Exploring Impact of Trust in Prisoner's Dilemma on Social Networks

ECE 227

Karl Hernandez, *Department of Electrical and Computer Engineering*
UC San Diego
kphernan@ucsd.edu
Achyut Pillai, *Department of Electrical and Computer Engineering*
UC San Diego
apillai@ucsd.edu
Alexis Yu, *Department of Electrical and Computer Engineering*
UC San Diego
ahyu@ucsd.edu

***Abstract*—In this paper, we investigate the dynamics of cooperation in the Prisoner's Dilemma (PD) when played on two real world social networks, the undirected Facebook "circles" network and the directed Epinions network with trust signing. For each network, we implement two classical payoff-based update rules, *imitate best neighbor* and *Fermi*, to observe how cooperation evolves under standard PD dynamics. The fraction of cooperators and the number of iterations to convergence over three bernoulli initial distributions ($p \in \{0.25, 0.50, 0.75\}$) are recorded. We then introduce two trust aware update rules on the Epinions network, one where agents only consider trusted neighbors when updating strategy and another where agents aggregate trust from all neighbors. Our results show that in the dense Facebook network, the standard update rules converge rapidly and to moderate levels only when there is high initial cooperation. The sparse Epinions network, however yields low cooperation and converges slightly slower. The pairwise trust boosts cooperation significantly, across all initial conditions, while the all-neighbor trust rule drives cooperation above 90% and converges very quickly. These results demonstrate that embedding trust into the update rules can drastically enhance both the level and speed of convergence to cooperation, especially in more sparse networks.**

I. Introduction

The Prisoner's Dilemma is a two-player game in which each participant can either cooperate or defect. The outcomes are either mutual cooperation, yielding a moderate reward for both; unilateral defection, giving the defector a high payoff and the cooperator a low payoff; and mutual defection, leaving both with a low payoff.

In this paper, we examine the Prisoner's Dilemma's dynamics on two empirically derived social networks: the Facebook Social Circles (from SNAP) and the Epinions graph (also from SNAP). On each network, we first employ a standard "imitate the best" update rule, whereby each node plays PD with all neighbors, counts payoffs, and then takes the highest payoff neighbor's strategy. Then, on the Epinions network, we suggest a trust-sensitive version: after each round, agents only consider neighbors that they trust when determining whom to copy.

II. Related Works

A substantial body of past research has explored how network structure and trust relationships influence cooperation in an iterative Prisoner's Dilemma setting. Cameron and Cintrón-Arias revisit the PD on real social networks Cameron and Cintrón-Arias (2013), demonstrating that empirical topologies, characterized by heterogeneous degree distributions and community structure, play a crucial role in the spread and stability of cooperative strategies. High-degree nodes and clustering sometimes facilitate and sometimes obstruct cooperation, depending on the payoffs and update rules. Dynamic social networks facilitate cooperation in the N-player Prisoner's Dilemma by Rezaei and Kirley Rezaei and Kirley (2012) and they build on this by introducing dynamic rewiring in the N-player PD. This allows agents to adjust ties based on past payoffs so that defectors lose links and cooperators attract new ones, which leads to defectors being isolated while cooperators cluster together. Meanwhile, Trust-induced cooperation under the complex interaction of networks and emotions by Xie et al. Xie et al.

(2024) looks at "trust-induced cooperation" under complex network and emotional interactions by embedding a continuous, weighted trust metric into the agent's decisions. This stabilized cooperation under harsher payoff conditions and produces more accurate patterns of cooperation compared to real social behavior.

III. Model

In this paper, we analyze the evolution of the proportion of cooperators in the network over time when playing the prisoner's dilemma game. The initial state assigns the node's initial strategy with a Bernoulli (p) distribution, giving a p chance of cooperation and $1-p$ chance of defection. Each node will play the game with each of its neighbors, will then update its strategy (cooperator or defector), and then repeat the process. We tested this cycle out with different update strategy rules, Prisoner's Dilemma rules, and p values.

A. Strategy Update Rules

Strategy revision is handled by the generic `update_strategies` routine, which takes as input the network G, a payoff map $\pi : V \to \mathbb{R}$, an `UpdateRule` callback, and a random seed. At each iteration, every node computes a new strategy in one of four ways:

Imitate–Best–Neighbor. Each node u compares its own accumulated payoff π_u against those of its neighbors. It adopts the strategy of the neighbor (or itself) with the highest payoff, breaking ties uniformly at random.

Trust–Aware Update. On graphs with signed edges ($s_{uv} \in \{+1, -1\}$), each node u and its neighbors v compute an *effective payoff* $s_{uv}\pi_v$. The candidate with the maximum effective payoff is chosen; if it is trusted ($s_{uv} = +1$), u copies its strategy, and if it is distrusted ($s_{uv} = -1$), u adopts the opposite strategy. Self-comparison preserves the current strategy.

Fermi Update. Each node u selects a random neighbor v and adopts v's strategy with probability

$$P(u \to v) = \frac{1}{1 + \exp\left[(\pi_u - \pi_v)/K\right]},$$

where $K > 0$ is a temperature parameter. Otherwise, u retains its current strategy.

All–Neighbors Trust–Aware Update. Each node u computes the weighted sum

$$S_u = \pi_u + \sum_{v \in N(u)} s_{uv}\pi_v .$$

If $S_u > 0$, u cooperates; if $S_u < 0$, it defects; and if $S_u = 0$, it flips or retains its strategy with equal probability.

These four rules capture a spectrum from pure payoff maximization to locally aggregated, trust-mediated imitation, and are used interchangeably to study the emergence of cooperation under different network and behavioral assumptions.

B. Game Rules

The first implementation of the game models a standard Prisoner's Dilemma on a network. In the game a node will iterate over all edges (u,v) (ensuring each unordered pair is considered just once) and looks up the current strategies of u and v. Using a fixed payoff matrix, where mutual cooperation yields (3,3), defection against cooperation yields (5,0), mutual defection yields (1,1), and cooperation against defection yields (0,5); it computes and stores the pairwise payoffs. Once all neighbor-to-neighbor payoffs are collected, the payoffs for each node will be summed up, giving us the final payoff matrix.

The second implementation was only implemented on the Epinions' data set and had the nodes consider if they trusted the node they were playing the game with to determine their strategy. Once paired up, if the node trusts the other node, there is a 70% chance that the node will switch its strategy to cooperation. This also applies conversely, with the node having a 70% chance that the node will switch to defect if it doesn't trust the other node.

IV. Dataset Description

The first network we examined was the Facebook "circles" network Leskovec et al. (2010), which is constructed by aggregating the ego-networks of ten anonymized survey participants, each of whom contributes an "ego" node linked to all their friends. The resulting graph ends up with 4,039 nodes and 88,234 edges, where each edge represents a friend connection. The network is highly connected and strongly clustered, with the average node degree being 43.7 and the cluster coefficient being 0.6055. Furthermore, the network has a small-world nature where the diameter of the network is 8, meaning the longest shortest path between any two users is 8 hops.

The second network we examined was the Epinions social network McAuley and Leskovec (2012), which is a directed graph where each node represents a user on the Epinions.com review website, and each edge will either state whether the user trusts or distrusts another user. The network is

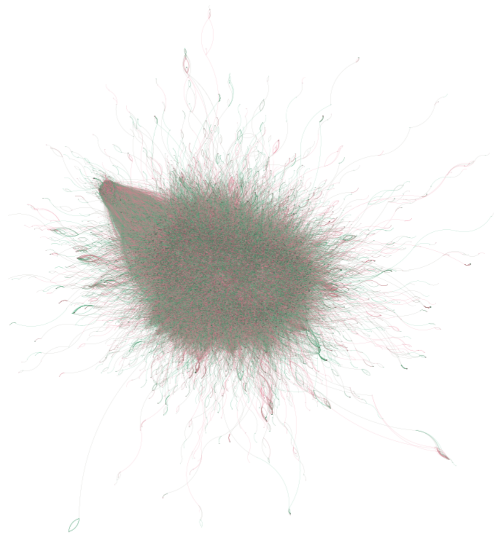


Figure 1: Facebook Network

comprised of 131,828 nodes and 841,372 edges. Although the entire users form a large strong weakly connected component of 119,130 nodes (90.4% of the graph) with 833,695 edges (99.1% of the total). The strongly connected component is relatively small, i.e., just 41,441 users (31.4%) who mutually trust one another with 693,737 edges (82.5%). The average clustering coefficient of the network is 0.1279, which indicates moderate local clustering. Lastly, this network also has a small-world nature, with the diameter being 14.

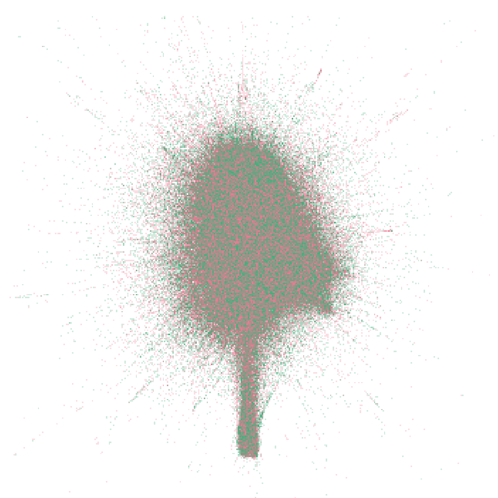


Figure 2: Epinions Network

V. Simulations and Numerical Experiments

We conducted evolutionary Prisoner's Dilemma simulations on both networks, varying the initial state of Bernoulli for the starting strategies $p \in \{0.25, 0.50, 0.75\}$. For each combination of network and p, we compared two standard payoff-based update rules (imitate-best-neighbor and Fermi). Then for every combination of the Epinions' data set, we compared the trust update rules (trust aware update and all neighbor trust). In every experiment, we recorded the final fraction of cooperators and the number of iterations required for the dynamics to converge.

A. Intuition

Prior to conducting our simulations, we formulated the following expectations. First, we hypothesized that under standard payoff-based update rules (imitate-best-neighbor and Fermi), the dense, highly clustered Facebook network would sustain higher levels of cooperation than the sparser, more hierarchical Epinions network. Second, we expected that introducing trust-awareness into the update process would uniformly boost cooperative behavior. Finally, we expected that an all-neighbor trust aggregation rule—by leveraging the full spectrum of positive and negative ties—would produce the most robust and rapid convergence to cooperation across both network topologies.

VI. Results

A. Cooperation Levels

Table I presents the final fraction of cooperators under three initial distribution parameters p (initial cooperation probability) for both the Facebook and Epinions networks across four update rules. Under standard *imitate-best-neighbor* and game, Facebook converges to 0% cooperation at $p = 0.25$ and to 95.8% at $p = 0.75$, whereas Epinions peaks at 27.4% cooperation when $p = 0.75$. This shows that for both the Facebook and Epinions networks, cooperation is highly sensitive to the initial cooperation distributions under the *imitate-best-neighbor* update rule. It is also likely that there exists a critical p for the Facebook network that determines whether it will converge to majority cooperate or defect.

The Fermi rule yields vastly different behavior: Facebook reaches only 5.94% cooperation at high p, while Epinions remains below 31%. In both networks, the Fermi update rule leads to a more gradual increase in cooperation as p increases, with Facebook showing a slight increase from 1.51% at $p = 0.25$ to 5.94% at $p = 0.75$, while Epinions shows a similar trend from 8.05% to 30.47%.

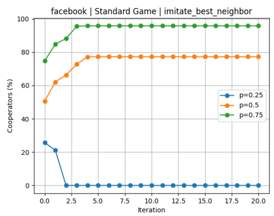
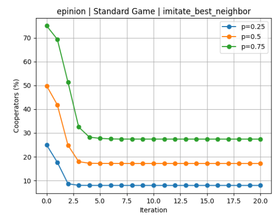
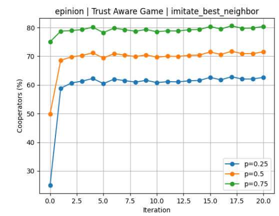
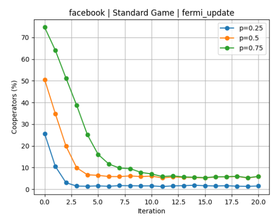
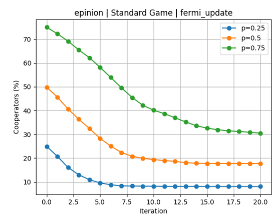
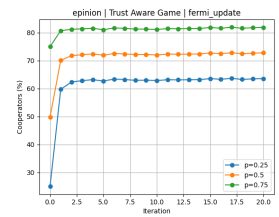
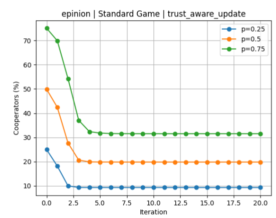
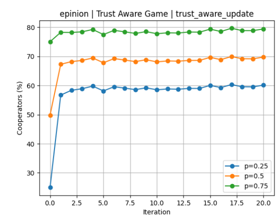
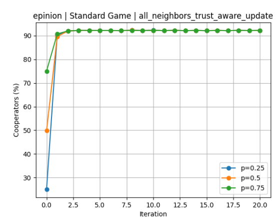
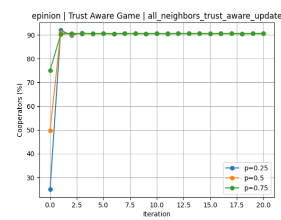
When we apply both imitation and Fermi updates on Epinions with the trust-aware game, final cooperation rises sharply into the 62–82% range across all

p. Unlike the standard game, the results of the trust-aware game show that cooperation is less sensitive to the initial distribution p and update rule, with all three p values yielding similar final cooperation levels. The *trust-aware update* rule achieves 9.32% cooperation at $p = 0.25$ and 31.5% at $p = 0.75$, while the *all-neighbor-trust* rule drives cooperation above 90% across all p values, demonstrating the power of trust in greatly enhancing cooperation. These cooperation results indicate that trust-aware updates generally outperform standard imitation and *Fermi* updates, especially in the sparse Epinions network.

B. Convergence Speed

Table I also presents the the number of iterations taken for each update rule to reach equilibrium with a tolerance of 0.05%. Under the standard game with the *imitate-best-neighbor* update rule, Facebook converges in 3 iterations at $p = 0.25$ and 5 iterations at both $p = 0.50$ and $p = 0.75$. On Epinions, the same rule requires 4, 6, and 8 iterations respectively, indicating that the sparse topology of Epinions leads to slower convergence compared to Facebook. The *Fermi* update rule exhibits a similar pattern; Facebook takes 10, 7, and 17 iterations at $p = 0.25$, $p = 0.50$, and $p = 0.75$ respectively, while Epinions requires 9, 16, and 20 iterations. However, the *Fermi* update rule is generally slower to converge as can be seen in the exponential shape of the convergence curves in the plots. The introduction of trust awareness significantly affects convergence on Epinions. The trust update rules including both pairwise trust updates and the *all-neighbor-trust* converge in under 10 iterations (often within 5–8), highlighting trust's ability to accelerate consensus. Thus, trust-aware mechanisms consistently accelerate convergence, reaching equilibrium in fewer iterations than standard imitation or *Fermi* update rules, regardless of initial conditions.

Table I: “Cooperator (%) vs. Iteration” plots, final-split percentages, and iterations to convergence, organized by update rule.

<table>
<tr><th rowspan="4">Update Rule</th><th colspan="4">Facebook</th><th colspan="8">Epinion</th></tr>
<tr><th colspan="4">Std. Game</th><th colspan="4">Std. Game</th><th colspan="4">Trust-Aware</th></tr>
<tr><th colspan="3">Cooperate %
Defect %
(Conv. Iter.)</th><th rowspan="2">Plot</th><th rowspan="2">Plot</th><th colspan="3">Cooperate %
Defect %
(Conv. Iter.)</th><th rowspan="2">Plot</th><th colspan="3">Cooperate %
Defect %
(Conv. Iter.)</th></tr>
<tr><th>$p = 0.25$</th><th>$p = 0.50$</th><th>$p = 0.75$</th><th>$p = 0.25$</th><th>$p = 0.50$</th><th>$p = 0.75$</th><th>$p = 0.25$</th><th>$p = 0.50$</th><th>$p = 0.75$</th></tr>
<tr><td>Imitate–Best–Neighbor</td><td>0.00%
100.00%
(3)</td><td>77.25%
22.75%
(5)</td><td>95.79%
4.21%
(5)</td><td>
</td><td>
</td><td>7.95%
92.05%
(4)</td><td>17.18%
82.82%
(6)</td><td>27.43%
72.57%
(8)</td><td>
</td><td>62.66%
37.34%
(19)</td><td>71.55%
28.45%
(19)</td><td>80.42%
19.58%
(20)</td></tr>
<tr><td>Fermi–Update</td><td>1.51%
98.49%
(10)</td><td>5.89%
94.11%
(7)</td><td>5.94%
94.06%
(17)</td><td>
</td><td>
</td><td>8.05%
91.95%
(9)</td><td>17.69%
82.31%
(16)</td><td>30.47%
69.53%
(20)</td><td>
</td><td>63.61%
36.39%
(9)</td><td>72.79%
27.21%
(9)</td><td>81.93%
18.07%
(9)</td></tr>
<tr><td>Trust–Aware–Update</td><td>N/A
N/A
N/A</td><td>N/A
N/A
N/A</td><td>N/A
N/A
N/A</td><td>N/A</td><td>
</td><td>9.32%
90.68%
(4)</td><td>19.82%
80.18%
(6)</td><td>31.53%
68.47%
(7)</td><td>
</td><td>60.15%
39.85%
(14)</td><td>69.78%
30.22%
(14)</td><td>79.43%
20.57%
(14)</td></tr>
<tr><td>All–Neighbor–Trust</td><td>N/A
N/A
N/A</td><td>N/A
N/A
N/A</td><td>N/A
N/A
N/A</td><td>N/A</td><td>
</td><td>92.24%
7.76%
(4)</td><td>92.25%
7.75%
(4)</td><td>92.27%
7.73%
(4)</td><td>
</td><td>90.54%
9.46%
(8)</td><td>90.57%
9.43%
(8)</td><td>90.57%
9.43%
(7)</td></tr>
</table>

VII. Conclusion

Our study confirms that there is a connection between network structure and the update rules used in the prisoner's dilemma, that determines the dynamics of cooperation. While the classical *imitate-best-neighbor* and *Fermi* update rules are effective enough in encouraging cooperation in highly clustered networks like Facebook, under favorable initial conditions, they are not sufficient to maintain cooperation in sparse networks like Epinions. By contrast, the *trust-aware* update rules, which incorporate trust relationships, substantially enhance cooperation levels and convergence speed in the Epinions network. Most notably, the *all-neighbor-trust* update rule achieves over 90% cooperation across all initial conditions with relatively fast convergence, thus demonstrating the power of trust in fostering cooperation. These findings suggest that real social networks, may require leveraging trust relationships to maintain cooperation as well as to overcome obstacles presented by sparse network structures.

VIII. Future works

Future work could proceed in several directions to further explore the dynamics of cooperation in the prisoner's dilemma on social networks. One potential direction is to introduce adaptive rewiring mechanisms where agents can change their connections and trust based on past interactions similar to the work by Rezaei and Kirley Rezaei and Kirley (2012). Next, we could investigate the impact of different trust models, such as those with weighted trust or trust decay based on length of connection and interaction history, to see how the dynamics of trust affect the dynamics of cooperation. Another direction is to see how a varying payoff matrix affects the dynamics, such as with different reward structures for cooperation and defection or a payoff dependent on a cost function that changes over time. Finally, we could explore the experiments on more real social networks, such as Twitter, Reddit, or GitHub, to see how different real-world structures affect the dynamics.

IX. Supplementary Materials

Code and Data Leskovec et al. (2010); McAuley and Leskovec (2012), available at https://github.com/alexishyu/ECE-227-Final-Project/tree/main

References

Cameron, S. M. and Cintrón-Arias, A. (2013). Prisoner's dilemma on real social networks: Revisited. *Mathematical Biosciences & Engineering*, 10(5&6):1381–1398.

Leskovec, J., Huttenlocher, D. P., and Kleinberg, J. M. (2010). Signed networks in social media. In *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI)*, pages 1361–1370, Atlanta, GA, USA.

McAuley, J. J. and Leskovec, J. (2012). Learning to discover social circles in ego networks. In *Advances in Neural Information Processing Systems*, volume 25, pages 548–556.

Rezaei, G. and Kirley, M. (2012). Dynamic social networks facilitate cooperation in the n-player prisoner's dilemma. *Physica A: Statistical Mechanics and its Applications*, 391(23):6199–6211.

Xie, Y., Bai, Y., Zhang, Y., and Peng, Z. (2024). Trust-induced cooperation under the complex interaction of networks and emotions. *Chaos, Solitons & Fractals*, 182:114727.

Karl Hernandez earned his B.S. degree in Computer Engineering at the University of California, San Diego in 2024, and is now pursuing his M.S. degree in Electrical and Computer Engineering with a specialization in Machine Learning and Data Science at the same institution. His research interests include privacy-preserving machine learning, and federated agentic AI.

Achyut Pillai earned his B.S. degree in Electrical Engineering from the University of California, San Diego in 2024. He is currently pursuing his M.S. degree in Electrical and Computer Engineering with a focus on Machine Learning and Data Science at the University of California, San Diego. His research interests include machine learning, network science, and large language models.

Alexis Yu earned her B.S. degree from University of California, San Diego in Electrical Engineering with a depth in Computer System Design in 2024. She is currently pursuing her Master's degree in Computer Engineering at the University of California, San Diego. She will be pursuing ASIC and FPGA design in the future.