Complex Network Dynamics, CS-484
Project Report
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 (Sybil Attacks). Platforms operating under full anonymity, such as the Bitcoin OTC trading network, cannot rely on external legal frameworks to enforce contracts. Instead, they are forced to develop inherent reputation mechanisms based exclusively on behavioral history. These mechanisms create complex, signed, and directed social networks.

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). By analyzing concepts such as Structural Balance, Centrality, and Robustness, we aim to understand how anonymity alters the fundamental topology of human trust.

Research Question and Hypotheses

The central research question driving this analysis is: *"How does anonymity (its presence or 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 rating patterns. The lack of social repercussions for negative behavior, combined with the need to strongly signal "scammers", will lead to a polarized distribution compared to the binary nature of the identified network.
II. *Hypothesis 2 (Robustness):* The anonymous network will be more

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

fragile, as we expect trust to accumulate in a very small number of Hubs (highly trusted intermediaries). Consequently, the removal of these few pillars will cause a rapid collapse of network connectivity compared to the identified network, where trust is expected to be more distributed.

III. *Hypothesis 3 (Structural Balance):* The anonymous network will display a significantly 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 a stable public user identity, leading to higher rates of structural inconsistency.

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)

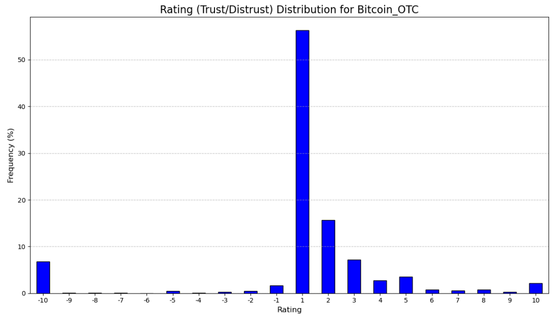
For the Bitcoin OTC dataset, the initial set (35,592 lines) was found to be extremely clean. For Epinions, 573 entries were removed as self-loops, leading to a final of 840,799 edges. This reduction of edges led to a decrease of the nodes, from the initial 131,828 to 131,580 (reduced by 248).

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.

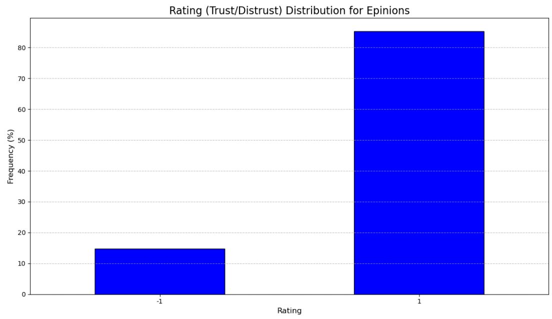
<u>Rating Distribution Analysis</u>

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



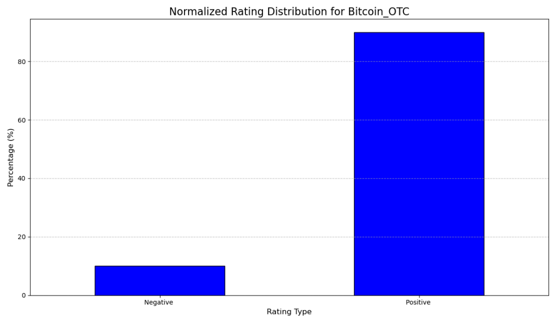
Graph 2: Rating Distribution of Epinions

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

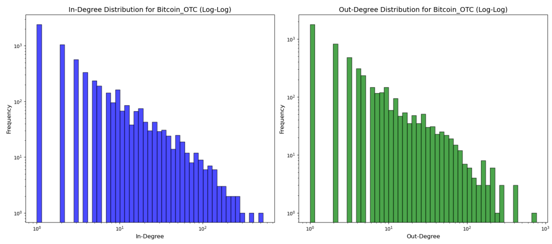
OTC ratings into {Negative, Positive}. Comparison with Graph 2 highlights the aforementioned difference in overall percentages.



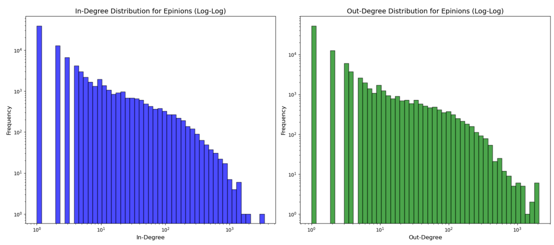
Graph 3: Normalized Rating Distribution of Bitcoin OTC

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

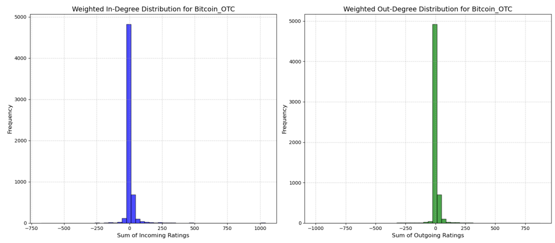
Both networks clearly follow a Power Law distribution, as expected for social networks:

- 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.

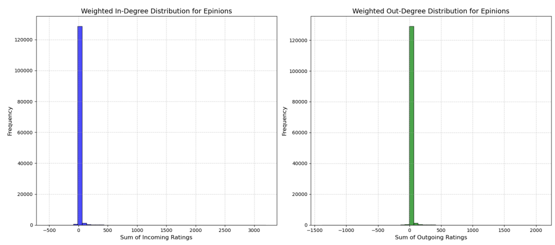
The main visual difference is that the "tail" in Bitcoin OTC (Graph 4) seems to drop more abruptly compared to the tail of Epinions (Graph 5), where the decrease in degrees is smoother and more extended. This suggests that Epinions, although less dense, manages to maintain more nodes with "medium" number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

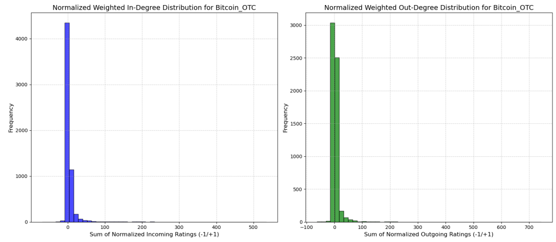
It is important to mention that the fact that bars are not visible at large (absolute) values of x does not mean they do not exist, but that the frequency is so small that their

[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).

Algorithms Implemented

To verify my hypotheses, the following network algorithms and metrics were implemented I provide the formal mathematical definitions used by the NetworkX library for clarity:

I. *Power Law Fitting (Least Squares):* I analyze the degree distribution P(k) by applying logarithmic binning to the degree frequencies. I assume the distribution follows $P(k) \sim k^{-a}$. By taking the logarithm of both sides, I perform a linear regression:

$$ln(P(k)) = -a \cdot ln(k) + c$$

The slope *a* of this line is the scaling exponent. An *a* between 2 and 3 typically indicates a Scale-Free network.

II. *Shortest Path Estimation (Small World):* I calculate the Average

Shortest Path Length to measure trust diffusion speed.

A. For small Graphs (Bitcoin OTC): I implemented an exact BFS computation on the Largest Strongly Connected Component (LSCC).
B. For large Graphs (Epinions): I implemented a sampling algorithm that computes paths from a random subset of five thousand source nodes to approximate L efficiently.

III. *Weighted PageRank:* To measure Global Reputation, I implemented a Weighted PageRank algorithm. For a node u, the PageRank *PR(u)* is calculated iteratively:

$$PR(u) = \frac{1-d}{N} + d \cdot \sum_{u \in B(u)} \frac{PR(u) \cdot w_{vu}}{\sum_{z \in N(v)} w_{vz}}$$

Where:

- d = 0.85 is the damping factor.
- B(u) is the set of nodes linking to u
- w_{vu} is the weight of the edge from v to u (in our case, the rating). This algorithm runs on the subgraph of positive edges only, simulating how trust flows transitively.

IV. *Trust Transitivity (Assortativity):* I measure the Pearson correlation coefficient between a node's PageRank score and the average PageRank of its neighbors. This determines if high-trust users preferentially interact with other high-trust users.

V. *Structural Balance Classification:* I classify all closed triads (triangles) in the underlying undirected signed graph into four categories based on Balance Theory:

A. Balanced: T_3(+++) and T_1(+--)
B. Unbalanced: T_2(++-) and T_0(---)

Optimization: I implemented a set-intersection algorithm to count triangles efficiently in large sparse graphs (Epinions), avoiding the memory overhead of clique enumeration.

VI. *Robustness Simulation (Targeted Attack):* I simulate a coordinated attack by sequentially removing the top N% of nodes based on Degree Centrality. After each batch removal, I recalculate the size of the Giant Connected Component (GCC) to quantify the rate of network collapse.

Tools and Methodology

The analysis was conducted using Python 3.13.5 and the libraries:

- NetworkX: Used for graph data structures, centrality calculations, and topology metrics.
- Pandas: Used for data cleaning, preprocessing, and timestamp management.
- NumPy/SciPy: Used for linear regression (Power Law fit) and correlation analysis.
- Matplotlib: Used for visualizing distributions (Log-Log plots) and robustness curves.

Methodological Workflow: The project follows three distinct phases:

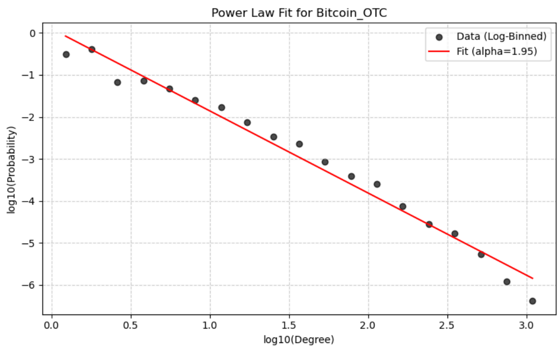
- Phase A: Structure & Connectivity (Power Laws Check, Clustering Coefficient, Average Path Length, Reciprocity).
- Phase B: Centrality & Trust Transitivity (PageRank, Trust Assortativity)
- Phase C: Signed Networks & Robustness (Structural Balance Check and Robustness Analysis)

Results

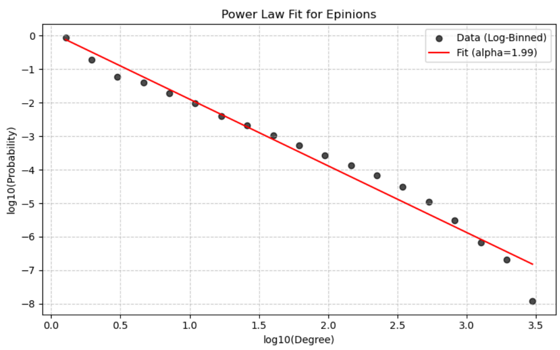
The analysis yielded distinct structural signatures for the anonymous (Bitcoin OTC) and pseudo-identified (Epinions) networks. The quantitative findings are presented below, categorized by the three analysis phases.

<u>Phase A: Structure & Connectivity</u>

- *Power Law (a):* Both networks exhibit Scale-Free properties.



Graph 9: Power Law Fit for Bitcoin OTC network



Graph 10: Power Law Fit for Bitcoin OTC network

Bitcoin OTC: $a \approx 1.95$ ($R^2 = 0.98$)
Epinions: $a \approx 1.99$ ($R^2 = 0.97$)

The exponents are remarkably low and similar (near the critical value of 2), confirming that both systems are dominated by a small number of hubs and that the degree distribution is extremely heavy-tailed. The Bitcoin OTC network ($a = 1.95$) demonstrates an even "fresher" power law than Epinions, suggesting that the anonymous environment may naturally foster the emergence of more extreme "super-hubs" to act as anchors of trust where identity is absent.

- *Clustering Coefficient:*
  - Bitcoin OTC: 0.151
  - Epinions: 0.096

Contrary to the expectation that anonymity prevents tight circles, the anonymous network is 57% more clustered than the identified one. This suggests that in high-risk anonymous markets, users form tight "trading cliques" for safety.

The Stanford Network Analysis Platform (SNAP) calculated Epinions' clustering coefficient to be 0.1279. However, even using this higher value, the Bitcoin OTC network remains 18% more clustered than Epinions, further solidifying the conclusion regarding the formation of tight anonymous cliques.

- *Small World Effect (Average Path Length):*
  - Bitcoin OTC: 3.6784 hops
  - Epinions: 4.2059 hops (using a sample of 5000 random nodes of the LSCC)

Trust diffuses faster in the anonymous network, likely due to its higher density.

- *Reciprocity:*
  - Bitcoin OTC: 0.7923
  - Epinions: 0.3083

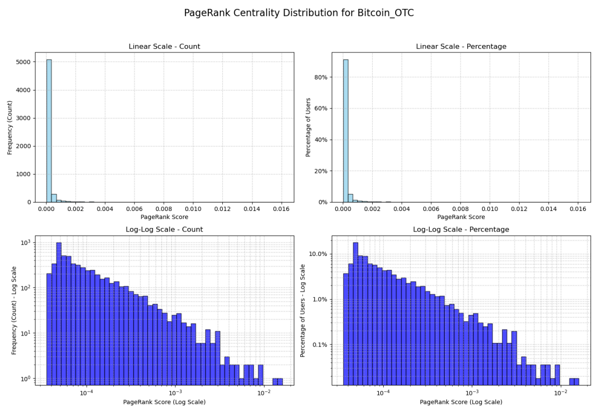
This is a critical finding. In the anonymous network studied, 79% of edges are reciprocal, compared to only 31% in Epinions. This indicates that trust in Bitcoin OTC is transactional and mutual ("I trust you because we traded"), whereas Epinions relies on one-way "fandom" or consumption of reviews.

<u>Phase B: Centrality & Trust Transitivity</u>

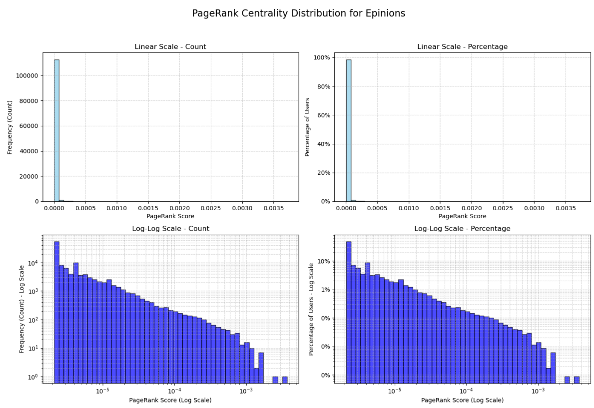
- *PageRank Distribution:*

The PageRank analysis confirms the "Rich-Get-Richer" effect. This is much more vivid in Epinions network, where the bottom

99% (approximately) of users have PageRank less than 0.005, while the top 1% of users accumulate exponential influence.

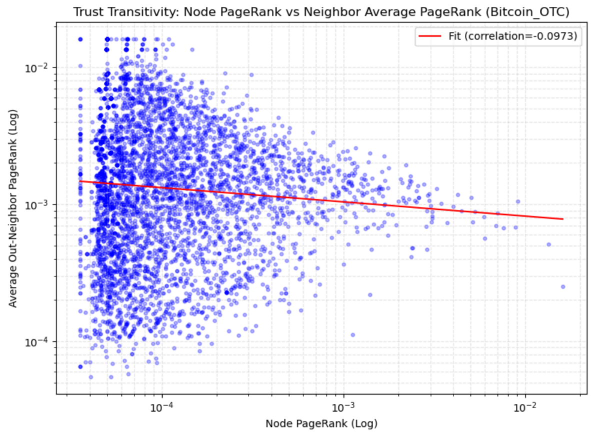


Graph 11: PageRank Centrality Distribution for the Bitcoin OTC network. Top Left: Linear Scale Bins showing frequency as count; Top Right: Linear Scale Bins showing frequency as a percentage; Bottom Left: Log-Log Scale Bins showing frequency as coun;, Bottom Right: Log-Log Scale Bins showing frequency as a percentage.

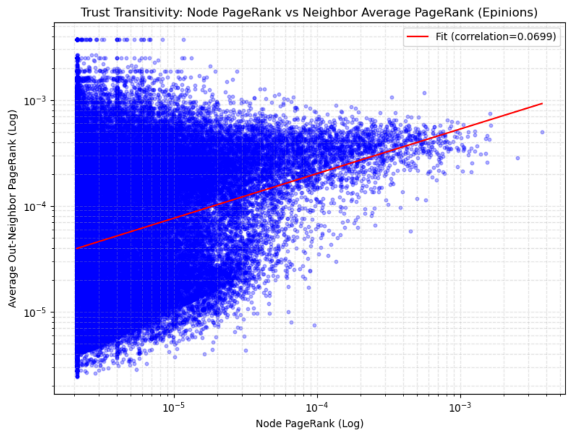


Graph 12: PageRank Centrality Distribution for the Epinions network. Top Left: Linear Scale Bins showing frequency as count; Top Right: Linear Scale Bins showing frequency as a percentage; Bottom Left: Log-Log Scale Bins showing frequency as coun;, Bottom Right: Log-Log Scale Bins showing frequency as a percentage.

- *Trust Assortativity (Correlation):*
  - Bitcoin OTC: r = -0.0973 (Disassortative)
  - Epinions: r = +0.0699 (Assortative)



Graph 13: Trust Transitivity of the Bitcoin OTC network (it is Disassortative, correlation is negative)



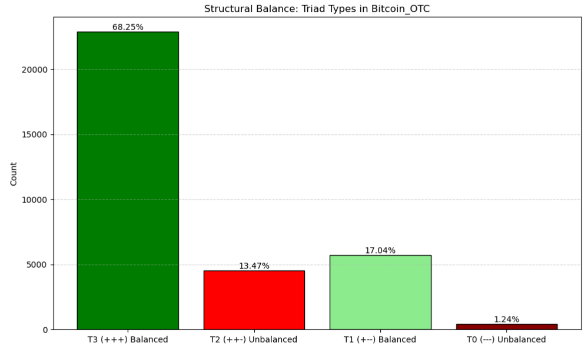
Graph 14: Trust Transitivity of the Epinions network (it is Assortative, correlation is positive)

In the identified network, high-reputation users tend to trust other high-reputation users (Homophily). However, in the anonymous network, high-reputation users frequently grant positive ratings[3] to low-reputation users. This suggests a "Vouching" mechanism: trusted pillars of the community actively test and validate new entrants, integrating them into the Web of Trust.

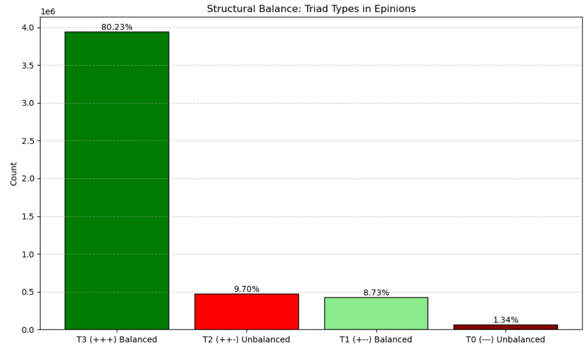
Phase C: Signed Network & Robustness

- *Structural Balance:*

[3] Trust Assortativity is calculated in my code using only the positive edges of the graph, not the negative ones.



Graph 15: Structural Balance of the triads of the Bitcoin OTC network.



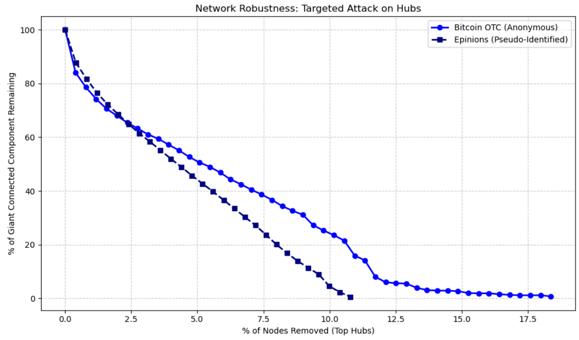
Graph 16: Structural Balance of the triads of the Epinions network.

Bitcoin OTC: 85.2924% Balanced Triads
Epinions: 88.9641% Balanced Triads

While both are balanced, the anonymous network has a higher ratio of Unbalanced triads (14.7% vs 11% of Epinions' Unbalanced triads). Specifically, the T_2 type (++-) is more prevalent, indicating higher "social friction" and disagreement in the anonymous setting.

- *Robustness (Targeted Attack):*

I defined the "Collapse Point" as the percentage of top hubs removed before the Giant Connected Component (GCC) shatters (size < 1%).



Graph 17: Network Robustness Analysis for both networks. Simulation of an attack in the top nodes based on their Total Degree (= In-Degree + Out-Degree).

Bitcoin OTC Collapse: At top 18.4% removal.
Epinions Collapse: At top 10.4% removal.

The observed data indicates that for those two platforms, the identified network is significantly more fragile than the anonymous one.

Discussion & Conclusions

Based on the empirical analysis of the Bitcoin OTC and Epinions datasets, I draw the following conclusions regarding these specific systems. By evaluating our results against the initial hypotheses, I draw the following conclusions:

I. *Polarization and Volatility (Hypothesis 1: Confirmed)*

As predicted, the anonymous network utilizes the full spectrum of ratings (-10 to +10) and exhibits drastically higher reciprocity (79%). In an environment without legal recourse, users must "shout louder" (make extreme ratings) and "bind tighter" (reciprocity) to survive. For the anonymous community, anonymity appears to turn trust into a bidirectional contract, whereas in the Epinions dataset, identity allows trust to function more as a uni-directional endorsement.

II. *Fragility (Hypothesis 2: Rejected)*

I hypothesized that Bitcoin OTC would be more fragile due to centralization. The results prove the opposite: Epinions collapsed twice as fast.

This means that Epinions relies heavily on a small core of "Super-Reviewers" to bridge the community. If they leave, the network fragments.

Bitcoin OTC, despite having Hubs, relies on a dense mesh of reciprocal Peer-To-Peer connections (High Clustering). If a hub is removed, the traders simply route trust through other partners in their "clique". Anonymity forced the network to build a redundant, resilient mesh rather than a fragile hierarchy.

This phenomenon can be explained by the concept of Topological Redundancy. In the identified network (Epinions), users rely on "Star" topologies where a central influencer connects many disconnected users. When the center of the star (the Hub) is removed, the spokes become isolated. In contrast, the Bitcoin OTC network exhibits a "Mesh" topology. Due to high risk of anonymous trading, users do not rely on a single point of failure (one voucher). Instead, they seek multiple, independent confirmations of trust. This creates alternate paths for information flow. When a major Hub is removed in the anonymous network, the high Clustering Coefficient (0.151) ensures that local bridges remain intact, maintaining the Giant Connected Component even under heavy attack.

III. *Structural Balance (Hypothesis 3: Confirmed)*

The anonymous network is indeed less structurally balanced (14.7% unbalanced). Furthermore, the negative assortativity found in Phase B (of my analysis) offers a profound insight: Trust in anonymous networks is not just about "liking" your peers (as in Epinions). It is about Vetting. The "pillars" of the Bitcoin OTC community do not just connect with other pillars; they actively vouch for the periphery (low-score users). This "Vouching" behavior is essential for the system's growth, allowing new, anonymous users to gain credibility through interaction with established hubs.

IV. *Final Conclusion*

Based on this empirical comparison, anonymity does not degrade the structure of trust; it transforms it.

The Identified Network (Epinions) evolves into a Hierarchy: assortative, stable, but fragile and dependent on elites.

The Anonymous Network (Bitcoin OTC) evolves into a Resilient Mesh: reciprocal, clustered, polarized, and robust against node failure.

These conclusions are specific to the dynamics of these two platforms and suggest that the removal of identity forces the system to replace "Who you are" with "Who you know" and "How much you risk", creating a trust topology that is locally more chaotic (unbalanced) but globally more resilient.

Limitations and Future Work

While this project provides significant insights into the topology of trust, it is subject to certain limitations that suggest avenues for future research:

I. *Temporal Aggregation:* This analysis treats the network as static snapshots, collapsing all time-stamped ratings into a single graph. This approach overlooks the evolution of trust. In reality, a "scammer" might build trust over time (positive ratings) before exiting with a fraud (negative ratings). A temporal analysis could reveal if the "policing" behavior we observed in Bitcoin OTC is a reaction to specific fraud events.

II. *Binary vs Weighted Simplification:* To compare the networks, I normalize Bitcoin OTC ratings. While necessary for direct comparison, this ignores the nuance of the weighted scale. Future work could analyze if the intensity of a rating (e.g. +1 vs +10) correlates with the longevity of the link.

III. *Sybil Attack Sophistication:* My structural balance analysis assumes that accounts act independently. In anonymous networks, a single actor may control multiple nodes (Sybil attack) to artificially inflate Balanced Triads. Integrating Sybil-detection algorithms with the current structural balance check would provide a cleaner view of genuine social dynamics.

Despite these limitations, the results successfully highlight the structural trade-offs users make when identity is removed, providing a strong foundation for designing resilient anonymous reputation systems.

References

I. Easley D., & Kleinberg J. (2010). *Networks, Crowds, and Markets: Reasoning About a Highly Connected World.* Cambridge University Press.

II. Leskovec J., & Krevl A. (2014). *SNAP Stanford Large Network Dataset Collection*

III. Page L., Brin S., Motwani R., & Winograd T. (1998). *The PageRank Citation Ranking: Bringing Order to the Web.* Stanford InfoLab.

IV. Hagberg A. A., Schult D. A., & Swart P. J. (2008). *Exploring Network Structure, Dynamics, and Function using NetworkX.* In Proceedings of the 7th Python in Science Conference.