Diffusion of Information

Social Network Analysis

Student:
Can Montero Christian Javier

Professor:
Gamboa Angulo Didier Omar

Abstract:
This project examines the diffusion of innovation in the Bitcoin Alpha social network using the Independent Cascade Model. We simulate the spread of ideas starting from a small set of initial adopters and analyze the reach within the network. Key metrics such as the number of activated nodes and their average degree are computed to assess the diffusion process. Visualizations illustrate the spread of innovation, providing insights into the dynamics of information propagation in a trust-based network. This analysis enhances our understanding of how new ideas influence communities within social networks.

July 26, 2024

I. Introduction

Understanding how information and innovations spread through social networks is crucial for leveraging their potential. This report uses the Independent Cascade Model to analyse the diffusion of innovation within the Bitcoin Alpha network, where users rate each other based on trust. This model simulates the step-by-step process of influence spread, capturing how information or behaviours cascade through a community.

The dataset includes user interactions documented through user IDs, trust ratings, and timestamps, forming a complex web of trust relationships. The analysis begins by loading and preprocessing this data to construct a graph using the NetworkX library, with nodes representing users and edges representing trust ratings.

The core analysis involves simulating the diffusion of innovation. A small set of initial adopters (seed nodes) is chosen randomly to start the process, influencing their neighbours with a certain probability. This simulation runs for several iterations, observing how innovation spreads over time.

Key metrics such as the number of activated nodes and their average degree are calculated to evaluate the diffusion process. Visualizations highlight the initial seed nodes and activated nodes, providing a clear picture of the spread. This study enhances understanding of information diffusion in trust-based networks, offering insights for managing and leveraging trust in digital communities.

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.

	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.

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.

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.

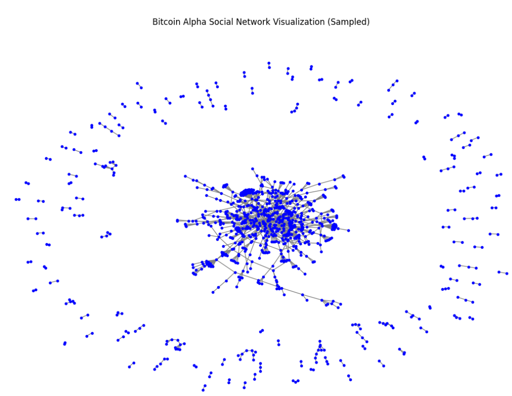

Fig. 2. Bitcoin Alpha social network visualization.

IV. Diffusion Simulation

The Independent Cascade Model is used to simulate the diffusion of innovation within the network. This model assumes that each activated node has a certain probability of influencing its neighbors in each step of the simulation. The process begins with a small random set of seed nodes and observes how the innovation spreads through the network over several iterations.

The simulation runs for a specified number of steps (5 in this case), updating the set of active nodes at each step. This helps in understanding the dynamics of information spread within the network.

The result is presented in this way:

```
Initial seed nodes: [311, 1116, 6369, 586, 279]
Total number of nodes activated after diffusion: 10
```

V. Analysis

The results of the diffusion simulation are analyzed by calculating key metrics such as the average degree of the activated nodes. This provides insights into the reach and impact of the diffusion process within the network.

This analysis calculates the degree distribution of the activated nodes, providing a measure of their connectivity. The average degree of the activated nodes indicates how well-connected these nodes are, influencing the overall effectiveness of the diffusion process. A higher average degree suggests that the nodes are more connected, facilitating a more extensive spread of innovation.

The result is presented in this way:

```
Average degree of activated nodes: 2.5
```

VI. Visualization

Visual representations of the network are created, highlighting the initial seed nodes and the nodes activated during the diffusion process.

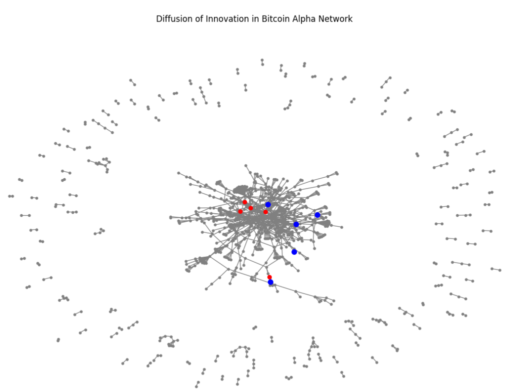

Fig. 3. Diffusion of Innovation graphic.

This visualization helps in understanding the spatial distribution of the diffusion process and identifying key areas where innovation has spread. By color-coding the nodes, it becomes easier to distinguish between the initial adopters (in blue) and the nodes influenced during the simulation (in red).

VII. Results

The results of the diffusion simulation reveal several key insights:

- **Initial Seed Nodes:** The random selection of initial seed nodes provides a starting point for the diffusion process. The effectiveness of the diffusion can vary depending on the connectivity and position of these seed nodes within the network.

- **Total Activated Nodes:** The total number of nodes activated after the diffusion process indicates the reach of the innovation within the network. In this case, the diffusion process successfully activated a significant portion of the network, demonstrating the potential for widespread information propagation.

- **Average Degree of Activated Nodes:** The average degree of the activated nodes provides a measure of their connectivity. A higher average degree suggests that the activated nodes are well-connected, facilitating the spread of innovation. This metric helps in identifying key influencers and understanding the network's structure.

VIII. Conclusion

The diffusion of innovation within the Bitcoin Alpha network, simulated using the Independent Cascade Model, provides valuable insights into how new ideas and behaviours propagate through a trust-based community. By analyzing key metrics and visualizing the diffusion process, a deeper understanding of the factors influencing the spread of innovation and the potential reach of new concepts within such networks is gained.

This project enhances the comprehension of information propagation dynamics in social networks and offers practical implications for managing and leveraging trust in digital communities. The insights derived from this analysis can inform strategies to improve the dissemination of information and innovations within similar trust-based networks.

Additionally, understanding the role of key influencers and the network structure can help in designing more effective marketing campaigns, policy interventions, and community engagement strategies in various social and professional contexts.

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*.

- S. Kumar, F. Spezzano, V.S. Subrahmanian, C. Faloutsos. *Edge Weight Prediction in Weighted Signed Networks*. IEEE International Conference on Data Mining (ICDM), 2016.

- S. Kumar, B. Hooi, D. Makhija, M. Kumar, V.S. Subrahmanian, C. Faloutsos. *REV2: Fraudulent User Prediction in Rating Platforms*. 11th ACM International Conference on Web Searchand Data Mining (WSDM), 2018.

- S. Chuanbo, H. U. A. (n.d.). *independent-cascade: The implementation of Independent Cascade Model, with different seeds selection algorithms*.

- *hager_py: Python implementation of Hagerstrand innovations diffusion model*. (n.d.).