Web of Trust "Bitcoin Over the Counter (OTC)" Community Network

Banandeep Singh
MSBA Spring 2020 – UIC
IDS 564 – Social Media and Network Analysis
UIN: 669265832

Ritu Gangwal
MSBA Spring 2020 – UIC
IDS 564 – Social Media and Network Analysis
UIN: 67064677

Abstract:

Social network analysis (SNA) is not a formal theory in sociology but rather a strategy for investigating social structures. Companies perform network analysis process to make decision thereby helping in visualizing and analyzing formal and informal relationships in any organization. It can help to shape business strategy that maximizes organic exchange of information, thereby helping the business become more sustainable and effective.

This research paper aims at providing meaningful insights about the OTC Bitcoin Community Platform by evaluating various network characteristics and connectivity measures.

Keywords: Graph Theory Metrics, Connectivity, Centrality Measures, Community Detection, Betweenness, Clustering, Clique census, Triadic closure

I. INTRODUCTION

Bitcoin-OTC is a peer to peer, over-the-counter marketplace for trading with bitcoin crypto-currency. To mitigate the risks of the unsupervised exchanges, the establishment of a reliable reputation system is needed: for this reason, a web of trust is implemented on the website [1]. We analyze the structure of this network to show how reputation and trustworthiness is influenced by connectivity. Which, hypothetically speaking, can help in more trading volume and more gains within the network.

To have a complete understanding about the OTC Bitcoin Network Community, following studies are done:
Step 1: Simplification of network
Step 2: Network Visualization
Step 3: Network Characteristics
Step 4: Centrality Measures
Step 5: Community Detection
Step 6: Trust-level Detection

II. PURPOSE OF STUDY

People can be friends but also foes, this idea has given rise to Weighted Signed Networks (WSN) where edges are labeled as positive or negative. Supposedly, if node i has a positive opinion for node j the edge will be positive, and if node i have a negative opinion the node j, the edge will be regarded as negative. However, in real no-one neither trust completely nor distrust completely. Thus, the trust will vary on a scale depending on the network.

The Bitcoin - OTC is one such network of who-trusts-whom. Since Bitcoin users are anonymous, there is a need to maintain a record of users' reputation to prevent transactions with fraudulent and risky users. Members of Bitcoin OTC rate other members in steps of 1 on a scale of -10 (total distrust) to +10 (total trust).

In this paper we will study internal and external relationships between the users and the problem of predicting the weights of the edges by examining the bitcoin dataset. Predicting such weights is beneficial for many reasons. For instance, raising flags for BOTS or fraudulent users can save harmful intrusions. Since, people are anonymous, there is a lingering possibility of people impost -ring and duping novice traders.

This learning can be translated into other signed networks to identify people how strongly they are associated with a particular topic. Moreover, edge weight prediction may be useful to improve traditional tasks in signed networks such as node ranking [2], community detection [3] or perhaps sentiment prediction [4], among others. Therefore, the prediction of edge weights in WSNs can be advantageous in various tasks.

III. DATA DESCRIPTION

Bitcoin-OTC is a peer-to-peer marketplace for crypto-currency trading with bitcoin. A web of trust is established to reduce the risk of fraudulent transactions, as this is an unsupervised platform. In this web of trust, users rate each other with an integer score on a scale of -10 to +10 with 90% of greater than zero edges. The dataset is obtained from webpage [5]. This dataset contains 5,881 nodes and 35,592 ratings given by the users.

There is a behavioral rule established on the website; it is suggested to assign a score of 1 if they interacted. This rule generates a high prevalence of $s_{ij} = 1$, which is evident from the histogram that high scores are assigned at the first interaction. Also, it is evident that 0-3 ratings are favorite trust values in the network. Fig. 1 shows the histogram of no. of users giving a particular rating.

There are 4 columns - rater, ratee, rating: the source's rating for the target, ranging from -10 to +10 in steps of 1, time stamp- the time of the rating, measured as seconds since Epoch. Since, we are not concerned about at what time the rating was given, timestamp dimension is dropped.

Since we have both positive and negative ratings, this graph can be divided into two parts. We have 89.98% of positive edges with 32029 edges and 5573 vertices and 10.01% of negative edges with 3563 edges and 1606 vertices in total.

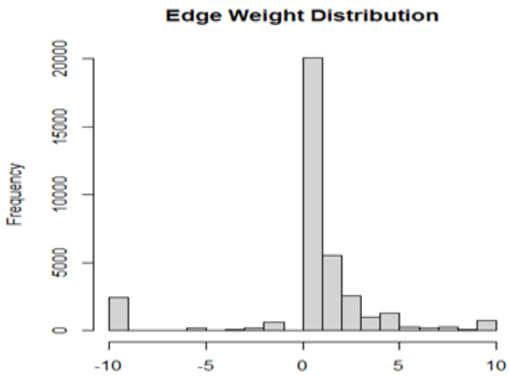

Fig. 1

IV. NETWORK LAYOUTS

Since a user cannot give the same rating again, there are no duplicates, so the resulting graph is simple.

Here are some popular layouts showing the simplified version of the OTC network. Layout DrL is a force-directed graph layout toolbox focused on real-world large-scale graphs depicted in Fig.2. Along with this, Kamada Kawai in Fig. 3 and Graphopt in Fig. 4 layouts are also produced.

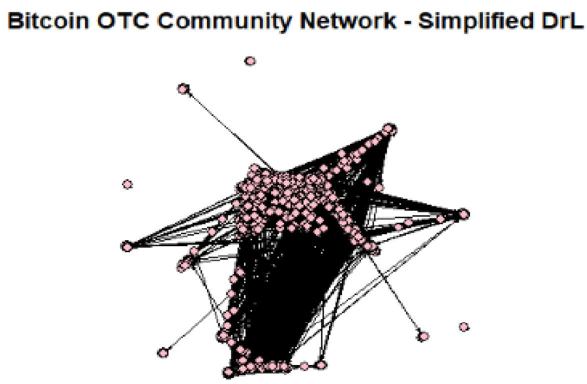

Fig. 2

An initial representation with Kamada Kawaii of the graph shows that there are inter-relationships amongst the raters.

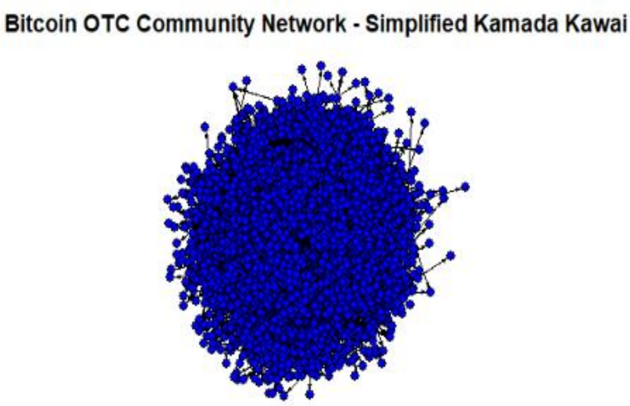

Fig. 3

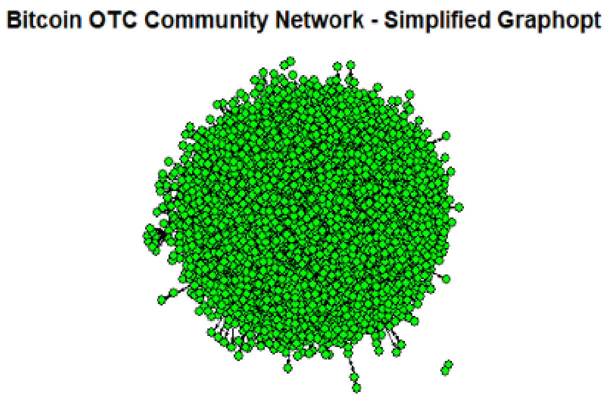

Fig. 4

V. NETWORK CHARACTERISTICS

Below are the basic network characteristics used to study the network. Some measures are calculated by finding log inverse of weights i.e., ratings. These weights are normalized by adding 12 to eradicate the effects of negative weights.

a. Density and average path length:
The density is an indicator of the level of connectedness in a network graph. It represents the ratio of connected ties over the total possible connections in the network between various users. The density of 0.001029 shows that there are numerous interactions, and this is in line with network diameter.

b. Average path length:
This refers to average number of steps covered in the shortest path to navigate through the network. The value of 3.718913 indicates that it takes 3.7-unit edges to travel from one node to another in the bitcoin network. This value is in sync with density value as less dense networks have high average path values.

c. Reciprocity:
Reciprocity refers to a measure of the likelihood of vertices in a directed network to be mutually linked i.e. situations where user X transacts with user Y and user Y transacts back with user X.

Reciprocity of 0.7923 indicates there are many dyads in bitcoin network. Alternatively, each user gives a rating to another node after trading, so this is in-line with the assumption of ratee rating users every time a trade happens.

d. Connectivity:
Low values of density and reciprocity implies that the graph is not very well connected i.e. neither weekly nor strongly; implying that not all ratee rates every person in the network. We see that the network decomposes into a total of 4721 SCCs, with the largest cluster having 4709 users in it. Most of the SCCs i.e., 1121 contain single node/vertex.

Cluster size (no. of raters)	1	2	3	6	4709
No. of clusters	1121	18	3	1	1

The table of weakly connected components (WCCs) shows that the network decomposes into total of 587 WCCs and the largest WCC consists of 5875 raters.

Cluster size (users)	2	5875
No. of clusters	3	1

e. Diameter:

The diameter of a graph is the greatest distance between any pair of nodes/users. Using the regular weights for edges, which we have normalized by adding 11 to weights, the diameter of the OTC network graph is computed as 147. When the inverse logarithm weight for edges is used, the diameter of graph reduces to 6.123736. This high value is in line with the fact that the OTC community is not linked.

f. Transitivity and clustering:

The transitivity – a measure relates three nodes in a sense that if node x is connected to node y, and node y is connected to node z, means node x will be connected to z – which suggests that there are only 5% of the node's transitive in nature. For a node to be transitive, if agent a trusts agent b, which in turn trusts agent c, then a will also trust c, to some extent. This makes trust propagation easier to analyze, while retaining its most intuitive properties.[6]

The global transitivity or the global clustering coefficient of the network is 0.05923844 i. e. about 6.0%, which means that not more than 6.0% of the connected triples are close to form a triangle. The mean of local clustering coefficient is 0.2882916 with a standard deviation of 0.3298311. Thus, there does not exist much triadic closure in the Bitcoin OTC network and it is sparse.

g. Clique Census:

Cliques are complete subgraphs and hence are subsets of vertices that are fully cohesive, in the sense that all users within the subset are connected by edges.

From the table below we see that in the OTC network there are 6405 edges (cliques of size 2), followed by 4223 triangles (cliques of size three), followed by 2636 cliques of size 4, and 2051 cliques of size 5. The largest clique of size 11 includes User with ID 35 (user with highest degree) in common.

Size of cliques	No. of cliques
2	6405
3	4223
4	2636
5	2051
6	1658
7	1122
8	645
9	355
10	68
11	24

The figure 5 shows 24 Cliques of Size 11. These cliques suggest that they establish positive trading trust within each other.

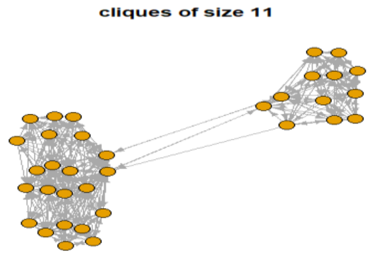

Fig. 5

The OTC network has a total of 19187 cliques with 4223 triadic closures. In an OTC transaction community people interact within themselves in small clusters where trust is high, but only within that cluster.

h. Embeddedness:

This measure is used to find the users which as structural holes in the network. They have lowest constraint value/ embeddedness. This OTC community has rater 35 with lowest embeddedness values of 0.0037 indicating this rater rates very less.

h. Hub Score and Authority Score:

A user is considered a hub if it points to many other users and is considered an authority if it has many nodes/users linking to it. Rater 1810 has highest hub score of 1 indicating that it is connected to many ratees, hence high outdegree. Also, rater 4531 has highest authority score of 1 implying that it has many ratees are connected to it i.e. maximum in-degree.

A further investigation of average weight of both suggests they don't have sufficiently desirable traits as their average rating is -9.3 and -2.3 respectively and user 4531 is a potential fraudulent trader.

VI. CENTRALITY MEASURES

Centrality measures attempt to locate "who is important in the network". They give a rough estimation of a social power of a user based on how well it is connected to other users.

In a magnanimous network like here, there often arises a question of figuring out what vertices are important? Who are the influential figures, and active within the network? The word influential is coincident with the definition of centrality.

"Influence" can alternatively be conceived as involvement in the cohesiveness of the network. This allows centralities to be classified based on how they measure cohesiveness.[7] A further conclusion is that a centrality which is appropriate for one category will often "get it wrong" when applied to a different category.[8]

When centralities are categorized by their approach to cohesiveness, it becomes apparent that most centralities inhabit one category. The count of the number of walks starting from a given vertex differs only in how walks are defined and counted. Restricting consideration to this group allows for a soft characterization which places centralities on a spectrum from walks of length one (degree centrality) to infinite walks (eigenvalue centrality).[9][10] The observation that many centralities share this familiar relationship perhaps explains the high rank correlations between these indices.

a. Degree Centrality:
Degree centrality measures the counts of how many neighbors a node has, especially in a directed network, it is evident to understand the interactions in both directions – in and out [11].

So, in our Bitcoin network, user 35, 2642, and 1810 are 3 most people rated by 535, 412, 311, respectively. Interestingly, they are the same ones having most outgoing edges, suggestive enough of their activeness and interactions within the network, with 763, 406, and 404 outgoing edges.

Counter-intuitively, users 35, 2642, and 2028 are those who give the most positive to others and 35, 2642, and 1810 are the one receiving positive feedback. However, users 2125, 1810, and 2266 figuring out bad/fraudulent traders and 3744, 1383, 2028 are users who received the most negative feedback.

b. Node Betweenness:
Betweenness centrality - measures the extent to which a node is located 'between' other pairs of vertices which is an efficient way to ask which ties in a social network are most important for the spread of, say, right or wrong answers[12]. 35, 2642, and 1810 are the top 3 users (or nodes) lying between interactions with 4912540, 2150220, and 1712299 nodes respectively. Alternatively, we can say that they are central to exchange of trades within the network. User 1810 gave an average of -2.3 rating meaning he is not able to predict non-fraudulent users effectively.

c. *Edge Betweenness:*
These are the edges through which maximum of information flows. Edge 17446 has highest betweenness value of 204161.24.

d. Closeness Centrality
The closeness centrality of a vertex in a graph is the inverse of the average shortest-path distance from the vertex to any other vertex in the graph. It can be viewed as the efficiency of each vertex (individual) in spreading information to all other vertices.[14] In our network we observed user 3665 has the highest closeness centrality measure of magnitude 5.106418e-06 suggesting this is a well spread-out graph. Also, the average weight associated with 3665 is -10 suggesting he is distant and sparsely connected.

e. Eigenvector Centrality
Eigenvector centrality considers the number of the neighbors of the node and their connectivity, which further identifies the influential nodes. In Bitcoin-OTC network user 2642 received an average of 1.88 ratings and gave 2.53 rating which is intuitively in line with eigenvector centrality.

VII. CORRELATION AMONG NETWORK MEASURES

All these measures are highly correlated to each other which is quite evident from the table shown.

	degree_OTC	nodebetweens_OTC	edgebetweens_OTC	close_OTC	eig_OTC
degree_OTC	1	0.9242	-0.0015	0.1421	0.8867
nodebetweens_OTC	0.9242	1	0	0.0695	0.7013
edgebetweens_OTC	-0.0015	0	1	-0.0169	-0.0059
close_OTC	0.1421	0.0695	-0.0169	1	0.1795
eig_OTC	0.8867	0.7013	-0.0059	0.1795	1
hub_OTC	0.3572	0.2679	0.0075	0.0673	0.3455
auth_OTC	0.2195	0.1342	-0.0047	0.0177	0.2663

There is a high correlation between degree, node betweenness and eigen vector centrality measures.

The "log-log degree distribution" shows that the intensity decreases with increase in degree. This implies that some raters are highly active, but majority participate very less in second transactions.

The k-NN degree measure plot instills the importance of connectivity as small cohesive user groups are also trading with highly dense user groups. These two plots are shown in figure 6.

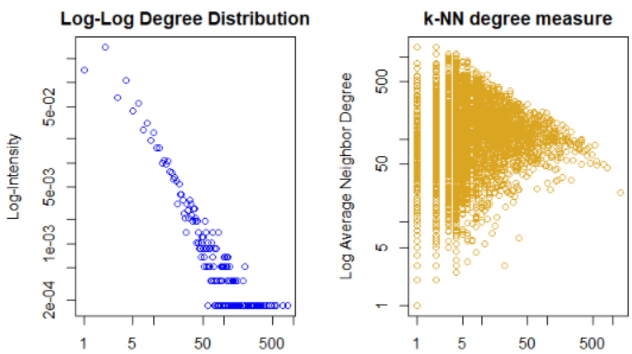

Fig. 6

These OTC network centrality measures are corelated with each other which is depicted in figures 7 and 8. Figure 7 depicts that there is positive correlation between average betweenness of edges and edge weights to some extent. On the right, we see a negative correlation between average betweenness of nodes and local clustering coefficients.

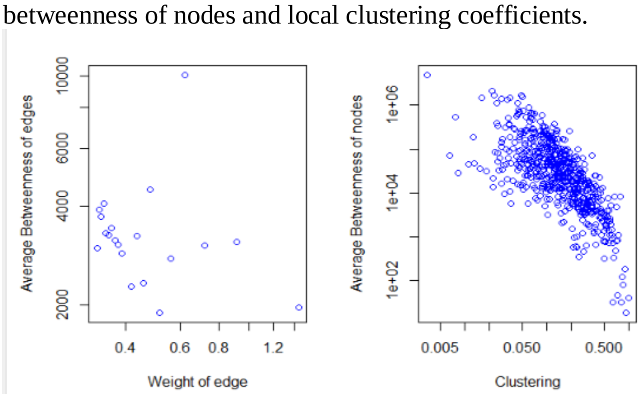

Fig. 7

Figure 8 shows more corelated plots. We can see that there is a negative correlation between individuals' average clustering coefficient and degree which implies that for the given OTC network the global clustering is lower than the average clustering. It is also observed that average betweenness of nodes increases as their degree increases. Further, the embeddedness decreases with increase in degree in bottom-right plot.

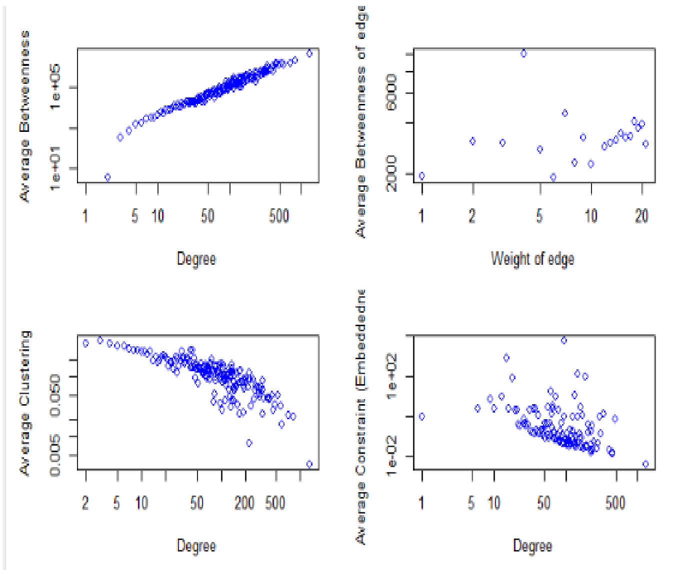

Fig. 8

All these plots show that all the network characteristics and centrality measures are inter-related and can be studied with the help of a network.

VII. COMMUNITY DETECTION

Most of the complex networks usually have modular or community structure and appear as a combination of groups that are fairly independent of each other. Vertices of the same community usually share some common behaviors. For instance people of the same community usually have a set of common properties such as having similar hobbies, working on a research with the same topic and so on. Thus, finding communities enables us not only to extract useful information from complex networks but also to understand how different groups or communities in a network evolve.

Community detection is defined as the process of discovering the cohesive groups or clusters in the network. It forms one of the key tasks of social network analysis. We did this analysis to find users of similar nature e.g. most active users, least active users and so on.

Moving forward to community detection we have used various algorithm to segregate communities in OTC network. Modularity measures the density of links inside these communities as compared to links between communities (groups/clusters/communities). Due to the presence of directed edges in our Bitcoin-OTC network we relied on walktrap and infomap community detection algorithms to understand the relations amongst the users. We have avoided fast-greedy and spin-glass algorithms as it is a directed network. Below figures show the network diagram of walktrap and infomap algorithms.

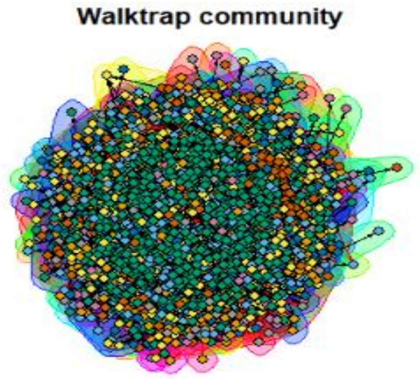

Fig. 9 – Walktrap Community

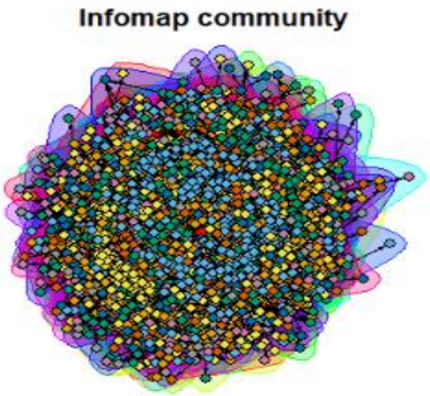

Fig. 10 – Infomap Community

The number of communities and modularity values are shown in the below table. Walktrap segregates the OTC network in 461 communities which is much less than Infomap.

Algorithms	No. of communities	Modularity
Walk-trap	329	0.4277861
Infomap	426	0.3555312

We have also segregated the communities into positive and negative network diagrams depending on their ratings.

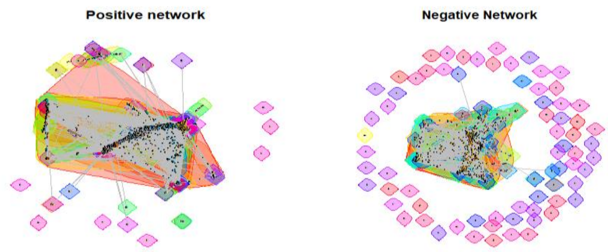

Fig. 10 – Positive and Negative Community

After analyzing all the network metrics of Bitcoin OTC Community, it is evident that majority of members are not directly connected with each other but broken down into several clusters.

VIII. TRUST-LEVEL DETECTION

To detect whether a particular member in OTC community is trustworthy or fraudulent, we used the average of hub scores and authority scores for a particular rater. The average weight of hub score denotes the average rating a particular user gives to others. Similarly, the average weight of authority score denotes the average rating a particular user receives from others.

	Average rating	
Member	**Received**	**Given**
2642	2.53	1.88
3665	-10.00	2.17
35	1.90	1.15
1810	0.74	2.32

Here, we observe that member 3665 has received extremely bad ratings, thus it is not advisable for a new user to have transaction with him as he can be tagged as a fraud. On the other hand, user 2642 has good reviews making him a suitable user to have transaction with.

IX. CONCLUSION

After analyzing all the network metrics of OTC Community it is evident that majority of members are not directly connected with each other but broken down into several clusters. Through this study we also understood the cohesiveness of the bitcoin-OTC network. Additionally, we figured a Naive approach to detect average rating of a user which could aid new users to predict if the transaction will be fruitful or fraudulent learning from the other user ratings.

X. LIMITATIONS AND FUTURE SCOPE

This research paper aims at analyzing the OTC community network overall without much diving into positive and negative networks. Hence, drawing inferences can be bit questionable.

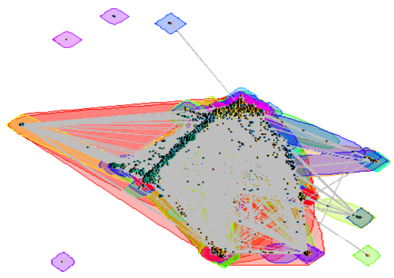

Fig. 11– Whole Community

Also, – these circled communities as shown in figure 11 will need to grow, and since the random linking will require a lot of time, whereas having a **link prediction/edge weight prediction** will help in connecting the right communities to the right users to increase the customer engagement in the network. This link prediction can be the future scope of this research study.

XI. REFERENCES

[1] Ilaria B., Sylvie H., Guillaume D., Floriana G. (2018) The Anatomy of a Web of Trust: The Bitcoin-OTC Market. In: Staab S., Koltsova O., Ignatov D. (eds) Social Informatics. SocInfo 2018. Lecture Notes in Computer Science, vol 11185. Springer, Cham.

[2] M. Shahriari and M. Jalili, “Ranking nodes in signed social networks,” Social Network Analysis and Mining, vol. 4, no. 1, 2014.

[3] M. Shafaei and M. Jalili, “Community structure and information cascade in signed networks,” New Generation Computing, vol. 32, no. 3-4, pp. 257–269, 2014.

[4] Y. Qian and S. Adali, “Foundations of trust and distrust in networks: Extended structural balance theory,” TWEB, vol. 8, no. 3, p. 13, 2014. https://snap.stanford.edu/data/soc-sign-bitcoinotc.html. Accessed 17 November 2020.

[5] Richters, Oliver, and Tiago P. Peixoto. "Trust transitivity in social networks." PloS one 6.4 (2011): e18384.

[6] Borgatti, Stephen P., and Martin G. Everett. "A graph-theoretic perspective on centrality." Social networks 28.4 (2006): 466-484.

[7] Borgatti, Stephen P. "Centrality and network flow." Social networks 27.1 (2005): 55-71.

[8] Bonacich, Phillip. "Power and centrality: A family of measures." American journal of sociology 92.5 (1987): 1170-1182.

[9] Benzi, Michele, and Christine Klymko. "A matrix analysis of different centrality measures." arXiv preprint arXiv:1312.6722 (2014).

[10] Srinivas, Amedapu, and R. Leela Velusamy. "Identification of influential nodes from social networks based on Enhanced Degree Centrality Measure." 2015 IEEE International Advance Computing Conference (IACC). IEEE, 2015

[11] Newman, Mark EJ. "A measure of betweenness centrality based on random walks." Social networks 27.1 (2005): 39-54.

[12] Okamoto, Kazuya, Wei Chen, and Xiang-Yang Li. "Ranking of closeness centrality for large-scale social networks." International workshop on frontiers in algorithmics. Springer, Berlin, Heidelberg, 2008.