Network Analysis of Weighted Signed Bitcoin Networks

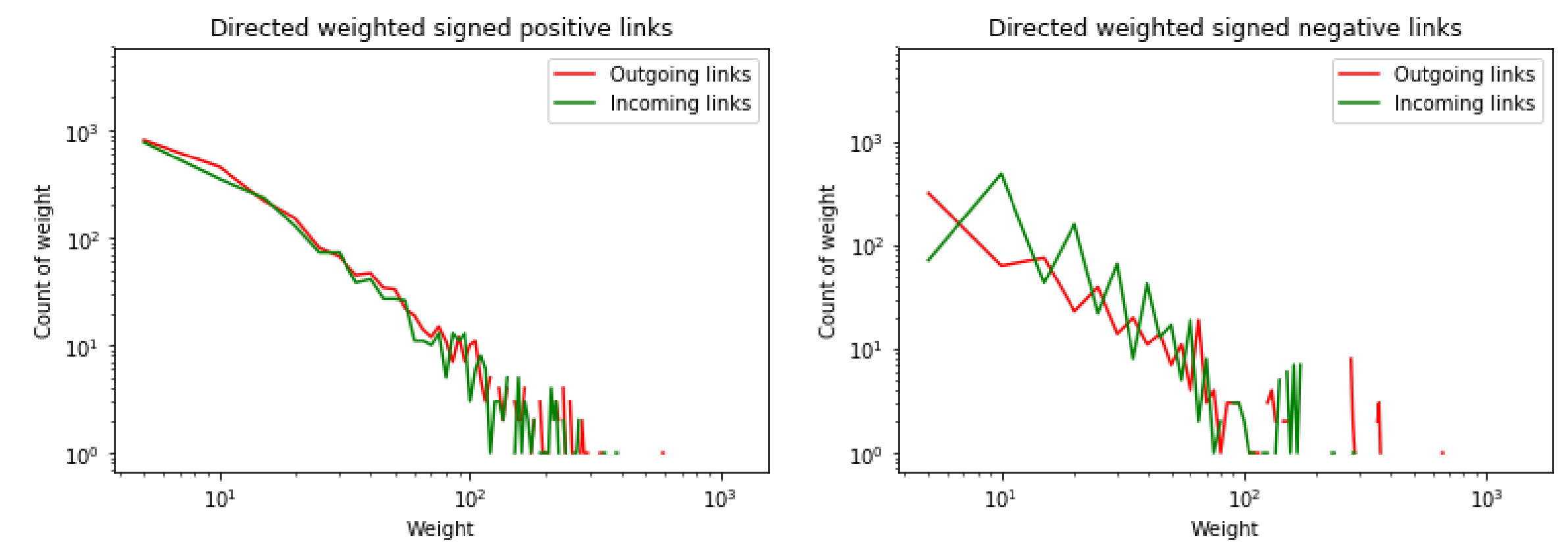
Akshay Kumar, Vincent Ying

CS 224W: Analysis of Networks Class Project

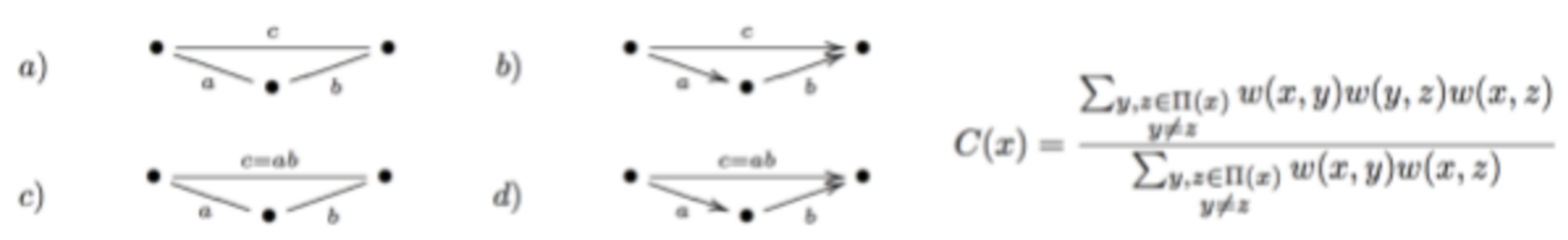
Introduction

This project involves the study of structural properties in bitcoin trust networks. Degree distribution, average clustering coefficient, balance & status theory, and sign & edge prediction are explored.

Degree and Weight Distribution



Clustering Coefficient

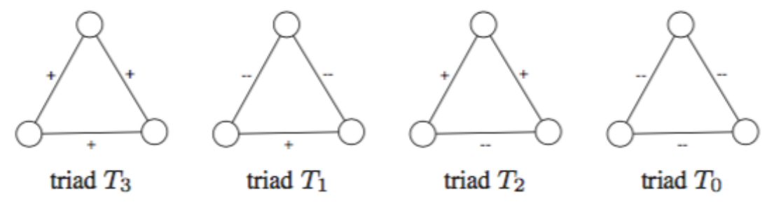


$$C(x) = \frac{\sum_{\substack{y,z\in\Pi(x)\\ y\neq z}} w(x,y)w(y,z)w(x,z)}{\sum_{\substack{y,z\in\Pi(x)\\ y\neq z}} w(x,y)w(x,z)}$$

- Bitcoin-otc) ¡ Bitcoin-alpha) (Fewer edges)
- Directed ¡ Undirected & Signed ¡ Unsigned
- Weighted ¡ Unweighted

Balance Theory

Edge sign possibilities for simple cycles of length 3.



A weakened form of the Balance Theory is where T_0 is considered balanced.

Status Theory

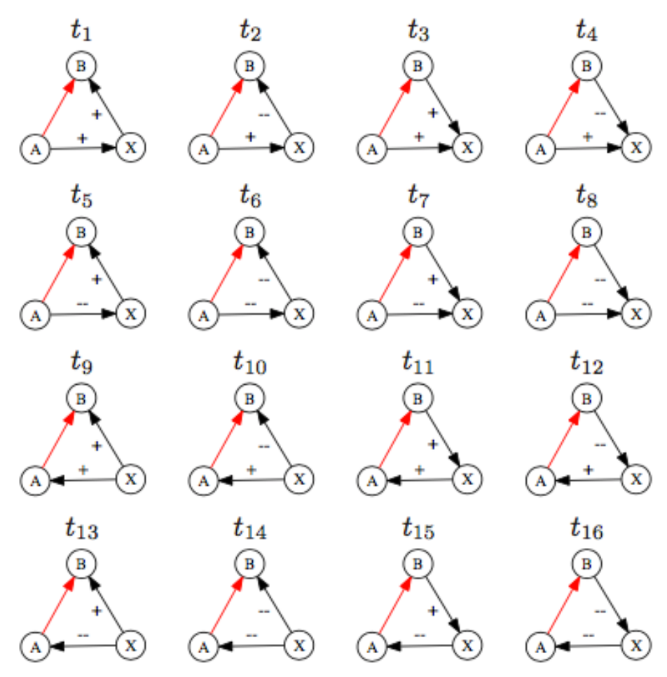


Figure 1: Triad possibilities for two signed edges.

	btc-alpha		btc-otc	
t_i	count	$p(t_i)$	count	$p(t_i)$
t_1	44946	0.68	61185	0.61
t_2	3763	0.06	6018	0.06
t_3	4320	0.07	5998	0.06
t_4	771	0.01	1592	0.02
t_5	3547	0.05	5307	0.05
t_6	1033	0.02	2124	0.02
t_7	403	0.01	2316	0.02
t_8	871	0.01	1988	0.02
t_9	4276	0.06	5961	0.06
t_{10}	573	0.01	1039	0.01
t_{11}	213	0.00	350	0.00
t_{12}	44	0.00	1476	0.01
t_{13}	1031	0.02	2340	0.02
t_{14}	485	0.01	2554	0.03
t_{15}	170	0.00	162	0.00
t_{16}	13	0.00	69	0.00

Sign Prediction with Local/Global Perspective

For the prediction of edge signs, we leveraged the local and global aspects of balance theory. The local aspect arises from the definition of structural balance of triads. A measure of social balance based on l-cycles (a path from a node to itself through l edges) is utilized in a simple prediction rule that we have adapted to weighted, signed networks. A global approach based on social balance is achieved by increasing the length of the l-cycle.

A low rank modeling approach was taken for the edge sign prediction. Since these networks have low rank adjacency matrices, they can be utilized to solve the sign prediction problem with a reduction to a low rank matrix completion problem. Although rank constraint of the low rank matrix completion problem is non-convex, convex relaxation of rank constraint is achieved through Singular Value Projection and Matrix Factorization using ALS.

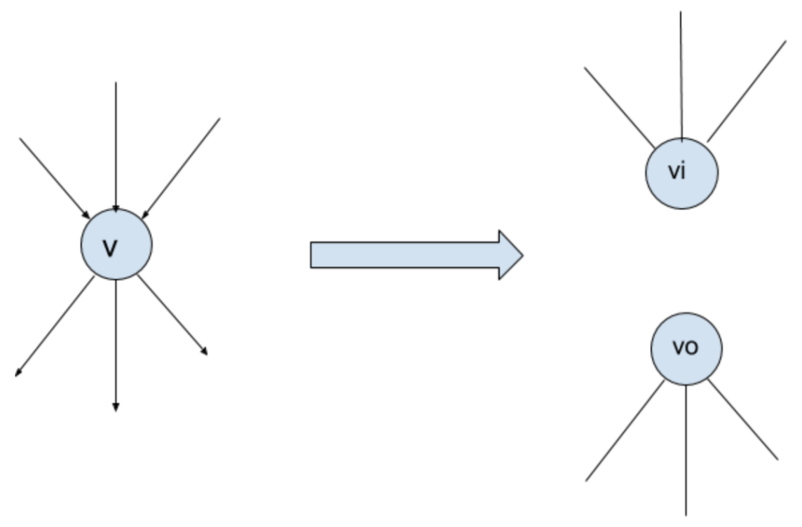
Local Methods:

1. Measures of Social Imbalance (MOI): This measure is based on structural balance.
2. Higher Order Cycles (HOC): This is the use of Status Theory in the prediction of edge sign.

Global Methods:

3. Singular Value Projection (SVP): From the use of MOI, the length of l-cycles is increased to provide a more global view of the balance structure for the network. Gradient descent update is projected onto the non-convex set of matrices using SVD.
4. Matrix Factorization with (ALS): This is a gradient based approach for edge sign prediction. Alternating Least Squares is used to solve the matrix completion problem by alternatively minimizing adjacency matrix.

Signed Weight Prediction



Feature vectors:

- Common Neighbors: $\sum_{z\in\Pi(x)\cap\Pi(y)} W_{xz} + W_{yz}$
- Jaccard Similarity: $\frac{\sum_{z\in\Pi(x)\cap\Pi(y)} W_{xz}+W_{yz}}{\sum_{a\in\Pi(x)} W_{xa}+\sum_{b\in\Pi(y)} W_{yb}}$
- Preferential Attachment: $\sum_{a\in\Pi(x)} W_{xa} \times \sum_{b\in\Pi(y)} W_{yb}$
- Adamic-Adar Coefficient: $\sum_{z\in\Pi(x)\cap\Pi(y)} \frac{W_{xz}+W_{yz}}{\log(1+\sum_{c\in\Pi(z)} W_{zc})}$
- Resource Allocation Index: $\sum_{z\in\Pi(x)\cap\Pi(y)} \frac{W_{xz}+W_{yz}}{\sum_{c\in\Pi(z)} W_{zc}}$
- Local Clustering Coefficient: $CC(x) + CC(y)$
  where $C(x) = \frac{1}{|\Pi(x)|(|\Pi(x)|-1)} \sum_{yz}(\hat{W_{xy}}\hat{W_{xz}}\hat{W_{yz}})$.
- We do linear regression using these features.

Results

Edge sign prediction:

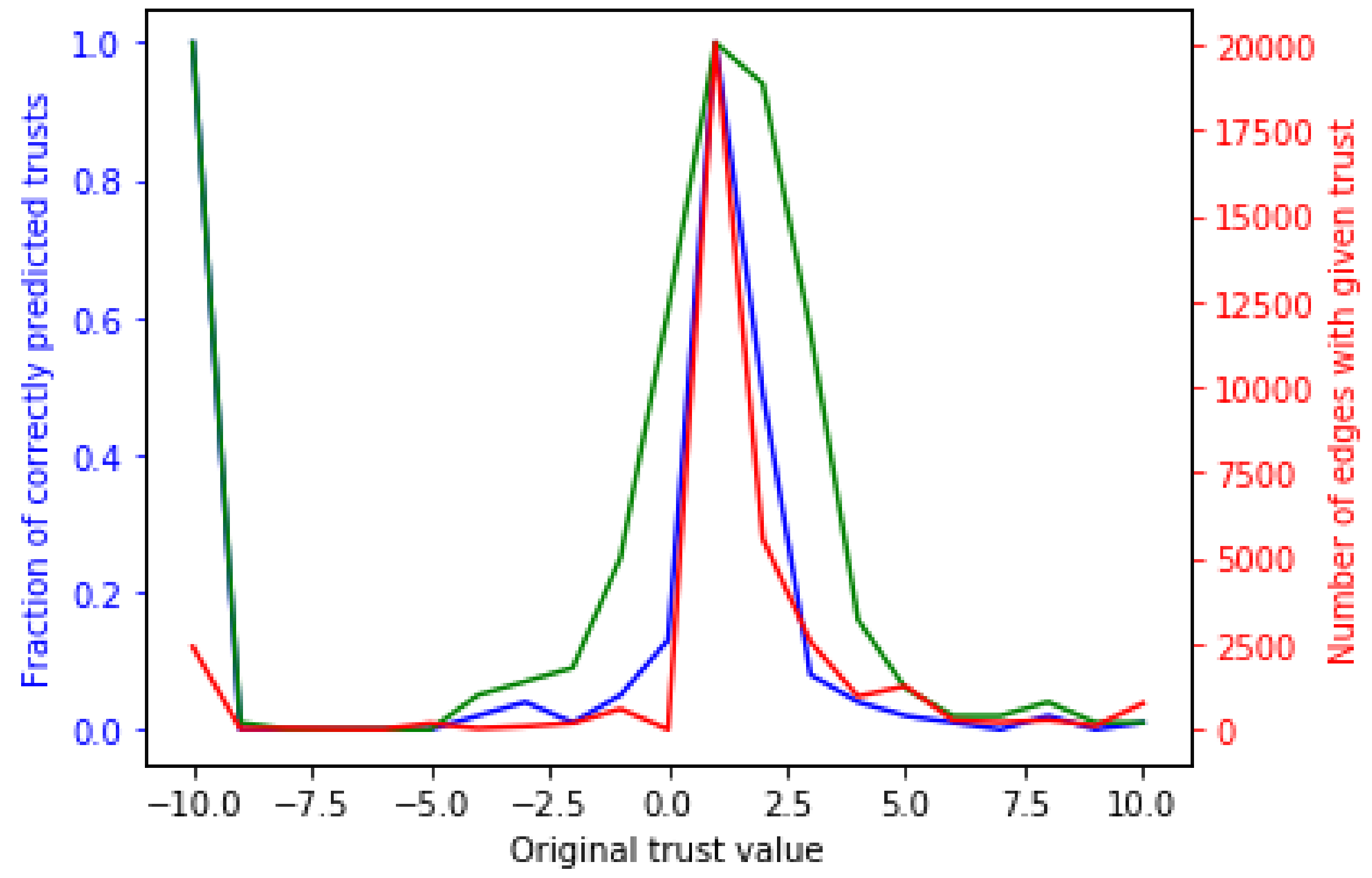
Edge sign prediction accuracy for local and global methods.

Edge Sign Prediction Accuracy.

		MOI	HOC	SVP	ALS
btc-alpha	Accuracy	0.9158	0.7631	0.9845	**0.9936**
	False Positive Rate	0.6009	0.2908	0.0694	0.0236
btc-otc	Accuracy	0.9086	0.7711	0.9265	**0.9901**
	False Positive Rate	0.3981	0.2429	0.2216	0.0249

Edge weight prediction:

Edge weight prediction with linear regression on various features.



Accuracy of linear regression = 72 % (Correct prediction of edge weight in 72 % of test set).

Conclusion

Working from current state of the art, we defined metrics (such as Clustering Coefficient, Adamic-Adar Coefficient, etc.) for weighted signed networks. From these metrics, we computed the clustering coefficients, predicted edge sign, and also predicted edge weights. The proposed algorithm for edge weight prediction is built on linear regression. This algorithm correctly computes weights (trust) of 72% of edges in the testing set.

Acknowledgements

This project was possible with the guidance of Srijan Kumar.

Edge sign prediction code adapted from Mark Heimann's implementation of "Prediction and Clustering in Signed Networks: A Local to Global Perspective", Chiang et al.