Networks Theory

Social Network Analysis

Student:
Can Montero Christian Javier

Teacher:
Gamboa Angulo Didier Omar

Abstract:
This project analyses a particular case study on the Bitcoin Alpha platform, where users evaluate one another according to trust. User interactions over time are documented by trust ratings in the dataset used in this project. To shed light on the dynamics of trust in the Bitcoin Alpha community, the main goal of this project is to use a variety of SNA techniques to analyse the temporal and structural elements of these interactions.

July 7, 2024

I. Introduction

Users rate each other's trustworthiness on the Bitcoin Alpha trust network, which shows the ties between users. Understanding trust distribution, identifying prominent nodes, and improving security in decentralised systems all depend on an analysis of this network.

The Bitcoin Alpha network, which has been used in numerous research projects, frequently displays a core of extremely reliable users and a large number of peripheral users. These results contribute to better trust management and utilisation in digital contexts.

Due to the directional nature of trust ratings and their variable value, this network is categorised as directed and weighted. It is also scant because there aren't many user-to-user trust ratings. Understanding the network's structure and trust distribution is made easier by classifying and evaluating the network.

With an emphasis on degree distribution, clustering coefficient, and shortest path length, the network will be visualized to compute fundamental metrics, analyse the findings, and to enhance the field of social network analysis, this analysis will shed light on user interactions and the network's connection.

II. Data Loading and Preprocessing

Data loading and preprocessing are the initial steps in the analytical process. The dataset, a CSV file `soc-sign-bitcoinalpha.csv.gz`, includes timestamps, user IDs, and trust ratings.

To make handling and analysis easier, the dataset is put into a pandas DataFrame in order to comprehend the structure and categories of data; the first step of the process is to look at the first few rows. Finding any possible problems or oddities in the dataset that could influence later stages of the analytic process depends on this step.

	SOURCE	TARGET	RATING	TIME
0	7188	1	10	1407470400
1	430	1	10	1376539200
2	3134	1	10	1369713600
3	3026	1	10	1350014400
4	3010	1	10	1347854400

Fig. 1. Bitcoin Alpha's dataframe.

III. Network Construction

The next stage is to build a network from the dataset. Every user is represented as a node in this context, and a directed edge is formed by the trust rating between two users. The trust rating's direction is shown by the edge's direction, and its magnitude is represented by the edge's weight.

This network is built using NetworkX, a Python toolkit for the construction, modification, and study of complex networks' dynamics, structure, and functions. With the help of this representation, different SNA approaches may be used to identify the underlying dynamics and trends in the user interactions.

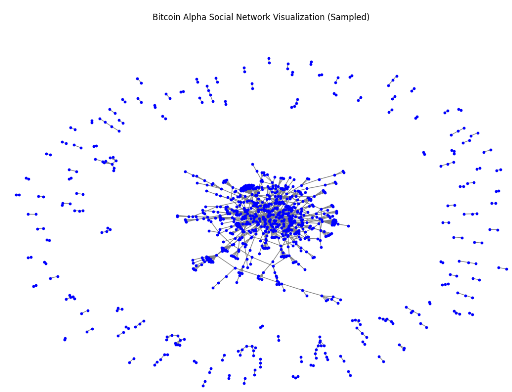

Fig. 2. Bitcoin Alpha social network visualization.

IV. Descriptive Network Analysis

After the network is built, a descriptive analysis is performed to comprehend the fundamental characteristics. This involves figuring out different metrics that offer a high-level summary of the connectivity and complexity of the network.

A. Network Characteristics

In order to comprehend the structure of the network, this section determines important network properties such as the number of nodes and edges, the average path length of the largest connected component, and the clustering coefficient of the network. These metrics reveal information about the size, connections, navigability, and clustering patterns of the network.

These features are the following:

```
Size (Number of Nodes): 977
Number of Links (Edges): 985
Average Path Length: 6.133620669954431
Clustering Coefficient: 0.003993739056767662
```

B. Distance Metrics

Key structural parameters such as diameter (longest shortest path) eccentricity (maximum distance from each node), radius (lowest eccentricity), center (nodes with eccentricity equal to the radius), and periphery (nodes with the highest eccentricity) are all found.

This also determines the average distance between nodes and the overall length of all shortest pathways.

The structure of the network is shown like this:

```
Diameter: 17
Eccentricity: {1: 11, 22: 12... 1248: 12, 601: 14}
Radius: 9
Center: [33, 16, 326, 130]
Periphery: [3386, 2104, 2991]
Average Distance: 2.8146991459301645
```

V. Centrality Measures

Different types of centrality measures are calculated, including pagerank centrality, betweenness centrality, and closeness centrality.

A. PageRank

PageRank is a measure of node relevance based on connections. It is calculated using `nx.pagerank(G_largest_cc)`, and the most influential nodes are highlighted by showing the top five nodes by PageRank. PageRank uses the arrangement of links between nodes to reveal which nodes have a lot of power within the network.

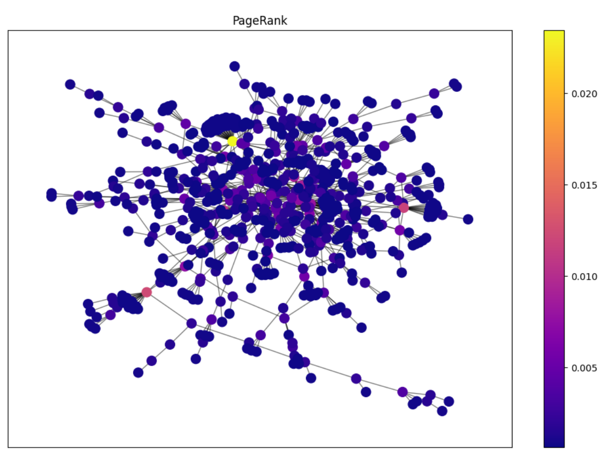

Fig. 3. PageRank visualization.

B. Betweenness Centrality

A significant one is betweenness centrality, which measures how frequently a node appears on the shortest pathways between other nodes, hence quantifying the value of nodes in facilitating communication across the network. Understanding the distribution and importance of core nodes throughout the network, especially those with high betweenness centrality, is easier with this visualization.

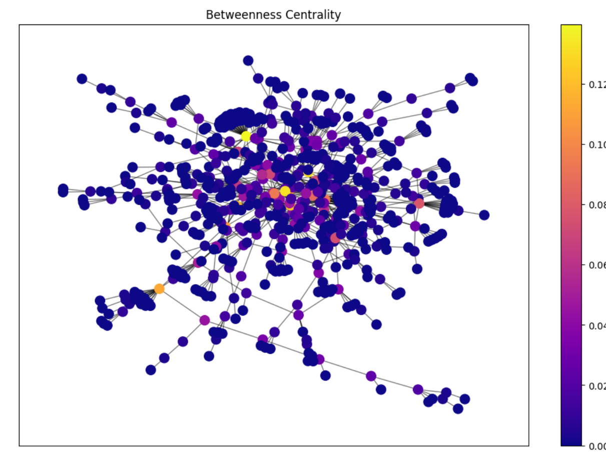

Fig. 4. Betweenness centrality plot.

C. Closeness Centrality

Closeness centrality, on the other hand, measures the speed at which a user can interact with every other user in the network. These measures are useful in identifying the most influential or significant nodes in the network.

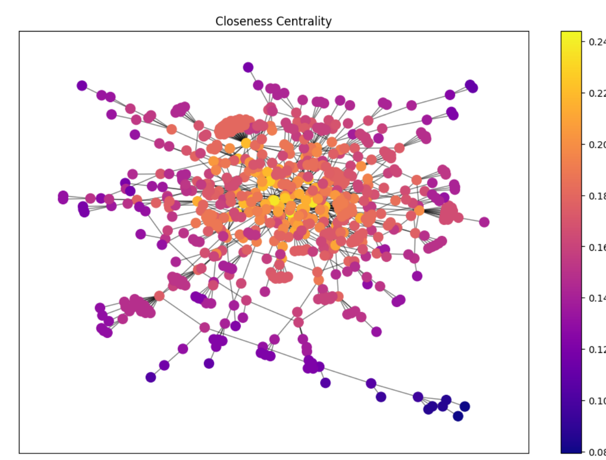

Fig. 5. Closeness centrality graphic.

VI. Degree Distribution

Understanding a network's structural aspects requires first analysing its degree distribution, which offers important information regarding node presence and connectivity patterns.

This approach helps in revealing the underlying dynamics of trust and locating important individuals who influence the connections of trust within the community. This information can be used to build tactics that strengthen the platform's trust mechanisms and enhance user interactions.

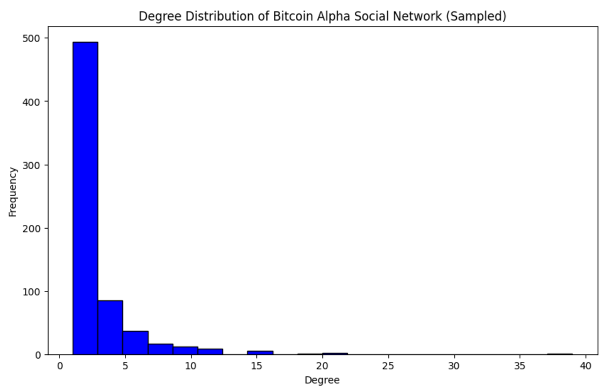

Fig. 6. Degree distribution plot

VII. Community Detection

Groups of nodes with stronger internal connections than external connections make up communities inside a network. Identifying these communities can shed light on user clustering patterns and the presence of more reliable subgroups. Several techniques, including the Girvan-Newman algorithm, are used to find communities on the Bitcoin Alpha network. Finding these communities is helpful to understand the social structure and how trust-based clusters are formed.

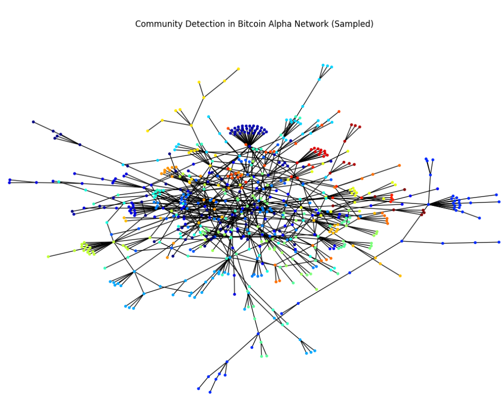

Fig. 7. Community detection in Bitcoin Alpha network.

VIII. Conclusion

In summary, the application of support vector analysis (SNA) methods on the Bitcoin Alpha dataset reveals significant trends and perspectives regarding user behaviour and trust relationships.

Considering comparable online platforms, where user engagement and community building are heavily reliant on reputation and trust, these findings carry more implications.

Platform administrators and developers can put these findings to use by putting methods into place that improve trust mechanisms, strengthen community ties, and ultimately create a more reliable and robust online environment.

References

- SNAP: Signed network datasets: Bitcoin Alpha web of trust network. (n.d.). Stanford.edu. Retrieved July 7, 2024, from http://snap.stanford.edu/data/soc-sign-bitcoin-alpha.html
- Newman, M. (2018). *Networks: The empirical study of Networks*. Oxford University Press.
- Zafarani, R., Abbasi, M. A., & Liu, H. (2014).
- Barabási, A.-L. (2016). *Network Science*.
- Menczer, F., Fortunato, S., & Davis, C. (2020). *A First Course in Network Science*.