

# Behaviour of Edge Signs in Economic and Non-Living Entities Networks

**Ravinithesh Annapureddy**

ravinithesh.annapureddy  
@epfl.ch

**Yasaman Haghighi**

yasaman.haghighi  
@epfl.ch

**Nikhil Mahtani**

nikhil.mahtani  
@epfl.ch

## Abstract

Signed network analysis enables us to evaluate the social psychological theories of balance and status in social networks. These theories are formulated and tested on humans and social networking sites. Apart from social media, humans form networks in various other settings like trading or gather a group identity. We analyze a financial network along with a non-living entity network maintained by living entities. We discover that the balance theory is able to predict network structure for the former, a mix of status and balance theories are need for the latter.

## 1 Introduction

Social computing is widespread and prevalent in society today. Understanding the intersection between computational systems and social behavior can enable the creation of meaningful and relevant digital social systems. Recently, interest has been garnered for the use of signed networks to examine complex social relationships. Leskovec et al. (2010) have studied networks created from Epinions, Slashdot, and Wikipedia and evaluated they evaluated the social psychology theories of *Balance* and *Status*.

As an extension, we assess how well those theories hold while studying networks created in an economic context and a network of non living entities. For these two cases, we use a bitcoin dataset (Kumar et al., 2016; Kumar et al., 2018) for economic context, and a Reddit dataset (Kumar et al., 2018community) for the latter.

### 1.1 Research Questions

In this extension, we want to discuss the generalization of the balance and status theory. Specifically, we want to answer the following questions

1. Whether the proposed theories can be extended to networks created based on monetary/profit relationships instead of relationships created as social connections in Slashdot data i.e. will the relations between people (users) still follow the balance and/or status theory even when there is a direct cash transaction involved?
2. Secondly, We verify if the collective behavior of humans in the entities created and maintained by them follow the same laws as themselves. Unlike the datasets that represent the opinion of one person on the other used originally to validate the theory, the Reddit dataset represents a network between non-living entities that are operated by humans and carry their collective opinion as a group.

## 2 Related theories for Signed Networks

In this section, the theories of balance and status are described along with an outline of conclusions drawn by (Leskovec et al., 2010).

### 2.1 Triadic Relation

Triadic relation is defined as the subnetwork of 3 nodes (triad) in a network. Although these subnetworks can be disconnected, we only consider connected triads in our work. Having edge weights with signs would make such triads, signed triads.

### 2.2 Structural Balance Theory

According to the balance theory, the triads with all positive edges and/or 2 negative edges are predominant when compared to others. Here, a positive edge can indicate notions like trust or friendships. A modified version, Davis' weak structural balance' theory, contrasts by claiming that only the arrangement of triads where there are two positive edges would not be prevalent in real networks.

	Bitcoin Alpha	Reddit Hyperlink
Nodes	3783	35776
Edges	24186	137821
+Edges %	93.6	94.4
-Edges %	6.4	5.6
Triads	22153	406391

Table 1: Dataset statistics for Bitcoin alpha dataset and Reddit Hyperlink dataset.

### 2.3 Status Theory

The status theory stipulates that a positive edge from one node to the other would indicate the former node expressing that the latter node has higher status. The inverse is true in the case of a negative link. Hence if a positive link is directed from a node A to B and is subsequently formed from B to C, a negative link from C to A would be predominantly found to stabilize the triad.

### 2.4 Conclusions on Balance and Status theories

The findings in (Leskovec et al., 2010) reveals that the undirected network structure fits the weak balance theory while status explains better the directed networks. Lastly, they suggest that balance theory explains the reciprocation in directed edges better than status.

## 3 Datasets

### 3.1 Description

For this extension, we considered two datasets where relationships between entities can be represented with signed links. First dataset is a network created by users trading Bitcoins (referred to as bitcoin dataset), who rate each other on a scale of -10 (total distrust) to +10 (total trust) to prevent transactions with fraudulent and risky users. The second one is the Reddit Hyperlink network (referred to as Reddit dataset) with connections between two subreddits (a community on Reddit). Each connection represents the sentiments of one community towards the other in one post.

### 3.2 Preprocessing

In the bitcoin dataset, the trust values range from -10 to +10. However, the network analysis requires the sign of the edge rather than its numerical value. Thus all values less than 0 are considered as -1 and the remaining as +1.

For the Reddit dataset, we merge all the connections from one subreddit to another and take the majority sign as the sign of the edge between them. Moreover, the last timestamp when the link is created between two subreddits is considered as the timestamp for the aggregated link.

### 3.3 Statistics

From Table 1, we observe that the Reddit dataset is large the tens of thousands nodes and the Bitcoin dataset is small with slightly under four thousand nodes. Both datasets have similar proportions of signed edges with around 94% of positive edges. The edges are counted on an undirected graph.

## 4 Analysis of Undirected Networks

Primarily, We evaluate the balance theory on our datasets in an undirected regime as described in 8.1. We consider 4 different types of triads based on the sign of triads' edges as shown in Figure 1. We can see in Table 2 that the triads  $T_1$  and  $T_0$  are overrepresented and  $T_3$  and  $T_2$  are underrepresented in both datasets. However, the  $P(T_i)$  and  $P_0(T_i)$  are only slightly different for each type of triad. In both datasets, surprise values are mostly larger than 10 and the observations could be considered statistically significant. The results reveal that neither Heider's theory nor Davis's weaker notation of balance theory hold. This can be interpreted based on the differences between social networks and the networks that are created in an economic context or the networks of nonliving entities like Reddit. When we consider our networks to be undirected, we are assuming that the behavioral symmetric holds in our networks. But our results show that this is not a suitable assumption for these two types of networks that we are studying.

## 5 Analysis of Evolving Directed Networks

In analyzing the datasets as directed evolution graphs, we examine the contextualized links (Leskovec et al., 2010). There are 16 different types of triads possible when considering signed directed networks as shown in Figure 2 and the results of the analysis are summarized in Table 4 and Table 5 in the appendix. The details of the analysis are presented in 8.2. In short, this analysis involves how a common neighbor would affect the link creation between two nodes. We observe the

Type	$ T_i $	$P(T_i)$	$P_0(T_i)$	$S(T_i)$
<b>Bitcoin</b>				
$T_3$	17014	0.768	0.825	-22.4
$T_1$	1667	0.075	0.011	93.4
$T_2$	3333	0.15	0.164	-5.4
$T_0$	139	0.006	0	46.3
<b>Reddit</b>				
$T_3$	328817	0.809	0.842	-57.7
$T_1$	8662	0.021	0.009	86.6
$T_2$	68461	0.168	0.149	34.9
$T_0$	451	0.001	0	41.5

Table 2: Analysis of networks as undirected.

triads with all positive edges in either direction are overrepresented and the remaining are underrepresented in both networks. For the bitcoin dataset, the balance theory explains the link formation better than the status theory. However, in the case of the Reddit dataset, there is no clear winner. It can be observed that even in cases where a negative edge is obvious, the theories fail to predict the sign. We explain this behavior by stating that the entities created by humans generally tend to broadcast positive feelings towards other entities. However, it should also be kept in mind that both of the networks analyzed chiefly consists of positive edges.

## 6 Reciprocation of Directed Edges

In the above analysis, we saw the effects of edge formation given a common neighbour. Here we study how one node connects to another node given that the inverse edge already exists.

From Table 3, we find that 41.60% and 9% of edges reciprocate in Bitcoin and Reddit datasets respectively. The ability of users to directly respond in the Bitcoin environment with equal strength to the links that they receive could be the reason for such high reciprocation.

In the Bitcoins dataset, mutually positive reciprocated links (+—+) and mutually negative reciprocated are present excessively. This functioning strongly corresponds to the Balance theory. For the Reddit dataset, we observe a very different scenario. Whether the first link is either positive or negative, the response link is by and large positive. The reasons follow from the directed analysis of the graph in that a group tends to exhibit positive sentiments when relating to other groups.

Type	<b>Bitcoin</b>		<b>Reddit</b>	
	count	Fraction	count	Fraction
P(+ +)	9678	0.98	12847	0.98
P(- +)	222	0.02	270	0.02
P(+ -)	26	0.16	345	0.92
P(- -)	136	0.84	29	0.08

Table 3: Edge Reciprocation Table.

## 7 Conclusion

In this work, we have studied the structure of social networks in two different settings. We find that status theory does not explain all the scenarios in a directed setting. The first is a financial environment where users trade bitcoins and rate each other for trust. We find that the assumption of structural symmetry does not hold when this network of users is considered as an undirected graph. At the same time, we see that balance theory fits the directed phenomenon instead of status. Lastly, we analyzed reciprocation of edges to see that the balance theory fits well where the user tends to reciprocate with the same sign they received. These patterns can be explained by the nature of the network where each user has an equal ability to directly respond, unlike other social networks. Also, the money involved and anonymous nature of this network forces users to make strict decisions without being affected by a third party.

The other type of network studied is a Reddit network where the nodes are subreddits and edges are the sentiments expressed towards each other. The peculiar nature of this network is that the nodes themselves do not express any opinions rather a group of people who are part of the subreddit express a collective opinion. While validating the laws of balance and status, we find out that balance theory could not explain the undirected network. Moreover, In a directed regime, both theories work only for half of the cases. This indicates a pattern that a subreddit usually receives a positive link from others irrespective of the interaction with a common neighbor. Nevertheless, it has to be noted that both the chosen networks have around 94% of positive edges. Lastly, while we look at the link reciprocation we see a profound number of positive link initiators in the network. These observations lead to the conclusion that when in a group, people tend to express positive sentiments towards other groups, both balance and status theory fails to explain the network.

## References

- Jure Leskovec, Daniel Huttenlocher, and Jon Kleinberg. 2010. *Signed networks in social media*. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 1361–1370.
- Srijan Kumar, Francesca Spezzano, VS Subrahmanian and Christos Faloutsos. 2016. *Edge weight prediction in weighted signed networks*. IEEE International Conference on Data Mining (ICDM), 221–230.
- Srijan Kumar, Bryan Hooi, Disha Makhija, Mohit Kumar, Christos Faloutsos and VS Subrahmanian. 2018. *Rev2: Fraudulent user prediction in rating platforms*. Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, 333–341.
- Srijan Kumar, William L Hamilton, Jure Leskovec and Dan Jurafsky. 2018. *Community interaction and conflict on the web*. Proceedings of the 2018 World Wide Web Conference on World Wide Web, 933–943.

## 8 Appendix

### 8.1 Un-Directed Analysis

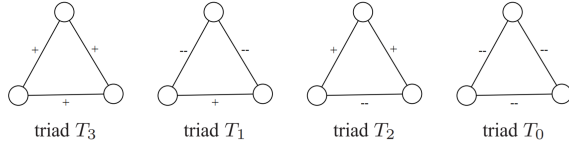


Figure 1: List of 4 possible triads in an undirected graph from Leskovec (2010)

In the analysis of the network as undirected graph, we consider  $|T_i|$  to be the number of triads type  $i$  and  $P(T_i)$  as the fraction of triad  $T_i$ . We also consider  $P_0(T_i)$  to be the fraction of each triad  $T_i$  if we assign the sign of edges randomly while keeping the percentage of positive and negative edges stay the same. We consider  $T_i$  to be overrepresented if the  $P(T_i) > P_0(T_i)$  and underrepresented otherwise. We evaluate our observation based on surprise  $s(T_i)$  that is calculated by  $\frac{(|T_i| - E[T_i])}{\sqrt{\Delta P_0(T_i)(1 - P_0(T_i))}}$  and shows the number of standard deviations by which the  $|T_i|$  differs from the expectation when we assign the sign of edges randomly.

### 8.2 Directed Analysis

To analyze the network as it grows, we do not build the entire graph at once rather we build edge by edge based on the time they were created. We

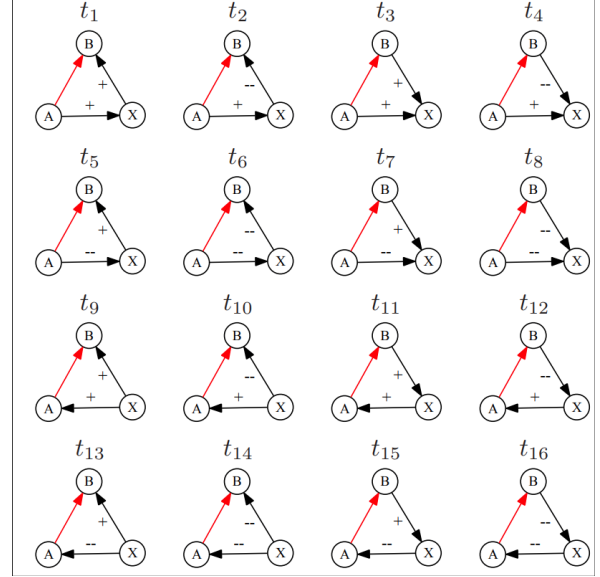


Figure 2: List of 16 possible triads in a directed graph from (Leskovec et al., 2010)

go over each edge in the data based on its time of creation and get the common neighbors between the starting node and receiving node. Then, we get the type of edges between the starting and receiving node to the common neighbor and determine the type of triad. Based on the type of triad we increment the counter and add the generative baseline of the starting (receiving) node to the generative (receptive) baseline of the contextualized link. update the respective count, generative, and receptive base. Using the edge to/from the common neighbor we assign the status nodes. A node receiving a positive edge or directing a negative link to another node will receive a status of 1 and -1 in the opposite case.

After obtaining the above metrics, we obtain the generative surprise of a conceptualized link as the number of standard deviations above or below the generative baseline of the link are the actual number of positive edges in that type of link. Similarly, the receptive surprise is defined with a receptive baseline.

Also, we say that generative (receptive) surprise for the triad type is consistent with status if receiving (starting) node status has the same (opposite) sign as the generative (receptive) surprise. In the same lines, we say that generative (receptive) surprise for the triad type is consistent with balance if the sign of the generative (receptive) surprise is equal to the sign of the edge as predicted by balance. The surprise is calculated in a similar way to

$t_i$	count	$P(+)$	$s_g$	$s_r$	$B_g$	$B_r$	$S_g$	$S_r$
$t_1$	26910	0.92	-23.7	6.1	False	True	False	True
$t_2$	1105	0.22	-98.7	-91.1	True	True	True	False
$t_3$	27368	0.9	-37.2	-15.3	False	False	True	False
$t_4$	709	0.67	-21.2	-13.9	True	True	False	False
$t_5$	1272	0.67	-36.3	-30.3	True	True	False	True
$t_6$	201	0.61	-16.9	-12.4	False	False	True	True
$t_7$	950	0.57	-41	-38	True	True	True	True
$t_8$	900	0.83	-9.3	0.2	False	True	False	False
$t_9$	27563	0.92	-30.2	3	False	True	False	False
$t_{10}$	1121	0.21	-100.6	-92.6	True	True	True	True
$t_{11}$	26497	0.9	-43.6	-19.3	False	False	True	True
$t_{12}$	1020	0.75	-22.2	-10.5	True	True	False	True
$t_{13}$	475	0.59	-9.8	-24	True	True	False	False
$t_{14}$	179	0.63	6.7	-11.8	True	False	False	False
$t_{15}$	475	0.56	-10.4	-28.1	True	True	True	False
$t_{16}$	159	0.81	2.2	1.9	True	True	True	True

Table 4: Analysis of Bitcoin network as a directed evolving network.

the directed graphs except that the expected number is already obtained through baselines of each triad type.

$t_i$  indicates one of the 16 types of triads shown above,  $P(+)$ , and is the probability that the closing red edge is positive.  $s_g$  gives the surprise of edge initiator giving a positive edge;  $s_r$  gives the surprise of edge destination receiving a positive edge.  $B_g$ ,  $B_r$ ,  $S_g$ , and  $S_r$  show if the balance and status theories are consistent with the generative surprise and receptive surprise respectively.

$t_i$	count	$P(+)$	$s_g$	$s_r$	$B_g$	$B_r$	$S_g$	$S_r$
$t_1$	189574	0.96	95.2	131	True	True	True	True
$t_2$	12164	0.93	18.8	38.1	False	False	False	True
$t_3$	178217	0.95	85.9	27.8	True	True	False	True
$t_4$	11257	0.93	19.4	37.3	False	False	True	True
$t_5$	16870	0.89	12.9	19.8	False	False	True	False
$t_6$	1626	0.87	0.1	4.9	True	True	False	False
$t_7$	13780	0.88	4.8	-15.6	False	True	False	True
$t_8$	1714	0.89	4.6	11.1	True	True	True	False
$t_9$	190700	0.96	37.8	122.7	True	True	True	False
$t_{10}$	12216	0.93	4.1	38.5	False	False	False	False
$t_{11}$	88097	0.97	24.1	23.2	True	True	False	False
$t_{12}$	4323	0.93	3.4	24.7	False	False	True	False
$t_{13}$	12370	0.93	26.3	33.5	False	False	True	True
$t_{14}$	2319	0.92	11	14.2	True	True	False	True
$t_{15}$	4491	0.93	15.6	3.6	False	False	False	True
$t_{16}$	434	0.89	1.8	3.9	True	True	True	True

Table 5: Analysis of Reddit Hyperlink network as a directed evolving network.

### 8.3 Github Repository

The github repository containing a note with the data can be found at [https://github.com/epfl-ada/ada-2020-project-milestone-p3-p3\\_data-nuage](https://github.com/epfl-ada/ada-2020-project-milestone-p3-p3_data-nuage)