Polarization Game over Social Networks

Xilin Zhang
System Science and Industrial Eng.
Binghamton University
Binghamton, NY, United States
xzhan176@binghamton.edu

Emrah Akyol
Electrical and Computer Engineering
Binghamton University
Binghamton, NY, United States
eakyol@binghamton.edu

Zeynep Ertem
System Science and Industrial Eng.
Binghamton University
Binghamton, NY, United States
zeynep@binghamton.edu

***Abstract*—This paper quantitatively analyzes a game over a social network of agents (nodes), some of which are controlled by two players who maximize and minimize polarization over this network. Agents' opinions evolve according to the Friedkin-Johnsen model, and players can change only the innate opinion of an agent. Polarization is measured as the sample variance of the agents' steady-state opinions. The game is considered void if the players choose the same agents, which constitutes a zero-sum game. We first analyze the functional properties of this game and characterize the best response for each player given the agent. Fictitious play is used to obtain the equilibria in real social networks. Subsequently, the paper delves into an exploration of the properties of the Nash equilibrium (NE).**

***Index Terms*—Social Networks, Game Theory, Polarization**

I. Introduction

Amidst the challenges posed by contemporary political polarization [1]–[3], *modeling* polarization dynamics over social networks remain largely unexplored. Several factors that impact network polarization, such as network size, structure, agents' innate opinions, behavioral/cognitive features, and the influence of exogenous campaigns, etc., all contribute to the complexity of this modeling problem. In the case where external players with opposing objectives, as witnessed in the 2016 US elections, targeted specific voter groups to deepen political divides [4], [5], further complicating theoretical modeling.

In light of these challenges, this paper conducts an empirical analysis of a polarization game between two players, Maximizer and Minimizer, each seeking to respectively maximize and minimize polarization by manipulating the innate opinion of a chosen agent within a social network. Agents, initially holding innate opinions on a known weighted network, engage in opinion exchanges with their network neighbors following the well-known Friedkin-Johnsen (FJ) opinion evolution model [6]. The game unfolds as players strategically position agents and update agents' innate opinions to impact network polarization. If players choose the same agent, rendering the agent's opinion unchanged, the game is considered void. In such competition, one side's gain equals the other's loss, resulting in a net benefit of zero, constituting a zero-sum game [7]. Drawing on established game theory frameworks, we explore the Nash equilibrium, representing a steady state where neither the Maximizer nor the Minimizer deviates from their current strategy.

This research is supported in part by the NSF via CAREER #2048042.

The practical implications of this game introduce a compelling dimension. Unlike conventional zero-sum games, such as rock-paper-scissors, where outcomes are immediately evident, the real-world competition involving Maximizers and Minimizers is characterized by uncertainty regarding the immediate payoffs of their actions, due to the network polarization hinges on the interplay of both players' actions and subsequent outcomes of opinion dynamics. The outcome of opinion dynamics is unpredictable for every party in this game. In an effort to unravel this intricate relationship, our problem formulation builds on a few recent prior works, including those by [8] and [9]. These studies individually focus on maximizing [8] and minimizing [9] polarization through the controlling of the innate opinions and network parameters in the presence of the Johnson-Friedkin opinion evolution model [6]. Our unique contribution lies in the integration of both perspectives, examining the polarization dynamic in the coexistence of both Maximizer and Minimizer influences. Hence, this study represents, to the best of our knowledge, the inaugural exploration of opinion polarization games within a social network context.

II. Related Works

A principal part of opinion dynamics is how nodes update their opinions over time. There are multiple models developed for opinion dynamics [10]. In the DeGroot model [11] each node updates its opinion using a weighted convex combination of its neighbors' opinions. The model developed by Friedkin and Johnsen [6] which is the model we base our work upon, considers that, in addition to its neighbors' opinions, a node also gives certain weightage to its innate opinion.

Previous research has extensively addressed the concept of influence maximization over networks, exemplified by notable contributions such as [12], [13]. This problem involves identifying nodes within a network whose adoption of a product or action will have maximal reach in the social network. Existing studies in this domain employ probabilistic propagation models, such as the independent cascade or linear-threshold model, to delineate how actions disseminate among individuals in the social network. These models categorize network individuals into active or inactive classes, assuming that active individuals influence their neighbors to become active based on specific probabilistic rules, thereby causing cascading activity spread in the network.

978-1-7281-9054-9/24/$31.00 ©2024 IEEE

Given the prevalence of multiple entities seeking to disseminate specific information, such as for marketing or propaganda purposes, a natural extension is the introduction of competition among parties aiming to propagate their preferred information. This competitive scenario finds representation within a game-theoretic framework, leading to the exploration of equilibria in such games. Consequently, scholars have sought to characterize the set of equilibria in these games, as evidenced by studies such as [14]–[16].

However, in this paper, we focus on the novel competition among players who aim to maximize and minimize the polarization of opinion in the social network. For such purposes, it is essential to model the opinion dynamics and formulate the polarization game.

III. Preliminaries

A. Notation and Network Model

Let $\mathcal{N}$ represent the set of the positive integers, and $[N] = \{1, 2, \ldots, N\}$. We respectively denote I and $\mathbf{1}$ the identity matrix and the vector of all ones with proper dimensions. The superscript '$\top$' stands for the matrix transposition. $\mathbf{c}_N \triangleq \frac{1}{N}\mathbf{1}_N$ and $\mathbf{1}_N$ denotes the $N \times 1$ 1s vector.

Indices:	
i	Node (agent) index ($i \in \{1, \ldots, N\}$)
t	Time index - for opinion dynamics ($t \in \{1, 2, ...\}$)
m	Player index ($m \in \{1, 2\}$)
T	Time index - for the game dynamics ($T \in \{1, 2, \ldots\}$)
l	Historical action index
κ_m	The number of nodes/agents player m can choose for each action
Network Parameters:	
w_{ij}	Weight between nodes i and j
α_m	The action of player m including selected agent (i_m) and changed opinion (s_i^m) $\leftarrow$ ($\alpha_m = (i_m, s_i^m)$).
G_m	The past history of $\mathcal{P}_m$, i.e., $G_m = [\alpha_m^1, \ldots, \alpha_m^l]$
$\Delta(n)$	The simplex in $\mathbb{R}^n$, i.e. $\{\mathbf{s} \in \mathbb{R}^n \vert s_i \geq 0\}, \forall i$ and $\mathbf{1}^\top \mathbf{s} = 1$.
$\mathcal{H}_m$	Action space for player m
$\mathcal{BR}_m$	The player m best pure strategy according to player $-m$ history
$\mu_m^l(T)$	The empirical frequency of player m that the other player choosing action α_{-m}^l at time T
J_m	The payoff according to player m's action
$\overline{J_m}$	The expected payoff after player m's actions

TABLE I: Nomenclature

We consider a social system with N agents, indexed by $i \in [N]$. The state/opinion of agent i at time t is denoted by $z_i(t) \in [0, 1]$. The network is characterized by the triplet $(\mathcal{V}, \mathcal{E}, \mathcal{W})$ where $\mathcal{V}$, $\mathcal{E}$ and $\mathcal{W}$ represent the set of nodes (vertices), edges and weights. There are no assumptions imposed on the structure of the network. Table I presents notations employed in this paper.

B. Opinion Dynamics Model

Throughout this paper, we use the well-known Friedkin-Johnsen model of opinion dynamics [6], where each agent updates their opinions by averaging between their innate opinion and neighbors' opinions. For example, agent i with an innate opinion $s_i \in [0, 1]$ forms its updated opinion at time t, the updated opinion $z_i(t)$, as a linear combination of its innate opinion and network neighbors' opinions is defined as follows:

$$z_i(t) = \beta_i s_i + (1 - \beta_i) \sum_{j \in \mathcal{V}} w_{ij} z_j(t - 1) \tag{1}$$

where $\beta_i = 1/\left(1 + \sum_{j \in V} w_{ij}\right)$ is a constant that ensures $z_i(t) \in [0, 1]$. The limit (in time) of this process is referred to in this work as steady-state[1] opinions, represented by $\tilde{z}_i = \lim_{t \to \infty} z_i(t)$. We let $\tilde{\mathbf{z}}$ to represent the vector of steady-state opinions, i.e., $\tilde{\mathbf{z}} = [\tilde{z}_1, \tilde{z}_2, \ldots, \tilde{z}_N]^\top$. The following theorem[2] states that the steady-state opinions ($\tilde{\mathbf{z}}$) can be obtained from the vector of innate opinions via a linear transformation:

Theorem 1. *Given the dynamics in (1),*

$$\tilde{\mathbf{z}} = H\mathbf{s}, \tag{2}$$

where $H = (I + L)^{-1}$ is a doubly-stochastic matrix, and L is the Laplacian of the network.

C. Polarization Metrics

There are several metrics used in the literature to measure the polarization of opinions in a social network. In [8], [9], two different metrics were employed to differentiate between "polarization" and "disagreement". Here, we adopt the network polarization metric in [8], [9], which is simply the sample variance of steady-state opinion, defined as follows:

$$J(\tilde{\mathbf{z}}) \triangleq (\tilde{\mathbf{z}} - z)^\top (\tilde{\mathbf{z}} - z) \tag{3}$$

where $z = \frac{1}{n} \sum_{v \in \mathbb{V}} \tilde{z}_v$ is the mean opinion averaged over the entire network. We note in passing that our results hold for a generic polarization function that is convex in $\tilde{\mathbf{z}}$.

D. Game Model

We consider a two-player game with players $\mathcal{P}_1$ (Maximizer) and $\mathcal{P}_2$ (Minimizer), who intend to respectively maximize and minimize the polarization of the social network with N agents/nodes. We will often use the index $m \in \{1, 2\}$ for one of the players and $-m$ for the other player. Assuming both players can reach every node in the network, each player has a discrete-continuous hybrid action space $\mathcal{H}_m$:

$$\mathcal{H} = \left\{(i_m, s_i^m) \mid s_i^m \in [0, 1] \text{ for all } i_m \in [N]\right\},$$

where the node index that is chosen by $\mathcal{P}_m$ is denoted by $i_m \in [N]$, and the new value of the innate option is denoted by $s_i^m \in [0, 1]$. An action pair $[i_m, s_i^m]$ is denoted by α_m.

[1] We reserve the term "equilibrium" to denote the game-theoretic notion of Nash equilibrium.

[2] This result is well-known in the literature, see e.g., [9].

Each player $\mathcal{P}_m$ selects a strategy σ_m randomized over the action space $\mathcal{H}$ that consists of a set of actions $\mathcal{A}_m = \{\alpha_m\}$ selected from $\mathcal{H}$ and a probability vector over the selected actions $\mathbf{p}_m \in \Delta(|\mathcal{A}_m|)$. Let us denote the set of all possible such strategies $\sigma_m = [\mathbf{p}_m, \mathcal{A}_m]$ as $\mathcal{G}_m$.

Each player $\mathcal{P}_m$ receives a real-valued reward according to the utility function $U_m : \mathcal{G}_m \times \mathcal{G}_{-m} \to \mathbb{R}^+$ as $U_m(\sigma_m, \sigma_{-m})$.

A Nash equilibrium (NE) consists of a pair of strategies σ_1^*, σ_2^* in $\mathcal{G}_1 \times \mathcal{G}_2$ such that, for all $\sigma_m \in \mathcal{G}_m$

$$U_m(\sigma_m, \sigma_{-m}^*) \leq U_m(\sigma_m^*, \sigma_{-m}^*).$$

In other words, each player has no incentive to deviate from an NE provided that the other player maintains her NE strategy. The special case where A_m is a singleton is a "pure" strategy.

We take payoff functions as $U_1 = -U_2 = P(\tilde{\mathbf{z}})$ where P is the steady-state polarization as given in (3). In the scenario where players choose the same node (with possibly different s_i^m values), we impose $s_i^m = s_i^{-m} = s_i$ for $i_m = i_{-m}$, hence polarization is unaltered via the players. We also note that this critical assumptions renders this game a zero-sum game. We refer the polarization for this "unaltered" case as J_0.

We next define the pure best response mappings as $\mathcal{BR} : \Delta(N) \times \mathcal{F}^N \to \mathcal{H}_m$, as follows:

$$\mathcal{BR}_m(\sigma_{-m}) = \underset{\alpha_m \in \mathcal{H}_m}{\arg\max}\, J_m(\alpha_m, \sigma_{-m}).$$

For the ease of exposition, we limit the number of agents that can be manipulated by each player to unity.

E. Fictitious Play

We utilize (discrete-time) fictitious play (FP) as a computational tool to determine the equilibria in our game variations. Here, we give an overview of the FP dynamics, where two players in a strategic form game with payoffs $J_m(\alpha_m, \alpha_{-m})$. At each iteration $T = 1, 2, \ldots$, $\mathcal{P}_m$ generates the pure best response to $\mathcal{P}_{-m}$ with her belief $\mu_{-m}^l(T)$ forms at time T, here the belief of opponent's stationary strategy is based on the empirical frequency of the past play (history) as follows:

$$\mu_{-m}^l(T) = \frac{f_{-m}^l(T)}{\sum_{l'} f_{-m}^{l'}(T)} \tag{4}$$

where $f_{-m}^l(T)$ represents the number of times player $-m$ takes action α_{-m}^l before time T, with a given initial condition $f_{-m}^l(0)$. l' represents all $\mathcal{P}_{-m}$ historical actions that observed by $\mathcal{P}_m$.

It is well understood that this procedure converges to a Nash equilibrium for several game classes including zero-sum, potential, 2×2 games and ones solvable by iterated strict dominance, see e.g., [17], Proposition 2.3 in [18]. We reproduce this well-known result in the following theorem.

Theorem 2. *For zero-sum games, the limit of beliefs computed via (4) converge to Nash equilibrium strategies, i.e.,*

$$\lim_{T \to \infty} \mu_{-m}^l(T) \to \sigma_m(l)$$

IV. Algorithm

This section introduces algorithms employed to identify Nash Equilibria (NE). The fictitious play algorithm is utilized for zero-sum games where both players lack perfect information about each other as outlined in Section III-E, where at each round $T = 1, 2, \ldots$, players go through each node to find the optimal pure best response to the entire history of the opponent. At each round T, we compute the best response for each player, with the beliefs described in Section III-E. We proceed through rounds $T = 1, 2 \ldots$ until the beliefs do not change. To compute the best response, we go through each node in the network.

In the following, we present the optimal values, i.e., s_i^m for a given selected node i_m and $\mathcal{P}_m$. $\mathbf{h_i}$ is the ith column of the matrix H, including the edge weight between node i and all other nodes in the network, recall $H = (I + L)^{-1}$ from Theorem 1.

Theorem 3. *For $\mathcal{P}_1$, the Maximizer, the best response to $\mathcal{P}_2$ is to set s_i^1 to either 0 or 1, regardless of i_1. When $\mathcal{P}_2$, the Minimizer, plays fully rationally, considering the Maximizer's past history, in response to $\mathcal{P}_1$, the optimal value s_i^{2*} for the innate opinion with given i_2 can be determined through Karush-Kuhn-Tucker (KKT) conditions.*

Proof. The adversary's optimization problem for the above-established polarization function in (3), is convex in $\mathbf{z}$ and hence in $\mathbf{s}$ [8]. Hence, the Maximizer's optimal choice for the innate opinion (s_i) of the selected agent i_1 is limited to the boundary value of s_i, which is 0 or 1. Meanwhile, the Minimizer's optimal strategy for minimizing polarization, responsive to each Maximizer's past action, can be derived via a straight-forward application of KKT optimality conditions.

By Theorem 1, J-F dynamics lead to $z = Hs$ as a steady-state point where $H = (I + L)^{-1}$ is a doubly-stochastic matrix. Plugging (2) into (3), we obtain polarization as a function of innate opinions: $J(s) = (Hs - c^\top Hs)^\top (Hs - c^\top Hs)$, where we define $\mathbf{c}_n \triangleq \frac{1}{n}\mathbf{1}_n$ and $\mathbf{1}_n$ refers to the $n \cdot 1$ 1s vector. We next note that $\mathbf{c}^\top Hs = \mathbf{c}^\top s$, due to H being column-stochastic, i.e., $c^\top H = c^\top$, hence we have

$$J(s) = (Hs - c^\top s)^\top (Hs - c^\top s) \tag{5}$$

To consider to learning dynamic of players, we obtain $(\overline{J(s)})$ as a function of innate opinions. Denote r_1 actions of Maximizer that have been observed in the game. s^l is the updated innate opinion after Maximizer's action. To reflect the influence of Maximizer's action(α^l), we rewrite (6) by replacing innate opinion (s) with the updated innate opinion (s^l). Recalling the probability of $\mathcal{P}_1$ (Maximizer) playing the action α^l ($l \in G_1$) is μ^l, hence we have:

$$\overline{J(s)} = \sum_{l=1}^{r_1} \mu^l (Hs^l - c^\top s^l)^\top (Hs^l - c^\top s^l) \tag{6}$$

Towards computing $\frac{d\overline{J}}{ds^l}$, we have $\overline{J}(s^l+\epsilon e_i)-\overline{J}(s^l)$

$$
\begin{aligned}
&= \sum_{l=1}^{r_1} \mu^l [(Hs^l - c^\top s^l + \epsilon(He_i - c^\top e_i))^\top \\
&\quad (Hs^l - c^\top s^l + \epsilon(He_i - c^\top e_i)) \\
&\quad - (Hs^l - c^\top s^l)^\top (Hs^l - c^\top s^l)] \\
&= \sum_{l=1}^{r_1} \mu^l [\epsilon^2 (He_i - c^\top e_i)^\top (He_i - c^\top e_i) \\
&\quad + \epsilon((Hs^l - c^\top s^l)^\top)(He_i - c^\top e_i) \\
&\quad + (He_i - c^\top e_i)^\top (Hs^l - c^\top s^l)]
\end{aligned}
$$

We next compute $\frac{d\overline{J}}{ds^l} = \lim_{\epsilon\to 0} \frac{\overline{J}(s+\epsilon e_i)-\overline{J}(s)}{\epsilon} = 2(Hs - c^\top s)^\top (He_i - c^\top e_i)$. Noting that $He_i = \mathbf{h}_i$ and $\mathbf{c}^\top e_i = \frac{1}{n}$, we obtain the locally optimal point s_i^* as the solution of

$$
\sum_{l=1}^{r_1} \mu^l \left(s_i \mathbf{h_i} - s_i \mathbf{c}_n + \sum_{j\neq i} (s_j^l \mathbf{h_j} - c_j s_j) \right)^\top (\mathbf{h}_i - \mathbf{c}_n) = 0.
$$

Using the column-stochasticity of H, we obtain s_i^{2*} as

$$
s_i^{2*} = \frac{-\sum_{l=1}^{r_1} \mu^l \sum_{j\neq i} {s^l}_j (\mathbf{h_j} - \mathbf{c}_n)^\top (\mathbf{h_i} - \mathbf{c}_n)}{(\mathbf{h_i} - \mathbf{c}_n)^\top (\mathbf{h_i} - \mathbf{c}_n)}.
$$

□

Algorithm 1 Proposed Algorithm

Input: innate opinion vector ($\mathbf{s}$)
Hyper-parameter:K - maximum # of iterations in the game, the simulation stops if game round T equals K;
Output:Nash Equilibrium

```
Given T = 0 T: count of game the algorithm runs
while T ≤ K do
    T = T + 1
    for i^1 ∈ [N] do
        for s_i^1 ∈ {0, 1} do
            Compute J̄_1^k(z) and store it to J_1
        end for
    end for
    (i^{1*}, s_i^{1*}) = arg max J_1
    Store (i^{1*}, s_i^{1*}) to history G_1^T
    for i^2 ∈ S_2(T) do
        calculate s_i^2 ← (see theorem. (2))
        compute J̄_2(z) and store it to J_2
    end for
    (i^{2*}, s_i^2) = arg max J_2
    Store (i^{2*}, s_i^2) to Minimizer history G_2^T
end while
return G_1^T, G_2^T
```

Therefore, for each given i_m, the optimal action for $\mathcal{P}_m$ is determined. In a single game iteration, $\mathcal{P}_m$ evaluates the polarization change with each agent i in the network, selecting agent i_m and its optimal innate opinion value that yields the maximum/minimum polarization. This pairing, denoted as $\mathcal{P}_m$'s optimal action (i_m, s_i^m), is chosen based on the comparison of all options. Subsequently, $\mathcal{P}_m$ iteratively employs the optimal action, adjusting the innate opinion of node i_m to the optimal value s_i^m, until the game converges or reaches the maximum specified iterations. The Nash Equilibrium is then derived from $\mathcal{P}_m$'s stationary strategy at convergence, as stated in Theorem 2.

V. Numerical Results

A. Data Description

We leverage two real social networks, as detailed in Table II. The Karate club network comprises 34 nodes, while the Reddit network[3] encompasses a larger scale, 553 nodes. The adjacency matrices for both networks are obtained from the respective datasets. In the case of Reddit, innate opinions are also available [19]. We study the Karate club network with generated innate opinion vector $\mathbf{s}_{karate}$, each entry randomly drawn from the range [0,1], similar to the experiment in [20].

Data	Size	Network	Polariz.	Source
Karate club	34	undirected	0.16	[21]
Reddit	553	undirected	0.01	[19]

TABLE II: Data Description

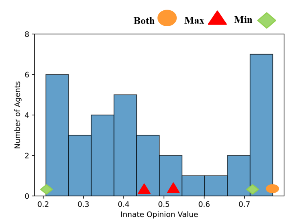


Fig. 1: Karate opinion distr.

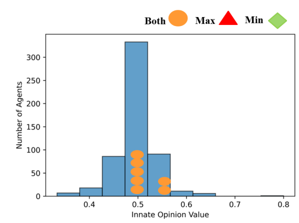


Fig. 2: Reddit opinion distr.

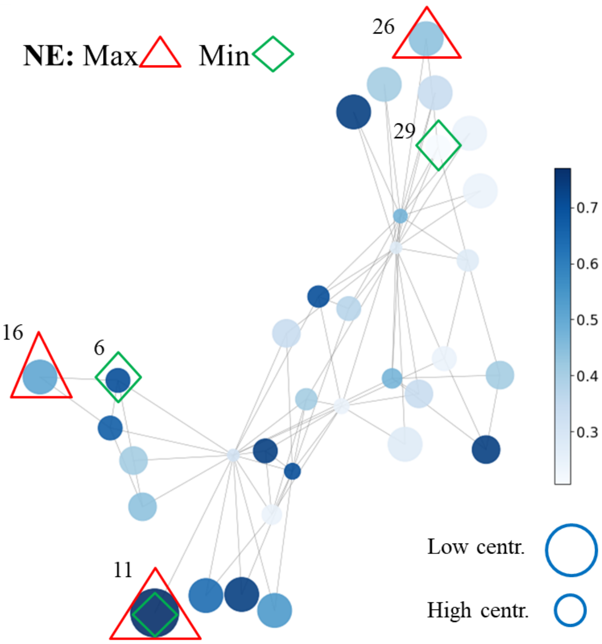
B. Game Results

In this section, we introduce our running example and simulation configurations, operating under certain assumptions consistently applied across all experiments. The datasets and associated codes are publicly accessible on the author's GitHub repository (https://github.com/xzhan176/Network_Node_Selection/tree/main). Figure 3 shows the NE of polarization games observed in two real social networks. In the legend at the upper right, "NE" denotes the players' NE choices in a zero-sum game, where the order of moves has no impact on the outcome. We are particularly interested in the change of polarization, as we define

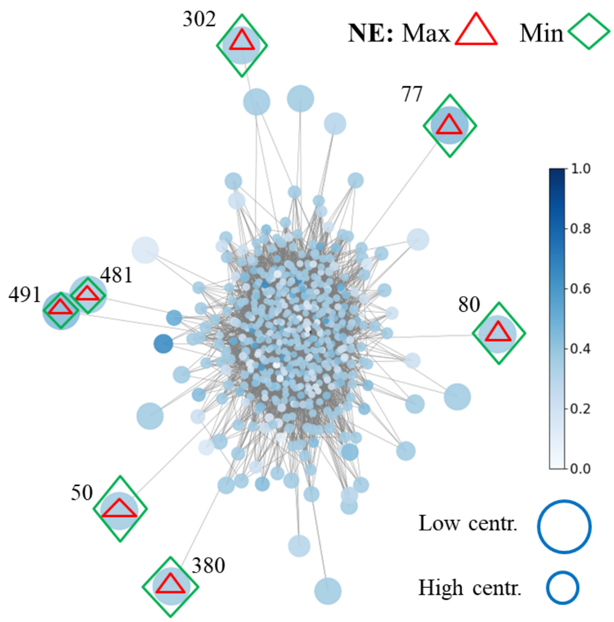
$$
\eta = \frac{J - J_0}{J_0}
$$

where J_0, as defined earlier, denotes the steady-state polarization of the unaltered network.

[3]The original Reddit data includes 556 nodes, with 3 isolated nodes. As these isolated nodes remain unaffected by opinion dynamics, we excluded them from our experiment.



(a) Karate Network Zero-sum Result



(b) Reddit Network Zero-sum Result

Fig. 3: The node color intensity represents the node's opinion value in the range [0,1], and the node size represents the node's degree centrality. At NE, nodes covered by a triangle, and a diamond are selected by the Maximizer and the Minimizer, respectively.

For the Karate club network, $J_0 = 0.16$. At NE, both players select node 11, additionally, the Minimizer opts for nodes 6, 29, while the Maximizer selects nodes 26 and 16. The equilibrium polarization $\overline{P} = 0.18$, yielding $\eta = 1/8$.

For the Reddit network, we have $J_0 = 0.005$ The nodes 50, 77, 80, 302, 380, 481, 491 are selected by both players with identical probabilities. The equilibrium polarization substantially increases here, $\overline{J} = 0.06$ yielding $\eta = 11$.

In Figure 1 and Figure 2, the inherent opinion distributions are depicted. In the Karate club network, agents' opinions exhibit a broader range from 0 to 1, whereas in the Reddit network, agents' innate opinions are more concentrated around 0.5. The red triangle is employed to denote the innate opinions of agents selected by the Maximizer within this distribution, while the green diamond indicates agents selected by the Minimizer.

	Maximizer				Minimizer				
	i^1	s_i	s_i^1	p	i^2	s_i	s_i^2	p	η
NE	11	0.77	1	0.6	11	0.77	0.39	0.11	1/8
	26	0.45	0	0.13	6	0.72	0	0.86	
	16	0.51	1	0.27	29	0.21	1	0.03	

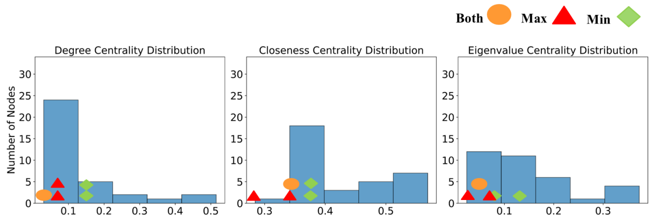
TABLE III: Games on Karate Network.

	Maximizer				Minimizer				
	i^1	s_i	s_i^1	p	i^2	s_i	s_i^2	p	η
NE	50	0.5	1	0.14	50	0.5	0.5	0.14	11
	77	0.5	1	0.14	77	0.5	0.5	0.14	
	80	0.56	1	0.15	80	0.56	0.5	0.15	
	302	0.5	0	0.14	302	0.5	0.5	0.14	
	380	0.5	1	0.14	380	0.5	0.5	0.14	
	481	0.5	1	0.15	481	0.5	0.5	0.15	
	491	0.56	1	0.14	491	0.56	0.5	0.14	

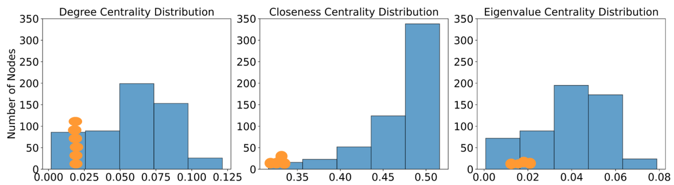
TABLE IV: Games on Reddit Network.

In Figure 4, we depict the centrality distribution in two distinct networks: the Karate club Network (Figure 4(a)) and the Reddit Network (Figure 4(b)). Three significant centrality metrics are presented: Degree Centrality, Closeness Centrality, and Eigenvalue Centrality [22]. Each subfigure illustrates the distribution of node centralities within its respective network and the position of the selected node in the overall centrality distribution. The circle denotes the centrality of nodes selected by both players, providing a shared perspective. Additionally, the red triangle highlights nodes selected by the Maximizer, while the green diamond represents nodes chosen by the Minimizer. This visualization offers insights into the structural importance of nodes within each network, emphasizing the nodes selected by different strategic players. Notably, the Maximizer's agent exhibits a relatively low centrality, nearly ranking as the smallest among all nodes in the network.

A few observations are in order. First, we observe that for both networks, polarization increases, i.e., $\eta > 0$, as a result of the polarization game. This is theoretically expected since opinion dynamics, which is essentially an averaging operation) makes the steady-state opinions narrowly accumulate around



(a) Karate Network



(b) Reddit Network

Fig. 4: Nodes centrality distribution in network. The circle indicates the centrality of the node selected by both players. The red triangle indicates the centrality of the Maximizer-selected node, and the green diamond indicates the centrality of the Minimizer-selected node.

mean. Hence, the Minimizer does not have a significant opportunity to decrease the already-low polarization. However, for the Maximizer there is significant room to increase polarization via moving the selected agents' opinions to extremal values, i.e., 0 or 1. Secondly, the Maximizer is expected to choose "less connected" agents (those with low centrality) to minimize the impact of averaging operator over the steady-state opinion. As a response, the Minimizer tends to select the same agents to render the impact of Maximizer minimal. Finally, we observe that for the Reddit network where the innate opinions are more centered around mean, compared to the Karate network, the Maximizer has more room to change the polarization, as we numerically observe as a larger η value.

VI. Conclusion

In this work, we have considered a novel problem formulation of polarization games with significant practical implications. We characterized the functional properties of a zero-sum game, including the optimal response for each player given the agent, and the properties of the Nash equilibria. Simulation results on real networks provided valuable insights into the players' behavior in the zero-sum game.

Our research makes notable contributions to the empirical modeling of polarization dynamics over networks, analysis of equilibrium strategies for external players seeking to maximize/minimize polarization, and the derivation of associated equilibrium results grounded in real networks. Moving forward, our focus will extend to exploring different game forms, such as scenarios where players can select more than one node and the dynamics of a non-zero-sum game where players are constrained to choose the same agent. Furthermore, we aim to develop prediction methodologies through the incorporation of advanced deep learning models, particularly those based on convolutional neural networks, to achieve predicting accurate equilibrium strategies on large networks.

References

[1] E. J. Finkel, C. A. Bail, M. Cikara, P. H. Ditto, S. Iyengar, S. Klar, L. Mason, M. C. McGrath, B. Nyhan, D. G. Rand *et al.*, "Political sectarianism in america," *Science*, vol. 370, no. 6516, pp. 533–536, 2020.

[2] S. Iyengar, Y. Lelkes, M. Levendusky, N. Malhotra, and S. J. Westwood, "The origins and consequences of affective polarization in the united states," *Annual review of political science*, vol. 22, pp. 129–146, 2019.

[3] H. Allcott, L. Boxell, J. Conway, M. Gentzkow, M. Thaler, and D. Yang, "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic," *Journal of public economics*, vol. 191, p. 104254, 2020.

[4] D. R. Coats, "Statement for the record, worldwide threat assessment of the us intelligence community, senate select committee on intelligence," *Office of the Director of National Intelligence. United States*, vol. 28, 2017.

[5] P. Howard, B. Ganesh, D. Liotsiou, J. Kelly, and C. François, *The IRA, Social Media and Political Polarization in the United States, 2012-2018*. Project on Computational Propaganda, 2018.

[6] N. E. Friedkin and E. C. Johnsen, "Social influence and opinions," *Journal of Mathematical Sociology*, vol. 15, no. 3-4, pp. 193–206, 1990.

[7] J. Von Neumann and O. Morgenstern, "Theory of games and economic behavior," in *Theory of games and economic behavior*. Princeton university press, 2007.

[8] M. F. Chen and M. Z. Racz, "An adversarial model of network disruption: Maximizing disagreement and polarization in social networks," *IEEE Transactions on Network Science and Engineering*, 2021.

[9] C. Musco, C. Musco, and C. Tsourakakis, "Minimizing polarization and disagreement in social networks," ser. Proceedings of the World Wide Web Conference, WWW 2018, Apr. 2018, pp. 369–378.

[10] A. V. Proskurnikov and R. Tempo, "A tutorial on modeling and analysis of dynamic social networks. part i," *Annual Reviews in Control*, vol. 43, pp. 65–79, 2017.

[11] M. H. DeGroot, "Reaching a consensus," *Journal of the American Statistical association*, vol. 69, no. 345, pp. 118–121, 1974.

[12] M. Richardson and P. Domingos, "Mining knowledge-sharing sites for viral marketing," in *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2002, pp. 61–70.

[13] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, 2003, pp. 137–146.

[14] S. Dhamal, W. Ben-Ameur, T. Chahed, and E. Altman, "Optimal investment strategies for competing camps in a social network: A broad framework," *IEEE Transactions on Network Science and Engineering*, vol. 6, no. 4, pp. 628–645, 2018.

[15] S. Eshghi, V. M. Preciado, S. Sarkar, S. S. Venkatesh, Q. Zhao, R. D'Souza, and A. Swami, "Spread, then target, and advertise in waves: optimal budget allocation across advertising channels," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 750–763, 2018.

[16] J. Gaitonde, J. Kleinberg, and E. Tardos, "Adversarial perturbations of opinion dynamics in networks," in *Proceedings of the 21st ACM Conference on Economics and Computation*, 2020, pp. 471–472.

[17] J. Robinson, "An iterative method of solving a game," *Annals of mathematics*, pp. 296–301, 1951.

[18] D. Fudenberg, F. Drew, D. K. Levine, and D. K. Levine, *The theory of learning in games*. MIT press, 1998, vol. 2.

[19] A. De, S. Bhattacharya, P. Bhattacharya, N. Ganguly, and S. Chakrabarti, "Learning a linear influence model from transient opinion dynamics," in *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, 2014, pp. 401–410.

[20] E. Biondi, C. Boldrini, A. Passarella, and M. Conti, "Dynamics of opinion polarization," *arXiv preprint arXiv:2206.06134*, 2022.

[21] W. W. Zachary, "An information flow model for conflict and fission in small groups," *Journal of anthropological research*, vol. 33, no. 4, pp. 452–473, 1977.

[22] F. A. Rodrigues, "Network centrality: an introduction," *A mathematical modeling approach from nonlinear dynamics to complex systems*, pp. 177–196, 2019.