Evolution of swing voters under strategic campaigns in presidential elections

Ziqian Liu [1,2,3] Xin Wang [1,2,3,4,5,6,*] Junyu Lu,[1,3] Longzhao Liu [1,2,3,4,5]
Hongwei Zheng [7] and Shaoting Tang [1,2,3,4,5,8,9,10,†]

[1]*School of Artificial Intelligence, Beihang University, Beijing 100191, China*
[2]*Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing, Beihang University, Beijing 100191, China*
[3]*Key Laboratory of Mathematics, Informatics and Behavioral Semantics, Beihang University, Beijing 100191, China*
[4]*Zhongguancun Laboratory, Beijing 100094, China*
[5]*State Key Laboratory of Complex & Critical Software Environment, Beihang University, Beijing 100191, China*
[6]*State Key Laboratory of General Artificial Intelligence, BIGAI, Beijing, China*
[7]*Beijing Academy of Blockchain and Edge Computing, Beijing 100085, China*
[8]*Hangzhou International Innovation Institute, Beihang University, Hangzhou 311115, China*
[9]*Institute of Trustworthy Artificial Intelligence, Zhejiang Normal University, Hangzhou 310012, China*
[10]*Institute of Medical Artificial Intelligence, Binzhou Medical University, Yantai 264003, China*

(Received 30 October 2025; accepted 29 January 2026; published 23 March 2026)

Political polarization, fueled by public discourse and echo chambers, has profound implications for the functioning of democratic elections. However, traditional one-dimensional opinion models—assuming "support for one party equals opposition to another"—fail to capture the nuanced dynamics of swing voters (including neutrals, left leaners, and right leaners), who are critical for the final election outcomes. This study introduces a two-dimensional opinion model that classifies voters into five groups, enabling precise characterization of the swing group's interactive behaviors. Importantly, we introduce antagonism effect to describe the intensity with which the two camps incite opposition and exert voting pressure in the run-up to the election, typically via Us-versus-Them framing. By integrating the open-mindedness of voters, the stubbornness of opinion interactions, and the antagonism effect arising from strategic campaigns, we systematically explore the intricate interplay between top-down political campaigns and bottom-up interpersonal opinion dynamics, unveiling their nonlinear coupling impacts on the emergence and evolution of swing voters. Counterintuitively, we find that extreme antagonism effects might backfire in presidential election: When both parties adopt intense antagonistic strategies, the party that polarizes more strongly risks alienating swing voters, thereby enabling its ostensibly weaker opponent to prevail. These insights are also illustrated on the core retweet networks during the 2020 U.S. presidential election. Building upon multidimensional opinion model, our results highlight the possibility of mobilizing swing voters and shaping electoral outcomes through antagonistic strategies of political parties. Our work also provides a nuanced and generalizable framework for analyzing opinion dynamics in other polarized public discourse.

DOI: 10.1103/jdyh-t4rp

I. INTRODUCTION

Political polarization has emerged as a widespread feature of contemporary democratic politics, with empirical research consistently documenting its deepening roots in online information ecosystems [1]. Substantial volumes of online network data afford scholars invaluable insights into the investigation of political polarization across diverse sociopolitical contexts, such as the US presidential elections [2,3], the impeachment of the former Brazilian President Dilma Roussef [4], COVID-19 pandemic [5,6], vaccination [7], and environmental protection [8]. Homophilous interactions force individuals to increasingly engage with like-minded peers, leading to the formation of echo chambers and information cocoons [9,10]. Conversely, polarized information ecosystems restructure social networks via information cascades, causing users to lose cross-ideological ties at rates exceeding random chance [11]. Empirical analyses further highlight how algorithmic recommendation systems amplify ideological segregation [12], while the spread of false news reveals that misinformation disproportionately reaches and mobilizes partisan groups, marginalizing the middle ground [13,14].

Amidst this backdrop of escalating political polarization, the swing groups—comprising neutrals and leaners—emerge as pivotal yet paradoxical actors in democratic discourse [15–19]. The importance of swing groups resides in their dual role as bridges for cross-ideological dialogue and critical decision-maker of political mobilization. Empirical analyses reveal that these groups serve as "bridge builders" in networked environments, maintaining connectivity between polarized camps and mediating cross-ideological information flow [20,21]. Moreover, their positional centrality in networks

*Contact author: wangxin_1993@buaa.edu.cn
†Contact author: tangshaoting@buaa.edu.cn

like Twitter during election debates highlights their capacity to shift political trends [22,23]. Yet this importance is matched by profound complexity, which stems from their multidimensional attitudinal configurations and context-dependent behaviors [8,24–26]. Recent empirical work has reconstructed high-dimensional political opinion spaces from online social networks and revealed structurally cohesive and weakly interacting political communities, providing direct evidence for multidimensional polarization in real-world digital platforms [27]. Moreover, in the context of presidential elections, the diversity of political issues (e.g., immigration, tax reform, and education) and the complexity of partisan ideologies create conditions for ideological uncertainty among swing voters [28–30].

Despite the empirical evidence highlighting the significant role of swing groups, modeling frameworks historically oversimplify their complexity, often treating them as homogeneous "undecided" blocs rather than dynamic, multidimensional actors [31]. As an effective modeling approach, opinion dynamics plays an important role in theoretically understanding how microscopic interaction rules among agents give rise to macroscopic patterns of group phenomena, such as consensus, polarization, or fragmentation [32,33]. Based on the classic DeGroot model of opinion dynamics, which assumes that individuals update their opinions by taking weighted averaging opinions of their networked neighbors [34–36], a large class of extensions of this mechanism have been proposed by introducing additional assumptions, such as Friedkin-Johnse model with initial prejudice [37] and Hegselmann-Krause model with bounded confidence [38]. In social networks, homophilic interactions and preferential engagement with like-minded peers have given rise to another class of models designed to capture the escalation of opinion polarization and the emergence of echo chambers [24,39].

Existing research on opinion dynamics has predominantly relied on one-dimensional continuous models, which represent attitudes as points on a linear spectrum. While these frameworks effectively capture simple polarization and fragmentation, they are constrained by the implicit binary-opposition hypothesis that support for one party is assumed to be inherent opposition to the other [40–42], exhibiting fundamental limitations in capturing the complexity of swing group evolution. Furthermore, the top-down dynamics—such as elite-driven strategies [43,44], algorithm curation [45–47], and media framing [48]—remain underexplored in modeling swing group evolution, despite their pervasive impact.

Therefore, this paper focuses on more precisely modeling the swing group and studying how the strategic campaigns of political parties affect their evolution and final election outcomes. We develop a mesoscale multicognitive model based on two-dimensional opinions, which explicitly incorporates open-mindedness of voters [49–51], the stubbornness of opinion interactions [52], and the antagonism effect from external political campaigns to capture the emergence and evolution of swing voters. First, by separately considering the different views of voters toward the two parties, we assign each voter a two-dimensional opinion vector and categorize voters into five groups—Party A supporters, Party A leaners, neutrals, Party B leaners, and Party B supporters—providing a more nuanced representation of the voter population [53–55]. We then integrate this multicognitive framework with traditional models to describe a cross-scale group opinion dynamics. Our results show the dual role for antagonism in the political system: While it enhances voter mobilization, it simultaneously erodes the diversity of ideological interaction and fosters two-dimensional echo chambers, trapping voters' opinion updating within two homogeneous networks. Our model also reveals the complex interaction between the top-down antagonistic political campaign and bottom-up individual stubbornness, triggering the emergence and evolution of neutral groups. Counterintuitively, competitive antagonism between parties has a complex impact on electoral outcomes: Under weak to moderate antagonism, the party with more intense antagonistic strategies secures a voting advantage by mobilizing more swing voters; while with both parties deploying extreme antagonism, the party with weaker antagonism gains more support from neutrals, leading to a reversal of voting advantage. Finally, we apply our proposed model to retweet networks from the 2020 U.S. presidential election, showing that its qualitative dynamics remain consistent in the real world.

II. MODEL

We now first introduce this multicognitive framework based on the two-dimensional coupling opinions to achieve the measurement and characterization of multiple cognitive groups, including party supporters, party leaners, and the neutrals [see schematic in Fig. 1(a)].

A. Agent properties

Each voter is characterized by a two-dimensional opinion vector $o_i(t) = (x_i(t), y_i(t)) \in [0, 1]^2$, where $x_i(t)$ and $y_i(t)$ represent the cognitions toward Party A and Party B, respectively, at time t. A clear preference for Party A corresponds to $x_i \approx 1$ and $y_i \approx 0$, while strong support for Party B results in $y_i \approx 1$ and $x_i \approx 0$. The opinion difference $z_i = x_i - y_i$ serves as the attitudinal indicator, whose symbolic value can measure the attitude and voting tendency of the voter i toward the two parties. A positive $z_i(t)$ ($\mathrm{sign}(z_i) = 1$) indicates Party A preferences, whereas a negative value ($\mathrm{sign}(z_i) = -1$) reflects Party B inclination. The initial opinions $(x_i(0), y_i(0))$ are independently distributed with uniform probability.

To further characterize the complexity of political spectrum, especially the differences within swing voters, we divide the population into five types based on z_i and two thresholds z_o (outer boundary for strong partisans) and z_v (inner boundary for neutrals): Party A supporters ($z_i(t) > z_o$), Party A leaners ($z_v < z_i(t) \leqslant z_o$), neutrals ($|z_i(t)| \leqslant z_v$), Party B leaners ($-z_o \leqslant z_i(t) < -z_v$), and Party B supporters ($z_i(t) < -z_o$). Among them, leaners and neutrals compose the swing voter population. The interactive classification parameter $\sigma_o(i, t)$ for opinion dynamics is defined as

$$\sigma_o(i,t) = \begin{cases} A, & z_i(t) > z_o, \\ C, & z_o \geqslant z_i(t) \geqslant -z_o, \\ B, & z_i(t) < -z_o, \end{cases} \tag{1}$$

Here, A, B, and C, respectively, designate Party A supporters, Party B supporters, and swing voters.

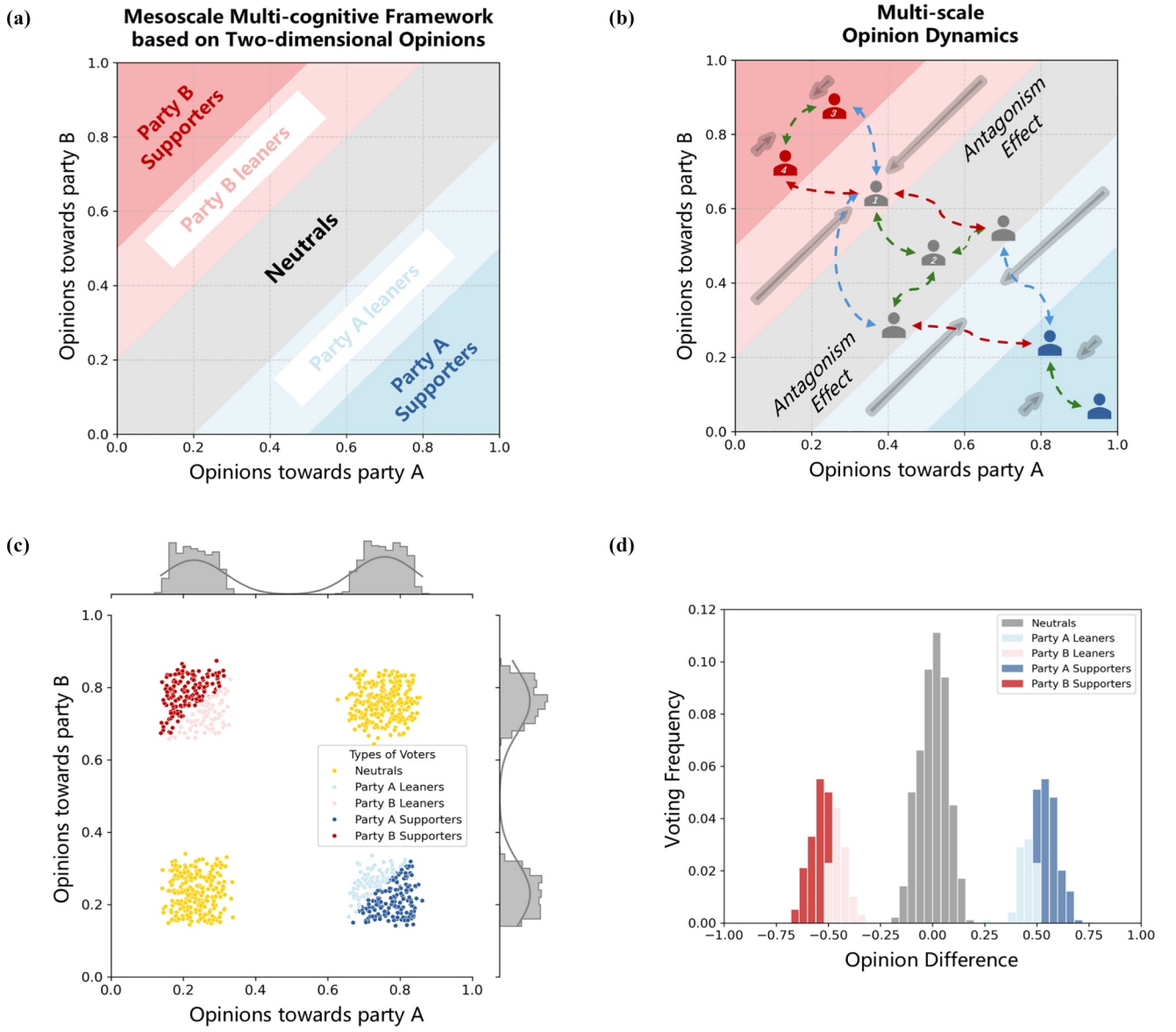

FIG. 1. (a) Schematic of the mesoscale multicognitive framework based on two-dimensional coupling opinions. Voters are divided into five categories: Party A supporters, Party A leaners, neutrals, Party B leaners, and Party B supporters according to the disparities of their opinions toward Parties A and B. (b) Schematic of the multiscale opinion dynamics, which incorporates voters' initial preferences, network interactions, and antagonism effect. The blue, red, and green arrows represent the interactions between opinions toward Party A, Party B, and both parties, respectively. Among them, voter 1 interacts with voter 3 and voter 4, respectively, on opinions toward Party A and Party B, while voter 1 and voter 2 interact with each other on opinions toward both parties. Furthermore, the antagonism effect forces the opinions of all voters to converge in the direction of one-dimensional opposition. (c), (d) An example of distributions of two-dimensional coupling opinions and voting outcomes. Polarization emerges in the group opinions of the two parties, and the voting outcomes show a trimodal distribution, with the neutral nonvoting group appearing. Parameters: The example begins from an ER graph with $N = 10^3$ and $\langle k \rangle = 40$, $z_o = 0.5$, and $z_v = 0.2$. The fixed parameters are $\epsilon = 0.2$, $\lambda = 0.3$, and $\rho = 0$.

The evolution of voters' opinions culminates in the election vote, and the total duration of opinion interactions and updates is denoted by T. All simulations in the main text are performed with $T = 100$, which is demonstrated in Appendix A to be sufficient for the opinion dynamics to reach a stationary state.

It is noteworthy that at election time T, only neutrals ($|z_i(T)| \leqslant z_v$) abstain from voting, while leaners and supporters cast votes according to their final $z_i(T)$.

B. Network structure

The model employs a graph $G = (V, E)$ with N nodes representing voters. Each node v_i features a self-loop $(v_i, v_i) \in E$, enabling self-opinion updating independent of external interactions. The edge $(v_i, v_j) \in E$ represents the path of interaction between the voter i and the neighboring voter j.

Based on their current state $\sigma_o(j, t)$, neighbor nodes v_j of given v_i are dynamically categorized into three mesoscale interaction groups:

(1) v_j is the Party A supporter. The sets of x-opinion and y-opinion neighbors for v_i at time t are as follows:

$$\mathcal{N}_i^{AX}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = A, \quad |x_i(t) - x_j(t)| < \epsilon\}$$
$$\mathcal{N}_i^{AY}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = A. \quad |y_i(t) - y_j(t)| < \epsilon\} \tag{2}$$

(2) v_j is the Party B supporter. The sets of x-opinion and y-opinion neighbors for v_i at time t are as follows:

$$\mathcal{N}_i^{BX}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = B, \quad |x_i(t) - x_j(t)| < \epsilon\}$$
$$\mathcal{N}_i^{BY}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = B. \quad |y_i(t) - y_j(t)| < \epsilon\} \tag{3}$$

(3) v_j is the swing voter. The sets of x-opinion and y-opinion neighbors for v_i at time t are as follows:

$$\mathcal{N}_i^{CX}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = C, \quad |x_i(t) - x_j(t)| < \epsilon\}$$
$$\mathcal{N}_i^{CY}(t) = \{j|(v_j, v_i) \in E, \sigma_o(j,t) = C. \quad |y_i(t) - y_j(t)| < \epsilon\} \tag{4}$$

Here, the open-mindedness ϵ represents the interaction threshold between nodes and their neighbors. Only when their opinions are close enough will they interact and influence each other.

C. Updating rules

Building upon the mesoscale multicognitive framework, we now establish a multiscale opinion dynamics mechanism [see schematic in Fig. 1(b)]. This model explicitly incorporates the open-mindedness of voters, the stubbornness of opinion interactions, and, importantly, the antagonism effect induced by external political campaigns to capture the realistic dynamics of partisan competition and swing voter dynamics in presidential elections.

The evolution of opinions balances the initial preferences and neighbor interactions. In the absence of antagonism, the heterogeneous updating rules for node v_i are defined as

(1) Party A supporters: Voters identified as Party A supporters update their x opinions by interacting with Party A supporters and swing voters whose x opinions lie within the interaction threshold, and update their y opinions by interacting with Party A supporters and swing voters whose y opinions lie within the interaction threshold.

$$\hat{x}_i(t+1) = \lambda_i x_i(0) + (1-\lambda_i)\left(w_{AX}\frac{1}{|\mathcal{N}_i^{AX}(t)|}\sum_{j\in\mathcal{N}_i^{AX}(t)} x_j(t) + w_{CX}\frac{1}{|\mathcal{N}_i^{CX}(t)|}\sum_{k\in\mathcal{N}_i^{CX}(t)} x_k(t)\right),$$
$$\hat{y}_i(t+1) = \lambda_i y_i(0) + (1-\lambda_i)\left(w_{AY}\frac{1}{|\mathcal{N}_i^{AY}(t)|}\sum_{j\in\mathcal{N}_i^{AY}(t)} y_j(t) + w_{CY}\frac{1}{|\mathcal{N}_i^{CY}(t)|}\sum_{k\in\mathcal{N}_i^{CY}(t)} y_k(t)\right). \tag{5}$$

(2) Party B supporters: Voters identified as Party B supporters update their x opinions by interacting with Party B supporters and swing voters whose x opinions lie within the interaction threshold, and update their y opinions by interacting with Party B supporters and swing voters whose y opinions lie within the interaction threshold.

$$\hat{x}_i(t+1) = \lambda_i x_i(0) + (1-\lambda_i)\left(w_{BX}\frac{1}{|N_i^{BX}(t)|}\sum_{j\in N_i^{BX}(t)} x_j(t) + w_{CX}\frac{1}{|N_i^{CX}(t)|}\sum_{k\in N_i^{CX}(t)} x_k(t)\right),$$
$$\hat{y}_i(t+1) = \lambda_i y_i(0) + (1-\lambda_i)\left(w_{BY}\frac{1}{|N_i^{BY}(t)|}\sum_{j\in N_i^{BY}(t)} y_j(t) + w_{CY}\frac{1}{|N_i^{CY}(t)|}\sum_{k\in N_i^{CY}(t)} y_k(t)\right). \tag{6}$$

(3) Swing voters: Voters identified as swing voters update their x opinions by interacting with all neighbors whose x opinions lie within the interaction threshold and update their y opinions by interacting with all neighbors whose y opinions lie within the interaction threshold.

$$\hat{x}_i(t+1) = \lambda_i x_i(0) + (1-\lambda_i)\left(w_{AX}\frac{1}{|\mathcal{N}_i^{AX}(t)|}\sum_{j\in\mathcal{N}_i^{AX}(t)} x_j(t) + w_{BX}\frac{1}{|\mathcal{N}_i^{BX}(t)|}\sum_{k\in\mathcal{N}_i^{BX}(t)} x_k(t) + w_{CX}\frac{1}{|\mathcal{N}_i^{CX}(t)|}\sum_{l\in\mathcal{N}_i^{CX}(t)} x_l(t)\right),$$
$$\hat{y}_i(t+1) = \lambda_i y_i(0) + (1-\lambda_i)\left(w_{AY}\frac{1}{|\mathcal{N}_i^{AY}(t)|}\sum_{j\in\mathcal{N}_i^{AY}(t)} y_j(t) + w_{BY}\frac{1}{|\mathcal{N}_i^{BY}(t)|}\sum_{k\in\mathcal{N}_i^{BY}(t)} y_k(t) + w_{CY}\frac{1}{|\mathcal{N}_i^{CY}(t)|}\sum_{l\in\mathcal{N}_i^{CY}(t)} y_l(t)\right). \tag{7}$$

TABLE I. Possible influence weight for i: $\sigma_o = A$.

Influence weight	w_{AX}	w_{CX}	w_{AY}	w_{CY}
If $\mathcal{N}_i^{CX} = \varnothing$	1	0	–	–
If $\mathcal{N}_i^{CX} \neq \varnothing$	0.6	0.4	–	–
If $\mathcal{N}_i^{CY} = \varnothing$	–	–	1	0
If $\mathcal{N}_i^{CY} \neq \varnothing$	–	–	0.6	0.4

Opinions of voters evolve through independent updates for each party dimension, incorporating contributions from initial preferences, where voters retain inherent stubbornness quantified by parameter $\lambda \in [0, 1]$ that controls the weight of initial opinions $x_i(0)$, $y_i(0)$, and network interactions, which involve weighted averages of mesoscale group opinions with the strength of intergroup influence being quantified by the weight w.

In the main text, we fix all the influence weight for simplicity (see Tables I–III) and further verify their robustness in Appendix B.

In addition, the stubbornness of all voters is equally quantified by $\lambda_i = \lambda \in [0, 1]$, thus the opinion dynamics can be expressed as

$$\begin{aligned} \hat{x}_i(t+1) &= \lambda x_i(0) + (1-\lambda)\bar{x}_i(t), \\ \hat{y}_i(t+1) &= \lambda y_i(0) + (1-\lambda)\bar{y}_i(t). \end{aligned} \tag{8}$$

Here, $\bar{x}_i(t)$, $\bar{y}_i(t)$ denote weighted means of x opinions and y opinions from neighboring voters, respectively.

The opinion updating rule, mathematically equivalent to a convex combination of initial and neighbor-averaged opinions, can be reformulated as a convex optimization problem seeking the global minimum of the following objective function:

$$\begin{aligned} \hat{\mathcal{L}}(x_i, y_i) = {} & \lambda(x_i - x_i(0))^2 + (1-\lambda)(x_i - \bar{x}_i(t))^2 \\ & + \lambda(y_i - y_i(0))^2 + (1-\lambda)(y_i - \bar{y}_i(t))^2. \end{aligned} \tag{9}$$

Further, to describe the intensities with which the two camps incite opposition and exert voting pressure in the run-up to the election, we introduce the antagonism effect that couples the two-dimensional opinion updates through a zero-sum framing. Here, "zero-sum" does not imply a strict conservation law, but rather an adjustable framing mechanism in which gains in support for one party are implicitly presented as losses for the opponent. In the two-dimensional opinion space (x_i, y_i), this antagonistic framing geometrically attracts agents toward the manifold $x_i + y_i = 1$, corresponding to the binary-opposition structure between the two parties.

TABLE II. Possible influence weights for i: $\sigma_o = B$.

Influence weight	w_{BX}	w_{CX}	w_{BY}	w_{CY}
If $\mathcal{N}_i^{CX} = \varnothing$	1	0	–	–
If $\mathcal{N}_i^{CX} \neq \varnothing$	0.6	0.4	–	–
If $\mathcal{N}_i^{CY} = \varnothing$	–	–	1	0
If $\mathcal{N}_i^{CY} \neq \varnothing$	–	–	0.6	0.4

TABLE III. Possible influence weights for i: $\sigma_o = C$.

Influence weight	w_{AX}	w_{BX}	w_{CX}	w_{AY}	w_{BY}	w_{CY}
If $\mathcal{N}_i^{AX} = \varnothing$ and $\mathcal{N}_i^{BX} = \varnothing$	0	0	1	–	–	–
If $\mathcal{N}_i^{AX} = \varnothing$ and $\mathcal{N}_i^{BX} \neq \varnothing$	0	0.5	0.5	–	–	–
If $\mathcal{N}_i^{AX} \neq \varnothing$ and $\mathcal{N}_i^{BX} = \varnothing$	0.5	0	0.5	–	–	–
If $\mathcal{N}_i^{AX} \neq \varnothing$ and $\mathcal{N}_i^{BX} \neq \varnothing$	0.3	0.3	0.4	–	–	–
If $\mathcal{N}_i^{AY} = \varnothing$ and $\mathcal{N}_i^{BY} = \varnothing$	–	–	–	0	0	1
If $\mathcal{N}_i^{AY} = \varnothing$ and $\mathcal{N}_i^{BY} \neq \varnothing$	–	–	–	0	0.5	0.5
If $\mathcal{N}_i^{AY} \neq \varnothing$ and $\mathcal{N}_i^{BY} = \varnothing$	–	–	–	0.5	0	0.5
If $\mathcal{N}_i^{AY} \neq \varnothing$ and $\mathcal{N}_i^{BY} \neq \varnothing$	–	–	–	0.3	0.3	0.4

The resulting function is as follows:

$$\begin{aligned} \mathcal{L}(x_i, y_i) = {} & (1-\rho_i)[\lambda(x_i - x_i(0))^2 + (1-\lambda)(x_i - \bar{x}_i(t))^2 \\ & + \lambda(y_i - y_i(0))^2 + (1-\lambda)(y_i - \bar{y}_i(t))^2] \\ & + \rho_i(x_i + y_i - 1)^2. \end{aligned} \tag{10}$$

Equation (10) displays the complex interplay of three key factors: initial partisan preferences, mesoscale-mediated interactions, and antagonism coupling between party opinions, where $\rho_i \in [0, 1]$ tunes the strengths of antagonism effect.

Intuitively, when $\rho = 1$, the antagonism reaches its maximum, which means the binary hypothesis holds strictly ($x_i + y_i = 1$), and the formula degenerates into a one-dimensional case; while when $\rho = 0$, the formula decomposes into two independent updating rules of classical opinion dynamics in Eq. (8).

Solving the optimization problem yields the following explicit expression for opinion dynamics:

$$\begin{aligned} x_i(t+1) &= \frac{\hat{x}_i(t+1) - \rho_i\hat{y}_i(t+1) + \rho_i}{1+\rho_i}, \\ y_i(t+1) &= \frac{\hat{y}_i(t+1) - \rho_i\hat{x}_i(t+1) + \rho_i}{1+\rho_i}. \end{aligned} \tag{11}$$

The detailed convexity proof and the solution process for the objective function are provided in Appendix C. The python implementation of the model simulations is available [56].

Without considering the antagonism effect ($\rho = 0$), Figs. 1(c) and 1(d), respectively, illustrate examples of the two-dimensional opinion distribution and the voting outcomes of the voter groups emerging at the macro level. Under moderate open-mindedness, the group opinions toward both parties exhibit polarization, which is consistent with classic opinion dynamics, but the final voting behavior presents a three-peak distribution. Besides the voting groups of the two parties, there are a large number of neutrals that do not participate in the voting. The opinions of this nonvoting community toward two parties are quite similar. They may hold both supportive or both opposing opinions simultaneously, reproducing the complexity of swing groups that cannot be described by traditional cognitive models based on one-dimensional opinions and binary hypothesis.

III. RESULTS

A. Opinion dynamics under varying open-mindedness

In the scenario where no antagonism effect is operative, the opinion-updating processes of the two parties unfold independently. Herein, the degree of open-mindedness plays a crucial role in shaping the evolution of opinions.

We begin from the exploration of the distribution of group opinions and voting outcomes across a spectrum of open-mindedness degrees. For a given open-mindedness, we not only characterize the two-dimensional distribution of group opinions but also analyze the marginal distributions of partisan attitudes. We explore the number of opinion clusters by employing the kernel density estimation curve of the x-opinion and y-opinion distributions and computing the number of its local maxima [57–59]. We utilize the average number of x-opinion and y-opinion peaks to represent the number of opinion clusters. Taking polarization (with a cluster number of 2) as the boundary, we classify the opinion clusters under different open-mindedness into four distinct regimes: fragmentation, polarization, multipolarization, and consensus, and present the voting outcomes in Fig. 2(a).

At the extreme end of the spectrum, when ϵ is exceedingly low, the opinions of both parties exhibit a multiclustered ideological partitions, while the ideological stances of both parties gradually converge to a neutral position with an exceedingly high ϵ. Under a low open-mindedness, voters are highly resistant to different ideas, leading to the formation of variegated opinion distribution and a balanced proportion of neutrals and actual voters [Fig. 2(b)]. Conversely, under a particularly high open-mindedness, the free flow of ideas and willingness of voters to consider alternative perspectives facilitate the formation of broad consensus and the emergence of a dominating neutral group [Fig. 2(e)].

When the degree of open-mindedness is at a moderate level, the well-known phenomenon of ideological polarization emerges distinctly within each party [Fig. 2(c)]. Consequently, the voting outcomes exhibit distinct tripolar characteristics with a significant proportion of the neutral group. However, the transition from polarized group opinion to neutral consensus involves an intermediate state of multipolarization [Fig. 2(d)]. Under a wider open-mindedness, some swing voters absorb political discourses from both parties, leading to the emergence of multiple scattered swing blocs. Notably, the multipolarization of partisan opinions also results in diversified voting outcomes, as the spread of ideological subgroups encourages broader electoral participation.

During competitive presidential elections, the opinion distributions of fragmentation and neutral consensus are unlikely to occur. Therefore, we focus on the dynamics of group polarization, with particular emphasis on the evolution of swing groups under strategic campaigns. In the subsequent investigation, we constrain the parameter of open-mindedness within a moderate regime.

B. Voting polarization and two-dimensional echo chambers under antagonism

At the moderate level of open-mindedness in the absence of antagonism effects, the ideological stances of both parties exhibit pronounced polarization, which turn into tripolarization in voting behavior. This state arises from the independent evolution of each party's opinion-updating system, unmoderated by cross-ideological interaction or external pressure. The neutral nonvoting group reflects the ideological ambivalence or strategic abstention of voters between the two polarized extremes.

In this section, we incorporate the antagonism effect that couples the two-dimensional opinions and explore how such external partisan rivalry reshapes opinion dynamics and voting behavior through mobilizing previously inactive voters. Here, all voters are subject to a homogeneous level of antagonism, allowing us to evaluate how the competition between two parties influences swing voter behaviors.

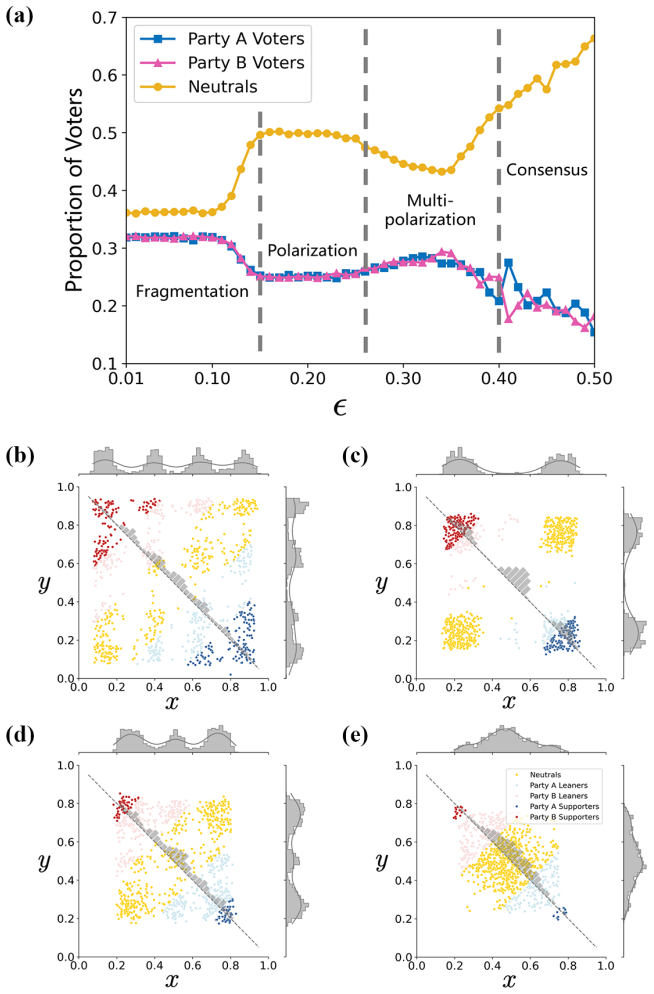

FIG. 2. Open-mindedness and opinion evolution. (a) The open-mindedness (characterized by the interaction threshold) of voters significantly influences the distribution of two-dimensional opinions. (b) Small open-mindedness leads to opinion segmentation. (c) Under a moderate level of open-mindedness, the ideological stances of both parties emerge polarization. (d), (e) The continuously enhanced open-mindedness goes through multipolarization and gradually fosters a broad consensus around neutrality. Simulation results are averaged over 100 independent runs. Parameters: Simulations in panel (a) begin from an Erdős-Rényi (ER) graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, $z_v = 0.2$, $\rho = 0$, and $\lambda = 0.3$. We change (b) $\epsilon = 0.1$. (c) $\epsilon = 0.25$. (d) $\epsilon = 0.35$. (e) $\epsilon = 0.45$.

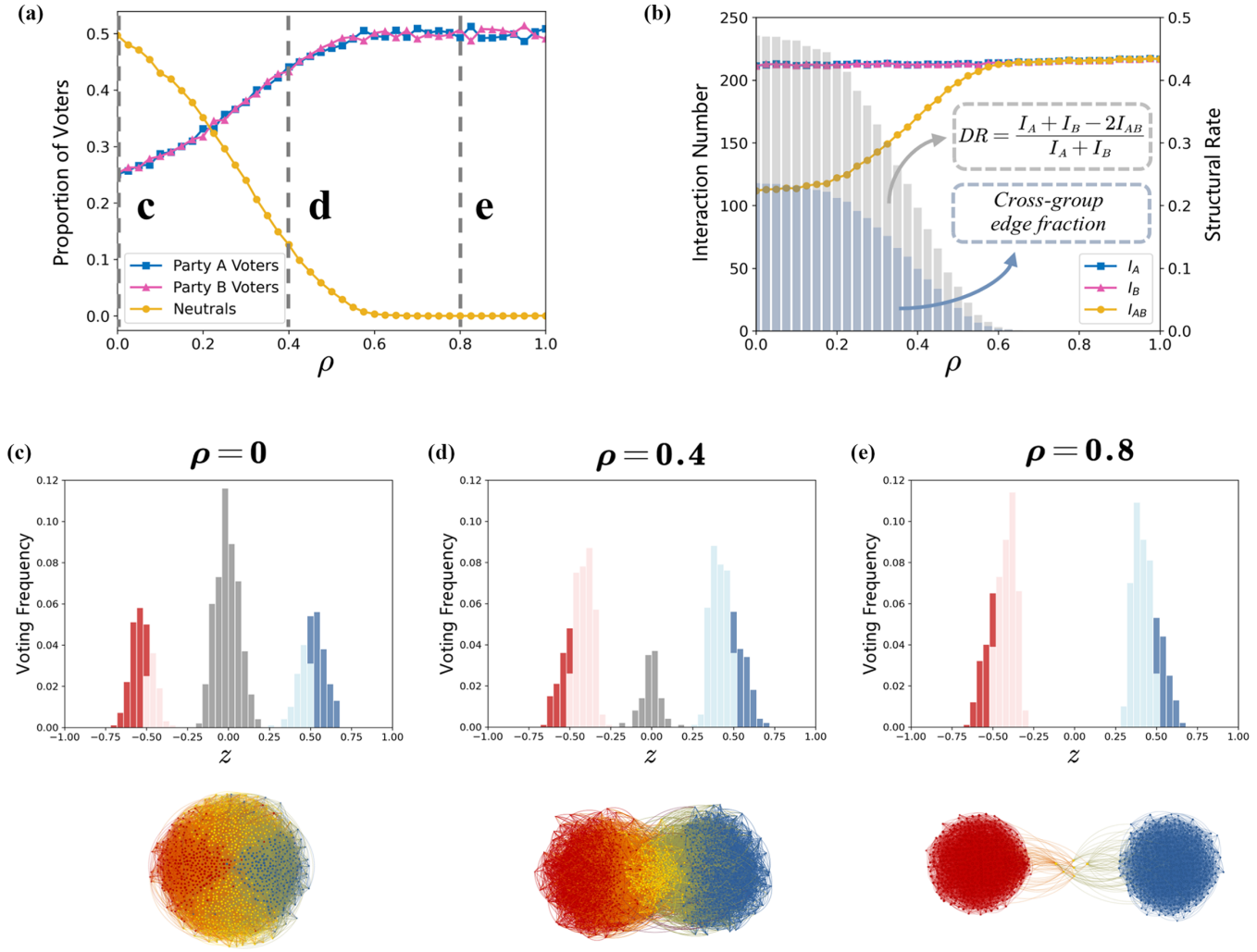

FIG. 3. The emergence of two-dimensional echo chambers under antagonism effect. Increasing the level of antagonism is sufficient to (a) promote the voting participation of swing voters and (b) reduce the diversity of opinion interactions and the cross-group edges. In panel (b), I_A, I_B, and I_{AB}, respectively, denote the average number of interacting neighbors regarding x opinion, y opinion, and both opinions. Panels (c)–(e) show stable voting distributions and network structures corresponding to different strengths of antagonism effects, illustrating the elimination of the neutrals and the emergence of two-dimensional echo chambers. Parameters: simulations begin from an ER graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, $z_v = 0.2$, $\epsilon = 0.2$, and $\lambda = 0.3$. We change (c) $\rho = 0$. (d) $\rho = 0.4$. (e) $\rho = 0.8$.

As the antagonism effect becomes stronger, Fig. 3(a) presents a clear trend in voting outcomes: The proportion of neutrals decreases monotonically, while the voting rates of both parties increase. This reflects the mobilizing effect of heightened partisan rivalry, which incentivizes swing voters—particularly neutrals—to commit to a certain party rather than remain inactive. Under extremely strong antagonism, the electorate reaches full participation, with voting outcomes exhibiting complete polarization where no neutral bloc remains. This transition underscores the role of antagonism as a catalyst for electoral engagement.

We further investigate whether the coupling effect between partisan opinions gives rise to opinion segregation and the emergence of two-dimensional echo chambers. First, let I_A, I_B, and I_{AB} denote the average number of interacting neighbors regarding x opinion, y opinion, and both opinions, respectively, which are defined as follows:

$$I_A = \frac{1}{N}\sum_i |\mathcal{N}_i^x|, \quad I_B = \frac{1}{N}\sum_i |\mathcal{N}_i^y|,$$
$$I_{AB} = \frac{1}{N}\sum_i |\mathcal{N}_i^x \cap \mathcal{N}_i^y|. \tag{12}$$

Here, $\mathcal{N}_i^x = \{j \in \mathcal{N}_i \,|\, |x_i - x_j| < \epsilon\}$ denote the interaction neighbors of each voter i under x opinion and analogously $\mathcal{N}_i^y$ for y opinion. Based on these definitions, we introduce the diversity rate defined as $DR = (I_A + I_B - 2I_{AB})/(I_A + I_B)$ to characterize the structural redundancy of opinion interactions in the two-dimensional opinion space. Intuitively, $DR = 0$ means voters' perception of both parties are almost entirely shaped by the same interaction networks, indicating the maximal redundancy where the diversity of opinion interaction

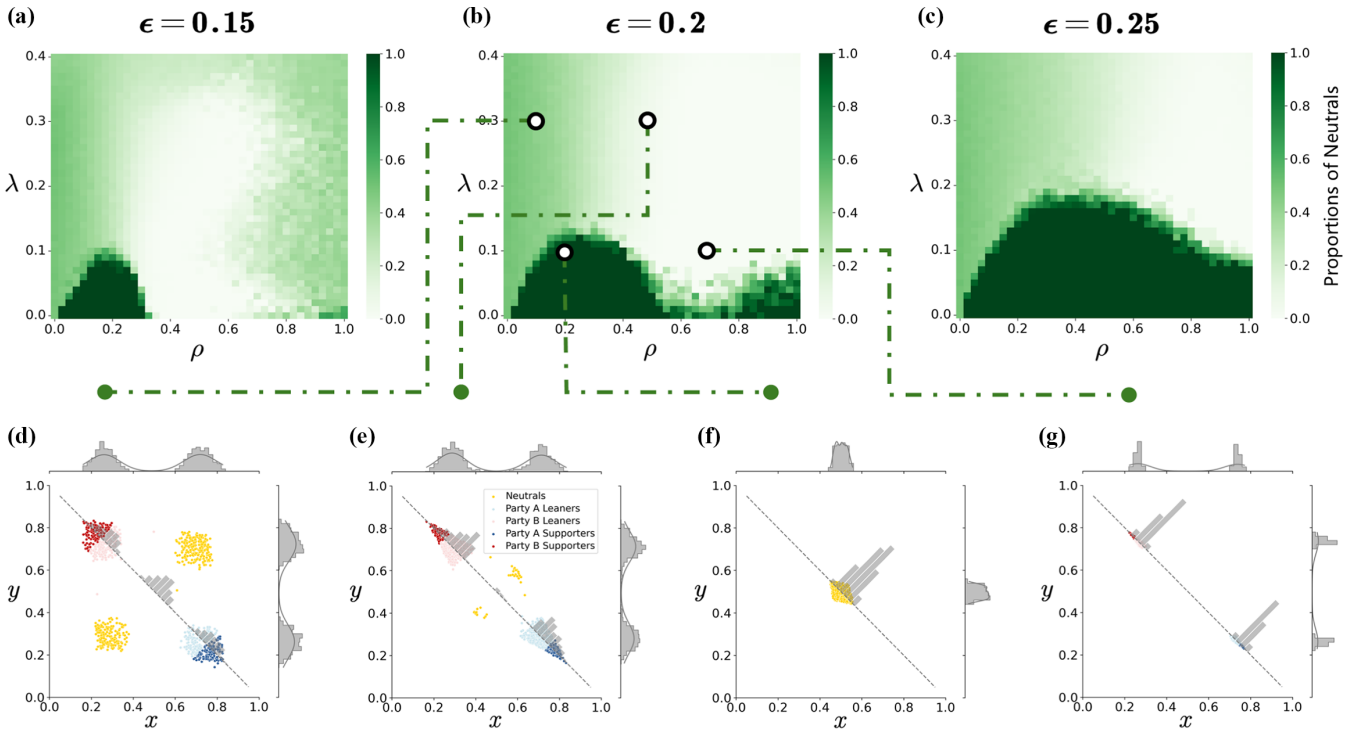

FIG. 4. Nonlinear coupling between stubbornness and antagonism effect: evolution of the swing group. We show the number of neutrals at low [(a), $\epsilon = 0.15$)], moderate [(b), $\epsilon = 0.2$)], and high [(c), $\epsilon = 0.25$)] open-mindedness. A relatively high value of stubbornness maintains the polarization of partisan opinions, while low stubbornness can cause complex phase transitions between polarization and consensus. (d)–(g) We present the distribution of two-dimensional opinions and voting outcomes under four typical cases. Simulation results are averaged over 50 independent runs. Parameters: simulations begin from an ER graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, and $z_v = 0.2$. We change (d) $\epsilon = 0.2$, $\lambda = 0.3$, $\rho = 0.1$. (e) $\epsilon = 0.2$, $\lambda = 0.3$, $\rho = 0.5$. (f) $\epsilon = 0.2$, $\lambda = 0.1$, $\rho = 0.2$. (g) $\epsilon = 0.2$, $\lambda = 0.1$, $\rho = 0.7$.

reaches its lowest point. When $DR = 1$, in the contrary, the interaction networks of both parties are totally independent.

In Fig. 3(b), we show that DR also decreases monotonically, together with the cross-group edge fraction. We classify voters into three groups based on their voting outcomes and calculate the fraction of cross-group edges. At low antagonism, over 40% of information transmission remains party-specific, allowing independent ideological exposure. As antagonism levels exceed 0.5, both the diversity rate and the cross-group edge fraction approach zero simultaneously. This homogenization of opinion interaction suppresses cross-ideological exchange, driving the formation of isolated echo chambers where voters are increasingly insulated from opposing views. These insights can also be observed from the stable voting distributions and network structures under different antagonism effects shown in Figs. 3(c)–3(e).

Our findings reveal a dual role for antagonism in the political system: While it enhances voter mobilization, it simultaneously erodes the diversity of ideological interaction. The vanishing of DR under high antagonism demonstrates how partisan competition can create self-reinforcing feedback loops, where voters' opinions become trapped within homogeneous networks—a hallmark of the two-dimensional echo chamber.

C. Nonlinear coupling between stubbornness and antagonism

In Fig. 4, we further investigate the intricate interplay between the top-down antagonism (ρ) exerted by external political campaigns and the bottom-up individual stubbornness (λ) applying to opinion dynamics.

Figure 4(b) clearly illustrates the complex nonlinear influence of the coupling effects on voting behavior under moderate open-mindedness. For moderate and relatively high values of λ, the neutral voter bloc is small, revealing that an adequate level of individual stubbornness promotes the formation and maintenance of political polarization. Under this circumstance, an increase in ρ drives a monotonic decrease in neutral voters, which is consistent with our prior observations [Figs. 4(d) and 4(e)]. Nevertheless, when the group's stubbornness is extremely high, voters are more likely to adhere to their initial opinions and resistant to external antagonism. As a result, the neutral group maintains a certain size and cannot be completely eliminated.

Conversely, when λ is relatively small, voters' opinions become highly malleable and easily swayed by external factors, making it challenging to maintain a stable polarized state. When ρ is small, the effect of antagonism in reducing ambiguous opinions is relatively slow. Meanwhile, it provides broader groups of voters with more opportunities to communicate and interact, leading the opinion evolution to shift from polarization to consensus [Fig. 4(f)]. As ρ continues to increase, the antagonism effect redominates the coupling system, with ambiguous opinions rapidly move toward the two extremes and even stronger polarization has emerged [Fig. 4(g)].

In the extreme case, where λ is extremely small and ρ is extremely large, population consensus achieves again. This

counterintuitive phenomenon arises from the interplay between the high plasticity of voters and the externally imposed pressure to take political sides. In this scenario, the two-dimensional opinions of all voters rapidly converge toward the line $x + y = 1$ under a strong antagonism effect. This convergence facilitates more cross-group interactions, making even extreme supporters susceptible to the opinions of leaners. Through extensive intergroup exchanges, neutral voters bridge the partisan divide, ultimately driving the system to global consensus.

For completeness, we also explore the detailed coupling effects under lower and higher open-mindedness in Figs. 4(a) and 4(c), respectively. The overall trends in $\rho - \lambda$ phase plane largely align with the behavioral patterns in Fig. 4(b). Besides, under a larger ϵ, voters are more susceptible to a wider range of interpersonal influence, driving the system to a broader consensus. In contrast, a smaller ϵ restricts large-scale communications, facilitating the emergence of stronger polarization.

In summary, our results have underscored the complex interplay between top-down political campaign and bottom-up interpersonal opinion dynamics. The coupling of antagonism, stubbornness, and open-mindedness exerts significant nonlinear effects on the emergence and evolution of swing groups, particularly the neutrals, which has a profound impact on the final election results.

D. Strategic campaigns on swing voters: effects of heterogeneous antagonism

Although the preceding analysis is restricted to symmetric and homogeneous parameter settings, our results have clearly suggested that election outcomes can be influenced through swing voter dynamics under party-level antagonistic interactions. In this section, we further examine how heterogeneous antagonism, exerted by different parties according to their own strategies, influences the election outcomes.

We divide the population into two groups according to their instantaneous partisan preference toward the two parties, characterized by the sign of opinion difference $z_i(t)$, and apply heterogeneous antagonism effects to these two groups. Specifically, the antagonism level assigned to voter i is defined as

$$\rho_i(t) = \begin{cases} \rho_A, & \text{if } \operatorname{sign}(z_i(t)) = 1, \\ \rho_B, & \text{if } \operatorname{sign}(z_i(t)) = -1, \end{cases} \tag{13}$$

where ρ_A and ρ_B denote the antagonism intensities exerted by Party A and Party B, respectively.

Here, the dependence of ρ_i on $z_i(t)$ is intended to represent targeted antagonistic exposure that adapts to voters' current partisan leanings rather than being fixed by their initial affiliations. Since this state-dependent assignment may raise concerns regarding endogeneity or regime switching during opinion evolution, a detailed clarification of this model choice, together with robustness checks under fixed antagonism assignments, is provided in Appendix D.

In addition, we define the relative voting rate p as the proportion of Party A voters among all actual voters that exclude the subgroup of neutrals. Clearly, $p > 0.5$ indicates that Party A holds a voting advantage, while $p < 0.5$ implies that Party B secures the upper hand.

Figure 5(a) demonstrates that, under low and moderate levels of antagonism, political parties with higher intensity of antagonism achieve a higher voting turnout. This can be attributed to the fact that increasing the strength of antagonism effect can effectively attract swing voters who lean toward a party and convert them into actual voters of that party. Nonetheless, when both parties adopt extreme campaigning strategies and exert strong antagonism, an intriguing phenomenon emerges: the party applying weaker antagonism attains a higher voting share, resulting in a dramatic reversal of voting advantage. To further examine the influence of initial opinion configurations on this reversal phenomenon, we conduct additional simulations under polarized initial conditions in Appendix F. We find that the symmetry of initial opinions plays a crucial role in the emergence of reversal: while reversal remains robust under symmetric polarized initialization, asymmetric polarization fundamentally reshapes the reversal dynamics. Under extremely strong antagonism, the party endowed with an initial partisan advantage completes and maintains electoral dominance, whereas the initially disadvantaged party is no longer able to reverse the outcome.

The robustness of these observations with respect to different levels of individual stubbornness λ is verified in Figs. 5(c)–5(e). We find that higher levels of stubbornness help the opinion system resist the influence of external antagonism, reducing the voting advantage of the dominant party. Of particular interest, as λ increases, the area of the parameter space where reversal occurs decreases. To further quantitatively assess the probability and parameter region of reversal phenomenon, we implement a dynamic square window of size 0.4×0.4, whose center slides along the diagonal of the ρ_a–ρ_b phase plane, and calculate the frequency of reversal within each window. Results in Fig. 5(b) show that the reversal rate grows monotonically as the intensity of antagonism increases. Here, we mark, with the threshold $p = 0.5$, the regions in which reversal emerges relatively stably and with higher probability. We show that lower levels of stubbornness not only facilitate the onset of reversal but also yield a higher reversal rate under identical antagonism intensities, thereby increasing the instability of electoral outcomes.

Further, to gain an in-depth understanding of the mechanism underlying the emergence of reversal, we characterize the temporal evolution process of group opinions under a set of fixed antagonism parameters, as illustrated in Fig. 5(f). We find that the reversal of voting advantage is rooted in the delicate transformation dynamics of the swing voters. When Party B exerts more extreme antagonism, its leaners quickly side with Party B supporters, breaking the communication bridge between Party B and neutral voters. In contrast, Party A leaners gradually absorb these neutrals through sustained ideological interactions, which eventually induces the macroscale reversal: The party adopting a weaker antagonistic strategy attracts more swing voters and, paradoxically, prevails in the fierce competition.

E. Model performance on core networks during the 2020 U.S. presidential election

To illustrate the model behavior under real-world topology, we apply our modeling framework to a representative political

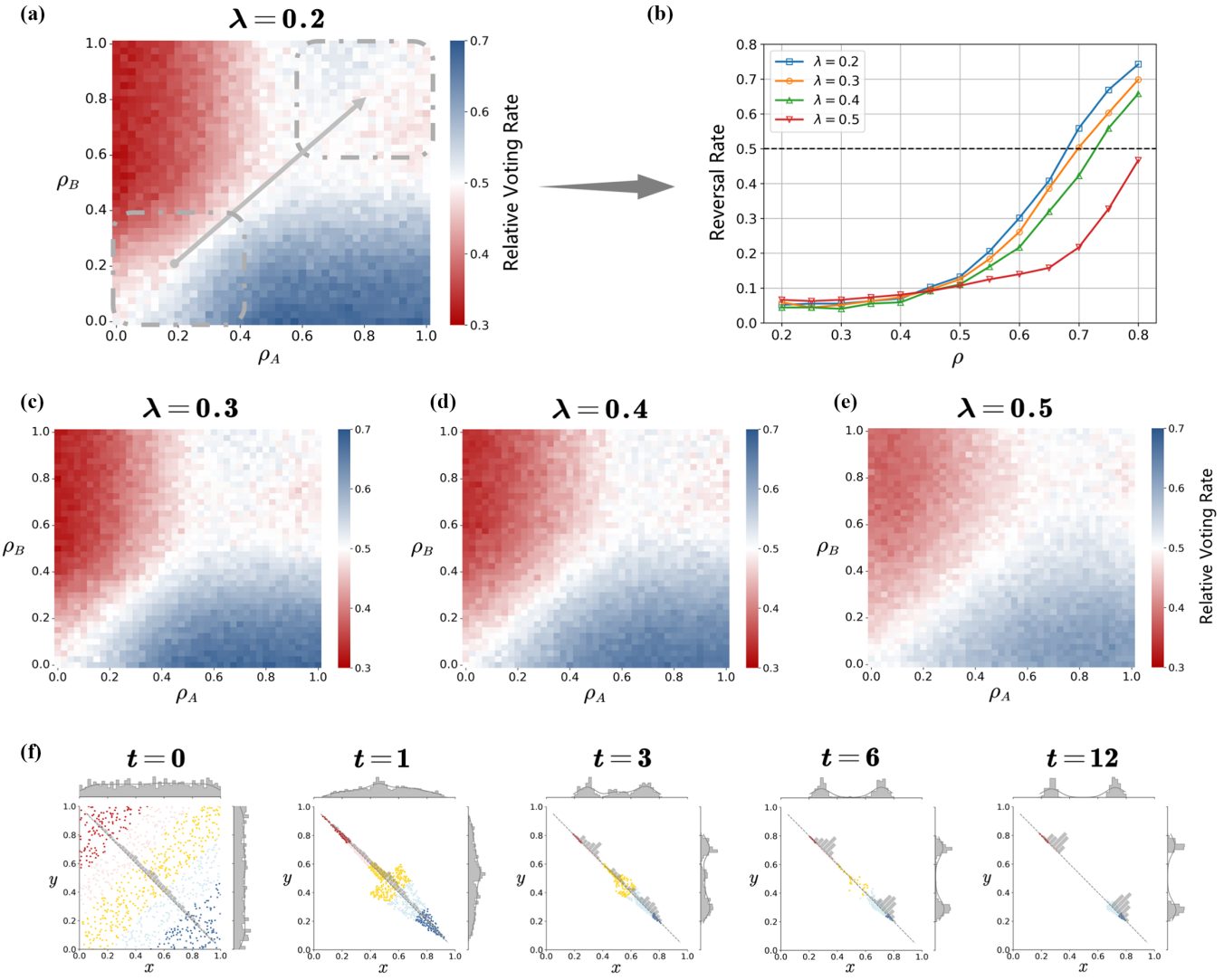

FIG. 5. Strategic influence on swing voters and reversal of voting advantage. We apply heterogeneous antagonism effects to voters with different party preferences and show phase diagrams (a), (c), (d), (e) at different stubbornness levels. Under low and moderate antagonism level, the party that imposes stronger antagonism can obtain higher vote support, and this advantage weakens as the group's stubbornness level increases. When both parties create extremely strong antagonism, the party with weaker antagonism strategy may obtain more votes from swing voters and surprisingly achieve a reversal of voting advantage. We verify the existence of such reversal through Panel (b). We implement a square box with a length and width of 0.4 on Panel (a), letting its center slide along the diagonal direction. The horizontal axis is the antagonism level corresponding to the center of the rectangular box, and the vertical axis is the proportion of times the party with weaker antagonism reverse the voting advantage. Panel (f) shows the evolution snapshots of voting reversal under fixed parameters. Simulation results are averaged over 50 independent runs. Parameters: simulations begin from an ER graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$. Other parameters (a) $\lambda = 0.2$. (c) $\lambda = 0.3$. (d) $\lambda = 0.4$. (e) $\lambda = 0.5$. (f) $\lambda = 0.2$, $\rho_A = 0.6$, $\rho_B = 0.9$.

election. We utilize Twitter (now called X) data from the 2020 US presidential election spanning from November 1 to November 30. User IDs are modeled as network nodes, and directed retweet relationships are represented as directed edges from retweeting users to the users being retweeted. In this data-driven network topology, retweeting users are assumed to be influenced by the users they retweet, so that the direction of influence flows opposite to the edge direction.

Here, we filter it using the keyword "vaccine" to perform the typical case study. Complying with previous studies, we extract the giant connected component and employ the k-core method to construct the core retweet network, deriving a clearer structure of the core groups while eliminating the interference from scattered and peripheral nodes (see data processing details in Appendix G).

Given that real-world retweet networks generally exhibit pronounced polarization, we identify two major partisan communities corresponding to the Democratic and Republican camps via the community detection algorithm. Users are thus categorized into two groups, and their initial opinions are assigned based on their community affiliations—distinct from the random initialization used in artificial networks before. Specifically, users in the Democratic community are set with $x(0) > y(0)$, whereas those in the Republican community satisfy $x(0) < y(0)$.

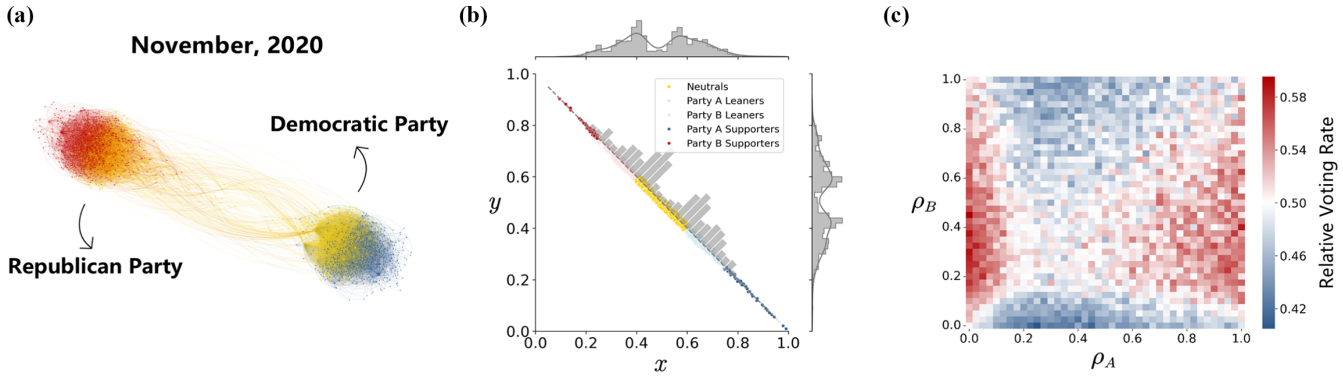

FIG. 6. Model performance in real-world networks and the robustness of key findings. We use Twitter (now X) data during the 2020 US election. In November 2020, we filter the retweet network using the k-core algorithm with $k = 9$. (a) The core retweet network after model simulations exhibits a symmetric echo-chamber structure, in which nodes represent users and edges denote effective retweet relationships. The blue, red, and yellow nodes represent Democratic voters, Republican voters, and neutral nonvoters, respectively. The final network contains 1981 nodes and 15 737 edges. (b) The opinion distributions of both parties and voting outcomes reveal a balanced polarization pattern between two parties. (c) We also verify the emergence of advantage reversal in this network structure. Simulation results are averaged over 50 independent runs. Parameters: $z_o = 0.5$, $z_v = 0.2$, $\epsilon = 0.2$, and $\lambda = 0.2$. We fix (b) $\rho = 0.8$.

In November, as the election approached its final stage, the monthly retweet network displayed balanced polarization between the two parties based on the keyword. We extract the core network using the k-core algorithm with $k = 9$ and initialize opinions based on the detected communities. According to the model's final state—representing voting outcomes—users voting for the Democratic Party, Republican Party, and abstaining neutrals are visualized in blue, red, and yellow, respectively. The resulting network structure and opinion distributions, shown in Figs. 6(a) and 6(b), indicate that the two partisan communities display symmetric structures and opinion distributions, with the neutral voter proportion accounting for 15%–20%. These intermediary neutrals facilitate limited cross-chamber interaction, acting as bridges between the polarized voting coalitions.

Of particular interest, we observe the emergence of the reversal phenomenon in this case. In specific, we show the relative voting advantage in ρ_A–ρ_B-phase plane under the echo-chamber structure in Fig. 6(c). Compared with ER random networks, real retweet networks exhibit an even broader parameter region for generating advantage reversal. Notably, the prepolarized network structure shows higher sensitivity to the modulation of antagonism effects, making it more prone to inducing large-scale reversals of voting advantage.

Overall, we demonstrate that in real election networks, there is an even greater likelihood of voting advantage reversal induced by extreme antagonism, highlighting the delicate balance through swing group dynamics in actual elections. The robustness of voting reversal regarding various network topologies is shown in Appendix H.

IV. CONCLUSIONS AND DISCUSSIONS

Under the furious circumstances of contemporary presidential elections, the behavior of swing voters—those navigating the ideological spectrum between polarized parties—represents a pivotal nexus where democratic outcomes are determined. These individuals embody the intricate interplay of ideological flexibility, social influence, and external political campaigns. Yet existing frameworks often oversimplify them as "undecided" and anchors in one-dimensional assumptions of binary opposition (support for one party as inherent opposition to another), failing to capture the multidimensionality of their attitudes. Swing voters may hold issue-specific preferences that defy strict partisan alignment or adopt neutrality as a strategic response to hostile political climates, a complexity that traditional models obscure. Moreover, these frameworks overlook the nontrivial coupling effects of top-down campaigns and bottom-up opinion interactions. These oversights have limited our ability to explain how these groups solidify into decisive blocs, retreat into abstention, or even trigger electoral shocks—dynamics that define modern democratic elections.

This study addresses this gap with a sociologically rich framework that reimagines political attitude as a two-dimensional space, where stances for opposing parties evolve not in lockstep but through dynamic interactions between campaign strategies and psychological dispositions. By examining partisan antagonism—the degree to which campaigns frame politics as an adversarial "us vs them" contest, the research reveals several interconnected insights with profound societal relevance. First, the external antagonism effect systematically erodes the diversity of ideological interactions while enhancing voting mobilization, leading to the formation of two-dimensional echo chambers. This process not only mobilizes neutrals into partisan blocs but also creates self-reinforcing feedback loops where voters are increasingly insulated from opposing views, mirroring empirical patterns in social media networks [2,4,9]. Second, the study uncovers a complex interaction between top-down antagonistic strategies and bottom-up individual stubbornness, which exerts a nonlinear impact on the emergence and evolution of the swing voters. Last but not least, we find a nuanced relationship between competitive antagonism and electoral outcomes: Under weak to moderate antagonism, the party with more intense antagonistic strategies secures a voting advantage by mobilizing more swing voters. Paradoxically, when both parties deploy extreme antagonism, the resultant ideological competition can

backfire, triggering advantage reversal where the party with weaker antagonism gains more support from neutrals.

Analysis of retweet networks during 2020 U.S. presidential election reveals that real-world echo chambers—characterized by homogeneous partisan interaction and limited cross-ideological engagement—create fertile ground for the model's predicted advantage reversal under extreme antagonism. In such polarized environments, swing voters trapped in information silos may defect to the less aggressive party as a psychological response to cognitive overload [60,61]. This phenomenon, rooted in sociological patterns of homophily and psychological mechanisms of cognitive dissonance reduction, highlights how digital segregation amplifies the risk of unpredictable electoral shifts [62,63].

Our framework has certain limitations. We adopt the static network structure, ignoring the evolution of the network itself caused by widely used AI recommendation algorithms that can be considered in future research. Additionally, to simplify our model and emphasize the coupling influence from top to bottom and bottom to top, we only consider the heterogeneity of external campaigns and the dynamic interactions among different types of voters, without taking into account the heterogeneity of individual stubbornness and open-mindedness. Future research may attribute the heterogeneity of individual characteristics based on real data.

In an age of increasing polarization and democratic strain driven by the growing complexity of social media, these findings highlight the need to reconsider the roles of political strategies and their impact on civic life [64]. We show the possibility of balancing clarity with ambiguity, using antagonism to define choices without destroying the informational diversity that enables meaningful engagement. This work reminds us that the future of democracy hinges not on the power of partisan extremes, but on the capacity to honor the complexity of swing voter dynamics—a sociological reality that holds the key to inclusive governance and resilient democratic cultures. Furthermore, although based on election contexts, our modeling framework is, in fact, applicable to studying the evolution of opinions in all polarized environments, such as vaccines, gun control, abortion, immigration, and other issues.

ACKNOWLEDGMENTS

This work was supported by National Science and Technology Major Project (2022ZD0116800), Program of National Natural Science Foundation of China (62141605, 12425114, 12201026, 12301305), the Fundamental Research Funds for the Central Universities, Beijing Natural Science Foundation (Z230001), and Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing. This work was also supported by the Opening project of the State Key Laboratory of General Artificial Intelligence, BIGAI/Peking University, Beijing, China (Project No. SKLAGI2025OP16).

DATA AVAILABILITY

The data that support the findings of this article are openly available [56].

APPENDIX A: MODEL ROBUSTNESS TO FIXED ELECTION TIME T

To address the concern that the evolution time horizon T may influence the reported outcomes, we explicitly examine the temporal convergence of the opinion dynamics under different parameter settings.

We focus on the temporal evolution of voters' opinions towards Party A, $x_i(t)$, opinions toward Party B, $y_i(t)$, and their opinion difference $z_i(t) = x_i(t) - y_i(t)$. Three representative parameter configurations are considered: (1) $\lambda = 0.2$, $\rho = 0.1$, corresponding to weak antagonism; (2) $\lambda = 0.2$, $\rho = 0.5$, corresponding to relatively strong antagonism; and (3) $\lambda = 0.2$, $\rho_A = 0.6$, $\rho_B = 0.9$, where asymmetric antagonistic strategies are present. For each parameter set, we record the temporal evolution of opinions under the fixed time horizon $T = 100$.

As shown in Fig. 7, in all three cases, the opinion dynamics rapidly approaches a stable stationary state well before $T = 50$. Increasing the time horizon to $T = 100$ does not lead to any qualitative or quantitative changes in the final opinion distributions or voting outcomes. In addition, we have tested the time horizon over a broader region of parameter space and network structures, and obtained consistent convergence behavior. This confirms that the fixed horizon $T = 100$ used in the main text is sufficient to capture the asymptotic behavior of the system and that all reported results are robust with respect to the choice of T.

APPENDIX B: MODEL ROBUSTNESS TO STRUCTURAL PARAMETERS

In our framework, the mesoscale multicognitive classification of voters and intergroup interaction dynamics are governed by two critical structure parameters: opinion difference thresholds (z_o, z_v) and influence weights (ω). Here, we assess the model's robustness to variations in these parameters, which define the boundaries of partisan categorization and the strength of intergroup influence, respectively.

1. Robustness to multicognitive thresholds

The thresholds z_o and z_v partition the voter population into five behavioral classes by quantifying the salience of attitudinal polarization. To validate robustness, we systematically vary $z_o \in \{0.45, 0.55\}$ and $z_v \in \{0.15, 0.25\}$, spanning realistic ranges of attitudinal ambiguity.

Results in Fig. 8 show that while the size of mesoscale groups changes with threshold settings, the qualitative dynamics of opinion polarization and voting outcome tripolarization remain invariant.

Specifically, the inner threshold z_v defines the voting boundary but does not alter the mesoscale interaction mechanisms, which are governed by the outer threshold z_o. While z_v influences the size of the neutral population, the nonlinear coupling between stubbornness and antagonism and the establishment and reversal of heterogeneous antagonism advantage persist across z_v fine-tunings, confirming in Figs. 8(a) and 8(d) that the model's core behavioral predictions are insensitive to the precise definition of voting boundaries.

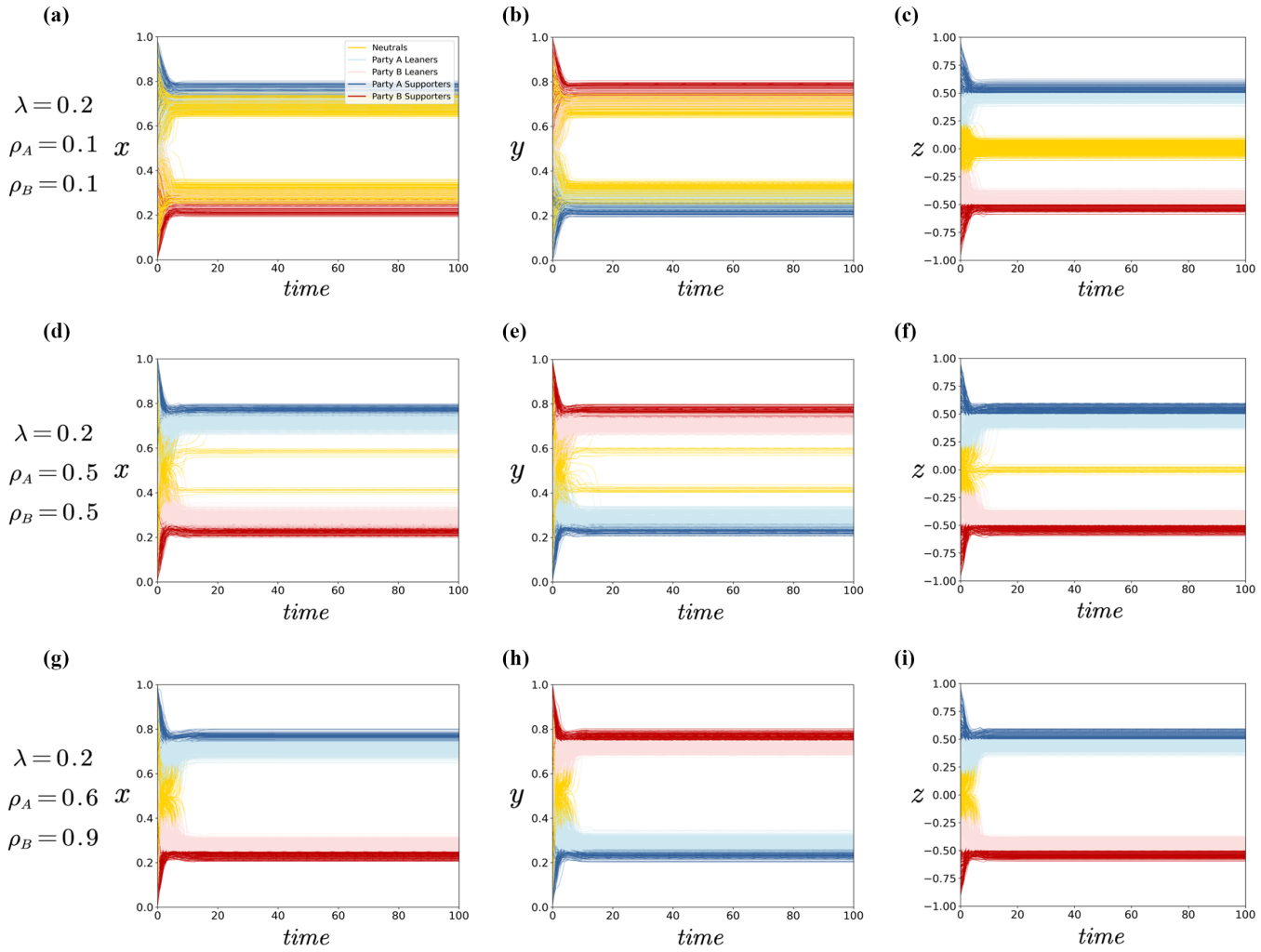

FIG. 7. Temporal evolution of group opinions toward Party A, Party B, and their opinion difference under three representative parameter settings. Results are shown for $T = 100$. In all cases, convergence to a stationary state occurs well before $T = 50$, indicating robustness with respect to the election time horizon. Parameters: simulations begin from networks with $N = 10^3$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$. We fix (a)–(c) $\lambda = 0.2$, $\rho = 0.1$, (d)–(f) $\lambda = 0.2$, $\rho = 0.5$, and (g)–(i) $\lambda = 0.2$, $\rho_A = 0.6$, $\rho_B = 0.9$.

The outer threshold z_o, in contrast, directly modulates the mesoscale composition of voter groups:

(1) Partisan polarization boundary. Contraction of z_o ($z_o \downarrow$) narrows the boundary for strong partisans, expanding the sizes of Party A/B supporters at the expense of swing voters. This shift reduces the pool of malleable leaners, inhibiting cross-ideological interactions and reinforcing intraparty homogeneity. Consequently, the system exhibits enhanced polarization stability across a wider parameter space [Fig. 8(b)], as the diminished swing population weakens the moderating effect of intergroup influence. Conversely, expanding z_o ($z_o \uparrow$) widens the threshold for strong partisans, increasing the proportion of swing voters and fostering broader ideological engagement [Fig. 8(c)]. Larger learner groups facilitate cross-cutting interactions and consensus-building mechanisms, making the model more prone to moderate, nonpolarized outcomes under higher z_o.

(2) Impactions for partisan advantage reversal. Lower z_o values, by shrinking swing voter pools and intensifying interparty segregation, impose structural barriers to advantage reversal under extreme antagonism [Fig. 8(e)]. The reduced availability of ideologically flexible neutrals and leaners limits the weak party's ability to attract support, even when confronting strong antagonism from the dominant party. Conversely, by preserving a sizable swing population, higher z_o maintains the conditions for threshold-dependent reversals, as moderate leaners remain susceptible to strategic antagonism and intergroup influence [Fig. 8(f)].

In summary, while z_v governs the magnitude of neutrals without altering core dynamics, z_o acts as a critical lever for balancing polarization and flexibility in partisan competition. The model's qualitative predictions—including tripolarization, antagonism-driven mobilization, and advantage reversal—remain robust to z_o adjustments, as the underlying mechanisms of intergroup interaction and convex opinion updating are preserved across reasonable threshold configurations. This robustness ensures the model's applicability to diverse political contexts, from highly polarized to moderately pluralistic electoral systems.

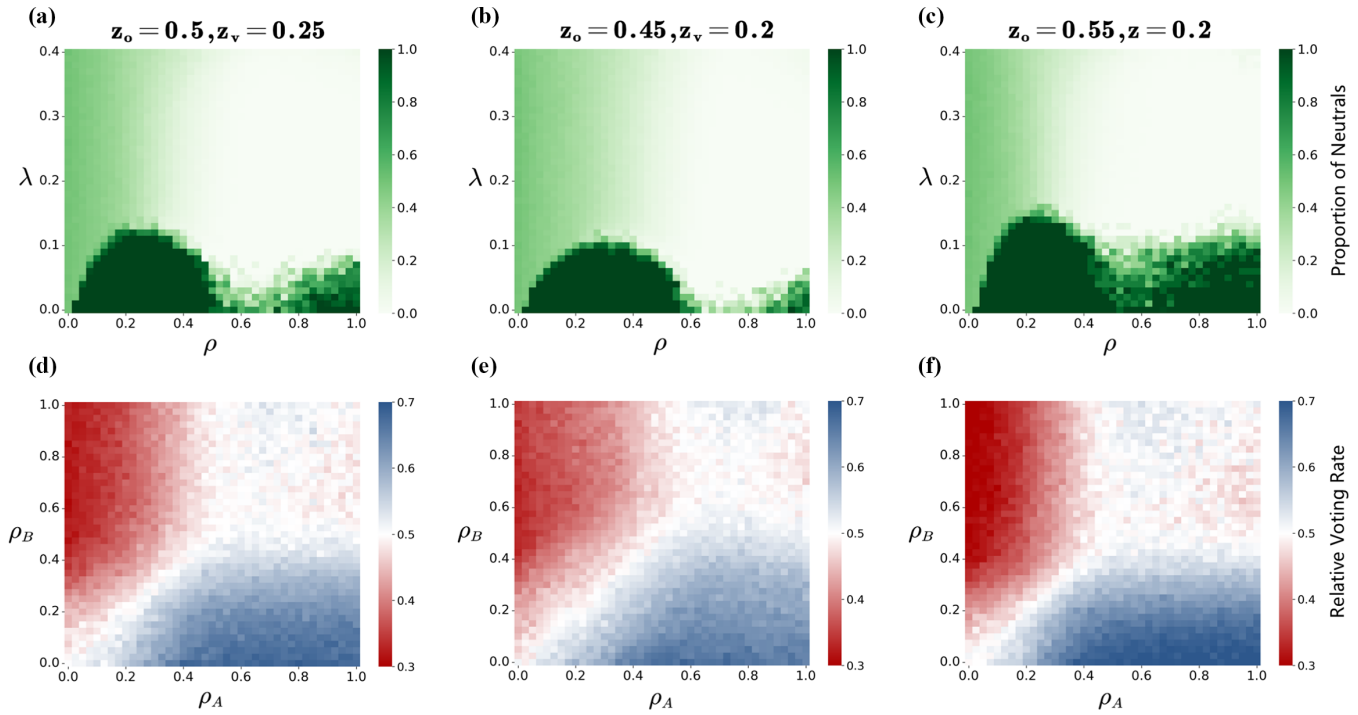

FIG. 8. Effects of nonlinear coupling between stubbornness and antagonism levels (a)–(c) and heterogeneous antagonism levels (d), (e) under different multicognitive thresholds. Results largely resemble findings from the original setting. The fine-tuning of z_o results in the shift of partisan polarization boundary. Simulation results are averaged over 50 independent runs. Parameters: simulations begin from an ER graph with $N = 10^4$, $\langle k \rangle = 40$, and $\epsilon = 0.2$. We change (a), (d) $z_o = 0.5$, $z_v = 0.25$. (b), (e) $z_o = 0.45$, $z_v = 0.2$. (c), (f) $z_o = 0.55$, $z_v = 0.2$. (d)–(f) $\lambda = 0.2$.

2. Robustness to influence weights

The influence weights w, which parametrized the strength of intergroup interactions (e.g., w_{AX}, w_{CX} in Table I), govern how mesoscale group opinions propagate through the network. We test robustness by perturbing these weights within theoretically justified ranges.

In light of the equilibrium of group interaction, we have meticulously configured two distinct sets of influence weights shown in Figs. 9(a) and 9(d): one favoring intragroup interaction dominance and the other emphasizing cross-group interaction dominance. Under the regime of intragroup interaction dominance, voters belonging to the same category assign a higher influence weight when accepting each other's opinions, while being less significantly influenced by opinions of voters from other categories. Conversely, the setting of cross-group interaction dominance presents an essentially opposite scenario.

Despite the alteration in the opinion convergence rate, results indicate that the model consistently exhibits stable macroscale outcomes. The observed differences in nonlinear phenomena are merely of an appropriate quantitative nature, without affecting the overall stability and integrity of the model's macroscopic manifestations.

(1) Partisan polarization boundary. When the interaction of voters' opinions is predominantly shaped by intragroup forces, the exchange of opinions between the extreme partisan supporters and the swing groups is severely restricted. This constraint poses a significant obstacle to the emergence of group consensus at the macro level [Fig. 9(b)]. In the phase diagram depicting the relationship between the stubbornness level and the antagonism level, polarization is prevalent and stable. Group consensus can only be achieved under conditions of extremely flexible initial preferences. In contrast, when greater consideration is given to the opinions of different types of voters, large-scale interactions become conducive to the generation of group consensus [Fig. 9(e)].

(2) Impactions for partisan advantage reversal. When the interaction of voters' opinions is mainly governed by intragroup influence, the limited exchange of opinions actually facilitates the establishment of partisan advantages at a low antagonism level. However, as the antagonism level rises, the dominant party experiences a more rapid disconnection from the neutral group, thereby triggering a reversal of the voting advantage [Fig. 9(c)]. On the contrary, in an environment that promotes cross-group interaction, the establishment and reversal of partisan advantages occur to a lesser degree [Fig. 9(f)].

Overall, the model demonstrates unwavering robustness in the face of structural parameter variations. Such parameter changes do not impinge upon the model's qualitative outcomes in any substantial way. Instead, the impact is confined to minor quantitative discrepancies, leaving the model's fundamental qualitative characteristic intact.

APPENDIX C: CONVEXITY AND GLOBAL MINIMUM OF THE OPINION UPDATING FUNCTION

In this Appendix, we provide a comprehensive analysis of the convexity of the simplified opinion updating function

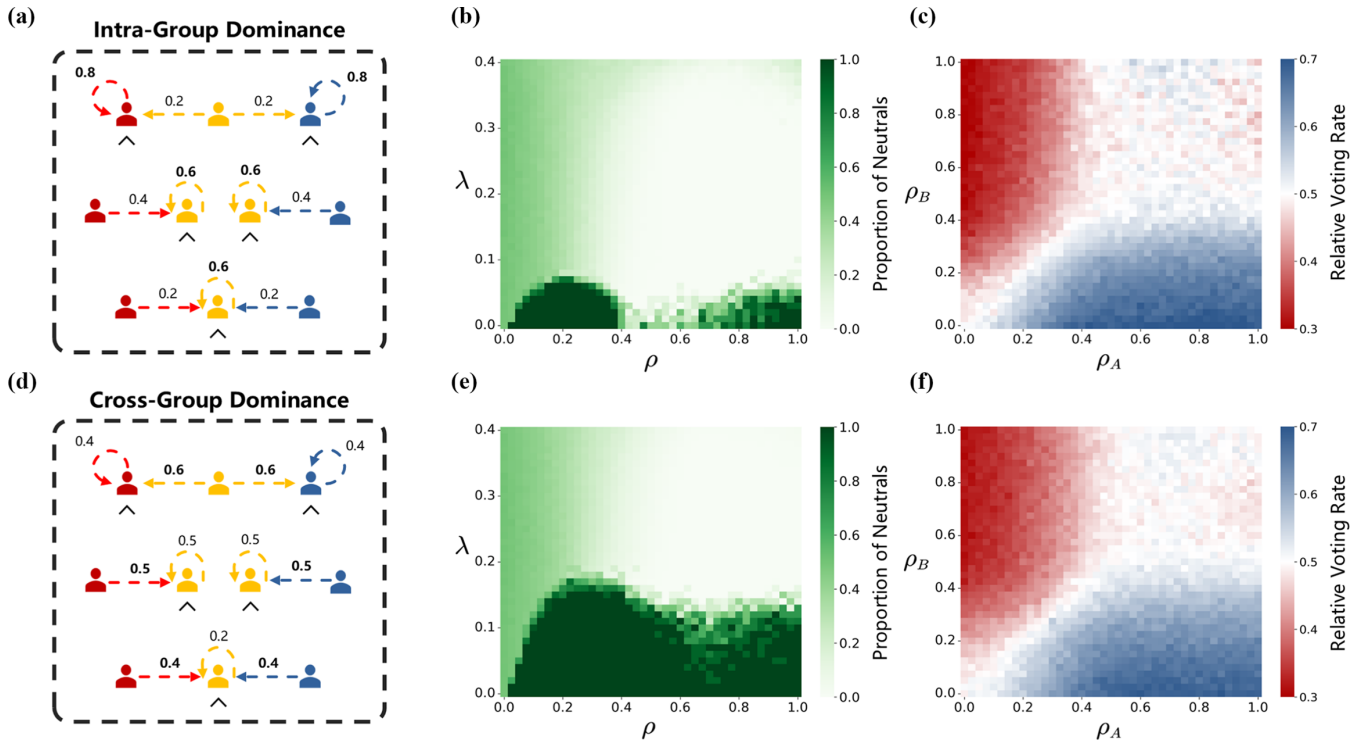

FIG. 9. Effects of nonlinear coupling between stubbornness and antagonism levels and heterogeneous antagonism levels under two sets of influence weights. We run simulations under the influence weights with intragroup dominance (a)–(c) and cross-group dominance (d)–(f). Results largely resemble findings from the original setting. Simulation results are averaged over 50 independent runs. Parameters: simulations begin from an ER graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$. We fix (c), (f) $\lambda = 0.2$.

$\mathcal{L}(x_i, y_i)$ and derive its unique global minimum. Demonstrating the convexity and finding the global minimum is crucial as it justifies the feasibility and correctness of using convex optimization methods for opinion updating.

1. Convexity proof of the function

The function $\mathcal{L}(x_i, y_i)$ is defined as

$$\begin{aligned}\mathcal{L}(x_i, y_i) = (1-\rho)[\lambda \cdot (x_i - x_i(0))^2 + (1-\lambda)(x_i - \bar{x}_i(t))^2 \\ + \lambda(y_i - y_i(0))^2 + (1-\lambda)(y_i - \bar{y}_i(t))^2] \\ + \rho(x_i + y_i - 1)^2, \end{aligned} \tag{C1}$$

where $\lambda \in [0, 1]$ represents a weighting parameter related to the influence of initial opinions and current neighborhood-averaged opinions, and $\rho \in [0, 1]$ is a parameter governing the strength of the antagonism.

We first calculate the first-order partial derivatives. The partial derivative with respect to x_i is

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial x_i} = 2[x_i - (1-\rho)\lambda x_i(0) - (1-\rho)(1-\lambda)\bar{x}_i(t) \\ + \rho y_i - \rho]. \end{aligned} \tag{C2}$$

The partial derivative with respect to y_i is

$$\begin{aligned}\frac{\partial \mathcal{L}}{\partial y_i} = 2[y_i - (1-\rho)\lambda y_i(0) - (1-\rho)(1-\lambda)\bar{y}_i(t) \\ + \rho x_i - \rho]. \end{aligned} \tag{C3}$$

Next, we find the second-order partial derivatives

$$\frac{\partial^2 \mathcal{L}}{\partial x_i^2} = 2[(1-\rho) + \rho] = 2, \tag{C4a}$$

$$\frac{\partial^2 \mathcal{L}}{\partial y_i^2} = 2[(1-\rho) + \rho] = 2, \tag{C4b}$$

$$\frac{\partial^2 \mathcal{L}}{\partial x_i \partial y_i} = \frac{\partial^2 \mathcal{L}}{\partial y_i \partial x_i} = 2\rho. \tag{C4c}$$

The Hessian matrix H of the function $\mathcal{L}(x_i, y_i)$ is

$$H = \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial x_i^2} & \frac{\partial^2 \mathcal{L}}{\partial x_i \partial y_i} \\ \frac{\partial^2 \mathcal{L}}{\partial y_i \partial x_i} & \frac{\partial^2 \mathcal{L}}{\partial y_i^2} \end{bmatrix} = \begin{bmatrix} 2 & 2\rho \\ 2\rho & 2 \end{bmatrix}. \tag{C5}$$

For a 2×2 matrix $A = \left[\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right]$ to be semipositive definite, we need $a \geqslant 0$, $d \geqslant 0$ and $ad - bc \geqslant 0$. In matrix H, $a = 2 \geqslant 0$, $d = 2 \geqslant 0$, and $ad - bc = 4 - 4\rho^2 = 4(1-\rho^2)$. Since $\rho \in [0, 1]$, then $1 - \rho^2 \geqslant 0$, so $ad - bc \geqslant 0$. Thus, the Hessian matrix H is semipositive definite, and the function $\mathcal{L}(x_i, y_i)$ is convex. This convexity justifies the use of convex optimization in our opinion updating model.

2. Finding the unique global minimum

Now we can proceed to find its unique global minimum, which occurs at the point where the first-order partial derivatives are equal to zero. That is, we need to solve the following

system of equations:

$$\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial \mathcal{L}}{\partial y_i} = 0. \tag{C6}$$

In the case when $\rho \neq 1$, we solve the system of equations for x_i, y_i:

$$\hat{x_i} = \frac{(1-\rho)[\lambda x_i(0) + (1-\lambda)\bar{x}_i] + \rho}{1-\rho^2} - \frac{\rho(1-\rho)[\lambda y_i(0) + (1-\lambda)\bar{y}_i] + \rho^2}{1-\rho^2}, \tag{C7}$$

$$\hat{y_i} = \frac{(1-\rho)[\lambda y_i(0) + (1-\lambda)\bar{y}_i] + \rho}{1-\rho^2} - \frac{\rho(1-\rho)[\lambda x_i(0) + (1-\lambda)\bar{x}_i] + \rho^2}{1-\rho^2}, \tag{C8}$$

which can be simplified to

$$\hat{x_i} = \frac{[\lambda x_i(0) + (1-\lambda)\bar{x}_i] - \rho[\lambda y_i(0) + (1-\lambda)\bar{y}_i] + \rho}{1+\rho}, \tag{C9}$$

$$\hat{y_i} = \frac{[\lambda y_i(0) + (1-\lambda)\bar{y}_i] - \rho[\lambda x_i(0) + (1-\lambda)\bar{x}_i] + \rho}{1+\rho}. \tag{C10}$$

When $\rho = 1$, the optimization objective $\mathcal{L}(x_i, y_i)$ simplifies to a constant value. Specifically, as shown in the previous derivations, when substituting $\rho = 1$ into $\mathcal{L}(x_i, y_i)$, we get $\mathcal{L}(x_i, y_i) = (x_i + y_i - 1)^2$. And from $\frac{\partial \mathcal{L}}{\partial x_i} = \frac{\partial \mathcal{L}}{\partial y_i} = 2x_i + 2y_i - 2 = 0$, we have $x_i + y_i = 1$ making $\mathcal{L}(x_i, y_i) = 0$.

In such a scenario, for the sake of consistency in the overall solution framework and without causing any contradictions in the theoretical context, we can artificially define the solution of the optimization problem when $\rho = 1$ to be identical to the form obtained in the general case when $\rho \neq 1$.

Consequently, in summary, by solving for the global minimum of the convex objective function, we obtain the following dynamic equations:

$$x_i(t+1) = \frac{[\lambda x_i(0) + (1-\lambda)\bar{x}_i] - \rho[\lambda y_i(0) + (1-\lambda)\bar{y}_i] + \rho}{1+\rho},$$
$$y_i(t+1) = \frac{[\lambda y_i(0) + (1-\lambda)\bar{y}_i] - \rho[\lambda x_i(0) + (1-\lambda)\bar{x}_i] + \rho}{1+\rho}, \tag{C11}$$

where $\rho \in [0, 1]$.

APPENDIX D: CLARIFICATION AND ROBUSTNESS OF STATE-DEPENDENT ANTAGONISM ASSIGNMENT

1. Interpretation of state-dependent heterogeneous antagonism

In the main model, the heterogeneous antagonism exposure ρ_i is assigned according to the instantaneous sign of the attitudinal indicator $z_i(t) = x_i(t) - y_i(t)$, as specified in Eq. (13). This formulation implies that the antagonistic influence exerted on a voter is conditional on their current partisan leaning rather than being fixed *ex ante*.

In contemporary political campaigns and online information environments, voters are not exposed to antagonistic messaging in a static manner; instead, exposure is dynamically adjusted through algorithmic recommendation systems, selective media consumption, and adaptive campaigning strategies that respond to users' evolving partisan preferences. As a result, voters may switch between antagonistic regimes as their attitudes evolve, reflecting a realistic form of endogenous feedback between opinion dynamics and information exposure.

Importantly, this state-dependent assignment does not impose an explicit bias toward either party. Rather, it allows antagonistic pressure to follow the voters' current leaning, capturing a regime-switching process that naturally emerges from adaptive political communication. Nevertheless, to ensure that the emergence of voting advantage reversal is not an artifact of this endogenous assignment rule, we perform an additional robustness check described below.

2. Robustness check: fixed antagonism based on initial affiliation

To examine whether the reversal phenomenon relies on state-dependent antagonism assignment, we conduct a robustness check in which each voter's antagonism level ρ_i is fixed according to their initial partisan leaning and remains unchanged throughout the entire evolution, regardless of subsequent opinion updates. Specifically, ρ_i is determined by the sign of $z_i(0)$ and does not respond to changes in $z_i(t)$ for $t > 0$. This setup corresponds to a scenario in which antagonistic exposure is determined by a stable group of community affiliation rather than by real-time ideological shifts.

The phase diagram of relative voting outcomes shown in Fig. 10(a) is fully consistent with the results reported in the main text, confirming that the establishment and reversal of voting advantage persist even when heterogeneous antagonism is fixed by voters' initial affiliations. Moreover, we verify the existence of reversal across different levels of individual stubbornness. As summarized in Fig. 10(b), the reversal rate exhibits a clear dependence on λ: Higher stubbornness systematically suppresses the likelihood of reversal, while moderate stubbornness allows reversal to emerge under strong antagonism strength.

The results demonstrate that the reversal phenomenon does not hinge on the endogenous updating of ρ_i with instantaneous opinion states. Instead, it is a robust collective outcome arising from the asymmetric interaction between antagonistic strategies and the dynamics of swing voters. Fixing antagonism exposure at the individual level merely suppresses regime switching at the microscopic scale, without altering the macroscopic emergence of advantage establishment and reversal.

APPENDIX E: HETEROGENEOUS ANTAGONISM AT LOW OPEN-MINDEDNESS

In the main text, we demonstrate that heterogeneous antagonism under moderate interaction open-mindedness drives the establishment and potential reversal of partisan voting advantages. Here, we characterize the dynamics under a low open-mindedness, where interaction constraints intensify

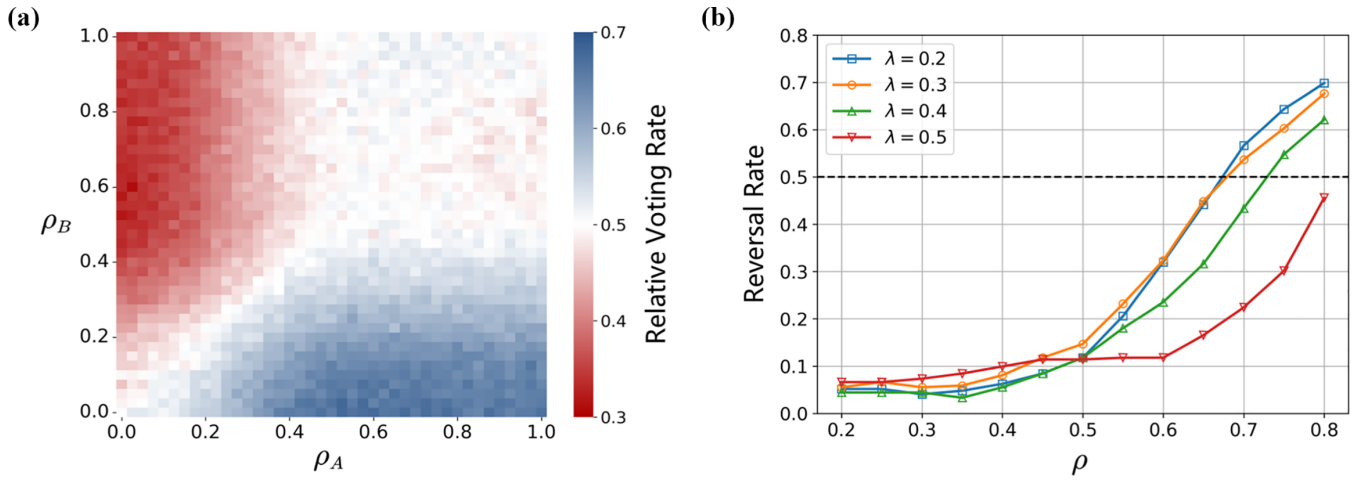

FIG. 10. Strategic influence on swing voters and reversal of voting advantage. We apply fixed antagonism based on initial affiliation. Panel (a) presents the phase diagram of relative voting rate under heterogeneous antagonism for $\lambda = 0.2$. Panel (b) further reports the reversal rate as a function of ρ for different levels of individual stubbornness $\lambda = 0.2, 0.3, 0.4$, and 0.5. Parameters: simulations begin from networks with $N = 10^4$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$.

ideological segregation, leading to distinct behavioral regimes compared to the moderate-threshold case.

A low ϵ imposes strict limits on cross-ideological engagement, confining voters to highly homophilous interactions—primarily with neighbors sharing nearly identical opinions. This structural isolation solidifies pre-existing opinions, inhibiting the diffusion of opposing opinions and fostering the emergence of tripolarized partisan configurations: cohesive blocs of Party A voters, Party B voters, and a sizable neutral bloc that remains ideologically detached from both extremes [Fig. 11(c)].

Under moderate open-mindedness, elevating the level of heterogeneous antagonism effectively drives the conversion of swing voters to a specific party, thereby boosting its voter turnout. Compared with Fig. 5(a), in regimes with severely constrained interactions as illustrated in Figs. 11(a) and 11(b), heightened antagonism exacerbates ideological polarization, amplifying the dominant party's voting advantage through the accelerated assimilation of marginally aligned individuals while curbing cross-ideological exchanges that might otherwise moderate partisan extremes.

Notably, when both parties deploy strong antagonism ($\rho_A, \rho_B > 0.5$) under low ϵ, a complex threshold phase transition emerges. The elimination of neutrals and the reversal of voting advantage, observed in moderate-threshold scenarios, becomes diverse at low ϵ:

(1) Reversal under near-polarization. When the neutral group is nearly eliminated (ρ_A, ρ_B both relatively high), the weak party can still attract residual swing voters, leading to reversals similar to the moderate-threshold case.

(2) Stable dominance with persistent neutrality (ρ_A, ρ_B both extremely high). Conversely, if a large neutral group persists due to extreme interaction restrictions, the dominant party's antagonism effectively freezes the electoral landscape, as neutrals neither participate nor shift allegiances, cementing the status quo.

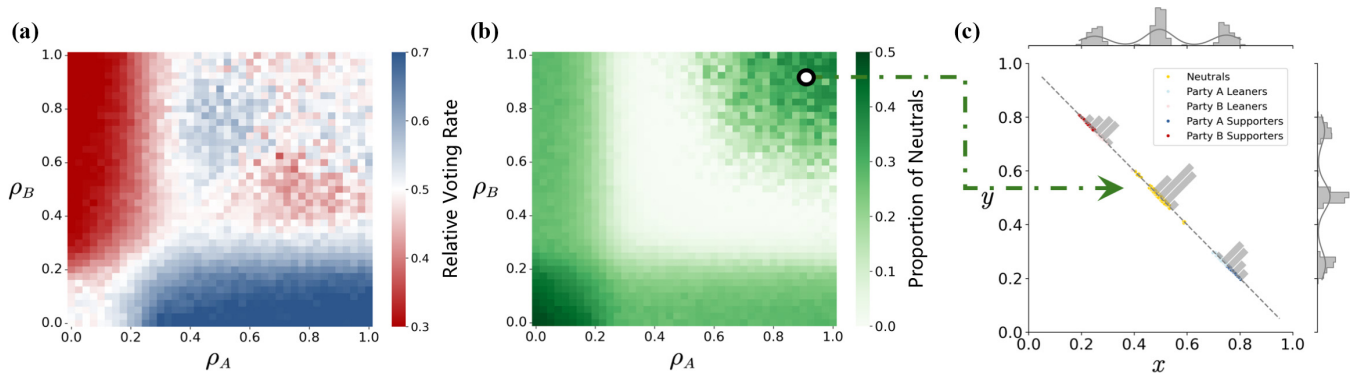

FIG. 11. The relative voting rate (a) and the number of neutrals (b) under heterogeneous antagonism at low open-mindedness. When both parties impose extremely intense levels of antagonism (c), a neutral group may emerge due to the tripolarization of ideological stances. Under such circumstances, neither the opposing party nor the weaker party manages to garner additional voting support from this middle group and the phenomenon of advantage reversal fails to materialize. Simulation results are averaged over 50 independent runs. Parameters: Simulations in panels (a) and (b) begin from an ER graph with $N = 10^4$ and $\langle k \rangle = 40$, $z_o = 0.5$, $z_v = 0.2$, $\epsilon = 0.15$, and $\lambda = 0.2$. We change (c) $\rho_A = \rho_B = 0.9$.

Mechanistically, the ability to reverse voting advantages hinges on the neutral bloc's connectivity to partisan groups: Low ϵ reduces the neutral bloc's susceptibility to extreme influence, making it a passive bystander rather than a swing constituency. Thus, advantage reversal occurs only when interaction constraints are sufficiently relaxed to allow limited cross-ideological signaling, even at low levels, enabling the weak party to gradually attract neutral voters through indirect ideological cues.

In summary, low open-mindedeness exacerbates the structural barriers to opinion diffusion, transforming heterogeneous antagonism from a driver of swing voter mobilization into a stabilizer of polarized or tripolarized states. These findings underscore the critical role of interaction topology in mediating the efficacy of antagonistic strategies, revealing how even subtle changes in engagement rules can fundamentally alter the dynamics of partisan competition and electoral outcomes.

APPENDIX F: MODEL EXPANSION TO POLARIZED INITIAL OPINIONS

Intuitively, the evolution of voters' opinions and voting outcomes under the antagonism effect of external strategic campaigns should depend on the initial distribution of two-dimensional partisan opinions. In the main text, voters' initial partisan opinions are randomly drawn from a uniform distribution, and all results are averaged over multiple independent realizations with different random seeds. In this part, we conduct simulations under structured polarized initial conditions to examine the robustness of out results with respect to nonuniform opinion initialization. Specifically, we consider two polarized initialization schemes, both characterized by bimodal opinion distributions.

In each scheme, the initial opinions $(x_i(0), y_i(0))$ are drawn independently from bimodal probability density functions given by

$$\begin{aligned} f(x_i(0)) &= 0.5 \times \mathrm{Norm}(x_i(0), \mu_1, \sigma_1) \\ &\quad + 0.5 \times \mathrm{Norm}(x_i(0), \mu_2, \sigma_2), \\ f(y_i(0)) &= 0.5 \times \mathrm{Norm}(\mathrm{y_i}(0), \mu_1', \sigma_1') \\ &\quad + 0.5 \times \mathrm{Norm}(\mathrm{y_i}(0), \mu_2', \sigma_2'), \end{aligned} \tag{F1}$$

where $\mathrm{Norm}(x_i(0), \mu, \sigma)$ reflects the PDF of the normal distribution with mean μ and standard deviation σ, evaluated at value $x_i(0)$. In addition, $x_i(0), y_i(0)$ are truncated to the interval [0,1].

We consider two types of polarized distributions. In the symmetric condition, voters' opinions toward the two parties are polarized by statistically unbiased, with parameter $\mu_1 = \mu_1' = 0.2, \sigma_1 = \sigma_1' = 0.2$ and $\mu_2 = \mu_2' = 0.8, \sigma_2 = \sigma_2' = 0.2$. In the asymmetric condition, the bimodal distributions are conducted to introduce an initial partisan advantage toward Party A: Opinions favoring Party A are more concentrated on the supportive side along the x dimension, while opinions toward Party B are more concentrated on the opposing side along the y dimension. Specifically, we set $\mu_1 = 0.2, \mu_1' = 0.15, \sigma_1 = 0.2, \sigma_1' = 0.16$ and $\mu_2 = 0.85, \mu_2' = 0.8, \sigma_2 = 0.16, \sigma_2' = 0.2$.

We first examine the coupling effect between bottom-up voter stubbornness and top-down antagonistic campaigns under polarized initial conditions, which are shown in Figs. 12(b) and 12(e) for symmetric and asymmetric polarized initializations, respectively. Compared with the corresponding results obtained under uniform random initialization [Fig. 4(b)], the qualitative structure of the nonlinear coupling between λ and ρ remains unchanged.

In particular, for both polarized initial conditions, polarization of partisan opinion emerges over a broader region of $(\lambda - \rho)$ space, with the polarization boundary occurring at lower values of ρ. Voters' initially polarized opinions and their stubbornness toward initial beliefs together facilitate the persistence and amplification of partisan attitudes. Importantly, the phase structure and coupling pattern remain consistent with those observed under uniform initialization, thereby confirming the robustness of the model with respect to nonuniform initial opinion distributions.

We next turn to the most critical issue: the formation and possible reversal of voting advantage under heterogeneous antagonistic effect. Under symmetric polarized initial conditions, Fig. 12(c) reveals results that are nearly identical to those reported in Fig. 5. At low and intermediate antagonism levels, the party adopting stronger strategic campaigns gains a clear voting advantage. However, when both parties impose extremely strong antagonistic effects, the disadvantaged party may attract a larger fraction of swing voters, leading a reversal of voting advantage.

Under asymmetric polarized initial conditions, the reversal phenomenon exhibits a fundamentally asymmetric character when both parties adopt extremely strong antagonistic strategies. In contrast to the symmetric initialization case, the system transitions to a regime of complete dominance by Party A, as shown in Fig. 12(f). Specifically, when both parties exert strong antagonism, Party A—endowed with a first-mover advantage—persistently maintains its voting advantage, while the initially disadvantaged Party B is no longer able to overturn the electoral outcome through relatively moderate antagonistic strategies. Party B can only obtain a marginal voting advantage within a narrow parameter regime where Party A applies relatively weak antagonism and Party B adopts a much stronger one. These results indicate that the nonuniform initialization and first-mover partisan advantage exert a strongly asymmetric influence on electoral outcomes in extreme antagonism regimes. In particular, initial asymmetry suppresses the reversal of the disadvantaged party by preventing the dominant party from becoming disconnected from swing voters when antagonism intensifies.

More broadly, these findings highlight the important role of voters' stubbornness toward initial opinions in shaping long-term opinion evolution and electoral outcomes. While reversal effects remain robust under symmetric polarized initial conditions, sufficiently strong asymmetry in initial partisan support can fundamentally alter the system's collective response to antagonistic campaigning.

APPENDIX G: TWITTER DATA PROCESSING

All original Twitter datasets are publicly available as Twitter IDs [56]. We first download the tweets using

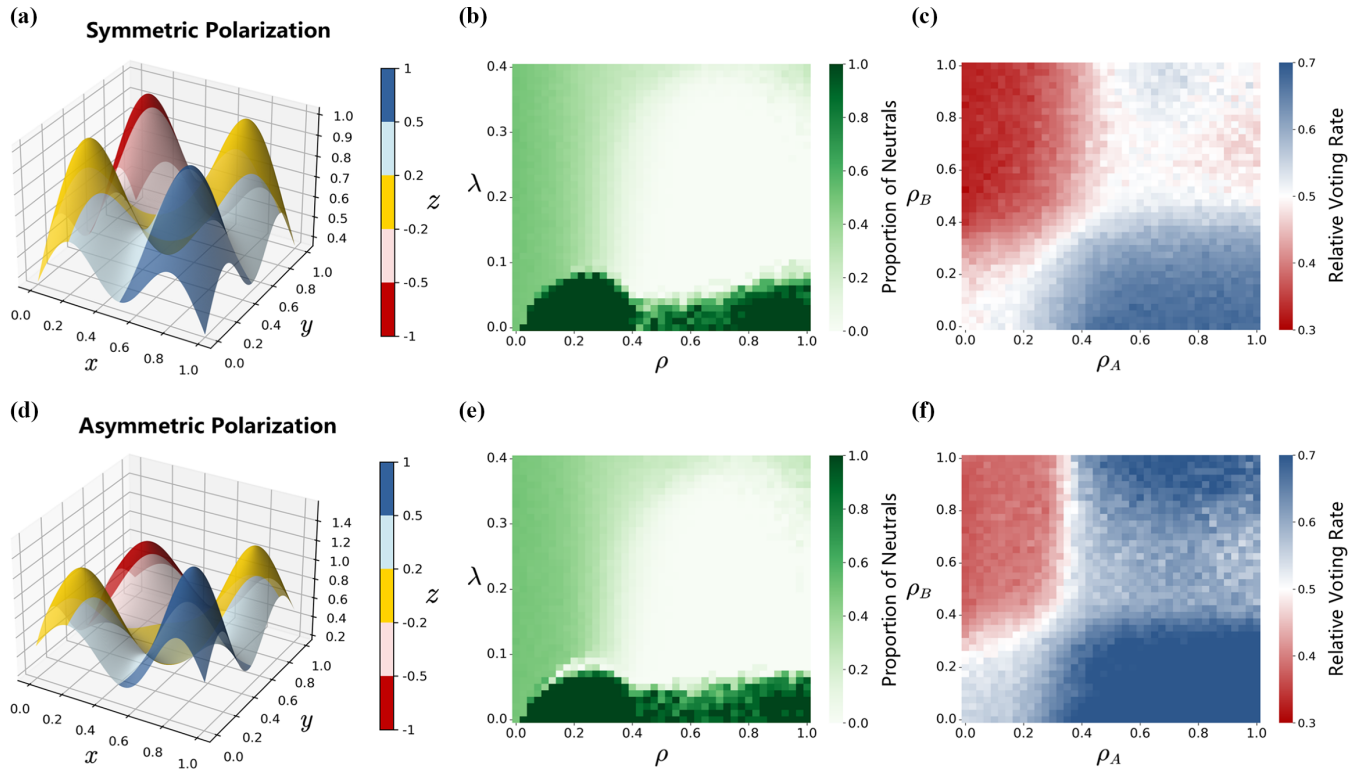

FIG. 12. Effects of nonlinear coupling between stubbornness and antagonism levels (b), (e) and heterogeneous antagonism levels (c), (f) under symmetric and asymmetric polarized initial opinions. Symmetric polarized initialization preserves the reversal phenomenon under strong antagonism, whereas asymmetric initialization with initial partisan advantage locks the system into persistent dominance by the initially advantaged party. Parameters: simulations begin from networks with $N = 10^4$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$. We fix (c), (f) $\lambda = 0.2$.

Twitter's API and obtain user ID (nodes), retweeting relationship (edges), and political orientation of tweets by month. Due to the computational constraints, we do not simulate the full network. Instead, we filter the networks with several keywords (e.g., vaccine, abortion, and covid). For visualization purposes, network layouts are generated using the Force Atlas 2 algorithm implemented in GEPHI, with node sizes weighted by their retweet frequencies. To extract the core structure of the network for our analysis, we apply the k-core algorithm to remove nodes with fewer retweet relationships, which leads to a more simplified network. Furthermore, by quantifying the political leanings reflected in partial retweets, we infer the political preferences of the core nodes, thereby characterizing partisan orientation to each community.

All the network visualizations in this paper are generated using the Force Atlas 2 layout algorithm implemented in GEPHI. The color of each node is determined by the voting outcomes obtained from model simulations.

APPENDIX H: MODEL ROBUSTNESS TO NETWORK TOPOLOGY

In the main text, the model network is initialized as an ER random graph with $N = 10^4$ and connection probability $p = 0.004$, corresponding to an average degree of $\langle k \rangle = 40$. To examine the robustness of our results with respect to network topology, we further consider ER networks with $p = 0.002$ ($\langle k \rangle = 20$) and $p = 0.008$ ($\langle k \rangle = 80$), as well as a Barabási-Albert (BA) network with $m = 40$ ($\langle k \rangle = 40$) and a Watts-Strogatz (WS) network with $k = 40$ and rewiring probability $p = 0.1$ ($\langle k \rangle = 40$). The results in Fig. 13 indicate that, despite the variations in the network topology and the average degree, the qualitative dynamics and macroscale results remain invariant.

The average degree directly determines the overall connectivity of the network, thereby influencing the threshold conditions for partisan polarization. Within the same ER topology, varying the average degree reveals that higher connectivity leads to a broader onset of polarization, as evidenced by the shift of the polarization boundary in Figs. 13(a) and 13(c). When polarization occurs under moderate and high stubbornness, the antagonism effect on promoting the voting participation of swing voters remains comparable. In cases where the two parties engage in competitive antagonism, no significant difference is observed in the resulting voting outcomes in Figs. 13(b) and 13(d).

We further verify that these outcomes still hold across heterogeneous and clustered network structures by replacing the ER topology with BA and WS networks. In particular, under the BA network with the same average degree as the ER network, we observe the nonlinear coupling pattern in Fig. 13(e), almost identical to that shown in Fig. 4(b). This consistency strongly demonstrates the model robustness against variations in the network topology. In contrast, the $\lambda - \rho$ phase diagram of the WS network in Fig. 13(g) is similar to that of the

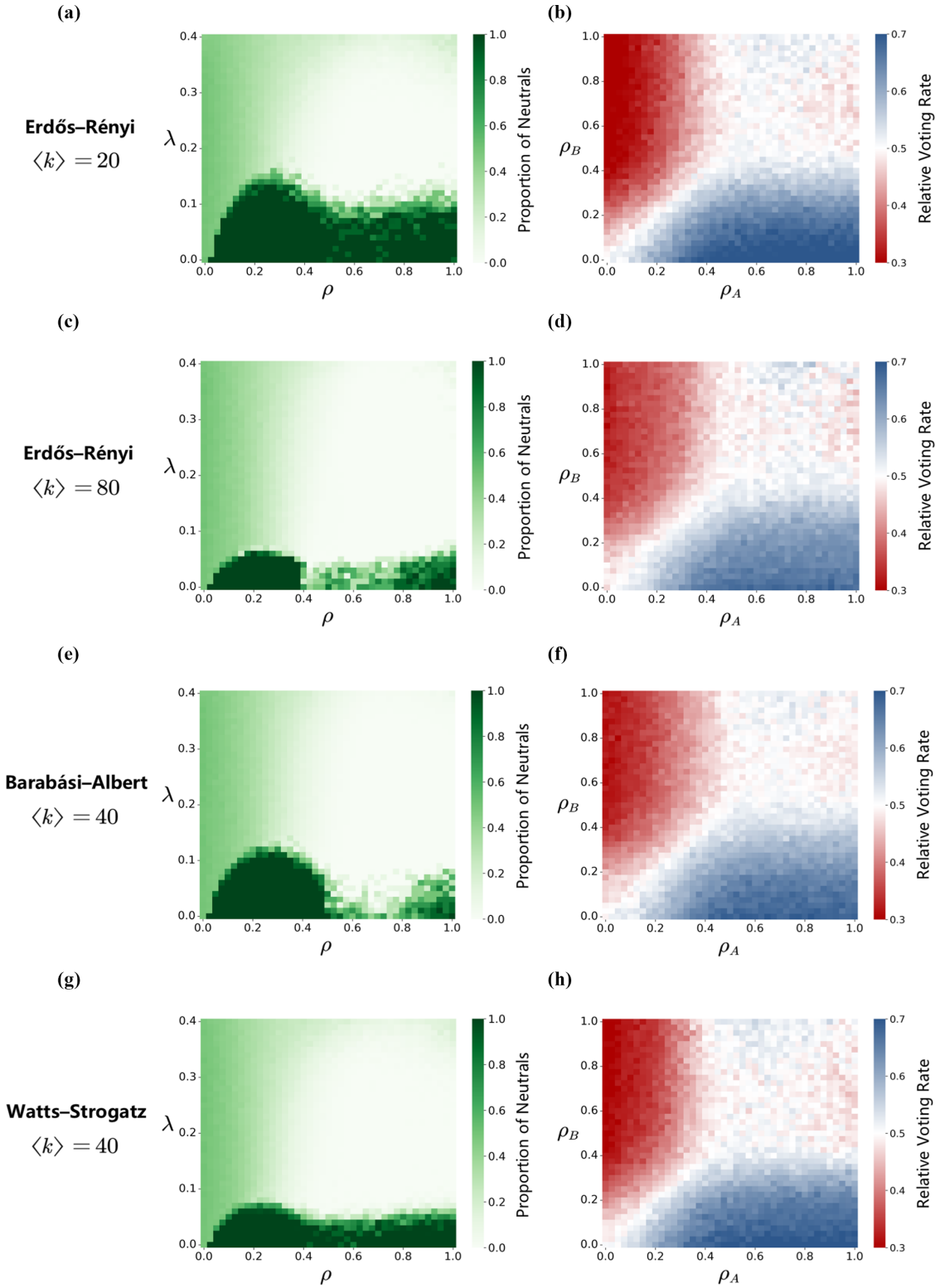

FIG. 13. Effects of nonlinear coupling between stubbornness and antagonism levels and heterogeneous antagonism levels under four different network topologies. We run simulations under ER random networks with $\langle k\rangle = 20$ (a), (b) and $\langle k\rangle = 80$ (c), (d), under a BA network with $\langle k\rangle = 40$ (e), (f) and under a WS network with $\langle k\rangle = 40$ (g), (h), respectively. Results largely resemble findings from the original setting. Simulation results are averaged over 50 independent runs. Parameters: simulations begin from networks with $N = 10^4$, $z_o = 0.5$, $z_v = 0.2$, and $\epsilon = 0.2$. We fix (b), (d), (f), (h) $\lambda = 0.2$.

ER network with $\langle k \rangle = 80$. Furthermore, we find that the qualitative establishment and reversal of voting advantages under antagonistic strategies remain robust, with only slight differences in the extent of reversal, as shown in Figs. 13(f) and 13(h).

These results demonstrate that the core dynamics of the model are structurally robust across both homogeneous and heterogeneous topologies, with only slight quantitative shifts under different degree distributions and clustering coefficients.

[1] E. Kubin and C. Von Sikorski, The role of (social) media in political polarization: A systematic review, Ann. Int. Commun. Assoc. **45**, 188 (2021).

[2] J. Flamino, A. Galeazzi, S. Feldman, M. W. Macy, B. Cross, Z. Zhou, M. Serafino, A. Bovet, H. A. Makse, and B. K. Szymanski, Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections, Nat. Hum. Behav. **7**, 904 (2023).

[3] S. González-Bailón, D. Lazer, P. Barberá, M. Zhang, H. Allcott, T. Brown, A. Crespo-Tenorio, D. Freelon, M. Gentzkow, A. M. Guess, *et al.*, Asymmetric ideological segregation in exposure to political news on Facebook, Science **381**, 392 (2023).

[4] W. Cota, S. C. Ferreira, R. Pastor-Satorras, and M. Starnini, Quantifying echo chamber effects in information spreading over political communication networks, EPJ Data Sci. **8**, 35 (2019).

[5] S. Jungkunz, Political polarization during the COVID-19 pandemic, Front. Polit. Sci. **3**, 622512 (2021).

[6] J. Kerr, C. Panagopoulos, and S. Van Der Linden, Political polarization on COVID-19 pandemic response in the United States, Person. Individ. Differ. **179**, 110892 (2021).

[7] N. F. Johnson, N. Velásquez, N. J. Restrepo, R. Leahy, N. Gabriel, S. El Oud, M. Zheng, P. Manrique, S. Wuchty, and Y. Lupu, The online competition between pro-and anti-vaccination views, Nature (London) **582**, 230 (2020).

[8] M. Falkenberg, A. Galeazzi, M. Torricelli, N. Di Marco, F. Larosa, M. Sas, A. Mekacher, W. Pearce, F. Zollo, W. Quattrociocchi, *et al.*, Growing polarization around climate change on social media, Nat. Clim. Change **12**, 1114 (2022).

[9] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, and M. Starnini, The echo chamber effect on social media, Proc. Natl. Acad. Sci. USA **118**, e2023301118 (2021).

[10] R. Interian, R. G. Marzo, I. Mendoza, and C. C. Ribeiro, Network polarization, filter bubbles, and echo chambers: An annotated review of measures and reduction methods, Int. Trans. Oper. Res. **30**, 3122 (2023).

[11] C. K. Tokita, A. M. Guess, and C. E. Tarnita, Polarized information ecosystems can reorganize social networks via information cascades, Proc. Natl. Acad. Sci. USA **118**, e2102147118 (2021).

[12] F. P. Santos, Y. Lelkes, and S. A. Levin, Link recommendation algorithms and dynamics of polarization in online social networks, Proc. Natl. Acad. Sci. USA **118**, e2102141118 (2021).

[13] S. Vosoughi, D. Roy, and S. Aral, The spread of true and false news online, Science **359**, 1146 (2018).

[14] N. Grinberg, K. Joseph, L. Friedland, B. Swire-Thompson, and D. Lazer, Fake news on Twitter during the 2016 US presidential election, Science **363**, 374 (2019).

[15] W. G. Mayer, *The Swing Voter in American Politics* (Rowman & Littlefield, Washington, DC, USA, 2008).

[16] S. Kosmidis and G. Xezonakis, The undecided voters and the economy: Campaign heterogeneity in the 2005 British general election, Elect. Stud. **29**, 604 (2010).

[17] E. Mahieux, An investigation into the psychological, cognitive and neural correlates of swing voting, Ph.D. thesis, UCL, University College London, 2024.

[18] A. Mathur and G. P. Moschis, How do information sources shape voters' political views? Comparing mainstream and social-media effects on democrats, republicans, and the undecided, J. Advert. Res. **62**, 176 (2022).

[19] D. Schill and R. Kirk, Angry, passionate, and divided: Undecided voters and the 2016 presidential election, Am. Behav. Sci. **61**, 1056 (2017).

[20] G. W. Cox, 13 - swing voters, core voters, and distributive politics, in *Political Representation*, edited by I. Shapiro, S. C. Stokes, E. J. Wood, and A. S. Kirshner (Cambridge University Press, 2009), pp. 342–357.

[21] E. Throsby, The deciders are undecided: Undecided voters, election campaigns, political media, and democracy in Australia, Ph.D. thesis, UNSW Sydney, 2018.

[22] A. F. Siegenfeld and Y. Bar-Yam, Negative representation and instability in democratic elections, Nat. Phys. **16**, 186 (2020).

[23] M. Pratelli, M. Petrocchi, F. Saracco, and R. De Nicola, Online disinformation in the 2020 US election: Swing vs safe states, EPJ Data Sci. **13**, 25 (2024).

[24] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Emergence of polarized ideological opinions in multidimensional topic spaces, Phys. Rev. X **11**, 011012 (2021).

[25] R. Kwon, W. J. Scarborough, and R. Gallardo, Configurations of attitudes toward immigration in Europe: Evidence of polarization, ambivalence, and multidimensionality, Comp. Migr. Stud. **12**, 25 (2024).

[26] E. Andrade, G. Seoane, L. Velay, and J.-M. Sabucedo, Multidimensional model of environmental attitudes: Evidence supporting an abbreviated measure in Spanish, Int. J. Environ. Res. Public Healths **18**, 4438 (2021).

[27] A. F. Peralta, P. Ramaciotti, J. Kertész, and G. Iñiguez, Multidimensional political polarization in online social networks, Phys. Rev. Res. **6**, 013170 (2024).

[28] S. Feldman and C. Johnston, Understanding the determinants of political ideology: Implications of structural complexity, Polit. Psychol. **35**, 337 (2014).

[29] S. Walgrave and J. Lefevere, Ideology, salience, and complexity: Determinants of policy issue incongruence between voters and parties, J. Elect., Public Opin. Parties **23**, 456 (2013).

[30] R. Hamill, M. Lodge, and F. Blake, The breadth, depth, and utility of class, partisan, and ideological schemata, Am. J. Polit. Sci. **29** 850 (1985).

[31] X. Wang, A. D. Sirianni, S. Tang, Z. Zheng, and F. Fu, Public discourse and social network echo chambers driven by sociocognitive biases, Phys. Rev. X **10**, 041042 (2020).

[32] M. Galesic and D. L. Stein, Statistical physics models of belief dynamics: Theory and empirical tests, Physica A **519**, 275 (2019).
[33] A. Nowak and R. R. Vallacher, Nonlinear societal change: The perspective of dynamical systems, British J. Soc. Psychol. **58**, 105 (2019).
[34] A. Baronchelli, The emergence of consensus: A primer, R. Soc. Open Sci. **5**, 172189 (2018).
[35] M. H. DeGroot, Reaching a consensus, J. Am. Stat. Assoc. **69**, 118 (1974).
[36] J. R. French Jr, A formal theory of social power, Psychol. Rev. **63**, 181 (1956).
[37] N. E. Friedkin and E. C. Johnsen, Social influence and opinions, J. Math. Sociol. **15**, 193 (1990).
[38] G. Deffuant, D. Neau, F. Amblard, and G. Weisbuch, Mixing beliefs among interacting agents, Adv. Complex Syst. **03**, 87 (2000).
[39] F. Baumann, P. Lorenz-Spreen, I. M. Sokolov, and M. Starnini, Modeling echo chambers and polarization dynamics in social networks, Phys. Rev. Lett. **124**, 048301 (2020).
[40] S. K. Rahmawati, The dynamics of binary oppositions in Arundhati Roy's The God of Small Things, Ph.D. thesis, Universitas Islam Negeri Maulana Malik Ibrahim, Malang, Indonesia, 2009.
[41] T. Herdin, Deconstructing typologies: Overcoming the limitations of the binary opposition paradigm, Int. Commun. Gazette **74**, 603 (2012).
[42] S. Martinek, Right and left or binary opposition as a cognitive mechanism, *Further Insights into Semantics and Lexicography*, edited by U. Magnusson K. Henryk, and G. Adam (Lublin, Wydawnictwo, UMCS), pp. 191–205.
[43] J. E. Rothschild, Elites and identities: The interactive effects of top-down cues and group memberships on political attitudes, Ph.D. thesis, Northwestern University, Evanston, Illinois, USA, 2020.
[44] Y. Li, X. Qin, A. Sullivan, G. Chi, Z. Lu, W. Pan, and Y. Liu, Collective action improves elite-driven governance in rural development within China, Humanit. Soc. Sci. Commun. **10**, 600 (2023).
[45] P. Liu, K. Shivaram, A. Culotta, M. A. Shapiro, and M. Bilgic, The interaction between political typology and filter bubbles in news recommendation algorithms, in *Proceedings of the Web Conference* (ACM, New York, NY, USA, 2021), pp. 3791–3801.
[46] J. Cho, S. Ahmed, M. Hilbert, B. Liu, and J. Luu, Do search algorithms endanger democracy? An experimental investigation of algorithm effects on political polarization, J. Broadcas. Electronic Media **64**, 150 (2020).
[47] H. Ibrahim, N. AlDahoul, S. Lee, T. Rahwan, and Y. Zaki, YouTube's recommendation algorithm is left-leaning in the United States, PNAS Nexus **2**, pgad264 (2023).
[48] R. M. Entman, Media framing biases and political power: Explaining slant in news of campaign 2008, Journalism **11**, 389 (2010).
[49] V. Pansanella, V. Morini, T. Squartini, and G. Rossetti, Change my mind: Data driven estimate of open-mindedness from political discussions, in *International Conference on Complex Networks and Their Applications* (Springer, Cham, Switzerland, 2022), pp. 86–97.
[50] H. Schawe and L. Hernández, When open mindedness hinders consensus, Sci. Rep. **10**, 8273 (2020).
[51] S. Y. Dolbier, M. C. Dieffenbach, and M. D. Lieberman, Open-mindedness: An integrative review of interventions, Psychol. Rev. **132**, 204 (2024).
[52] J. Ghaderi and R. Srikant, Opinion dynamics in social networks with stubborn agents: Equilibrium and convergence rate, Automatica **50**, 3209 (2014).
[53] W. G. Mayer, The swing voter in American presidential elections, Am. Polit. Res. **35**, 358 (2007).
[54] S. J. Hill, Changing votes or changing voters? How candidates and election context swing voters and mobilize the base, Elect. Stud. **48**, 131 (2017).
[55] C. Chang and C.-l. Wu, Active vs passive ambivalent voters: Implications for interactive political communication and participation, Commun. Res. **50**, 828 (2023).
[56] Z. Liu, https://github.com/ziqianliu2001-hub/Emergence-Evolution-and-Manipulation-of-Swing-Voters-in-Presidential-Election, Github (2025).
[57] R. A. Davis, K.-S. Lii, and D. N. Politis, Remarks on some nonparametric estimates of a density function, in *Selected Works of Murray Rosenblatt* (Springer, New York, USA, 2011), pp. 95–100.
[58] B. W. Silverman, *Density Estimation for Statistics and Data Analysis* (Routledge, Boca Raton, FL, 2018).
[59] D. W. Scott, *Multivariate Density Estimation: Theory, Practice, and Visualization* (John Wiley & Sons, Hoboken, NJ, 2015).
[60] C. S. Haynes and T. J. Hayes, Cognitive load and candidate evaluation: Evidence from an experimental design, in *Proceedings of the APSA 2010 Annual Meeting Paper* (Elsevier, Amsterdam, 2010).
[61] J. Metag and G. Gurr, Too much information? A longitudinal analysis of information overload and avoidance of referendum information prior to voting day, Journal. Mass Commun. Q. **100**, 646 (2023).
[62] P. Törnberg, How digital media drive affective polarization through partisan sorting, Proc. Natl. Acad. Sci. USA **119**, e2207159119 (2022).
[63] A. Antelmi, L. La Cava, and A. Pera, Characterizing swing voters in online social media: The case of the 2022 Italian elections, Soc. Netw. Anal. Min. **15**, 77 (2025).
[64] S. Aral and D. Eckles, Protecting elections from social media manipulation, Science **365**, 858 (2019).