Complex Network Dynamics, CS-484
Project Report, Phase A
Papadakis Georgios (4975)
Supervisor: Dr. Lionakis Panagiotis

"Structure and Dynamics of Trust in Anonymous and Identified Online Networks"

Motivation

In the "highly connected world"[1] we live in, the concept of trust constitutes a cornerstone for the success of online interactions, whether they concern e-commerce, the creation of social bonds, or the exchange of information. In traditional networks, trust is often founded on repeated interaction and the existence of a stable and verifiable identity.

The central motivation of this work stems from a fundamental question: *What happens to the structure of trust when the factor of identity is removed?*

Anonymity, while providing freedom - and in many cases protection - simultaneously introduces significant risks, such as fraud and malicious behaviors. Many platforms, especially in the cryptocurrency sector like Bitcoin, operate under a regime of full anonymity. To ensure security and stability within their network, they are forced to develop inherent reputation mechanisms which are not based on external user identities, but exclusively on their behavior within the network.

These mechanisms create complex, signed social networks, which we can analyze using the theories of Network Dynamics. Concepts such as Structural Balance and Link Analysis for reputation calculation (PageRank) cease to be theoretical ideas and become critical tools for understanding the sustainability of these systems.

The purpose of this work is to empirically study these mechanisms by comparing the structure of a purely anonymous network (Bitcoin OTC) with that of a traditional, identified social network (Epinions).

Research Question and Hypotheses

The central question of this work is: *"How does anonymity (its presence and absence) affect the structure, stability, and diffusion patterns of trust in online signed networks?"*

To provide a reliable answer to the above question, we will examine the following:

I. *Polarization:* Is the distribution of trust and distrust more polarized and extreme in the anonymous network compared to the identified one?
II. *Structure:* Does the anonymous network form fewer balanced local structures (triangles) compared to the identified one?
III. *Robustness:* Is the anonymous network more fragile? That is, does the total reputation of the network depend on a disproportionately small number of nodes, the removal of which would lead to a collapse of connectivity?
IV. *Reciprocity:* Does anonymity reduce the probability of mutual trust between users?

I formulate the following initial hypotheses:

I. *Hypothesis 1 (Polarization):* The anonymous network Bitcoin OTC will exhibit higher volatility and more extreme ratings. The -10 to +10 scale allows for intense signaling, which we assume users utilize to warn about

---

[1] Easley & Kleinberg, 2010, "Networks, Crowds and Markets: Reasoning about a **Highly Connected World**

scams or reward the most trustworthy users, in contrast to the binary approach of Epinions (-1/+1).

II. *Hypothesis 2 (Robustness):* The Bitcoin OTC network will be more fragile, as trust will accumulate in fewer nodes (Rich-Get-Richer), making the network vulnerable to their removal. Conversely, in the Epinions network, trust is expected to be more distributed.

III. *Hypothesis 3 (Structural Balance):* The anonymous network will display a lower percentage of balanced triangles. The "social pressure" to comply with rules that lead to balance (i.e., "the friend of my friend is my friend" and "the enemy of my enemy is my friend") is expected to be visibly weaker in the absence of stable and public user identity.

Dataset Description

For the empirical analysis of the above hypotheses, two publicly available datasets from the *Stanford Network Analysis Project (SNAP)* will be used. Each of these networks constitutes a directed and signed graph, where edges represent trust relationships.

<u>Bitcoin OTC Trust Weighted Network (Anonymous):</u>

- *Source and Context:* This network records trust ratings between users of the P2P *(peer-to-peer)* trading platform Bitcoin OTC *(Bitcoin Over-The-Counter)*. Since Bitcoin users are anonymous, the platform relies on the Web of Trust system to protect members from fraudulent transactions.
- *Structure:* Each node is an anonymous user. Each directed edge (u, v) represents a rating given by user *u* to user *v* after a transaction.
- *Data Format:* Each entry is in the format: SOURCE, TARGET, RATING, TIME
- *Basic Statistics (pre-cleaning):*
    - Nodes (Users): 5.881
    - Edges (Ratings): 35.592
    - Rating Scale: Integers in the range [-10 (total distrust), +10 (total trust)].
    - Percentage of Positive Edges: ~89%

<u>Epinions Social Network (Identified):</u>

- *Source and Context:* The network comes from the consumer review platform Epinions.com (today known as Shopping.com). Users maintained public, pseudo-identified profiles and could declare which other users they trust or distrust.
- *Structure:* Each node is a user. Each directed edge (u, v) represents the declaration of trust (or distrust) by *u* towards *v*.
- *Data Format:* Each entry is in the format: SOURCE, TARGET, RATING.
- *Basic Statistics (pre-cleaning):*
    - Nodes (Users): 131.828
    - Edges (Ratings): 841.372
    - Rating Scale: -1 (distrust), +1 (trust)
    - Connectivity (WCC): 90.4% of nodes belong to the Largest Weakly Connected Component.
    - Connectivity (SCC): 31.4% of nodes belong to the Largest Strongly Connected Component.
    - Average Clustering Coefficient: 0.1279

Preliminary Analysis

The Preliminary Analysis includes two stages: (a) data cleaning and extraction of final basic statistics, and (b) visualization of initial distributions to draw the first conclusions regarding our hypotheses.

<u>Data Cleaning and Final Statistics</u>

Prior to analysis, the data underwent a rigorous cleaning process to ensure integrity. For both datasets, the following steps were performed:

I. *Removal of Missing Data:* Rows with NaN values were removed.
II. *Removal of Self-Loops:* Entries where SOURCE == TARGET were removed.
III. *Removal of Duplicates:* Fully duplicate entries were removed.
IV. *Rating Validation:*
    A. For Bitcoin OTC, we kept only entries with integer RATING in [-10,10].
    B. For Epinions, we kept only entries with RATING equal to -1 or +1.

The results of this process are shown in Table 1.

Metric	Bitcoin OTC (Anonymous)	Epinions (Identified)
Total Nodes	5.881	131.580
Total Edges	35.592	840.799
Density	0.001029	0.000049

Table 1: Basic Network Statistics (after cleaning)

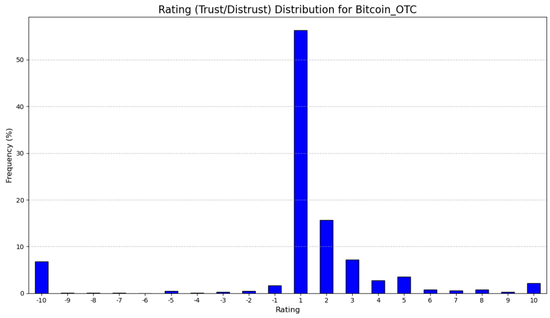
For the Bitcoin OTC dataset, the initial set (35,592 lines) was found to be extremely clean. For Epinions, 573 entries were removed as self-loops, leading to a final of 840,799 edges. This reduction of edges led to a decrease of the nodes, from the initial 131,828 to 131,580 (reduced by 248).

We observe that the Epinions network is significantly larger (approximately 22 times more nodes and 23 times more edges), but the Bitcoin OTC network is 21 times more dense. This suggests that, although smaller, the Bitcoin OTC community is more "tight-knit", with a higher ratio of interactions per user.

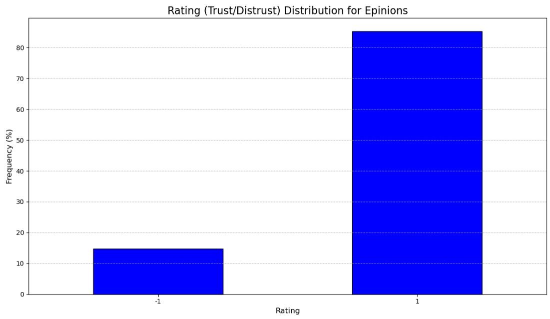
<u>Rating Distribution Analysis</u>

An interesting initial finding is that the anonymous network displays a smaller percentage of negative ratings compared to the identified one. Specifically:

- Bitcoin OTC: 89.99% Positive (>0) and 10.01% Negative (<0)
- Epinions: 85.29% Positive (+1) and 14.71% Negative (-1)



Graph 1: Rating Distribution of Bitcoin OTC



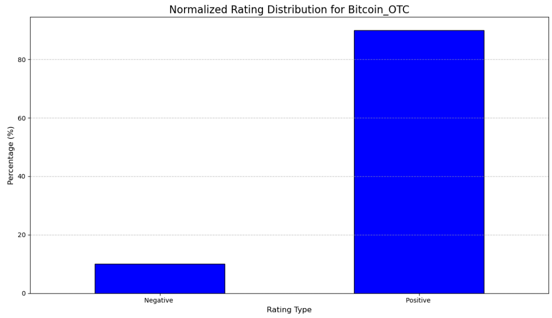
Graph 2: Rating Distribution of Epinions

Observations:

- Bitcoin OTC (Graph 1): While the majority of positive ratings are +1, there is strong signaling at the extremes. The -10 rating (total distrust) accounts for about 6% of the total, and +10 (total trust) about 3%. This confirms my hypothesis that users utilize the full scale to strongly warn about scams or reward trustworthiness.
- Epinions (Graph 2): The graph is strictly binary.

For a direct comparison, I created Graph 3, where I "normalized" the Bitcoin

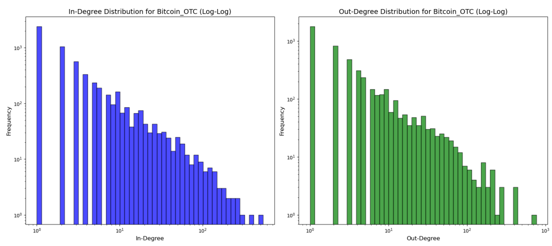
OTC ratings into {Negative, Positive}. Comparison with Graph 2 highlights the aforementioned difference in overall percentages.



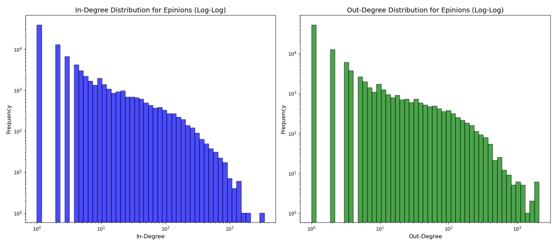
Graph 3: Normalized Rating Distribution of Bitcoin OTC

Degree Distribution Analysis

Degree Distribution shows us how the "popularity" (number of connections) is distributed among the nodes.



Graph 4: In/Out-Degree Distribution (Log-Log) of Bitcoin OTC



Graph 5: In/Out-Degree Distribution (Log-Log) of Epinions

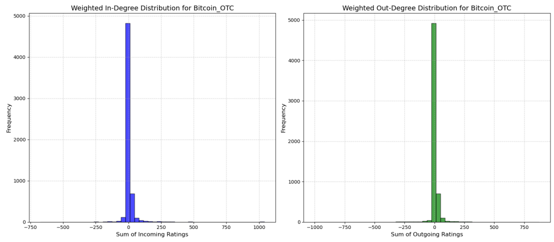
Both networks clearly follow a Power Law distribution, as expected for social networks:

- The overwhelming majority of nodes have very low In-Degree (received few ratings) and Out-Degree (did few ratings).
- A minority of nodes functions as Hubs, possessing hundreds of connections.

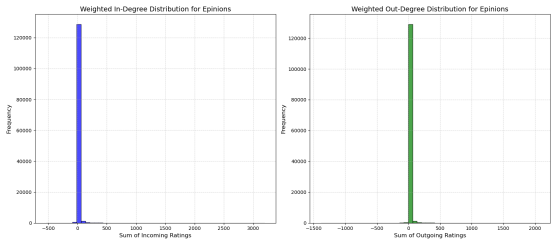
The main visual difference is that the "tail" in Bitcoin OTC (Graph 4) seems to drop more abruptly compared to the tail of Epinions (Graph 5), where the decrease in degrees is smoother and more extended. This suggests that Epinions, although less dense, manages to maintain more nodes with "medium" number of connections[2], indicating more active users overall.

Weighted Degree Distribution Analysis

Here we measure net reputation, i.e., the sum of rating values a node received (In-Degree) or gave (Out-Degree).



Graph 6: Weighted In/Out-Degree Distribution (with real weights) of Bitcoin OTC



Graph 7: Weighted In/Out-Degree Distribution (with real weights) of Epinions

We observe that the x-axes in Graph 6 for Bitcoin OTC take values from -750 to 1,000 (approximately) for In-Degree, meaning there were users who received a cumulative rating near these values, and values from -1,000 to 800 (approximately) for Out-Degree, meaning there were users who gave a cumulative rating near these values.

Similarly, in Graph 7 for Epinions, we observe that the x-axis for In-Degree takes values from -600 to 3,000 (approximately), and the x-axis for Out-Degree takes values from -1,500 to 2,000 (approximately).

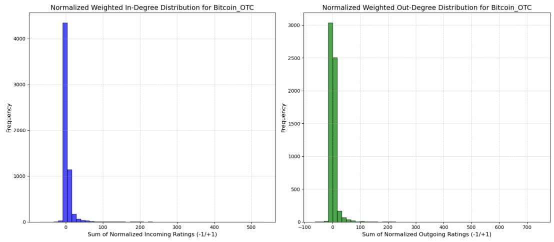
It is important to mention that the fact that bars are not visible at large (absolute) values of x does not mean they do not exist, but that the frequency is so small that their

[2] A larger percentage of nodes compared to Bitcoin OTC, because in absolute numbers it has more nodes anyway.

bars are invisible, especially in comparison to the column near x=0.

The finding is almost identical in both networks: a vast majority of users have a net reputation very close to zero. This suggests that most users either received minimal ratings, or the ratings they received canceled each other out.

To ensure a fair comparison, I normalized Bitcoin OTC so that it also uses -1/+1 weights like Epinions.



Graph 8: Weighted In/Out-Degree Distribution (with normalized weights) of Bitcoin OTC

In Graph 8 for Bitcoin OTC, we observe that the x-axis for In-Degree takes values from -100 to 600 (approximately), and the x-axis for Out-Degree takes values from -100 to 800 (approximately).

The comparison between Graph 7 (Epinions) and Graph 8 (Bitcoin OTC - Normalized) is revealing: users in the Epinions network seem to make more negative ratings. Comparing the pure number of weights, even with the normalized Bitcoin OTC, does not yield a sound conclusion because the users of the Epinions network are 22 times those of Bitcoin OTC. I make the following comparison:

- The Weighted In-Degree of the graphs takes values [-600, 3,000] in Epinions and [-100, 600] in Bitcoin OTC. If I multiply the Bitcoin OTC values by 6 to get the same lower bound, the upper bound becomes 3,600 (clearly more positive than 3,000). While if I multiply them by 5 to get the same upper bound, the lower bound becomes -500 (again more positive than -600).
- Similarly, the Weighted Out-Degree of the graphs takes values [-1,500, 2,000] in Epinions and [-100, 800] in Bitcoin OTC. If I multiply the Bitcoin OTC values by 15 to get the same lower bound, the upper bound becomes 12,000 (clearly more positive than 2,000). While if I multiply them by 2.5 to get the same upper bound, the lower bound becomes -250 (again more positive than -1,500).

Comparing our observations in the Preliminary Analysis chapter and the hypotheses I made in the Research Question and Hypotheses section, I conclude that I expected anonymous networks to be much more “negative” compared to identified ones (even pseudo-identified ones like Epinions). However, Hypothesis 1 was not entirely wrong, because by viewing Graphs 6 and 8 we see that users indeed utilize the -10 to +10 scale to warn about scams or reward the most trustworthy users, in contrast to the binary approach of Epinions (-1/+1).

Tools and Planned Methodology

<u>Software Tools</u>

Analysis is performed exclusively using Python, utilizing the following libraries:

I. *NetworkX*: For graph creation, manipulation, and metric calculation.
II. *Pandas*: For data management and preprocessing.
III. *Matplotlib* & *NumPy*: For visualization and numerical calculations (e.g., log-binning).

<u>Analysis Methodology</u>

I will divide the methodology of my analysis in this project into three distinct phases, which cover the structure, centrality, and dynamics of the signed and directed networks, linking the findings to the theory of Easley and Kleinberg.

Phase A: Structure & Connectivity

In this phase, I will examine if the trust networks exhibit "Small World" properties and how information/trust diffuses within them.

- *Power Law Check*: I will study the exponent *a* of the degree distributions (which appended to follow a Power Law in the Preliminary Analysis) to mathematically confirm the "Rich-Get-Richer" phenomenon.
    - A strict Power Law distribution indicates a "scale-free" network dominated by a few "hubs". This is the structural precursor of Hypothesis 2; if *a* is significantly different between the networks, it implies different mechanisms of reputation accumulation (centralized vs. distributed).
- *Clustering Coefficient*: I will measure the tendency of nodes to create closed triangles ("the friends of my friends become my friends").
    - In anonymous networks, we expect lower clustering if users are wary of forming tight circles without verification.
- *Average Path Length*: I will calculate the average distance between all pairs of nodes to determine how "small" the world is in these two Networks.
    - A smaller average path length implies that reputation (or distrust) spreads rapidly across the network.
- *Reciprocity*: I will measure the ratio of bidirectional edges to total edges.
    - To test if anonymity discourages mutual trust. Low reciprocity in Bitcoin OTC would suggest that trust is often one-way, whereas high reciprocity implies mutual social bonds.

Phase B: Centrality & Trust Transitivity

The goal, in this phase of the analysis, is to identify the most important nodes and examine if trustworthiness is a transitive property.

- *PageRank & Eigenvector Centrality*: I will apply Weighted PageRank algorithms in each graph.
    - These metrics are crucial because they measure Global Reputation. In an anonymous network, having many positive ratings is not enough (one could create fake users to rate themselves). PageRank ensures that a user is trusted only if they are trusted by other trusted users. This filters out local manipulation and identifies the true "pillars" of each community.
- *Trust Transitivity (Assortativity)*: I will examine the correlation between the reputation (e.g., Weighted In-Degree or PageRank) of a node and the average reputation of its neighbors.
    - The goal is to see if trustworthy users tend to connect with (rate) other trustworthy users (homophily regarding trustworthiness), i.e., if trustworthiness is "contagious".

Phase C: Signed Networks & Robustness

This is the core analysis of the work, focusing on the sign of the relationships.

- *Structural Balance:* I will count and classify all network triangles into the 4 types of balance theory (Chapter 5 of the book). I will check if balanced triangles (i.e., those with an even number of negative signs) outnumber unbalanced ones, and if this differs between anonymous and identified networks.
    - If Hypothesis 2 is correct, the anonymous Bitcoin OTC

network will have a significantly higher ratio of Unbalanced triads compared to Epinions, indicating that users feel less pressure to align their opinions with their peers due to the lack of identity.

- *Robustness Analysis*: I will simulate "attacks" on the network, sequentially removing nodes with the highest centrality (Hubs). I will monitor how the connectivity of the network collapses (size of the Giant Component) to determine which network is more fragile to the loss of its "pillars" of trust.
    - If a network is fragile, removing just a few key "hubs" (e.g., the top 1%) acts like cutting the main power lines. The Giant Component suddenly shatters into hundreds of tiny, isolated islands. Based on Hypothesis 2, I expect the anonymous network to be more fragile.

Work Plan and Next Steps

This project is already in an advanced stage of preparation (data has been cleaned and initial plots have been made). The next steps are organized as follows:

I. Implementation of Phases A, B and C of my Analysis.
II. Comparison of results between Bitcoin OTC and Epinions.
III. Synthesis of all findings and production of final charts.
IV. Interpretation of results in light of the initial hypotheses.
V. Writing and formatting of the Final Report and Final Presentation.

<u>Expected Outcomes</u>

I expect to show that anonymity in Bitcoin OTC leads to a network that is more fragile, less structurally balanced, and with more polarized rating patterns compared to the pseudo-identified Epinions network.