NETWORK SCIENCE: PROJECT 1

Neves, João
83405
Instituto Superior Técnico

Niewiński, Łukasz
ist195018
Instituto Superior Técnico

Santos, Pedro
83550
Instituto Superior Técnico

October 22, 2019

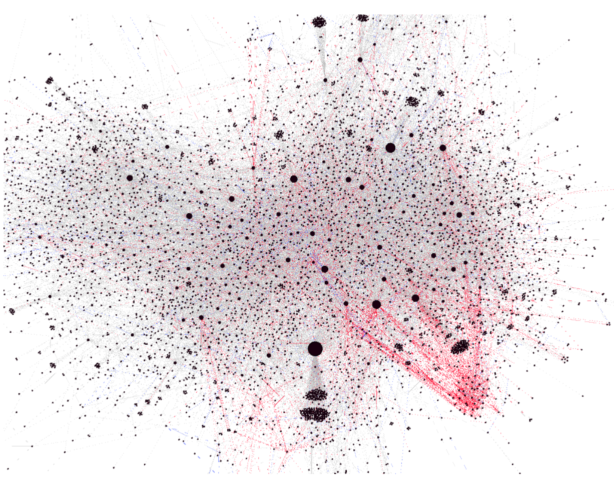
1 Introduction

Throughout this project we will focus our work on the analysis of the Bitcoin OTC trust weighted signed network [1, 2]. The network captures the rating-relations (who-trusts-whom) between people who trade Bitcoin on a platform called Bitcoin OTC. Each user u is represented as a node on the graph. If user u_1 rates user u_2 then a directed link between both nodes (from u_1 to u_2) is added to the network. Each link has a weight w assigned, ranging from -10 (total distrust) to $+10$ (total trust), which corresponds to the rating value. This is the first explicit weighted signed directed network available for research.

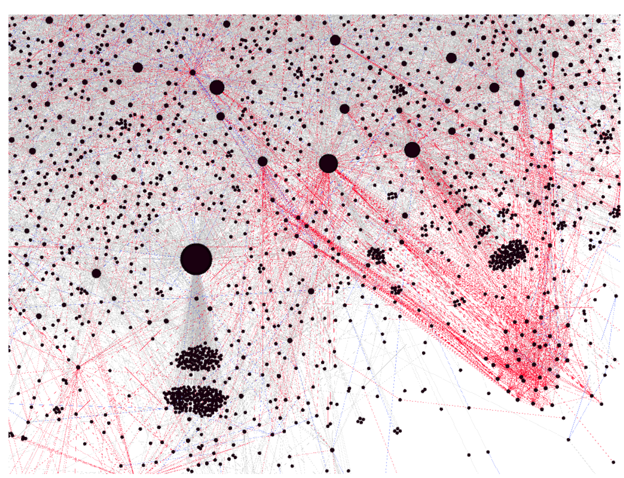
The graph is composed of $5,881$ nodes (users) and $35,592$ edges (ratings). It is worth noticing that the ratings distribution is skewed, as seen in Figure 3e (89% of the edges have a positive weight).

For the development of this project we used both GEPHI [3] and NetworkX [4] softwares. Most of the calculations were computed on both programs and their results compared for sanity-check. We used GEPHI to create a visualization of the network, seen in Figure 1.

On Section 2 we will analyze some properties of this network. We will start by investigating the degree centrality metric for both weighted and non-weighted versions of the graph. Afterwards, we take a look at the average in/out ratings for each node. We plot the distribution of the average in/out ratings for each node and the distribution of the absolute difference between average in/out ratings for each node. We also take a look at the strongly connected components (SCCs) of the graph. On Section 3 we will summarize our findings.



(a) Global



(b) Zoom

Figure 1: Network draw from GEPHI. Red and blue is used to, respectively, color -10 and $+10$ edges. The size of a node is proportional to its degree.

2 Findings & Results

The developed code is publicly available at: https://github.com/LukaszNiewinski/Network-Science

2.1 Degree centrality

The degree centrality quantifies the importance of a node by counting the number of edges connected to it. In our case, the network is directed so *in-* and *out*-degrees should be considered. To compute the *weighted degree* we summed up all the ratings of the edges connected to a given node; this parameter gives us a sense of the importance of a node. Note that, the network has negative weights because ratings range from -10 to 10. Degree centrality-related values are presented below on Table 1.

Figure 3a plots the log-scale in/out-degree distributions for the network. Figure 3b plots the weighted in-degree distribution on a linear scale. It is worth noticing that the distribution looks symmetric around zero. Figures 3c and 3f plot, respectively, the in and out weighted degree distributions.

	Degree	In-degree	Out-degree	Weighted in-degree	Weighted out-degree
Mean	12.10	6.05	6.05	6.12	6.12
Std	38.30	17.67	21.09	38.84	38.52
Min	1	0	0	-675	-999
Max	1,298	535	763	1,041	874

Table 1: Degree centrality metrics.

2.2 Average rating

The average in-rating μ_{in} for the node u is the average of the ratings directed towards u (incoming links). The average out-rating μ_{out} for the node u is the average of the ratings directed away from u (outgoing links). The network has 1067 nodes only with incoming edges. For these nodes, the average number of links per node is 1.7 and the average rating is -0.62. Moreover, the sum of the ratings for the aforementioned nodes is -1364. This indicates that those 'passive' nodes are rated by the other users negatively. On the other hand, there are only 23 nodes which only have out-ratings. For these nodes, the average number of links per node is 2.04, the sum of ratings is -121 and the average out-rating per node of 0.76.

		Average in-rating (μ_{in})	Average out-rating (μ_{out})
All nodes	Mean	0.73	1.35
All nodes	Std	2.82	2.21
Nodes only with in-links	Mean	-0.62	–
Nodes only with in-links	Std	16.59	–
Nodes only with out-links	Mean	–	0.76
Nodes only with out-links	Std	–	40.37

Table 2: Average (in & out)-rating metrics.

The analysis of received/given average ratings for each node might help to distinguish highly trusted people and frauds within the network. Figure 3d plots the projection on the 2d plane of the average out and in ratings for each node. As we can see, most of the users rate others well and receive positive ratings. Users who receive negative rates and rate others negatively might indicate fraudulent behaviour. The number of nodes per quadrant (for Figure 3d plot):

- Quadrant I: 4916
- Quadrant II: 353
- Quadrant III: 476
- Quadrant IV: 143

Moreover, the absolute difference between average in/out ratings for each node was computed. For a given node u it is calculated as the absolute difference between the average in-rating for node u (μ_{in}) and the average out-rating for node

u (μ_{out}), i.e. $|\mu_{in} - \mu_{out}|$. The resulting distribution can be seen in Figure 2. As we can see, how an user rates others is similar to how others rate the user.

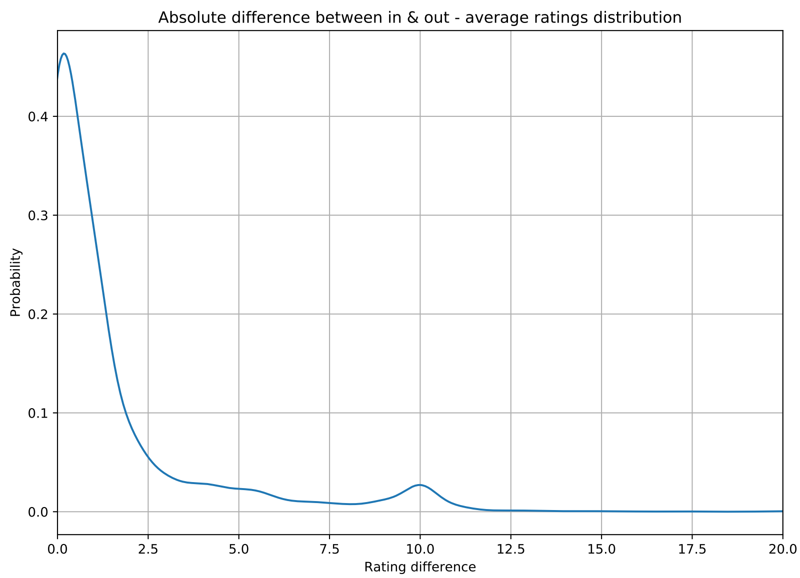


Figure 2: Absolute difference between in & out - average ratings distribution.

2.3 Strongly connected components

With respect to the SCCs, the network is composed of 1144 components. Eighty percent of the nodes belong to the same strongly connected component. Twenty two components have a size between 2 and 6 elements. All remaining SCCs consist of only one node (singletons).

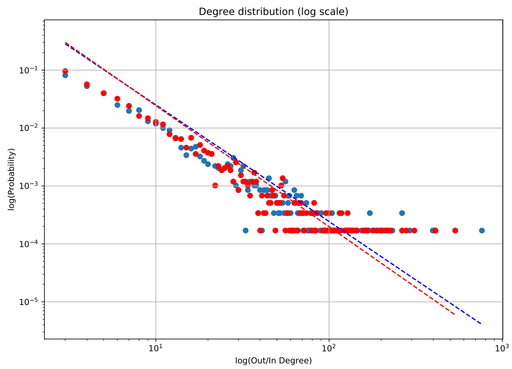
3 Conclusion

The network consists mainly of bi-directional links; the existence of a link in one direction indicates a high probability of existence of a returning link. This mutuality supports the fact that one SCC contains 80% of the nodes. We investigated this further and calculated the reciprocity of the network. Indeed, the network has a reciprocity of $\approx 80\%$, confirming our previously stated idea that most of the links are bi-directional. The high value of reciprocity (that allows to have one big SCC) suggests that the network posses the characteristics of a bow-tie network.

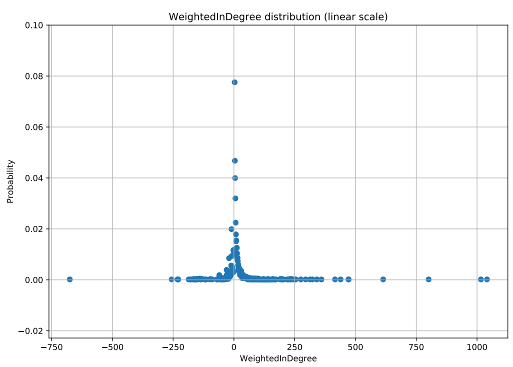
The majority of the ratings are slightly positive. If a user rates other members of the network positively, it is likely that they will rate this user positively too. The most popular rating is 1 which might indicate that users rate cautiously. The most used negative rating is -10, which might indicate that users rate negatively only when they are sure that the other user is untrustworthy (or they are angry).

Another interesting property of the network is the fact that there is a low number of nodes with only out-links but a high number of nodes with only in-links. Possessing only in-links suggests that the users are passive. The majority of these nodes are rated negatively by active users. Unweighted in- and out- degrees are highly correlated. The distributions and power laws are similar among the tested metrics, however, a bigger difference is noticeable when considering the ratings (weights) of the network. Computed gamma values indicate that the network is scale-free.

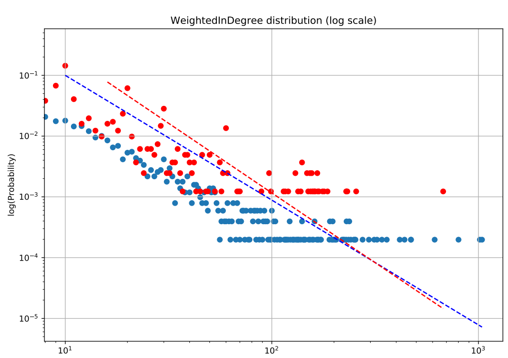
Collected information about the network can be used to identify frauds and to predict its growth - expansion of the network. Future work can include precise identification of fraud and collective fraudulent behaviour detection within communities (groups of friends), as those can have a big influence on the safety of the network.



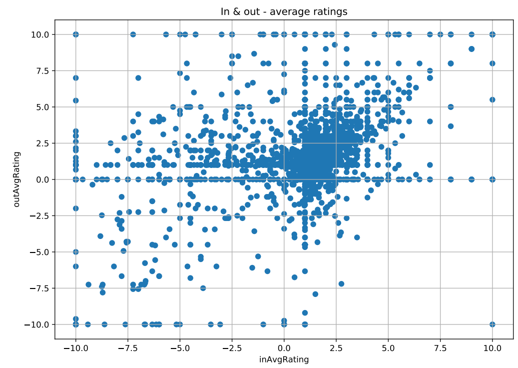
(a) Degree distribution and powerlaw fit. Blue instances represent Out-degree ($\gamma = 2.09$). In-degree is red ($\gamma = 2.01$).



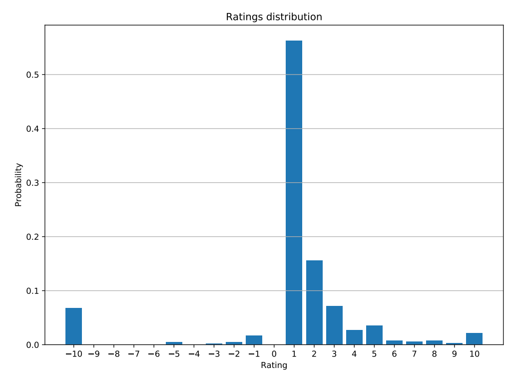
(b) Weighted In-degree, symmetry between positives and negatives is well visible on linear plot.



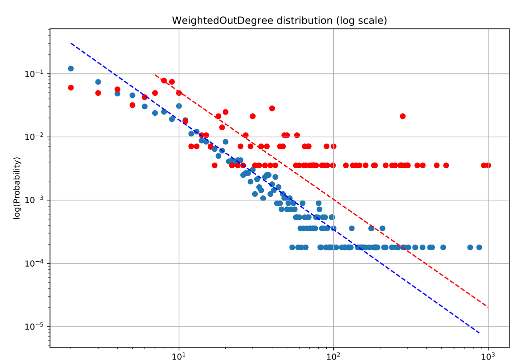
(c) Weighted In-degree distribution. Blue instances are positives ($\gamma = 2.05$). Red are the modulus (absolute value) of the negative degrees ($\gamma = 2.29$)



(d) Average In - Out ratings 2d projection map. Most users rate positively and are rated positively.



(e) Ratings distribution. Mostly ratings equals 1, negative ones are mostly -10.



(f) Weighted Out-degree distribution. Blue instances are positives ($\gamma = 1.73$). Red are the modulus (absolute value) of the negative degrees ($\gamma = 1.70$)

Figure 3: Network metrics plots.

References

[1] Srijan Kumar, Francesca Spezzano, VS Subrahmanian, and Christos Faloutsos. Edge weight prediction in weighted signed networks. In *Data Mining (ICDM), 2016 IEEE 16th International Conference on*, pages 221–230. IEEE, 2016.

[2] Srijan Kumar, Bryan Hooi, Disha Makhija, Mohit Kumar, Christos Faloutsos, and VS Subrahmanian. Rev2: Fraudulent user prediction in rating platforms. In *Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining*, pages 333–341. ACM, 2018.

[3] Mathieu Bastian, Sebastien Heymann, and Mathieu Jacomy. Gephi: An open source software for exploring and manipulating networks, 2009.

[4] NetworkX developer team. Networkx, 2014.