POLITECNICO DI TORINO
Repository ISTITUZIONALE

Coevolution of actions and opinions in networks of coordinating and anti-coordinating agents

Original
Coevolution of actions and opinions in networks of coordinating and anti-coordinating agents / Liang, H., Ye, M., Zino, L., Xia, W.. - In: IFAC JOURNAL OF SYSTEMS AND CONTROL. - ISSN 2468-6018. - ELETTRONICO. - 33:(2025), pp. 1-13. [10.1016/j.ifacsc.2025.100325]

Availability:
This version is available at: 11583/3001848 since: 2025-07-18T19:15:27Z

Publisher:
Elsevier

Published
DOI:10.1016/j.ifacsc.2025.100325

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

(Article begins on next page)

12 November 2026

Coevolution of Actions and Opinions in Networks of Coordinating and Anti-Coordinating Agents*

Hong Liang[a], Mengbin Ye[b], Lorenzo Zino[c], Weiguo Xia*[a]

[a] *Key Laboratory of Intelligent Control and Optimization for Industrial Equipment of the Ministry of Education and the School of Control Science and Engineering, Dalian University of Technology, Dalian, China*
[b] *School of Computer and Mathematical Sciences, University of Adelaide, Adelaide, SA, Australia*
[c] *Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy*

Abstract

In this paper, we investigate the dynamics of coordinating and anti-coordinating agents in a coevolutionary model for actions and opinions. In the model, individuals in a population interact on a two-layer network, sharing their opinions and observing others' actions, while revising their own opinions and actions according to a game-theoretic mechanism grounded in social psychology literature. First, we consider the scenario of coordinating agents, where convergence to a Nash equilibrium (NE) is guaranteed. We identify conditions for reaching consensus configurations and establish regions of attraction for these equilibria. Second, we study networks of anti-coordinating agents. Here, we prove that all trajectories converge to an NE by leveraging potential game theory. Then, we establish analytical conditions on the network structure and model parameters to guarantee the existence of consensus and polarized equilibria, characterizing their regions of attraction.

Keywords: Evolutionary game theory, network system, opinion dynamics, social network, influence network

1. Introduction

Over the past decades, extensive research has been conducted on dynamic models of opinion formation across multiple disciplines [1–8].These models have been instrumental in understanding opinion formation processes and social power, further analyzing complex social phenomena, such as the emergence of consensus, oscillations and disagreement, or polarization [9–15]. Moreover, they have shown significant potential for applications in different areas, e.g., marketing [16], advertising [17], recommendation systems [18, 19], and epidemic mitigation [20, 21]. Meanwhile, evolutionary game theory has emerged as a powerful framework for realistically predicting how individuals make decisions from a limited set of possible actions and how these actions are revised due to social interactions [22–28].

1.1. Literature Review

Research from social psychology, supported by empirical data, shows a strong intertwining between an individual's opinion and their actions. These studies highlight that within a group, sharing opinions can influence people's behaviors, and in turn, an individual's opinion may be shaped by observing the behaviors of others [29–31]. Mathematical models have been proposed to capture the nontrivial interdependency between opinion and actions, based on the so-called continuous opinion and discrete actions (CODA) model [32–36]. Despite being able to capture some important features of the presence of these two distinct aspects of human behavior, CODA models rely on the assumption that an agent's action is only directly (exogenously) influenced by their own opinion, neglecting the presence of any other endogenous mechanisms. In other words, they assume that actions are just expressions of the opinions. This does not allow for the capture of misalignment between actions and opinion due to, e.g., social pressure, which may lead to the well-known phenomena of the emergence of unpopular norms [37]. Recently, a novel paradigm has been proposed, in which actions and opinions coevolve on a two-layer network model [38–40].

In particular, Dehghani Aghbolagh *et al.* [40] proposed a coevolutionary model for opinions and actions in an evolutionary game-theoretic form. The proposed model accounts for social influence due to opinion sharing between individuals, social pressure due to observing the actions of others, and an individual's tendency to act in consistency with their opinion. Potential game theory was leveraged to prove the global convergence of the dynamics to an NE (NE) [41], and the existence and stability of polarized equilibria were studied. The model of Dehghani Aghbolagh *et al.* [40] relies on the assumption that individuals are always faced with a situation of strategic complements,

*This work was supported in part by the National Natural Science Foundation of China (61973051, 62122016), LiaoNing Science and Technology Program (2023JH2/101700361), and the Fundamental Research Funds for the Central Universities (DUT24ZD122).

*Corresponding author. Email: wgxiaseu@dlut.edu.cn

and thus tend to coordinate with others on the same action. However, this is not the case in many real-life scenarios, where strategic substitutes are often present [42]. For instance, when a similar service is offered by different providers that are subject to congestion (e.g, different paths to reach a destination), the provider that is less congested becomes more attractive. Hence, the observation of others' actions yields a negative feedback effect, whereby agents tend to anti-coordinate with others to reduce congestion.

1.2. Summary of Contribution

In this paper, we extend the coevolutionary model proposed in [40], to account for the presence of coordinating agents (capturing strategic complements) or anti-coordinating agents (capturing strategic substitutes). Our key contributions focus on consensus and polarization equilibria. For coordinating agents, we establish analytical conditions (on the network and agent parameters) for the existence of consensus equilibria and their basins of attraction. For anti-coordinating agents, we first establish a general result, showing convergence to an NE occurs under mild assumptions on the network and agent activation sequence. We then identify analytical conditions for the existence of consensus and polarization equilibria in networks of anti-coordinating agents. By numerically evaluating the various aforementioned conditions on the equilibria, we shed light on how the coevolutionary process hinders or facilitates the emergence of consensus and polarization, and how this differs for coordinating and anti-coordinating agents.

The rest of this paper is organized as follows. Section 2 provides the notational and mathematical preliminaries, introduce the model, and derive its closed-form update rule. The main results are presented in Section 3. Section 4 and Section 5 discuss the our findings and conclude the paper, respectively.

2. Methods

2.1. Preliminaries: Notation and Graph Theory

We use bold lowercase font to denote vectors (e.g., $\boldsymbol{v} \in \mathbb{R}^n$), while v_i denotes its ith entry. The n-dimensional vectors of all 1s and all 0s are denoted as $\mathbf{1}_n$ and $\mathbf{0}_n$, respectively. We use Roman uppercase font to denote matrices (e.g., $M \in \mathbb{R}^{n\times m}$), denoting by m_{ij} the jth entry of its ith row. A matrix M is said to be row stochastic if all entries are non-negative and each row sums to 1.

A graph is defined by $\mathcal{G} = (\mathcal{V}, \mathcal{E})$, where $\mathcal{V} = \{1, \dots, n\}$ is the node set and $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is the set of edges, such that $(i, j) \in \mathcal{E}$ if and only if (iff) there is a direct edge from i to j. In general, a graph is assumed to be directed, unless it holds that $(i, j) \in \mathcal{E} \Leftrightarrow (j, i) \in \mathcal{E}$, i.e., all connections are bi-directional. In this case, the network is said to be undirected. Given a real square matrix $M \in \mathbb{R}^{n\times n}$ with $m_{ij} \geq 0$, we can associate a weighted (directed) graph $\mathcal{G}[M] := (\mathcal{V}, \mathcal{E}_M, M)$ to M, where M is called the (weighted) adjacency matrix of this graph and $(i, j) \in \mathcal{E}_M$ iff $m_{ij} > 0$.

2.2. Preliminaries: Game Theory

A game $\Gamma = (\mathcal{V}, \mathcal{A}, \boldsymbol{f})$ is defined by three terms: 1) a set of players $\mathcal{V} = \{1, \dots, n\}$, 2) a set of strategies $\mathcal{A}$, and 3) a payoff function $\boldsymbol{f}$. Player i can choose their strategy from the set $\mathcal{A}$, which is denoted by $z_i \in \mathcal{A}$. Strategies are gathered in the vector $\boldsymbol{z} = [z_1, \dots, z_n]^\top$ which is called strategy profile. We further define the vector $\boldsymbol{z_{-i}} \in \mathcal{A}^{n-1}$ that represents the strategy profile of all agents except for agent i. For each player $i \in \mathcal{V}$, we define their payoff function $f_i : \mathcal{A} \times \mathcal{A}^{n-1} \to \mathbb{R}$, denoted as $f_i(z_i, \boldsymbol{z_{-i}})$, which associates to player i the payoff that i would receive by choosing strategy $z_i \in \mathcal{A}$ when the action profile of the others is $\boldsymbol{z_{-i}}$. Payoff functions are gathered in the vector function $\boldsymbol{f} = [f_i, \dots, f_n]^\top$. Given a game $\Gamma = (\mathcal{V}, \mathcal{A}, \boldsymbol{f})$, we can define for each player i the notion of the best response given the strategy profile of the others, $\boldsymbol{z_{-i}}$, as the strategy that they should play in order to maximize their payoff.

Definition 1 (Best response)**.** *Given a game $\Gamma = (\mathcal{V}, \mathcal{A}, \boldsymbol{f})$, the best response strategies for player $i \in \mathcal{V}$ is defined as*

$$\mathcal{B}_i(f_i(\cdot, \boldsymbol{z_{-i}})) \triangleq \{\zeta^* : \zeta^* \in \arg\max_{\zeta\in\mathcal{A}} f_i(\zeta, \boldsymbol{z_{-i}})\}. \quad (1)$$

A NE is a key concept in game theory, representing strategy profiles $\boldsymbol{z}$ where no player has any incentives to deviate from the current strategy, as formalized in the following.

Definition 2 (Nash equilibrium)**.** *The strategy profile $\boldsymbol{z}^*$ is an NE iff $f_i(z_i^*, \boldsymbol{z}_{-i}^*) \geq f_i(z_i^{'}, \boldsymbol{z}_{-i}^*)$, $\forall\, z_i^{'} \in \mathcal{A}$ and $\forall\, i \in \mathcal{V}$.*

In other words, at an NE, all players are playing their current best response to the other players' strategies, yielding the following classical result from [43].

Lemma 1. *The strategy profile $\boldsymbol{z}^*$ is an NE iff it holds true that $z_i^* \in \mathcal{B}_i\big(f_i(\cdot, \boldsymbol{z^*_{-i}})\big)$, for all $i \in \mathcal{V}$.*

A network game is a game $\Gamma = (\mathcal{V}, \mathcal{A}, \boldsymbol{f})$ where the payoff function is modulated by a network $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Specifically, the payoff that each agent i would receive depends only on the strategy of their neighbors $\mathcal{N}_i := \{j : (i, j) \in \mathcal{E}\}$. In the following, we present two examples of network games that will be used in this paper.

Example 1 (Network coordination and anti-coordination)**.** *Consider a population of $n \geq 2$ players, where each player has a binary strategy set $\mathcal{A} = \{-1, +1\}$. Player interactions are represented via a weighted graph $\mathcal{G}[A] = (\mathcal{V}, \mathcal{E}_A, A)$, where the weight a_{ij} represents the impact of the strategy of j on i's payoff. Player i can be either a coordinating agent or an anti-coordinating agent. A coordinating agent receives a positive contribution to their*

payoff from each of their neighbors with whom they coordinate (i.e., they adopt the same strategy). On the contrary, an anti-coordinating agent would receive a positive contribution to their payoff from each of their neighbors with whom they do not coordinate (i.e., they adopt different strategies).

In mathematical terms, given strategy profile $\boldsymbol{x}$*, each player* $i \in \mathcal{V}$ *is associated with the following payoff function:*

$$f_i\left(x_i, \boldsymbol{x}_{-\boldsymbol{i}}\right) = \frac{\varepsilon_i}{4} \sum\nolimits_{j \in \mathcal{V}} a_{ij} \left[\left(1 - x_j\right)\left(1 - x_i\right) + \left(1 + \alpha\right)\left(1 + x_j\right)\left(1 + x_i\right)\right] \tag{2}$$

with

$$\varepsilon_i = \begin{cases} +1 & \text{if } i \text{ is a coordinating agent,} \\ -1 & \text{if } i \text{ is an anti-coordinating agent.} \end{cases} \tag{3}$$

The constant $\alpha \geq 0$ *is a non-negative parameter that captures the relative advantage of strategy* $+1$ *with respect to strategy* -1*: a coordinating agent* i *would get a unit payoff for coordinating on action* -1 *and payoff* $1 + \alpha$ *for coordinating on action* $+1$*; an anti-coordinating agent would get the opposite payoffs.*

Observe that the best-response strategy for the payoff function in (2) *as defined in* (1) *yields a threshold such that* $+1 \in \mathcal{B}_i\left(f_i(\cdot, \boldsymbol{x})\right)$ *iff a discriminant quantity* $\delta_i(\boldsymbol{x})$ *satisfies*

$$\delta_i(\boldsymbol{x}) = \varepsilon_i \sum\nolimits_{j \in \mathcal{V}} a_{ij} \left[2x_j + \alpha\left(1 + x_j\right)\right] \geq 0. \tag{4}$$

In other words, action $+1$ *is the best response for a coordinating (anti-coordinating, respectively) agent if a sufficiently large (small, respectively) fraction of the neighbors adopts it (weighted by the entries of matrix* A *and by the relative advantage* α*).*

Remark 1. *Network coordination and anti-coordination games have been extensively used in the context of decision-making [22]. On the one hand, network coordination games model strategic complements, whereby a player choosing a strategy makes that strategy more appealing to other players. This captures, e.g., the adoption of social norms and conventions, or of new technological assets [23, 24, 27]. On the other hand, network anti-coordination games are representative of scenarios with strategic substitutes, whereby a player choosing a strategy makes that strategy less appealing to others. This is the case, e.g., of competition for resources that can be subject to congestion [42, 44]. In real-world scenarios, there could also be situations in which part of the population has a tendency to conform with others and part of the population aims at differentiation, yielding heterogeneous* ε_i *across the population [45, 46].*

Example 2 (Opinion dynamics)**.** *The French-DeGroot opinion dynamics model [1, 47] can be cast as a network game [48]. We consider a population of* $n \geq 2$ *players, where each player has a continuous strategy* $\mathcal{A} = [-1, +1]$*, representing the agent's opinion. We denote by* $y_i(t) \in \mathcal{A}$ *the opinion of* i *at discrete time* t*. agents share their opinions through interactions on a weighted graph* $\mathcal{G}[W] = (\mathcal{V}, \mathcal{E}_W, W)$*, where* W *is assumed to be row stochastic. The weight* w_{ij} *represents the social influence of* j *on* i *in its opinion formation process. The opinion formation process is captured by a network game, with a payoff function*

$$f_i\left(y_i, \boldsymbol{y}_{-\boldsymbol{i}}\right) = -\frac{1}{2} \sum\nolimits_{j \in \mathcal{V}} w_{ij} \left(y_i - y_j\right)^2. \tag{5}$$

The best-response strategy to (5) *yields*

$$\mathcal{B}_i\left(f_i(\cdot, \boldsymbol{y})\right) = \sum\nolimits_{j \in \mathcal{V}} w_{ij} y_j(t), \tag{6}$$

i.e., a weighted average of the opinions shared by the neighbors [47]. An additional term can be added to the payoff function to account for agents' attachment to their prejudices [49].

2.3. Model

In this section, we present the coevolutionary model of actions and opinions for networks of coordinating and anti-coordinating agents. We start by formulating the payoff function, building on [40], and discussing its main terms. Then, we present the model dynamics by defining the update rule and expressing a closed-form expression for it.

2.3.1. Payoff Function

Consider a population of $n \geq 2$ agents. According to the coevolutionary dynamics proposed in [40], at each discrete-time instant $t \in \mathbb{Z}_{>0}$, each agent $i \in \mathcal{V}$ is characterized by a binary state variable $z_i(t) = (x_i(t), y_i(t)) \in \mathcal{A} = \{-1, 1\} \times [-1, 1]$. The first entry, $x_i(t) \in \{-1, +1\}$, represents the agent's binary action, while the second entry, $y_i(t) \in [-1, 1]$, is the agent's opinion, viz. their orientation and preference towards the two action options. Namely, $y_i(t) > 0$ and $y_i(t) < 0$ indicate the agent i prefers action $+1$ and action -1, respectively, with the magnitude capturing the strength of preference. Agents' state variables are gathered in a n-by-2 matrix $\boldsymbol{z}(t) = [\boldsymbol{x}(t), \boldsymbol{y}(t)] = [z_1(t)^\top, z_2(t)^\top, \ldots, z_n(t)^\top]^\top$, which represents the state of the population at time t.

Agents interact on social networks, exchanging opinions on a weighted graph $\mathcal{G}[W]$ and observing the actions of others on a weighted graph $\mathcal{G}[A]$, with weighted adjacency matrices W and A that are both assumed to be row-stochastic. Hence, each entry in W and A represents the relative social influence of other players on opinions and actions, respectively.

In [40], agents revise their actions assuming that they engage in a game, with a payoff function that accounts for coordination with others' actions, the social influence of others' opinions, and self-consistency between their own actions and opinions. Here, we build on that model and extend it to allow agents to anti-coordinate with their

peers. Specifically, given a state $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y})$, we define the payoff function $f_i(z_i, \boldsymbol{z_{-i}})$ for agent i to adopt strategy $z_i = (x_i, y_i)$ as

$$f_i(z_i, \boldsymbol{z_{-i}}) = \frac{\varepsilon_i \lambda_i (1-\beta_i)}{4} \sum\nolimits_{j\in\mathcal{V}} a_{ij} \Big[(1-x_i)(1-x_j) + (1+\alpha)(1+x_j)(1+x_i)\Big] - \frac{1}{2}(1-\lambda_i)\beta_i \sum\nolimits_{j\in\mathcal{V}} w_{ij}(y_i - y_j)^2 - \frac{1}{2}\lambda_i\beta_i(y_i - x_i)^2. \tag{7}$$

This expression accounts for three terms. The first term involves only the action and coincides with the payoff for a coordination or anti-coordination game in (2). The second term depends only on the opinion and is associated with the payoff of a French-DeGroot opinion dynamics in (5). The third term accounts for the human tendency to behave consistently with their opinions, penalizing the adoption of actions that are distant from them. While the last two terms coincide with those in [40], the first one differs by the inclusion of the parameter $\varepsilon_i \in \{-1, +1\}$, accounting for whether agent i is a coordinating or an anti-coordinating agent. In particular, i engages in a pairwise coordination game if $\varepsilon_i = +1$ or a pairwise anti-coordination game if $\varepsilon_i = -1$. The payoff in (7) coincides with the one from [40] if $\varepsilon_i = +1$ for all $i \in \mathcal{V}$. Parameters $\lambda_i \in [0, 1]$ and $\beta_i \in [0, 1]$ weight the contribution of actions and opinions in determining agent i's payoff, respectively. These two parameters also regulate the coupling of opinion formation and decision-making in the third term in (7). Similarly to [40], a fourth term can be added to account for agents' prejudices. Here, we omit this term, since it is not relevant to our analyses.

2.3.2. Discussion on the model

The game with payoff function in (7) was originally proposed in [40] with $\varepsilon_i = 1$ for all $i \in \mathcal{V}$, as a combination of a network coordination game and the French-DeGroot opinion dynamics [49], coupled by the consistency term. Such a payoff function was proposed to capture the complex intertwining between opinion formation processes and decision-making in social networks. In fact, empirical evidence suggests that the exchange of opinions can affect agents' actions, while the way people form opinions can be influenced by observing the actions of others [29–31]. However, the model was proposed in the context of actions with strategic complements, whereby the decision-making is captured by a coordination mechanism, and both opinion sharing and observation of others' actions yield a positive feedback loop.

Here, we relax this assumption, and we allow to have strategic substitutes, and thus, anti-coordinating agents. In fact, for decisions concerning actions with strategic substitutes, opinion formation, and decision-making are also deeply intertwined. For instance, when a similar service can be offered by different providers who are subject to congestion (e.g, two different restaurants to have lunch at or two different streets to take in an infrastructure network), agents still share their opinions on which provider they believe is better/preferred, and this still has a (positive) influence on others' opinions, even though the observation of others' actions yields now a different feedback effect: the provider that is less congested becomes more attractive, yielding an anti-coordination game.

In general, complex decision-making may involve, to some extent, both strategic complements and substitutes, and agents can be heterogeneously more subject to one or the other. This is the case, e.g., of marketing decisions, where there are agents who prefer to conform with the majority, while others prefer to differentiate from it. For this reason, each agent $i \in \mathcal{V}$ is characterized by a parameter $\varepsilon_i \in \{-1, +1\}$, which determines the mechanism (anti-coordination or coordination) that affects the agent. Finally, it is worth noticing that opinions always provide a positive social influence, consistent with empirical observations on the limited impact of negative social influence [50].

2.3.3. Update Rule

Agents revise their strategy by maximizing their payoff function jointly with respect to their two decision variables (actions and opinions), as proposed in [40]. At each time t, each agent is either activated or inactive. An agent i that is activated updates their opinion and action simultaneously according to the best-response update rule:

$$z_i(t+1) = \mathcal{B}_i(f_i(\cdot, \boldsymbol{z_{-i}}(t))). \tag{8}$$

where $\mathcal{B}_i$ is the best-response strategy defined in (1) and $f_i(\cdot, \boldsymbol{z_{-i}}(t))$ is the payoff from (7), with the convention that, if $\mathcal{B}_i(f_i(\cdot, \boldsymbol{z_{-i}}(t)))$ comprises multiple elements, then it is chosen the one with $x_i(t+1) = x_i(t)$, which is unique for $\beta_i > 0$ (see [40]), to account for social inertia in human decision-making [27]. For all other agents who are inactive, their strategies are unchanged: $z_k(t+1) = z_k(t)$. In other words, active agents tend to maximize their current payoff, based on the current action of others, as supported by empirical evidence [51]. Following [40], we derive a closed-form expression for the update rule of the coevolutionary dynamics by explicitly writing the maximizers of (7).

Proposition 1. *Given a state $\boldsymbol{z} = (\boldsymbol{x}, \boldsymbol{y}) \in \mathcal{A}$ and an agent $i \in \mathcal{V}$, let us define the discriminant*

$$\delta_i(\boldsymbol{z}(t)) = \frac{\varepsilon_i \lambda_i (1-\beta_i)}{2} \sum_{j\in\mathcal{V}} a_{ij} \left[2x_j(t) + \alpha(1 + x_j(t))\right] + 2(1-\lambda_i)\lambda_i\beta_i \sum\nolimits_{j\in\mathcal{V}} w_{ij} y_j(t). \tag{9}$$

If $i \in \mathcal{V}$ is activated at time t, the state is updated to

$$x_i(t+1) = \mathcal{S}(\delta_i(\boldsymbol{z}(t))) \tag{10a}$$

$$y_i(t+1) = (1-\lambda_i) \sum\nolimits_{j\in\mathcal{V}} w_{ij} y_j(t) + \lambda_i \mathcal{S}(\delta_i(\boldsymbol{z}(t))), \tag{10b}$$

where

$$\mathcal{S}(\delta_i(\boldsymbol{z}(t))) = \begin{cases} +1 & \text{if } \delta_i(\boldsymbol{z}(t)) > 0, \\ x_i(t) & \text{if } \delta_i(\boldsymbol{z}(t)) = 0, \\ -1 & \text{if } \delta_i(\boldsymbol{z}(t)) < 0. \end{cases} \tag{11}$$

Proof. Case $\varepsilon = +1$ is proved in [40]. The proof for $\varepsilon = -1$ follows the same arguments, and is thus omitted. □

Next, we formalize two standing assumptions that will be enforced for the rest of the paper. First, we assume that the population is homogeneous, second, we pose some mild constraints on the activation sequence, which are quite standard in the related literature [38, 52].

Assumption 1 (Homogeneous population)**.** *The parameters are homogeneous, i.e., for all $i \in \mathcal{V}$, there holds $\lambda_i = \lambda \in (0,1), \beta_i = \beta \in (0,1)$, and $\varepsilon_i = \varepsilon \in \{-1,+1\}$. Moreover, the two graphs are undirected, i.e., $W = W^\top$ and $A = A^\top$.*

Assumption 2 (Activation sequence)**.** *There exists a finite positive integer T such that, for any $t \geq 0$, every agent activates at least once in the interval $t, t+1, \ldots, t+T-1$.*

Under Assumption 1, all agents have the same ε. Hence, we can identify two scenarios of interest: i) Networks of *coordinating agents*, where $\varepsilon = +1$; and ii) Networks of *anti-coordinating agents*, where $\varepsilon = -1$. In the rest of this paper, we present results on these two scenarios. First, in Section 3.1, we consider a population of coordinating agents, where [40] guarantees convergence to an equilibrium, and we focus on characterizing conditions under which the system converges to a consensus. Second, in Section 3.2, we consider a population of anti-coordinating agents and, after proving convergence to an equilibrium, we investigate consensus and polarized equilibria and their basins of attraction.

3. Results

3.1. Results for Coordinating Agents

In the presence of only coordinating agents ($\varepsilon_i = +1, \forall\, i \in \mathcal{V}$), the results from [40] guarantee that the coevolutionary dynamics under Assumptions 1–2 converges to an NE. However, in general, multiple equilibria may be present. In [40], some conditions have been established under which polarized equilibria exist and are stable. Here, instead, we focus on consensus equilibria, which are defined as follows.

Definition 3 (Consensus configuration)**.** *We say that the vector $\boldsymbol{z}(t) = [\boldsymbol{x}(t), \boldsymbol{y}(t)]$ is a consensus configuration at time t if $\boldsymbol{x}(t) = \boldsymbol{y}(t) = \mathbf{1}_n$ or $\boldsymbol{x}(t) = \boldsymbol{y}(t) = -\mathbf{1}_n$.*

We now prove that consensus configurations are always NEs of the game when all players are coordinating agents.

Theorem 1. *Consider the coevolutionary model in* (10)*. Suppose that Assumption 1 holds and $\varepsilon = 1$. Then the consensus configurations $\boldsymbol{z}^* = [\mathbf{1}_n, \mathbf{1}_n]$ and $[-\mathbf{1}_n, -\mathbf{1}_n]$ are NE.*

Proof. The proof is reported in Appendix A. □

In the following, we establish conditions under which the dynamics converge to such a consensus configuration.

3.1.1. Basin of Attraction of Consensus Configuration

Given a time instant $t \in \mathbb{Z}_+$, and $\boldsymbol{z}(t) = (\boldsymbol{x}(t), \boldsymbol{y}(t))$, we define the following mutually exclusive sets

$$\begin{aligned} \mathcal{V}_1^+(t) &:= \{j : \operatorname{sgn}(y_j(t)) \geq 0, x_j(t) = +1\}, \\ \mathcal{V}_{-1}^+(t) &:= \{j : \operatorname{sgn}(y_j(t)) > 0, x_j(t) = -1\}, \\ \mathcal{V}_1^-(t) &:= \{j : \operatorname{sgn}(y_j(t)) < 0, x_j(t) = +1\}, \\ \mathcal{V}_{-1}^-(t) &:= \{j : \operatorname{sgn}(y_j(t)) \leq 0, x_j(t) = -1\}, \end{aligned}$$

with $\operatorname{sgn}(0) = 0$. Intuitively, $\mathcal{V}_1^+(t)$ and $\mathcal{V}_{-1}^-(t)$ are the sets of agents whose actions and opinions are aligned (the agent's opinion indicates a preference for their current action). On the contrary, $\mathcal{V}_{-1}^+(t)$ and $\mathcal{V}_1^-(t)$ are the sets of agents who are misaligned, i.e., the sign of their opinion differs from their action. Before presenting our results, we report two graph-theoretic definitions.

Definition 4 (Cohesive set)**.** *Let $\mathcal{G}[M] = (\mathcal{V}, \mathcal{E}_M, M)$ be a graph with associated adjacency matrix M. Then, for a given constant $q \in (0,1)$, a set $S \subset \mathcal{V}$ is called q-cohesive if*

$$\frac{\sum_{j \in S} m_{ij}}{\sum_{j \in \mathcal{V}} m_{ij}} \geq q, \ \forall\, i \in S. \tag{12}$$

Conversely, a set $S \subset \mathcal{V}$ is called q-diffusive if

$$\frac{\sum_{j \in S} m_{ij}}{\sum_{j \in \mathcal{V}} m_{ij}} < q, \ \forall\, i \in S. \tag{13}$$

Theorem 2. *Consider the coevolutionary model in* (10)*. Suppose that Assumptions 1–2 hold and $\varepsilon = 1$, and that the initial condition satisfies $\boldsymbol{z}(0) \in \{-1,1\}^n \times (0,+1]^n$. If the set $\mathcal{V}_1^+(0)$ is $\frac{1}{\alpha+2}$-cohesive and $\mathcal{V}_{-1}^+(0)$ is $\frac{\alpha+1}{\alpha+2}$-diffusive, then $\lim_{t\to\infty} \boldsymbol{z}(t) = [\mathbf{1}_n, \mathbf{1}_n]$.*

Proof. The proof is reported in Appendix B. □

Remark 2. *When $\alpha = 0$, the condition reduces to requiring $\mathcal{V}_1^+(0)$ to be $\frac{1}{2}$-cohesive, and the complementary set to be $\frac{1}{2}$-diffusive. On the opposite limit, $\alpha \to \infty$, the condition on $\mathcal{V}_1^+(0)$ intuitively reduces to requiring non-zero cohesiveness.*

We introduce the following assumptions to facilitate our analysis of some (but not all) of the results in our paper.

Assumption 3. *We assume that $A = W$ and $\alpha = 0$.*

Under Assumption 3, the action and opinion update rules are simplified, because (9) reduces to

$$\delta_i(\boldsymbol{z(t)}) = 2\lambda\beta(1-\lambda)\sum_{j\in\mathcal{V}} w_{ij}y_j(t) + (1-\beta)\lambda\sum_{j\in\mathcal{V}} w_{ij}x_j(t). \tag{14}$$

We now explore the conditions that lead two initially polarized groups to reach a consensus equilibrium. To begin, we introduce the following definition.

Definition 5 (Polarized configuration)**.** *We say that a vector $\boldsymbol{z}(t)$ is a polarized configuration at time t if $\mathcal{V}_1^+(t) \cup \mathcal{V}_{-1}^-(t) = \mathcal{V}$, and both $\mathcal{V}_1^+(t) \neq \emptyset$ and $\mathcal{V}_{-1}^-(t) \neq \emptyset$. Specifically, we say that $\boldsymbol{z}(t)$ is polarized with respect to the partition $(\mathcal{V}_1^+(t), \mathcal{V}_{-1}^-(t))$.*

In other words, in polarized configurations, the opinions and actions of every agent are aligned, but they are not at a consensus. Note that in [40], a result is established for when a polarized initial state leads to convergence to a polarized equilibrium, whereas here, we identify conditions for when instead there is convergence to a consensus equilibrium.

Theorem 3. *Consider the coevolutionary model in (10). Suppose that Assumptions 1–3 hold and $\varepsilon = 1$. Suppose $\boldsymbol{z}(0)$ is polarized, with respect to a partition $(\mathcal{V}_1^+(0), \mathcal{V}_{-1}^-(0)) =: (\mathcal{V}_p, \mathcal{V}_n)$. If the below inequalities are satisfied,*

$$\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij} > \max\left\{\frac{1-2\lambda}{1-\lambda}, 1 - \frac{1-\beta}{2(1-\lambda\beta)}\right\}, i \in \mathcal{V}_p, \tag{15a}$$

$$\sum\nolimits_{j\in\mathcal{V}_{-1}^-(0)} w_{ij} < \min\left\{\frac{\lambda}{1-\lambda}, \frac{1-\beta}{2(1-\lambda\beta)}\right\}, i \in \mathcal{V}_n. \tag{15b}$$

Then, it holds $\lim_{t\to\infty} \boldsymbol{z}(t) = [\mathbf{1}_n, \mathbf{1}_n]$.

Proof. The proof is reported in Appendix C. □

Observe that if (15a) holds, then $\mathcal{V}_p$ is $\frac{1}{2}-$cohesive since $\frac{1-2\lambda}{1-\lambda}$ and $1 - \frac{1-\beta}{2(1-\lambda\beta)}$ are always greater than $\frac{1}{2}$.

We now present contour plots to explain the inequalities in (15) more intuitively. In this paper, we make repeated use of similar contour plots, and thus we first clarify the horizontal and vertical axes, as well as the colors used, which will remain consistent across all sections. The 2D contour plots help to visualize the relationship between β and λ (parameters in the model) and a function $f(\lambda, \beta)$ that arises for various inequalities identified in theorems. The X-axis plots β, the parameter that represents the contribution of the opinion in determining an agent's payoff, while the Y-axis plots λ, which represents the contribution of the action in determining an agent's payoff. The contour colors range from black to orange, corresponding to $f(\lambda, \beta)$ taking the smallest to largest value, respectively. Fig. 1 presents the two inequalities in (15). From these two contour plots, it appears that $f(\lambda, \beta)$ is minimized in Fig. 1a and maximized in Fig. 1b for high β values, and medium to high λ values. Put another way, given W and $\boldsymbol{z}(0)$, the inequalities in (15) are more likely to be satisfied with high β, and medium to high λ. This suggests that one can reach a consensus equilibrium from an initially polarized state, if the opinion dynamics strongly impact the coevolutionary process (and have a stronger impact on coordination).

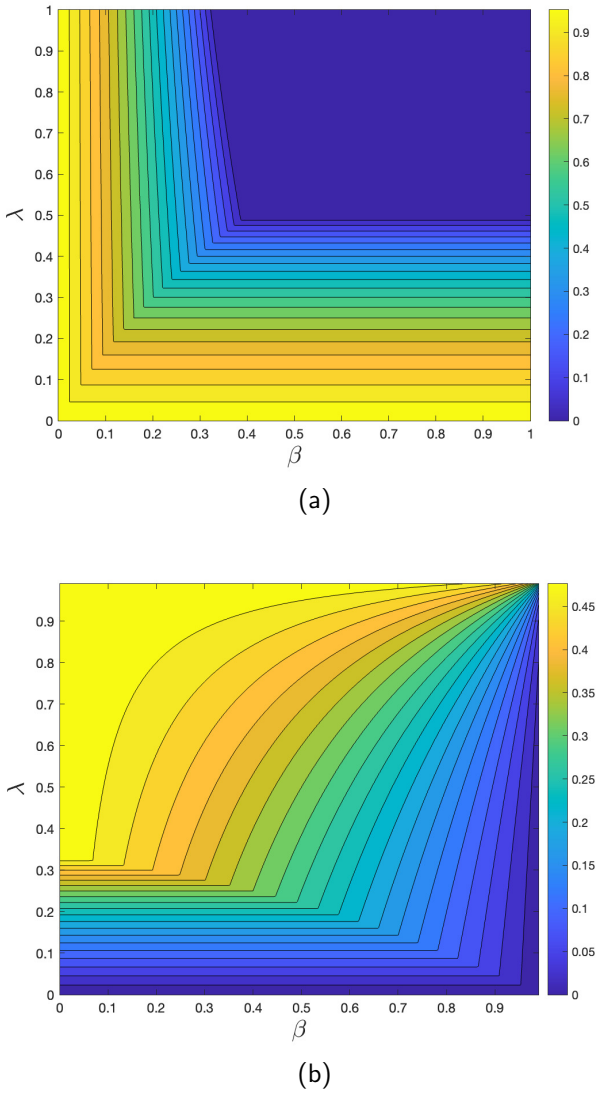

Figure 1: Contour plots of the conditions in (15) from Theorem 3. Panel (a): $f(\lambda, \beta)$ corresponds to the right-hand-side (RHS) of (15a). Panel (b): $f(\lambda, \beta)$ corresponds to the RHS of (15b).

Example 3. *Consider a network of $n = 30$ agents, which are partitioned into two groups, $\mathcal{V}_p = \{1, \ldots, 15\}$ and $\mathcal{V}_n = \{16, \ldots, 30\}$. Given $\boldsymbol{z}(0)$, we set weights uniformly at random and then re-scale them so that conditions in (15) are met with $\mathcal{V}_p = \mathcal{V}_1^+(0)$ and $\mathcal{V}_n = \mathcal{V}_{-1}^-(0)$, and weights are row-stochastic. The network obtained is illustrated in Fig. 2(a). The parameters are $\lambda = 0.8, \beta = 0.6$. As shown in Fig. 2(b) and 2(c), the final actions and opinions converge to a consensus.*

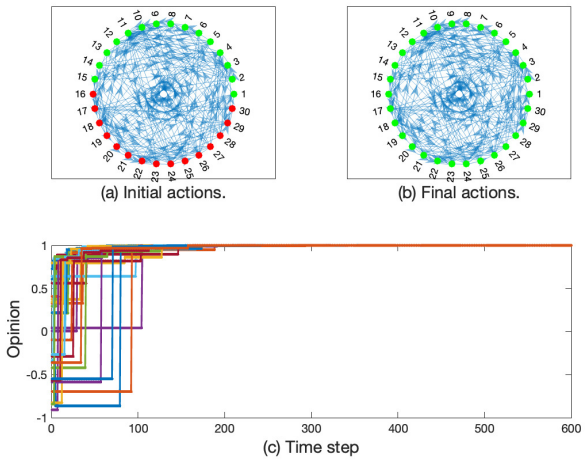

Figure 2: Simulation of Example 3, showing both the network view and the dynamics over time. The network is shown at time $t = 0$ (Panel (a)) and $t = 600$ (Panel (b)), with each node corresponding to an agent. Node i is colored green if agent i is such that $\text{sgn}(y_i) = x_i = +1$ (both opinion and action are positive), and in red if $\text{sgn}(y_i) = x_i = -1$ (both opinion and action are negative). In Panel (c), the temporal evolution of the opinions are shown, with each line corresponding to the opinion of an agent. As $t \to \infty$ we observe convergence to a consensus of $+1$ opinion.

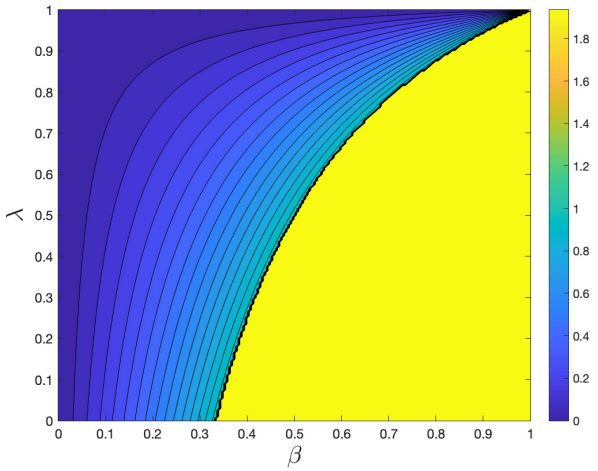

Figure 3: Contour plot of the RHS in (16) in Theorem 5 when $\alpha = 0$.

3.2. Results for Anti-coordinating Agents

3.2.1. Convergence

We start the analysis of the coevolutionary model with anti-coordinating agents ($\varepsilon_i = -1$, $\forall i \in \mathcal{V}$) by proving the following convergence result.

Theorem 4. *Consider the coevolutionary model in* (10). *Suppose that Assumptions 1 and 2 hold and* $\varepsilon = -1$. *Then, the state of the system* $\boldsymbol{z}(t)$ *converges to a steady state* $\boldsymbol{z}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*)$. *Specifically, the action vector* $\boldsymbol{x}(t)$ *converges to* $\boldsymbol{x}^*$ *in finite time, while the opinion vector* $\boldsymbol{y}(t)$ *converges to* $\boldsymbol{y}^*$ *asymptotically.*

Proof. The proof is reported in Appendix D. □

After having established convergence in the presence of anti-coordinating agents, we start analyzing some equilibria of interest, namely, consensus and polarized configurations.

3.2.2. Consensus Equilibria

We start by analyzing whether a consensus equilibrium exists, establishing the following result.

Theorem 5. *Consider the coevolutionary model in* (10). *Suppose that Assumptions 1–2 hold and* $\varepsilon = -1$. *Then,* $\boldsymbol{z}^* = [\mathbf{1}_n, \mathbf{1}_n]$ *is an equilibrium of the coevolutionary model on the two-layer network* $\mathcal{G}$ *iff, for any* $i \in \mathcal{V}$, *it holds*

$$\sum\nolimits_{j \in \mathcal{V}} a_{ij} < \frac{2(1-\lambda)\beta}{(1-\beta)(1+\alpha)}. \tag{16}$$

Proof. The proof is reported in Appendix E. □

We give a 2D contour plot of the RHS of the inequality in Theorem 5 to observe whether a given influence network $\mathcal{G}[A]$ easily satisfies the condition when λ and β take different values with $\alpha = 0$. Fig. 3 depicts the condition in (16) with $\alpha = 0$. From the contour plot, it appears that $\frac{2(1-\lambda)\beta}{(1-\beta)}$ is maximized for high β values, and low to medium λ values. In other words, (16) is more likely to be satisfied when β is large and λ is small. Intuitively, this suggests that it is easier to reach a consensus equilibrium when the decision-making process is more affected by the opinion exchange. We also see a clear distinction between coordinating and anti-coordinating agents. For the former, there is always a consensus equilibrium [53, Proposition 2]. For the latter, the existence of such equilibrium depends on the parameters. Now, we present an example that satisfies the conditions of Theorem 4.

Example 4. *Consider a network of* $n = 30$ *agents with weights sampled uniformly at random and then re-scale so that* $\boldsymbol{A}$ *and* $\boldsymbol{W}$ *are row-stochastic and the conditions in* (16) *are met with* $\alpha = 0$. *The network is illustrated in Fig. 4(a). The parameters are* $\lambda = 0.5, \beta = 0.8$. *As shown in Fig. 4(b) and 4(c), the final actions and opinions converge to a consensus.*

3.2.3. Polarized Equilibria

Here, we discuss how our model may lead to the emergence of stable polarized configurations (see Definition 5). In particular, the following result provides a sufficient condition for the existence of a polarized equilibrium in the presence of anti-coordinating agents, and complements a similar result for coordinating agents in [40, Theorem 2].

Theorem 6. *Consider the coevolutionary model in* (10). *Suppose that Assumptions 1–3 hold and* $\varepsilon = -1$. *Suppose*

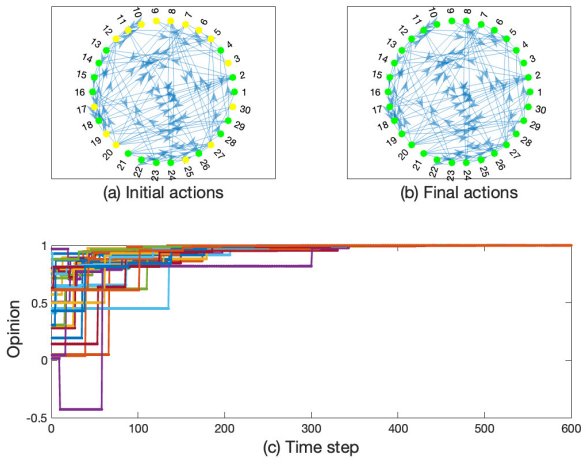

Figure 4: Simulation of Example 4, showing both the network view and the dynamics over time. The network is shown at time $t = 0$ (Panel (a)) and $t = 600$ (Panel (b)), with each node corresponding to an agent. Node i is colored green if agent i is such that $\text{sgn}(y_i) = x_i = +1$ (both opinion and action are positive), and in yellow if $\text{sgn}(y_i) \neq x_i$ (opinion and action have the different sign). In Panel (c), the temporal evolution of the opinions are shown, with each line corresponding to the opinion of an agent. As $t \to \infty$ we observe convergence to a consensus of $+1$ opinion.

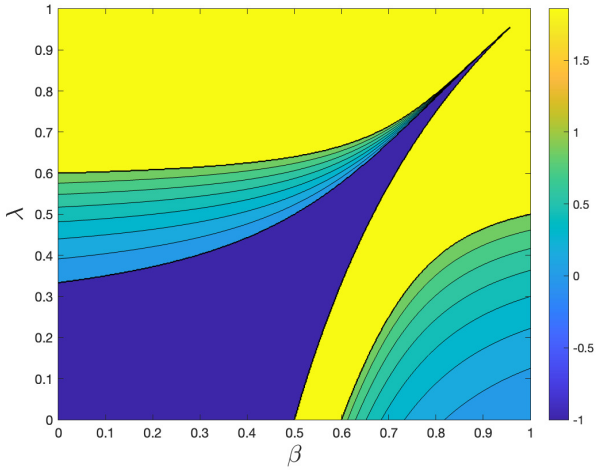

Figure 5: Contour plot of the difference between the RHS and left-hand-side (LHS) in the condition of (19) from Theorem 7 for agents in $\mathcal{V}_p$ or $\mathcal{V}_n$.

there exists a partitioning of the agents $(\mathcal{V}_p, \mathcal{V}_n)$ *with*

$$\begin{aligned} \underline{d}_p &:= \min_{i\in\mathcal{V}_p} \sum\nolimits_{j\in\mathcal{V}_p} w_{ij}, \bar{d}_p := \max_{i\in\mathcal{V}_p} \sum\nolimits_{j\in\mathcal{V}_p} w_{ij}, \\ \underline{d}_n &:= \min_{i\in\mathcal{V}_n} \sum\nolimits_{j\in\mathcal{V}_n} w_{ij}, \bar{d}_n := \max_{i\in\mathcal{V}_n} \sum\nolimits_{j\in\mathcal{V}_n} w_{ij}, \end{aligned} \tag{17}$$

such that the following three conditions are satisfied:

$$\lambda > \max\Big(\frac{\underline{d}_p - \bar{d}_n}{1 + \underline{d}_p - \underline{d}_n}, \frac{\bar{d}_n - \underline{d}_p}{1 + \bar{d}_n - \underline{d}_p}\Big), \tag{18a}$$

$$\frac{\lambda(\underline{d}_p + \bar{d}_n - 1) + \underline{d}_p - \bar{d}_n}{1 - (1-\lambda)(\bar{d}_p + \bar{d}_n - 1)} + \frac{(1-\beta)(1-2\bar{d}_p)}{2\beta(1-\lambda)} > 0, \tag{18b}$$

$$\frac{\lambda(\underline{d}_n + \bar{d}_p - 1) + \underline{d}_n - \bar{d}_p}{1 - (1-\lambda)(\bar{d}_n + \bar{d}_p - 1)} - \frac{(1-\beta)(1-2\bar{d}_n)}{2\beta(1-\lambda)} > 0. \tag{18c}$$

Then, the coevolutionary model in (10) *has a polarized equilibrium with respect to the partition* $(\mathcal{V}_p, \mathcal{V}_n)$.

Proof. The proof is reported in Appendix F. □

Theorem 6 establishes three conditions for the existence of a polarized equilibrium. Although the last two conditions may not be intuitive, all three are easy to check. Below, we study the stability of such equilibrium, establishing a more conservative sufficient condition.

Theorem 7. *Consider the coevolutionary model in* (10). *Suppose that Assumptions 1–3 hold,* $\beta < 1/(2-\lambda)$, *and* $\varepsilon = -1$. *Suppose* $\boldsymbol{z}(0)$ *is a polarized configuration with respect to a partition* $(\mathcal{V}_1^+(0), \mathcal{V}_{-1}^-(0)) =: (\mathcal{V}_p, \mathcal{V}_n)$. *If*

$$\frac{1-2\lambda}{1-\lambda} < \sum\nolimits_{j\in\mathcal{S}} w_{ij} < \frac{1}{2}\Big(1 + \frac{\beta(1-\lambda)}{2\beta - \beta\lambda - 1}\Big), \tag{19}$$

for all $i \in \mathcal{S}$, *holds separately for* $\mathcal{S} = \mathcal{V}_p$ *and* $\mathcal{S} = \mathcal{V}_n$, *then* $\boldsymbol{z}(t)$ *is polarized for any* t *with respect to the same partition* $(\mathcal{V}_p, \mathcal{V}_n)$. *Moreover,* $\lim_{t\to\infty} \boldsymbol{z}(t) = \boldsymbol{z}^*$, *where* $\boldsymbol{z}^*$ *is a polarized equilibrium with respect to the same partition.*

Proof. The proof is reported in Appendix G. □

Remark 3. *In [40, Theorem 3] a sufficient condition for existence and convergence to a polarized equilibrium for networks of coordinating agents is reported, paralleling Theorem 7. The sufficient condition reported in [40, Theorem 3] can be expressed in the notation of this paper as*

$$\sum\nolimits_{j\in\mathcal{S}} w_{ij} > \max\Big\{\frac{1-2\lambda}{1-\lambda}, \frac{1}{2}\Big(1 + \frac{\beta(1-\lambda)}{1-\beta\lambda}\Big)\Big\}, \tag{20}$$

for all $i \in \mathcal{S}$, *for, separately,* $\mathcal{S} = \mathcal{V}_p$ *and* $\mathcal{S} = \mathcal{V}_n$. *The set* S *in* (20) *are all* $\frac{1}{2}$*-cohesive sets since* $\max\{\frac{1-2\lambda}{1-\lambda}, \frac{1}{2}(1 + \frac{\beta(1-\lambda)}{1-\beta\lambda})\} \geq \frac{1}{2}$.

We now present 2D contour plots to explain the inequalities in (19) and (20) more intuitively. For the anti-coordination game, Fig. 5 presents the two inequalities in (19). From the contour plot, it appears that $f(\lambda, \beta) = \frac{1}{2}(1 + \frac{\beta(1-\lambda)}{2\beta-\beta\lambda-1}) - \frac{1-2\lambda}{1-\lambda}$ is maximized in Fig. 5 for low to medium β values, and high λ values. Specifically when λ is less than about 0.3 and β is less than 0.5, no network can be found to satisfy the conditions (19). Put another way, given W and $\boldsymbol{z}(0)$, the inequalities in (19) are more likely to be satisfied with low to medium β, and high λ. This suggests that an initially polarized state could still converge to a polarized state if the anti-coordinate action dynamics strongly impact the coevolutionary process (and have a stronger impact in the anti-coordination games component). For the coordination game, Fig. 6 presents the two inequalities in (20). From the contour plot, it appears that

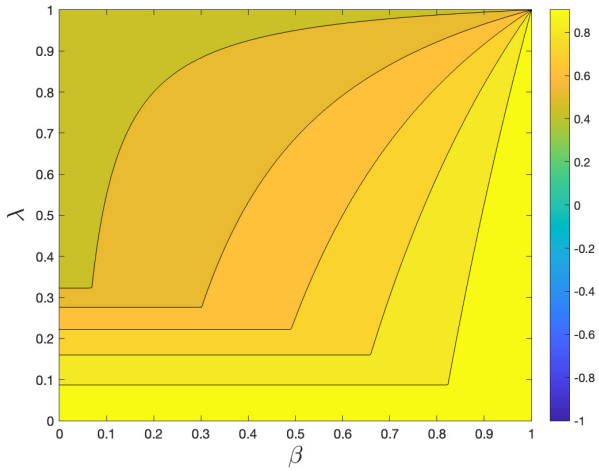

Figure 6: Contour plot of the RHS in (20) for agents in $\mathcal{V}_p$ or $\mathcal{V}_n$.

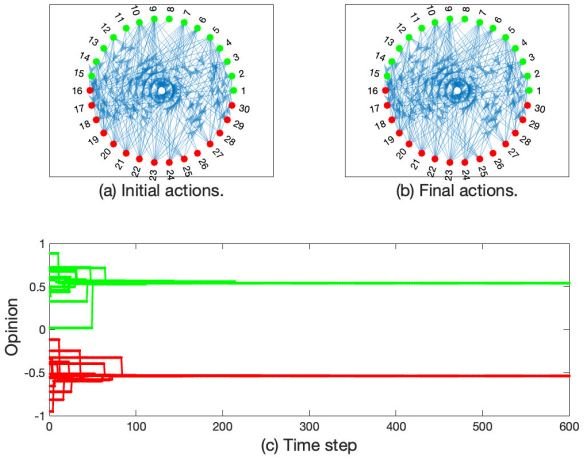

Figure 7: Simulation of Example 5, showing both the network view and the dynamics over time. The network is shown at time $t = 0$ (Panel (a)) and $t = 600$ (Panel (b)), with each node corresponding to an agent. Node i is colored green if agent i is such that $\text{sgn}(y_i) = x_i = +1$ (both opinion and action are positive), and in red if $\text{sgn}(y_i) = x_i = -1$ (both opinion and action are negative). In Panel (c), the temporal evolution of the opinions are shown, with each line corresponding to the opinion of an agent. As $t \to \infty$ the opinions converge to a polarized state.

$f(\lambda, \beta) = \max\{\frac{1-2\lambda}{1-\lambda}, \frac{1}{2}(1+\frac{\beta(1-\lambda)}{1-\beta\lambda})\} > 0.5$ regardless of the values of λ and β, which indicates that closer connections are required within the set S. We can thus see how different mechanisms (coordinating vs anti-coordinating) shape the existence and stability of polarized equilibria.

Example 5. *Consider a network of $n = 30$ agents with weights sampled at random and re-scaled so that, given $\boldsymbol{z}(0)$, the graph is a balanced complete bipartite graph with $\mathcal{V}_p = \mathcal{V}_1^+(0) = \{1, \dots, 15\}$ and $\mathcal{V}_n = \mathcal{V}_{-1}^-(0) = \{16, \dots, 30\}$. Specifically, There are no edges within sets $\mathcal{V}_1^+(0)$ or $\mathcal{V}_{-1}^-(0)$, but there are edges between two sets $\mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(0)$. It is easy to prove that the conditions in (20) are met. The network obtained is illustrated in Fig. 7(a). We set $\lambda = 0.7, \beta = 0.6$, which satisfy (19). As shown in Fig. 7(b) and 7(c), the final actions and opinions converge to a bipartite consensus, consistent with Theorem 7. In this case, the NE reached is not only polarized but also a bipartite consensus as depicted in Fig.7(b) and 7(c). Notice that the above scenario is quite extreme and it is presented for the sake of having a simple illustrative example, while the conditions in (19) can be satisfied also by less extreme network structures.*

4. Discussion

In this paper, we proposed a generalization of the co-evolutionary model of opinion and actions initially proposed in [40] that accounts for incorporating not only strategic complements, but also strategic substitutes [42]. Through such a generalization, the co-evolutionary model is now able to reproduce a broader range of congestion phenomena, that are typical of real-world complex systems, such as infrastructure networks and logistic systems, as extensively discussed in Section 2.3.2.

Our results highlight several different aspects of the complex intertwining between opinion formation processes and decision making in complex social systems. First, in the presence of strategic complements, whereby agents tends to coordinate with others, we showed that the network structure (especially its cohesiveness) critically determines whether a social system reaches a collective consensus, even starting from a polarized initial state. This observation is relevant for many real-world applications, whereby public authorities may be interested in reducing the polarization of a social system by acting on the structure of social contacts, e.g., by incentivizing discussion across different social bubbles (thereby increasing their diffusiveness, as in Definition 4). This is consistent with what observed in experimental studies [54].

Second, in the presence of strategic substitutes, agents' decision making follows an anti-coordination game. In this scenario, our analysis proved that the system always converges to an NE, and we derived explicit conditions under which consensus or polarization arises. These conclusions provide potential insights into the interplay between social influence and strategic decision-making, and offer theoretical foundations for understanding real-world patterns of consensus and polarization in social systems. In fact, also in this case, we derived conditions for the emergence of the two phenomena in terms of the structure of the social network. These conditions can be very useful to understand which social networks are more subject to the emergence of polarization and echo chambers, providing further model-based insight into existing studies on the topic, e.g., [55].

Despite the generality of our model and the novelty of the phenomena studied in this study, there are some limitations to this study. First, we have not fully characterized

all the Nash equilibria, and the total number of equilibria remains unknown — we conjecture that the network structure and agent parameters play a nontrivial role in this matter. While analytical expressions might be obtainable for highly specialized graph classes (e.g. line or star networks), we may need to employ numerical methods, such as evolutionary algorithms [56] or homotopy continuation [57], for general networks. Second, our model relies on the DeGroot update mechanism to handle opinion dynamics, which is a relatively simple (and linear) model that does not account for more complex cognitive and behavioral biases and tendencies that people exhibit during opinion formation processes. Generalizing to more sophisticated opinion dynamics models should be considered when one wishes to more closely describe real-world scenarios. For example, one could consider more specifically the possibility of replacing the DeGroot model with the Friedkin-Johnsen model [5], the Altafini model [10, 58], or other models [15, 59]. Those models which allow for a game-theoretic formulation may prove more feasible. Finally, although the coevolutionary model proposed in this paper is based on mature sociological and psychological theories and can qualitatively reproduce some emergent phenomena in social systems and real-world data (see [40]), it still lacks rigorous empirical testing, which is a key step towards practical application. Verification and validation, whether by experimental study or fitting to observed field data, is a long-term and important line of effort.

5. Conclusion

Building on the coordination game framework proposed by [40], we examined consensus equilibria and their domains of attraction. Then, we further extended the modeling framework and analysis to incorporate anti-coordinating agents, establishing a general convergence result, and identifying conditions for consensus and polarized equilibria to exist, as well as their regions of attraction. Finally, we compared our results on the configuration of equilibria and their basins of attraction in the presence of coordinating and anti-coordinating agents. In the Discussions section above, we have detailed several directions for future research, including complete equilibria characterization, improving the opinion dynamics model, and real-world validation. More broadly, we hope to use the modeling framework to describe and understand real-world social phenomena through the coevolutionary dynamics of actions and opinions with coordinating and anti-coordinating agents.

CRediT authorship contribution statement

Hong Liang: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Visualization, Writing — Original Draft

Mengbin Ye: Conceptualization, Methodology, Investigation, Supervision, Writing — Review & Editing

Lorenzo Zino: Conceptualization, Methodology, Investigation, Supervision, Writing — Review & Editing

Weiguo Xia: Conceptualization, Methodology, Investigation, Resources, Supervision, Project administration, Funding Acquisition, Writing — Review & Editing

Declaration of Competing Interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

No data was used for this study.

References

[1] J. R. P. French Jr, A formal theory of social power, Psychol. Rev. 63 (3) (1956) 181–194.

[2] R. Hegselmann, U. Krause, Opinion dynamics and bounded confidence models, analysis, and simulation, J. Artific. Soc. Soc. Simul. 5 (3) (2002) 1–30.

[3] S. Biswas, P. Sen, Model of binary opinion dynamics: Coarsening and effect of disorder, Phys. Rev. E 80 (2) (2009) 4–7.

[4] F. Ding, Y. Liu, B. Shen, X.-M. Si, An evolutionary game theory model of binary opinion formation, Physica A 389 (8) (2010) 1745–1752.

[5] N. E. Friedkin, E. C. Johnsen, Social Influence Network Theory: A Sociological Examination of Small Group Dynamics, Cambridge University Press, 2011.

[6] A. V. Proskurnikov, R. Tempo, A tutorial on modeling and analysis of dynamic social networks. Part I, Annu. Rev. Control 43 (2017) 65–79.

[7] C. Ravazzi, F. Dabbene, C. Lagoa, A. V. Proskurnikov, Learning hidden influences in large-scale dynamical social networks: A data-driven sparsity-based approach, in memory of Roberto Tempo, IEEE Control Syst. 41 (5) (2021) 61–103.

[8] N. Loy, M. Raviola, A. Tosin, Opinion polarization in social networks, Philos. Trans. R. Soc. A 380 (2224) (2022).

[9] N. E. Friedkin, E. C. Johnsen, Social influence and opinions, J. Math. Sociol. 15 (3-4) (1990) 193–206.

[10] C. Altafini, Consensus Problems on Networks with Antagonistic Interactions, IEEE Trans. Automat. Contr. 58 (4) (2013) 935–946.

[11] P. Bolzern, P. Colaneri, G. De Nicolao, Opinion influence and evolution in social networks: A markovian agents model, Automatica 100 (2019) 219–230.

[12] D. Acemoglu, G. Como, F. Fagnani, A. Ozdaglar, Opinion fluctuations and persistent disagreement in social networks, in: 50th IEEE Conf. Decis. Control, 2011, pp. 2347–2352.

[13] P. Cisneros-Velarde, K. Chan, F. Bullo, Polarization and fluctuations in signed social networks, IEEE Trans. Automat. Contr. 66 (8) (2020) 3789–3793.

[14] G. De Pasquale, M. E. Valcher, Consensus for clusters of agents with cooperative and antagonistic relationships, Automatica 135 (2022) 110002. doi:10.1016/j.automatica.2021.110002.

[15] A. Bizyaeva, A. Franci, N. E. Leonard, Nonlinear opinion dynamics with tunable sensitivity, IEEE Trans. Autom. Control. 68 (3) (2023) 1415–1430.

[16] D. Godes, D. Mayzlin, Firm-created word-of-mouth communication: Evidence from a field test, Mark. Sci. 28 (4) (2009) 721–739.

[17] J. Candia, Advertising and irreversible opinion spreading in complex social networks, Int. J. Mod. Phys. C 20 (06) (2009) 799–815.

[18] J. Castro, J. Lu, G. Zhang, Y. Dong, L. Martínez, Opinion dynamics-based group recommender systems, IEEE Trans. Syst. Man Cybern.: Syst. 48 (12) (2017) 2394–2406.
[19] J. Tang, X. Hu, H. Liu, Social recommendation: a review, Soc. Netw. Anal. Min. 3 (2013) 1113–1133.
[20] B. She, J. Liu, S. Sundaram, P. E. Paré, On a networked sis epidemic model with cooperative and antagonistic opinion dynamics, IEEE Trans. Control Netw. Syst. 9 (3) (2022) 1154–1165.
[21] Q. Xu, H. Ishii, On a discrete-time networked siv epidemic model with polar opinion dynamics, IEEE Trans. Netw. Sci. Eng. 11 (6) (2024) 6636–6651.
[22] M. O. Jackson, Y. Zenou, Games on Networks, in: Handbook of Game Theory with Economic Applications, Vol. 4, Elsevier, 2015, Ch. 3, pp. 95–163.
[23] H. P. Young, The dynamics of social innovation, Proc. Natl. Acad. Sci. USA 108 (Supplement 4) (2011) 21285–21291.
[24] A. Montanari, A. Saberi, The spread of innovations in social networks, Proc. Natl. Acad. Sci. USA 107 (47) (2010) 20196–20201.
[25] J. Riehl, P. Ramazi, M. Cao, A survey on the analysis and control of evolutionary matrix games, Annu. Rev. Control 45 (2018) 87–106.
[26] A. Flache, M. Mäs, T. Feliciani, E. Chattoe-Brown, G. Deffuant, S. Huet, J. Lorenz, Models of Social Influence: Towards the Next Frontiers, J. Artif. Soc. S. 20 (4) (2017).
[27] M. Ye, et al., Collective patterns of social diffusion are shaped by individual inertia and trend-seeking, Nat. Comm. 12 (2021) 5698.
[28] B. Mroué, A. Couthures, S. Lasaulce, I.-C. Morărescu, On the impact of cross-country imitation on climate change: A game-theoretical analysis, IFAC J. Syst. Control 32 (2025) 100309.
[29] G. T. Henry, M. M. Mark, Beyond use: Understanding evaluation's influence on attitudes and actions, Am. J. Eval. 24 (3) (2003) 293–314.
[30] J. Haidt, The emotional dog and its rational tail: a social intuitionist approach to moral judgment., Psychol. Rev. 108 (4) (2001) 814.
[31] S. Gavrilets, P. J. Richerson, Collective action and the evolution of social norm internalization, Proc. Natl. Acad. Sci. USA 114 (23) (2017) 6068–6073.
[32] A. C. Martins, C. d. B. Pereira, R. Vicente, An opinion dynamics model for the diffusion of innovations, Physica A 388 (15-16) (2009) 3225–3232.
[33] A. C. Martins, Continuous opinions and discrete actions in opinion dynamics problems, Int. J. Mod. Phys. C 19 (04) (2008) 617–624.
[34] N. R. Chowdhury, I.-C. Morărescu, S. Martin, S. Srikant, Continuous opinions and discrete actions in social networks: A multi-agent system approach, in: IEEE 55th Conf. Decis. Control, 2016, pp. 1739–1744.
[35] F. Ceragioli, P. Frasca, Consensus and disagreement: The role of quantized behaviors in opinion dynamics, SIAM J. Control Optim. 56 (2) (2018) 1058–1080.
[36] V. S. Varma, I.-C. Morărescu, S. Lasaulce, S. Martin, Opinion dynamics aware marketing strategies in duopolies, in: IEEE 56th Conf. Decis. Control, 2017, pp. 3859–3864.
[37] D. Centola, R. Willer, M. Macy, The emperor's dilemma: A computational model of self-enforcing norms, Am. J. Sociol. 110 (4) (2005) 1009–1040.
[38] L. Zino, M. Ye, M. Cao, A two-layer model for coevolving opinion dynamics and collective decision-making in complex social systems, Chaos 30 (8) (2020) 083107.
[39] L. Zino, M. Ye, M. Cao, A coevolutionary model for actions and opinions in social networks, in: 59th IEEE Conf. Decis. Control., 2020, pp. 1110–1115.
[40] H. D. Aghbolagh, M. Ye, L. Zino, Z. Chen, M. Cao, Coevolutionary dynamics of actions and opinions in social networks, IEEE Trans. Autom. Control. 68 (12) (2023) 7708–7723.
[41] D. Monderer, L. S. Shapley, Potential games, Games Econ. Behav. 14 (1) (1996) 124–143.
[42] Y. Bramoullé, Anti-coordination and social interactions, Games Econ. Behav. 58 (1) (2007) 30–49.
[43] D. Fudenberg, J. Tirole, Game theory, MIT Press, 1991.
[44] F. Alvarez, K. Ritzberger, Network formation and anti-coordination games, Int. J. Game Theory 32 (2004) 501–517.
[45] P. Ramazi, J. Riehl, M. Cao, Networks of conforming or nonconforming individuals tend to reach satisfactory decisions, Proc. Natl. Acad. Sci. USA 113 (46) (2016) 12985–12990.
[46] M. Vanelli, L. Arditti, G. Como, F. Fagnani, On games with coordinating and anti-coordinating agents, in: 21st IFAC World Cong., Vol. 53, Elsevier BV, 2020, pp. 10975–10980.
[47] M. H. DeGroot, Reaching a consensus, J. Am. Stat. Assoc. 69 (345) (1974) 118–121.
[48] J. R. Marden, G. Arslan, J. S. Shamma, Cooperative control and potential games, IEEE Trans. Syst. Man Cybern. B 39 (6) (2009) 1393–1407.
[49] N. E. Friedkin, E. C. Johnsen, Social influence and opinions, J. Math. Sociol. 15 (3-4) (1990) 193–206.
[50] K. Takács, A. Flache, M. Mäs, Discrepancy and disliking do not induce negative opinion shifts, PLOS ONE 11 (6) (2016) e0157948.
[51] M. Mäs, H. H. Nax, A behavioral study of noise in coordination games, J. Econ. Theory 162 (2016) 195–208.
[52] F. Xiao, L. Wang, State consensus for multi-agent systems with switching topologies and time-varying delays, Int. J. Control 79 (10) (2006) 1277–1284.
[53] R. Raineri, G. Como, F. Fagnani, M. Ye, L. Zino, On controlling a coevolutionary model of actions and opinions, in: IEEE 63rd Conf. Decis. Control, 2024, pp. 4550–4555.
[54] A. Combs, et al., Reducing political polarization in the United States with a mobile chat platform, Nat. Human Behav. 7 (9) (2023) 1454–1461.
[55] C. M. Valensise, M. Cinelli, W. Quattrociocchi, The drivers of online polarization: Fitting models to data, Inf. Sci. 642 (2023) 119152.
[56] S. G. Ficici, J. B. Pollack, A game-theoretic approach to the simple coevolutionary algorithm, in: International Conference on Parallel Problem Solving from Nature, Springer, 2000, pp. 467–476.
[57] P. J.-J. Herings, R. Peeters, Homotopy methods to compute equilibria in game theory, Economic Theory 42 (1) (2010) 119–156.
[58] W. Xia, M. Cao, K. H. Johansson, Structural balance and opinion separation in trust–mistrust social networks, IEEE transactions on control of network systems 3 (1) (2015) 46–56.
[59] W. Xia, M. Ye, J. Liu, M. Cao, X.-M. Sun, Analysis of a nonlinear opinion dynamics model with biased assimilation, Automatica 120 (2020) 109–113.

Appendix A. Proof of Theorem 1

According to Definition 1 and Lemma 1, we need to show $\forall i \in \mathcal{V}$ that $z_i = [1,1]$ is the best response when $\boldsymbol{z_{-i}} = [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]$. That is, $[1,1] \in \mathcal{B}_i(f_i(\cdot, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]))$. By Assumption 1, we write

$$\begin{aligned}
&\mathcal{B}_i(f_i(\cdot, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}])) = \arg\max_{z_i} f_i(z_i, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]),\\
&= \arg\max_{z_i} \frac{\lambda(1-\beta)}{2} \sum\nolimits_{j\in\mathcal{V}} a_{ij}(1+\alpha)(1+x_i)\\
&- \frac{1}{2}(1-\lambda)\beta \sum\nolimits_{j\in\mathcal{V}} w_{ij}(y_i-1)^2 - \frac{1}{2}\lambda\beta(y_i-x_i)^2.
\end{aligned}$$

Since x_i can only take values -1 or 1, the above equation is a discrete optimization problem. We can determine the optimal choice of x_i by substituting these two values into $f_i(z_i, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}])$. For $y_i \in [-1,1]$, we need to find the value that maximizes f_i in the interval $[-1,1]$.

Case $x_i = +1$**.** We compute $f_i(z_i, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]) = \lambda(1-\beta)(1+\alpha)\sum_{j\in\mathcal{V}} a_{ij} - \frac{1}{2}(1-\lambda)\beta\sum_{j\in\mathcal{V}} w_{ij}(y_i-1)^2 - \frac{1}{2}\lambda\beta(y_i-1)^2$. Taking the partial derivative of f_i with respect to y_i and setting it to zero yields

$$\frac{\partial f_i}{\partial y_i} = -(1-\lambda)\beta\sum\nolimits_{j\in\mathcal{V}} w_{ij}(y_i-1) - \lambda\beta(y_i-1) = 0,$$

which is maximized by $y_i = 1$, which yields the claim.

Case $x_i = -1$**.** We compute $f_i(z_i, [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]) = -\lambda(1-\beta)\sum_{j\in\mathcal{V}} a_{ij} - \frac{1}{2}(1-\lambda)\beta\sum_{j\in\mathcal{V}} w_{ij}(y_i+1)^2 - \frac{1}{2}\lambda\beta(y_i+1)^2$. The partial derivative of f_i with respect to y_i, set to zero, is

$$\frac{\partial f_i}{\partial y_i} = -(1-\lambda)\beta\sum\nolimits_{j\in\mathcal{V}} w_{ij}(y_i+1) - \lambda\beta(y_i+1) = 0,$$

which is maximized by $y_i = -1$.

In summary, when $\boldsymbol{z_{-i}} = [\mathbf{1}_{n-1}, \mathbf{1}_{n-1}]$, the best response is $[1, 1]$ or $[-1, -1]$, depending on x_i, verifying that $[\mathbf{1}_n, \mathbf{1}_n]$ and $[-\mathbf{1}_n, -\mathbf{1}_n]$ are NE, using Lemma 1. □

Appendix B. Proof of Theorem 2

First, our assumption on $\boldsymbol{z}(0)$ implies that $\mathcal{V}_1^+(0) \cup \mathcal{V}_{-1}^+(0) = \mathcal{V}$. Exploiting this, the key argument is to show that the following three statements hold true for all $t \geq 0$: i) $\delta_i(\boldsymbol{z}(t)) \geq 0$, for all $i \in \mathcal{V}_1^+(0)$, and $\delta_i(\boldsymbol{z}(t)) > 0$, for all $i \in \mathcal{V}_{-1}^+(0)$; ii) $\boldsymbol{x}(t+1) \geq \boldsymbol{x}(t)$; and iii) $\boldsymbol{y}(t+1) \geq \mathbf{0}_n$. Here, the inequalities are taken entry-wise. We prove the three statements jointly by (strong) induction. We start by proving that the statements hold true for $t = 0$. To prove i), we compute δ_i by splitting each sum into two parts, obtaining

$$\begin{aligned}\delta_i(\boldsymbol{z}(0)) = & p\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij}y_j(0) + p\sum\nolimits_{k\in\mathcal{V}_{-1}^+(0)} w_{ik}y_k(0) \\ & + \frac{q}{4}\sum\nolimits_{j\in\mathcal{V}_1^+(0)} a_{ij}\left[2x_j(0) + \alpha(1+x_j(0))\right] \\ & + \frac{q}{4}\sum\nolimits_{k\in\mathcal{V}_{-1}^+(0)} a_{ik}\left[2x_k(0) + \alpha(1+x_k(0))\right],\end{aligned}$$

where $p = 2\lambda\beta(1-\lambda)$ and $q = 2\lambda(1-\beta)$ are positive constants. For $i \in \mathcal{V}_1^+(0)$, we use the cohesiveness of $\mathcal{V}_1^+(0)$ and the fact that $\boldsymbol{y}(0) \geq 0_n$, to bound

$$\begin{aligned}\delta_i(\boldsymbol{z}(0)) &\geq \frac{q}{2}\sum\nolimits_{j\in\mathcal{V}_1^+(0)} a_{ij}(1+\alpha) - \frac{q}{2}\sum\nolimits_{k\in\mathcal{V}_{-1}^+(0)} a_{ik} \\ &\geq \frac{q}{2}\left[\frac{1}{\alpha+2}(1+\alpha) - \frac{\alpha+1}{\alpha+2}\right] \geq 0.\end{aligned} \tag{B.1}$$

Similarly, for $i \in \mathcal{V}_{-1}^+(0)$, we use the diffusiveness of $\mathcal{V}_{-1}^+(0)$ and the fact that $\boldsymbol{y}(0) \geq 0$, to bound

$$\begin{aligned}\delta_i(\boldsymbol{z}(0)) &\geq \frac{q}{2}\sum\nolimits_{j\in\mathcal{V}_1^+(0)} a_{ij}(1+\alpha) - \frac{q}{2}\sum\nolimits_{k\in\mathcal{V}_{-1}^+(0)} a_{ik} \\ &> \frac{q}{2}\left[\frac{1}{\alpha+2}(1+\alpha) - \frac{\alpha+1}{\alpha+2}\right] \geq 0.\end{aligned} \tag{B.2}$$

The strict inequality is obtained due to the $\frac{\alpha+1}{\alpha+2}$-diffusiveness of $\mathcal{V}_{-1}^+(0)$, which implies that $\sum_{k\in\mathcal{V}_{-1}^+(0)} a_{ik} < \frac{\alpha+1}{\alpha+2}$. To prove ii) and iii), let i_0 be the agent active at time $t = 0$. For all $i \neq i_0$, no update is performed, hence, it clearly holds true that $x_i(1) = x_i(0)$ and $y_i(1) = y_i(0) \geq 0$. Then, we distinguish two cases. If $i_0 \in \mathcal{V}_1^+(0)$, and because $\delta_i(\boldsymbol{z}(0)) \geq 0$ as established above, it follows that $x_i(1) = 1$ and $y_i(1) \geq \lambda \geq 0$, according to Proposition 1. If $i_0 \in \mathcal{V}_{-1}^+(0)$, and because $\delta_i(\boldsymbol{z}(0)) > 0$ as established above, it follows that $x_i(1) = 1$ and $y_i(1) \geq \lambda \geq 0$, according to Proposition 1, yielding the claim. This concludes the proof of the induction basis. We are now left to prove that, assuming i)–iii) hold up to time t, they necessarily hold at $t+1$. The proof of i) follows the same arguments used to prove the induction basis. In fact, due to strong induction assumption, we have $x_i(t) \geq x_i(0)$, for all $i \in \mathcal{V}$, implying

$$\begin{aligned}\delta_i(\boldsymbol{z}(t)) \geq & \frac{q}{4}\sum\nolimits_{j\in\mathcal{V}_1^+(t)} a_{ij}\left[2x_j(t) + \alpha(1+x_j(t))\right] \\ & + \frac{q}{4}\sum\nolimits_{k\in\mathcal{V}_{-1}^+(t)} a_{ik}\left[2x_k(t) + \alpha(1+x_k(t))\right] \\ \geq & \frac{q}{4}\sum\nolimits_{j\in\mathcal{V}_1^+(0)} a_{ij}\left[2x_j(0) + \alpha(1+x_j(0))\right] \\ & + \frac{q}{4}\sum\nolimits_{k\in\mathcal{V}_{-1}^+(0)} a_{ik}\left[2x_k(0) + \alpha(1+x_k(0))\right],\end{aligned} \tag{B.3}$$

which is nonnegative (or strictly positive) as per (B.1) and (B.2), respectively. Similar, ii) and iii) are obtained as consequences of the positivity of δ_i from (B.3).

Finally, if at a generic time t, agent $i_t \in \mathcal{V}_{-1}^+(0)$ activates, then $\delta_{i_t}(\boldsymbol{z}(t)) > 0$, implying that $x_{i_t}(t+1) = 1$. Due to Assumption 2 and the monotonicity property proved above in item ii), it necessarily holds true that $\boldsymbol{x}(t) = \mathbf{1}$, for all $t \geq T$. Finally, the fact that $[\mathbf{1}_n, \mathbf{1}_n]$ is the only NE with $\boldsymbol{x} = \mathbf{1}_n$ guarantees convergence [53, Theorem 1]. □

Appendix C. Proof of Theorem 3

For the sake of convenience in our proof, we define positive constants $p = 2\lambda\beta(1-\lambda)$ and $q = 2\lambda(1-\beta)$. The proof is based on the following set of arguments. If the partitions $\mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(0)$ satisfy

$$\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij} \geq 1 - \frac{q}{2(p+q)}, \qquad i \in \mathcal{V}_1^+(0), \tag{C.1a}$$

$$\sum\nolimits_{j\in\mathcal{V}_{-1}^-(0)} w_{ij} < \frac{q}{2(p+q)}, \qquad i \in \mathcal{V}_{-1}^-(0), \tag{C.1b}$$

then i) $\delta_i(\boldsymbol{z}(t)) \geq 0$, for all $i \in \mathcal{V}_1^+(0)$, and $\delta_i(\boldsymbol{z}(t)) > 0$, for all $i \in \mathcal{V}_{-1}^-(0)$; and ii) the action vector $\boldsymbol{x}(t+1) \geq \boldsymbol{x}(t)$. Then, we show that if there also holds

$$\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij} \geq \frac{1-2\lambda}{1-\lambda} \qquad i \in \mathcal{V}_1^+(0), \tag{C.2a}$$

$$\sum\nolimits_{j\in\mathcal{V}_{-1}^-(0)} w_{ij} < \frac{\lambda}{1-\lambda}, \qquad i \in \mathcal{V}_{-1}^-(0), \tag{C.2b}$$

we have iii) the opinion vector $\boldsymbol{y}(t+1)) \geq \mathbf{0}_n$. We start by proving that the statements hold for $t = 0$. To prove

i), we compute δ_i by splitting each sum into two parts, obtaining

$$\delta_i(\boldsymbol{z}(0))=p\sum\nolimits_{j\in\mathcal{V}_1^+(0)}w_{ij}y_j(0)+p\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik}y_k(0) + \frac{q}{4}\left[\sum\nolimits_{j\in\mathcal{V}_1^+(0)}w_{ij}2x_j(0)+\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik}2x_k(0)\right].$$

For $i \in \mathcal{V}_1^+(0)$, we use (C.1a), the fact that $\sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik} = 1 - \sum_{j\in\mathcal{V}_1^+(0)} w_{ij}$ because W is stochastic, and $y_k(0) \in [-1, 0)$, $\forall k \in \mathcal{V}_{-1}^-(0)$, $y_j(0) \in (0, +1]$, $\forall j \in \mathcal{V}_1^+(0)$ to bound

$$\begin{aligned}\delta_i(\boldsymbol{z}(0)) &\geq -p(1-\sum\nolimits_{j\in\mathcal{V}_1^+(0)}w_{ij})+q\Big[\sum\nolimits_{j\in\mathcal{V}_1^+(0)}w_{ij}-\frac{1}{2}\Big]\\ &\geq -(p+\frac{q}{2})+(p+q)\sum\nolimits_{j\in\mathcal{V}_1^+(0)}w_{ij}\\ &\geq -(p+\frac{q}{2})+(p+q)\big(1-\frac{q}{2(p+q)}\big)\geq 0.\end{aligned}$$

Similarly, for $i \in \mathcal{V}_{-1}^-(0)$, we use (C.1b), $\sum_{j\in\mathcal{V}_1^+(0)} w_{ij} = 1 - \sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}$, and $y_k(0) \in [-1, 0), \forall k \in \mathcal{V}_{-1}^-(0), y_j(0) \in (0, +1], \forall j \in \mathcal{V}_1^+(0)$ to bound

$$\begin{aligned}\delta_i(\boldsymbol{z}(0)) &\geq -p\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik}+q\Big[\frac{1}{2}-\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik}\Big]\\ &> -(p+q)\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik}+\frac{q}{2}\\ &> -(p+q)\frac{q}{2(p+q)}+\frac{q}{2}\geq 0.\end{aligned}$$

To prove ii) and iii), let i_0 be the agent active at time $t = 0$. For all $i \neq i_0$, it holds $x_i(1) = x_i(0)$ and $y_i(1) = y_i(0) \geq 0$. Then, we distinguish two cases. If $i_0 \in \mathcal{V}_1^+(0)$, being $\delta_i(\boldsymbol{z}(0)) \geq 0$, then $x_i(1) = 1$ and $y_i(1) \geq (1-\lambda)(-\sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}) + \lambda \geq 0$ where we used (C.2a) and $\sum_{j\in\mathcal{V}_1^+(0)} w_{ij} + \sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik} = 1$. If $i_0 \in \mathcal{V}_{-1}^-(0)$, being $\delta_i(\boldsymbol{z}(0)) > 0$, then $x_i(1) = 1$ and

$$y_i(1) \geq (1-\lambda)(-\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)}w_{ik})+\lambda > 0,$$

according to (C.2b), concluding the proof of the induction basis.

Finally, by using a strong induction argument, similar to the proof of Theorem 2, it can also be established that, assuming that the three properties hold up to time t, they necessarily hold true at time $t + 1$, yielding convergence. □

Appendix D. Proof of Theorem 4

The proof primarily uses game-theoretic methods. We first introduce the following function:

$$\begin{aligned}\Phi(z) = &-\sum\nolimits_{i\in\mathcal{V}}\sum\nolimits_{j\in\mathcal{V}\setminus\{i\}}\eta_i\frac{a_{ij}}{2}\Big[(1-x_j)(1-x_i)\\ &-(1+\alpha)(1+x_j)(1+x_i)\Big]-\frac{1}{2}\sum_{i\in\mathcal{V}}\sum_{j\in\mathcal{V}}\frac{w_{ij}}{2}(y_i-y_j)^2\\ &-\frac{1}{2}\sum\nolimits_{k\in\mathcal{V}}\frac{\lambda_i}{(1-\lambda_i)}(x_k-y_k)^2,\end{aligned}\tag{D.1}$$

with $\eta_i = \lambda_i(1-\beta_i)/4\beta_i(1-\lambda_i)$. Following an approach similar to [40, Lemma 1] (and hence details are omitted), one can prove that, under Assumption 1 and when $\varepsilon = -1$, the coevolutionary game is a generalized ordinal potential game [41] with the potential function in (D.1). Following the approach of [40, Theorem 1], we can prove that the generalized ordinal potential game necessarily converges to an equilibrium, yielding the claim. The details are omitted. □

Appendix E. Proof of Theorem 5

Our objective is to identify conditions for when the state $\boldsymbol{z}^* = (\boldsymbol{x}^*, \boldsymbol{y}^*) = [\mathbf{1}_n, \mathbf{1}_n]$ (or $-[\mathbf{1}_n, \mathbf{1}_n]$) is an equilibrium of the coevolutionary model on the two-layer network $\mathcal{G}$. If $[\mathbf{1}_n, \mathbf{1}_n]$ is an equilibrium, then from (10), there must hold

$$1 = x_i^* = \mathcal{S}(\delta_i(z^*)),\tag{E.1a}$$

$$1 = y_i^* = (1-\lambda)\sum\nolimits_{j\in\mathcal{V}}w_{ij}y_j^*+\lambda\mathcal{S}(\delta_i(z^*)).\tag{E.1b}$$

(E.1b) can be further simplified to $1 = (1-\lambda)\sum_{j\in\mathcal{V}} w_{ij}+\lambda$, which always holds since W is row-stochastic. Meanwhile, (9) evaluated at $[\mathbf{1}_n, \mathbf{1}_n]$ yields

$$\delta_i(\boldsymbol{z}^*) = -\frac{\lambda(1-\beta)}{2}\sum_{j\in\mathcal{V}}a_{ij}(2+2\alpha)+2(1-\lambda)\lambda\beta\sum\nolimits_{j\in\mathcal{V}}w_{ij}.$$

It follows that $\delta_i(\boldsymbol{z}^*) > 0$ (which ensures $x_i^* = \mathcal{S}(\delta_i(\boldsymbol{z}^*)) = 1$) is implied by $\sum_{j\in\mathcal{V}} a_{ij} < \frac{2(1-\lambda)\beta}{(1-\beta)(1+\alpha)}$. Thus, $[\mathbf{1}_n, \mathbf{1}_n]$ satisfies the equilibrium equations if (16) holds for all $i \in \mathcal{V}$. □

Appendix F. Proof of Theorem 6

The proof is divided into two parts. In the first part, we establish that given a fixed action vector $\boldsymbol{x}^*$ that is polarized concerning a partitioning $(\mathcal{V}_p, \mathcal{V}_n)$, and under the conditions in (18a), the unique vector $\boldsymbol{y}^*$ that satisfies the equilibrium conditions for (10b) is polarized. The proof that $\boldsymbol{y}^*$ is polarizing with respect to $(\mathcal{V}_p, \mathcal{V}_n)$ follows closely the proof of [40, Theorem 2], and it is thus omitted.

In the second part, with $\boldsymbol{y}^*$ being the polarized vector identified in part one, we show that conditions in (18b) and (18c), guarantee that the action vector $\boldsymbol{x}^*$ is invariant under (10a), which implies that $\boldsymbol{z}^*$ is a polarized equilibrium of the coevolutionary model. Given $\boldsymbol{y}^*$ which is polarized with respect to $(\mathcal{V}_p, \mathcal{V}_n)$, define $y_m^+ = \min_{i\in\mathcal{V}_p} y_i^*$ and $y_m^- = \min_{i\in\mathcal{V}_n}$ and $y_M^+ = \max_{i\in\mathcal{V}_p} y_i^*$ and $y_M^- = \max_{i\in\mathcal{V}_n} y_i^*$.

For any $i \in \mathcal{V}_p$ and $t \geq 0$, $x_i(t) = +1 = x_i(t+1) = +1$ if $\delta_i(\boldsymbol{z}(t)) \geq 0$. From (14), we obtain the following bound

$$\begin{aligned}\delta_i(\boldsymbol{z}) &= p\Big[\sum\nolimits_{j\in\mathcal{V}_p}w_{ij}y_j+\sum\nolimits_{k\in\mathcal{V}_n}w_{ik}y_k\Big]\\ &\quad-\frac{q}{2}\Big[\sum\nolimits_{j\in\mathcal{V}_p}w_{ij}x_j+\sum\nolimits_{k\in\mathcal{V}_n}w_{ik}x_k\Big]\\ &\geq p\Big[\underline{d}_p y_m^+ + (1-\underline{d}_p)y_m^-\Big]-\frac{q}{2}\Big[\bar{d}_p-(1-\bar{d}_p)\Big],\end{aligned}$$

where $p = 2\lambda\beta(1-\lambda), q = 2\lambda(1-\beta)$. We shall now prove that

$$2\lambda\beta(1-\lambda)(\bar{d}_p y_m^+ + (1-\bar{d}_p)y_m^-) + (1-\beta)\lambda(1-2\bar{d}_p) > 0, \tag{F.1}$$

which implies that (Appendix F) holds. Before proceeding, we state two important inequalities, whose derivations are found in the proof of [40, Theorem 2], and is obtained by leveraging the equilibrium expression for y_i^*, i.e., setting $y_i(t+1) = y_i(t)$ in (10). In particular, one can show that $y_m^+ \geq \lambda + (1-\lambda)[\underline{d}_p y_m^+ + (1-\underline{d}_p)y_m^-]$ and also that

$$y_m^+ \geq \frac{\lambda - \lambda(1-\lambda)(1-\underline{d}_p+\bar{d}_n)}{1-(1-\lambda)(\underline{d}_p+\bar{d}_n)-(1-\lambda)^2(1-\underline{d}_p-\bar{d}_n)}. \tag{F.2}$$

The first inequality can be used to show that $2\beta(y_m^+ - \lambda) - (1-\beta)(2\bar{d}_p - 1) > 0$ implies (F.1). Substituting y_m^+ from (F.2) into this latest inequality, and simplifying, yields (18b). Thus, (18b) implies (Appendix F), and thus $\delta_i(\boldsymbol{z}) > 0$. In other words, $x_i(t+1) = x_i(t) = +1$ for all $t \geq 0$ if (18b) holds.

For any $i \in \mathcal{V}_n$ and $t \geq 0$, $x_i(t+1) = x_i(t) = -1$ iff $\delta_i(x(t), y(t)) < 0$, which is implied by

$$\begin{aligned}
&\delta_i(\boldsymbol{z}) = p\sum\nolimits_{j\in\mathcal{V}_p} w_{ij}y_j + p\sum\nolimits_{k\in\mathcal{V}_n} w_{ik}y_k \\
&-\frac{q}{2}\Big[\sum\nolimits_{j\in\mathcal{V}_p} w_{ij}x_j + \sum\nolimits_{k\in\mathcal{V}_n} w_{ik}x_k\Big] \\
&\leq p\Big[\bar{d}_n y_M^- + (1-\bar{d}_n)y_M^+\Big] - \frac{q}{2}\Big[(1-\bar{d}_n) - \bar{d}_n\Big],
\end{aligned} \tag{F.3}$$

where $p = 2\lambda\beta(1-\lambda), q = 2\lambda(1-\beta)$. We omit the detailed arguments, which follow similarly to the above, one can show that (18c) implies the right hand side of (F.3) is strictly negative, and thus $\delta_i(\boldsymbol{z}) < 0$. □

Appendix G. Proof of Theorem 7

First, our assumption on polarized $\boldsymbol{z}(0)$ implies that $\mathcal{V}_1^+(0) \cup \mathcal{V}_{-1}^-(0) = \mathcal{V}$. For the sake of convenience in our proof, we define positive constants $p = 2\lambda\beta(1-\lambda)$ and $q = 2\lambda(1-\beta)$. The proof is by induction. To begin, we show that if the partitions $\mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(0)$ satisfy (19), then i) $\boldsymbol{x}(1)$ is polarized with respect to the same partition and ii) $\boldsymbol{y}(1)$ is polarized with respect to the same partition, i.e., $\mathcal{V}_1^+(1) = \mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(1) = \mathcal{V}_{-1}^-(0)$ Then, Theorems 4 and 6 are leveraged to guarantee convergence to equilibrium , and that $\mathcal{V}_1^+(t) = \mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(t) = \mathcal{V}_{-1}^-(0)$ for all $t \geq 0$ (implying the equilibrium is polarized), respectively.

We start from the induction basis. To prove i), we compute $\delta_i(\boldsymbol{z}(0))$ by splitting each sum into two parts, obtaining

$$\begin{aligned}
\delta_i(\boldsymbol{z}(0)) = &p\Big[\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij}y_j(0) + \sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}y_k(0)\Big] \\
&-\frac{q}{2}\Big[\sum\nolimits_{j\in\mathcal{V}_1^+(0)} w_{ij}x_j(0) + \sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}x_k(0)\Big].
\end{aligned}$$

For $i \in \mathcal{V}_1^+(0)$, we use the following facts: a) the inequalities in (19), b) $\beta < 1/(2-\lambda)$, c) $y_k(0) \in [-1, 0), \forall k \in \mathcal{V}_{-1}^-(0), y_j(0) \in (0, +1], \forall j \in \mathcal{V}_1^+(0)$, and d) $\sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik} = 1 - \sum_{j\in\mathcal{V}_1^+(0)} w_{ij}$ to bound

$$\begin{aligned}
\delta_i(\boldsymbol{z}(0)) &= p\left[\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij}y_j + \sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}y_k\right] \\
&\quad - q\Big(\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij} - \frac{1}{2}\Big) \\
&\geq -p\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik} - q\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij} + \frac{q}{2} \\
&\geq (p-q)\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij} - p + \frac{q}{2} \\
&\geq (p-q)\frac{1}{2}\big(1 + \frac{p}{p-q}\big) + \big(\frac{q}{2} + p\big) > 0.
\end{aligned} \tag{G.1}$$

For $i \in \mathcal{V}_{-1}^-(0)$, we use the inequalities in (19), the condition $\beta < 1/(2-\lambda)$, the fact that $y_k(0) \in [-1, 0), \forall k \in \mathcal{V}_{-1}^-(0), y_j(0) \in (0, +1], \forall j \in \mathcal{V}_1^+(0)$, and $\sum_{j\in\mathcal{V}_1^+(0)} w_{ij} = 1 - \sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}$ to bound

$$\begin{aligned}
\delta_i(\boldsymbol{z}(0)) &= p\left[\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij}y_j + \sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}y_k\right] \\
&\quad - q\Big(\frac{1}{2} - \sum\nolimits_{j\in\mathcal{V}_{-1}^-(0)} w_{ij}\Big) \\
&\leq p\sum\nolimits_{j\in\mathcal{V}_{+1}^+(0)} w_{ij} - q\Big(\frac{1}{2} - \sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}\Big) \\
&\leq -(p-q)\sum\nolimits_{k\in\mathcal{V}_{-1}^-(0)} w_{ik} + \big(p - \frac{q}{2}\big) \\
&\leq -(p-q)\frac{1}{2}\big(1 + \frac{p}{p-q}\big) + \big(p - \frac{q}{2}\big) < 0.
\end{aligned} \tag{G.2}$$

Let i_0 be the agent active at time $t = 0$. For all $i \neq i_0$, no update is performed, so those agents in $\mathcal{V}_{-1}^-(0)$ remain in $\mathcal{V}_{-1}^-(1)$ (and agents in $\mathcal{V}_{+1}^+(0)$ stay in $\mathcal{V}_{+1}^+(1)$). Concerning i_0, we distinguish two cases. If $i_0 \in \mathcal{V}_1^+(0)$, being $\delta_i(\boldsymbol{z}(0)) \geq 0$, then $x_i(1) = x_i(0) = 1$. From (10b), we obtain $y_i(1) \geq (1-\lambda)(-\sum_{k\in\mathcal{V}_{-1}^-(0)} w_{ik}) + \lambda > 0$, with the latter due to the left inequality of (19). If $i_0 \in \mathcal{V}_{-1}^-(0)$, being $\delta_i(\boldsymbol{z}(0)) < 0$, then $x_i(1) = -1$ and similarly, $y_i(1) \leq (1-\lambda)\sum_{j\in\mathcal{V}_{+1}^+(0)} w_{ij} - \lambda < 0$ according to (19). This concludes the proof of the induction basis. By using an induction argument, it can be shown by repeating the arguments used in (G.1) and (G.2) that if $\boldsymbol{z}(t)$ is polarized with respect to the same partitioning and (19) holds, then also $\boldsymbol{z}(t+1)$ is polarized with respect to that partitioning. Hence, the induction principle yields the first claim.

Finally, Theorem 4 guarantees convergence to an equilibrium: $\lim_{t\to\infty} \boldsymbol{z}(t) \to \boldsymbol{z}^*$. Now, $\boldsymbol{z}(0)$ is polarized with respect to $(\mathcal{V}_1^+(0), \mathcal{V}_{-1}^-(0))$ by hypothesis, and the above has established that $\mathcal{V}_1^+(t) = \mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(t) = \mathcal{V}_{-1}^-(0)$ for all $t \geq 0$. In other words, the elements in the two sets remain unchanged over time. Separately, the inequalities in (19) satisfy the conditions (18a)–(18c) of Theorem 6 (which can be applied since Assumptions 1–3 hold), and thus $\boldsymbol{z}^*$ must be polarized with respect to $(\mathcal{V}_1^+(0), \mathcal{V}_{-1}^-(0))$. For notational convenience, we denote $\mathcal{V}_1^+(0)$ and $\mathcal{V}_{-1}^-(0)$ as $\mathcal{V}_p$ and $\mathcal{V}_n$, which completes the proof. □