denser and edges are created both randomly and depending on mutual positive or negative edges.

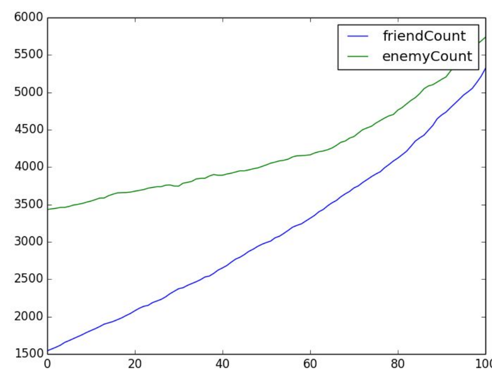
However, the behavior is vastly different when we consider *the number* of edges formed over the lifetime of our model. We observe that much more edges are created with PA as compared with ER. Examining our model, we see that having a mutual acquaintance translates to a higher probability of forming an edge between two previously disconnected nodes. Therefore, with a PA network, where we have central, high-degree nodes connecting a large portion of the graph and thus forming a very large number of triads, edges will be created much more rapidly. Intuitively, we have nodes that “know” everyone, and so introduction are much better facilitated.

4.2 Friends and Enemies Count Over Time

Friends and Enemies Count for PA (left) and ER (right) initial setup with the same set of parameters

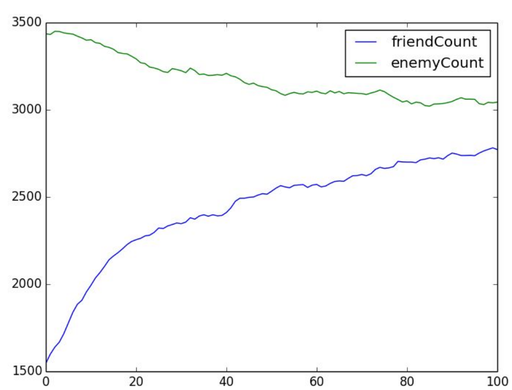


1.5

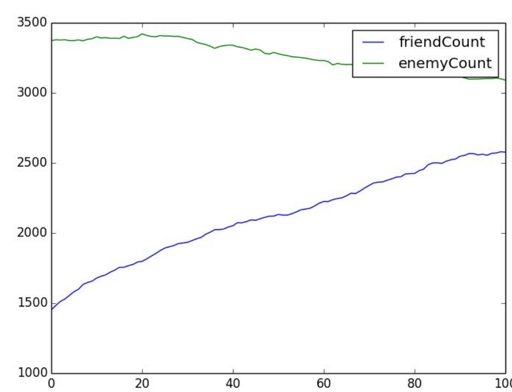


1.6

Same parameter setup with higher edge dying probability, PA(left) and ER(right)



1.7

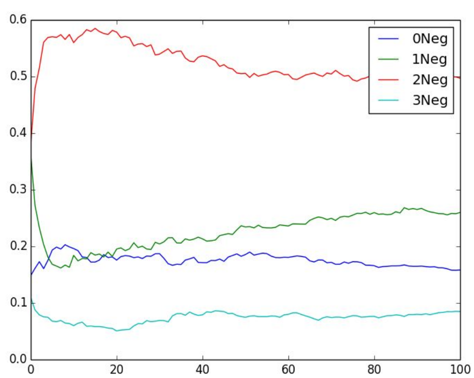


1.8

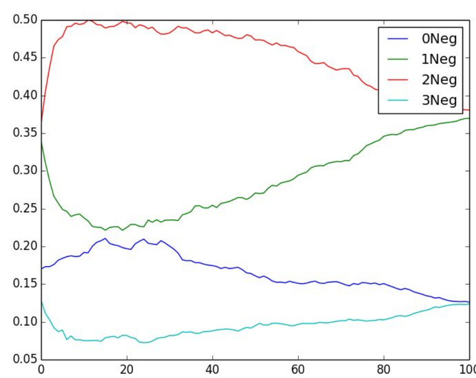
We analyzed the number of friends and enemies in our network, over time, for both ER and PA. For all tested configuration we see an exponential increase in number of positive and negative edges (plots 1.5, 1.6). Analyzing our model, this behavior is expected as the probability of creating an edge between two nodes is proportional to the number of shared neighbors. Hence, as the number of edges increases, the number of created edges increases exponentially. However when we introduce a high probability of edge death (plot 1.7), we see a more stable behavior. Another method we implemented to mitigate the exponential growth was "penalizing" the high degree nodes - that is a node with higher degree is less likely to create a new edge. However, going forward we may consider relating the probability of creating an edge to be inversely proportional to a node's degree, and so essentially cap a node's degree. This would be an interesting feature to integrate in a future model.

4.3 Triads Distribution

Triad distribution for ER (left) and PA (right) initial setup with the same set of parameters



1.9



1.10

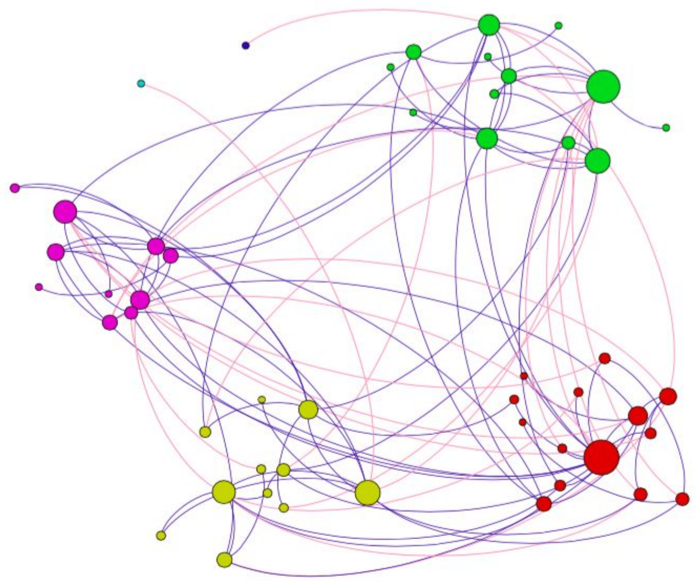
In Erdos-Renyi (plot 1.9), we see that triads with two negative edges are highly overrepresented and that triads with one negative edge as well as triads with three negative edges are underrepresented, which is consistent with Status Theory. On the other hand, with Preferential Attachment (plot 1.10), triad distribution converges to what the distribution would be if edges had been randomly assigned. We believe this is due to the fact that with Preferential Attachment, edge number grows exponentially and the number of positive edges and negative edges eventually end up being roughly the same. As a result, newly created edge signs are more or less random based on the way the sign of a newly created edge is decided. However, after we increase the probability of edge death to reach a stable graph (see section 4.2), the behavior is similar to the ER plot on the left.

4.4 Network Visualization

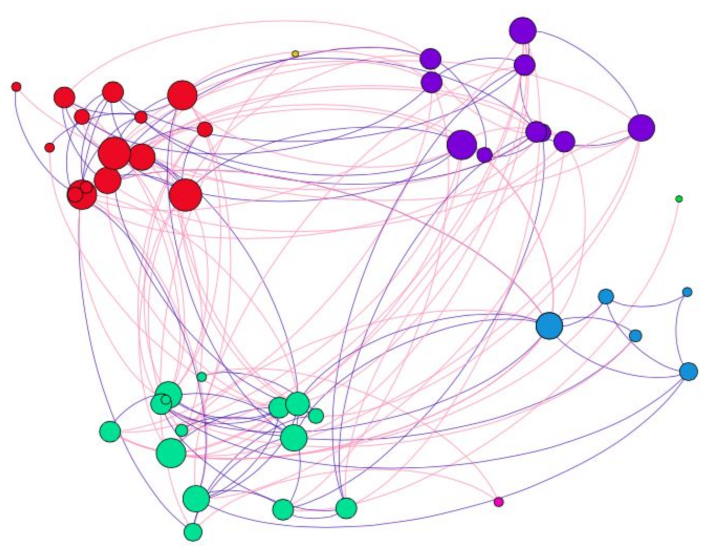
To more intuitively understand the network's evolution, as well as showcase our model and its features, we ran various configurations of our model and produced visual representations using Gephi. Below is a single network configuration over three stages of its evolution: initial setup, mid-point (after 50 rounds) and final (100 rounds). Particularly, this was set up as a Preferential Attachment network with an initial high degree of positive edges (0.8) and a relatively high degree of random edge sign change (0.2). To allow a clean visual representation, we increased edge death greatly to avoid exponential growth of edges, as noted in section 4.2.

Visual features include: node size proportional to degree, node color and position according to community detection, edge color of positive (blue) and negative (red). Note that community detection does not group positively linked edges, e.g. "friends", but *communities*. For example, if a node is negatively connected to a community, it could nevertheless be part of that community.

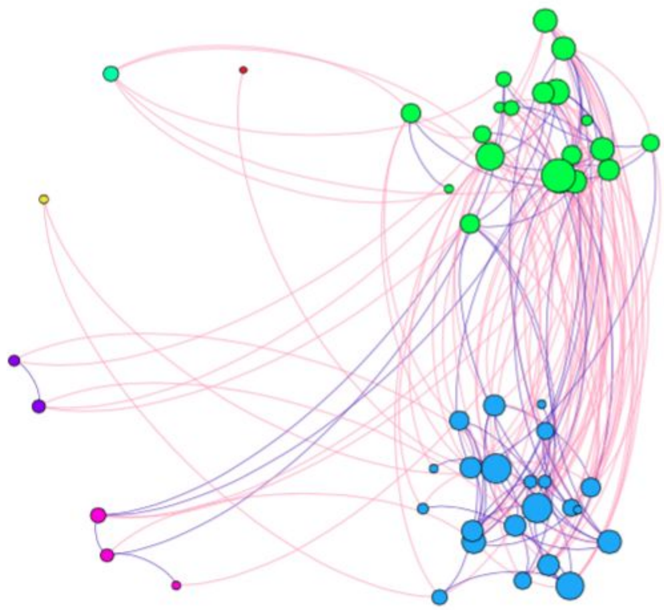
Round 0 - Initial Setup



Round 50 - Midpoint
We observe edge creation, more enemies and community growth



Final Round
We see community clustering and a few looners or small groups apart



5. Conclusion

Our model of friends and enemies, positive and negative edges, provides an interesting examination of the evolution of signed networks. We introduced a few key features - edge creation and sign change based on mutual acquaintances, edge death, and random sign change. We considered different initial conditions, Erdős–Rényi and Preferential Attachment, and a vast array of parameter configuration. We compared our model to popular psychology theories of human networks, Status Theory and Balance Theory, and analyzed where are model is similar and where it differs.

Going forward, we considered various possibility for extension. Our model does not have random edge generation. It would be interesting to see what effect random edge generation would have on the model. Additionally, we saw that different starting configurations had significant effect - more starting configurations could yield interesting results. Lastly, a different way of controlling for edge growth (e.g. inversely proportional to edge count as opposed to negatively proportional) would be interesting.

6. Work Cited

[1] Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010, April). Signed networks in social media. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1361-1370). ACM.

[2] Leskovec, J., Huttenlocher, D., & Kleinberg, J. (2010, April). Predicting positive and negative links in online social networks. In Proceedings of the 19th international conference on World wide web (pp. 641-650). ACM.

[3] Brzozowski, M. J., Hogg, T., & Szabo, G. (2008, April). Friends and foes: ideological social networking. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 817-820). ACM.

[4] Heider, F. (1958). The Psychology of Interpersonal Relation. John Wiley & Sons.

[5] Khanafiah, D. and Situngkir, H. (2004) Social balance theory. http://arxiv.org/pdf/nlin/0405041.pdf