

Article

Opinion Dynamics in Social Networks with Edge-Heterogeneous Confidence Bounds: Clustering, Polarization, and Implications for Online Platforms

Zumei Huang ¹, Zhuangzhuang Ma ^{2,*} , Lei Shi ² and Lulu Chen ³

¹ School of Finance, Henan Finance University, Zhengzhou 451464, China; huangzumei@hafu.edu.cn

² School of Artificial Intelligence, Henan University, Zhengzhou 450046, China; shilei910918@henu.edu.cn

³ School of Mechanical Engineering, Xihua University, Chengdu 610039, China; chenlulu@xhu.edu.cn

* Correspondence: zhuangma912@henu.edu.cn

Abstract

Opinion polarization, echo chambers, and the rapid formation of opinion clusters have become defining features of debates on contemporary online social platforms. To explain these phenomena from a control-theoretic perspective, this paper investigates opinion dynamics in social networks with edge-heterogeneous confidence bounds, focusing on clustering and polarization behaviors driven by pair-dependent trust and asymmetric influence. Two discrete-time models are proposed, including an unsigned bounded-confidence model and a more general signed model that incorporates both supportive and oppositional interactions. The interaction structures are described by time-varying unsigned and signed digraphs, respectively, in which heterogeneous interpersonal influence is characterized by edge-dependent confidence bounds that naturally encode platform-mediated trust. For the proposed models, rigorous sufficient conditions are established for invariant cluster consensus and structurally balanced polarization. Numerical simulations, including a case study on the Slashdot Zoo signed social network with 50 controversial users, illustrate the theoretical results and demonstrate their relevance for understanding opinion evolution on internet-scale platforms.

Keywords: opinion dynamics; bounded confidence; online social networks; network polarization; signed digraphs; edge heterogeneity

1. Introduction

The rapid growth of online social platforms such as Twitter (X), Reddit, Facebook, and Weibo has transformed how public opinions are formed, exchanged, and amplified. On these platforms, algorithmic recommendation, follower-based information delivery, and platform-driven content curation jointly shape the topology of interpersonal influence, often producing pronounced phenomena such as filter bubbles, echo chambers, and large-scale opinion polarization [1–3]. Understanding how heterogeneous and pair-dependent interpersonal trust drives the emergence of these collective patterns is therefore not only a fundamental scientific question, but also a practical concern for online community governance, misinformation containment, and the design of socially responsible recommendation systems. Motivated by these challenges in the modern internet ecosystem, this paper proposes an opinion-dynamics framework with edge-heterogeneous confidence bounds that captures the asymmetric and pair-dependent nature of trust on online platforms.



Academic Editors: Francesco La Barbera and Roberta Verso

Received: 2 June 2026

Revised: 3 July 2026

Accepted: 8 July 2026

Published: 9 July 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Over the past decade, opinion dynamics has received increasing attention at the intersection of network science, control theory, and social psychology. It offers a mathematical framework for analyzing how repeated interpersonal influence shapes the evolution of individual opinions and gives rise to collective phenomena in social networks [4,5]. Among various approaches, bounded-confidence models are particularly appealing because they incorporate a realistic feature of social interaction, namely that individuals tend to exchange opinions only with those whose views are sufficiently close to their own. This mechanism, typically represented by the Hegselmann–Krause (HK) model [6–8], combines clear social interpretability with rich nonlinear dynamics. Owing to the co-evolution of opinions and interaction topology, bounded-confidence models provide a natural framework for studying consensus, clustering and polarization behaviors in social networks.

The literature on bounded-confidence dynamics has grown rapidly in recent years. A substantial body of work has focused on fundamental properties of discrete-time and continuous-time models, including convergence, stability, and asymptotic behavior [1,9–14]. Other studies have broadened the classical framework by incorporating random pairwise communication, probabilistic influence, noise, and asynchronous updating [15–19]. Meanwhile, bounded-confidence ideas have been used to model a range of application-driven scenarios, such as leader–follower interactions, rumor propagation, political competition, and manipulative agents [2,3,20–22]. Despite this broad development, the majority of existing models adopt simplified confidence mechanisms, typically assuming a common confidence bound for all interactions or confidence parameters associated only with agents rather than edges. As a result, they are not well suited to describing pair-dependent influence patterns that are common in real social networks, where interpersonal trust, receptiveness, and even antagonism may vary from one relationship to another.

Another important direction in opinion dynamics is the incorporation of antagonistic and distrustful interactions. In realistic social settings, individuals may respond to discrepant opinions not only by disengaging from them, but also by resisting, rejecting, or reacting against them. Therefore, models based solely on attractive influence may be insufficient for describing opinion evolution in polarized or low-trust environments. This consideration motivates signed bounded-confidence dynamics, in which both supportive and oppositional interactions are taken into account [23,24]. Similar motivations can be found in recent studies on cognitive dissonance, echo chambers, and strategic influence, where the distinction between positive and negative interpersonal effects is essential for explaining the emergence of polarization and persistent disagreement [1–3].

Despite the substantial progress in the literature, most existing bounded-confidence models still rely on homogeneous or node-heterogeneous confidence mechanisms. Such formulations are often insufficient for capturing the pair-dependent nature of interpersonal trust in real social systems. In practice, individuals do not evaluate all neighbors in the same way: one may be more willing to accommodate a close friend, a family member, an expert, or a highly visible media account than a stranger, even when the expressed opinions are equally distant. Social judgment theory provides a natural behavioral explanation for this phenomenon by emphasizing that opinion acceptance depends jointly on message discrepancy and interpersonal relationship [25]. This pair-dependent mechanism is particularly relevant to online environments, where a user’s willingness to accept dissonant content often depends on the source’s identity (e.g., verified account, close friend, anonymous stranger, or institutional media) rather than purely on opinion proximity. Therefore, introducing edge-heterogeneous confidence bounds leads to a more realistic description of interpersonal influence in online discussions, political communication, and other social interaction networks.

This edge-level perspective also clarifies how the present framework departs from the heterogeneity already studied in the literature. In the homogeneous setting, a single confidence bound governs all interactions, whereas in the node-heterogeneous setting, each agent carries one bound that it applies to every neighbor alike [9,10,13,14]. Both make trust intrinsically symmetric: whether one agent listens to another is decided by a single threshold attached to the listener, regardless of who the speaker is. As a result, several behaviorally important situations cannot be represented. The first is one-way trust, in which one agent accepts another while the latter rejects the first at exactly the same opinion gap, as when an ordinary user follows a verified influencer; capturing this requires the tolerance an agent grants to a given partner to differ from the tolerance it receives in return. The second is relationship-specific tolerance, in which the same agent extends a wide bound to a trusted source but only a narrow bound to a stranger holding an equally distant opinion. The third is the persistently asymmetric interaction graph that such one-way relationships produce, whose lack of a symmetric weighting removes the doubly stochastic structure on which most convergence arguments for node-heterogeneous models rely. Allowing the confidence bound to depend on the ordered pair of interacting agents is the minimal extension that captures all three, and the analysis developed below is precisely what is needed to handle the asymmetric, discontinuous switching dynamics it produces.

The introduction of edge heterogeneity also gives rise to significant mathematical difficulties. When confidence bounds are pair-dependent, the interaction topology becomes not only state-dependent but also generally asymmetric at every time step. The resulting dynamics form a discontinuous switching system, whose asymptotic behavior depends delicately on both the evolving opinion state and the detailed edge-specific trust structure. As emphasized in the bounded-confidence literature, discontinuous confidence mechanisms must be treated with particular care, because continuous relaxations may fail to reflect the true final behavior of the original system [26]. These difficulties are especially pronounced in the analysis of fully asymmetric models.

Motivated by the above observations, this paper investigates two edge-heterogeneous bounded-confidence models, including an unsigned bounded-confidence model and a more general signed model. Furthermore, we establish rigorous sufficient conditions for cluster consensus, and complement the theoretical results with numerical experiments, including a case study on the well-known Slashdot Zoo signed online social network. The main contributions of this paper are summarized as follows.

1. Two edge-heterogeneous bounded-confidence models are proposed, in which the confidence bound is allowed to depend on each ordered pair of interacting agents. Compared with the homogeneous and node-heterogeneous settings commonly adopted in the literature, the proposed models provide a more flexible framework for capturing pair-dependent interpersonal trust in online environments.
2. Rigorous sufficient conditions for cluster behavior of the proposed models are established. For the edge-heterogeneous bounded-confidence model, once the opinions evolve into narrow and well-separated groups, the induced partition becomes invariant and each group reaches an internal consensus. For the signed model, a sufficient condition is derived under which the interaction pattern remains structurally balanced and the network converges to a polarized state, providing a mechanism-level explanation for echo chambers on adversarial online platforms.
3. Numerical examples are constructed for both synthetic and real-world signed networks. The synthetic experiments illustrate how randomized edge-dependent confidence bounds give rise to terminal clustering and polarization. A case study on the Slashdot Zoo dataset further illustrates the relevance of the proposed framework for understanding opinion evolution on real internet-scale signed networks.

The remainder of the paper is organized as follows. Section 2 reviews the required notation from linear algebra and graph theory. Section 3 introduces the edge-heterogeneous attractive and signed models. Section 4 presents the rigorous sufficient conditions. Section 5 discusses the implications for online social platforms. Section 6 provides five numerical examples, including the Slashdot case study. Section 8 concludes the paper.

2. Background Material

2.1. Linear Algebraic Notions

Denote the set of natural numbers and the set of real numbers by $\mathbb{N}$ and $\mathbb{R}$, respectively. Let $\mathbb{R}^{n \times m}$ be the set of all real $(n \times m)$ matrices. $I_n \in \mathbb{R}^{n \times n}$ and $\mathbf{1}_n \in \mathbb{R}^{n \times 1}$ denote the identity matrix and the column vector of all ones. The matrix of all zeros is denoted by $\mathbf{0}$. For a matrix $A \in \mathbb{R}^{n \times n}$, the (i, j) -th entry is denoted by A_{ij} . A matrix A is nonnegative (denoted by $A \geq 0$) if all its entries are nonnegative real numbers, and it is positive (denoted by $A > 0$) if all its entries are positive real numbers. A nonnegative matrix $A \in \mathbb{R}^{n \times n}$ is row-stochastic if $\sum_{j=1}^n a_{ij} = 1, i = 1, 2, \dots, n$. The number of elements in a set S is denoted by $|S|$.

2.2. Graph Theory

Given a real matrix $A \in \mathbb{R}^{n \times n}$, associate with it the digraph $\mathcal{G}[A] = \{\mathcal{V}, \mathcal{E}[A], A\}$, where $\mathcal{V} = \{v_1, v_2, \dots, v_n\}$ is the node set, $\mathcal{E}[A] \subseteq \mathcal{V} \times \mathcal{V}$ is the edge set, and A is the adjacency matrix. In particular, $\mathcal{G}[A]$ is called a *signed digraph* if matrix A contains both positive and negative entries, and it is called an *unsigned digraph* if A is nonnegative. An edge (v_j, v_i) is incoming to node v_i and outgoing from node v_j . Self-loops are excluded in $\mathcal{E}[A]$, i.e., $(v_i, v_i) \notin \mathcal{E}[A], v_i \in \mathcal{V}$. We have $A_{ij} \neq 0$ if and only if $(v_j, v_i) \in \mathcal{E}[A]$; otherwise $A_{ij} = 0$. In particular, $A_{ij} = 1$ and $A_{ij} = -1$ if the edge (v_j, v_i) has positive and negative weight, respectively. The degree matrix is denoted by D , in which $D_{ii} = \sum_{j=1}^n |A_{ij}|, i = 1, 2, \dots, n$, and $D_{ij} = 0, i \neq j$.

An alternating sequence of nodes and edges is called a *walk*, i.e., $v_1, e_1, v_2, e_2, \dots, v_{s-1}, e_{s-1}, v_s$, where $e_i = (v_i, v_{i+1}), i = 1, 2, \dots, s-1$. If there are no repeated edges in a walk, then the walk is further called a *path*. For convenience, the path from node v_j to node v_i is denoted by $\mathcal{P}(v_j \rightarrow v_i)$. A signed digraph is *strongly connected* if for any nodes $v_i \neq v_j$, there exist two paths $\mathcal{P}(v_j \rightarrow v_i)$ and $\mathcal{P}(v_i \rightarrow v_j)$. It is said that $\mathcal{G}[A]$ is a complete signed digraph if there exist edges $(v_i, v_j), (v_j, v_i) \in \mathcal{E}[A]$ for any pair of nodes v_i, v_j .

Structural balance is a basic concept in the study of signed graphs, motivated by social, interpersonal, and economic networks. A signed digraph $\mathcal{G}[A]$ is structurally balanced if there exists a partition of the node set into $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$, where $\mathcal{V}_1$ and $\mathcal{V}_2$ are nonempty and disjoint, such that for any $(v_j, v_i) \in \mathcal{E}[A]$, $A_{ij} > 0$ if v_i and v_j belong to the same subset, and $A_{ij} < 0$ if v_i and v_j belong to different subsets.

3. Model Formulation

In this section, we first briefly review the node-heterogeneous bounded-confidence model available in the literature. Based on this framework, we formulate two edge-heterogeneous models with pair-dependent confidence bounds.

3.1. Existing Node-Heterogeneous Bounded-Confidence Model

Consider a group of agents discussing the same issue. At each discrete-time instant t , agent v_i holds a scalar opinion $x_i(t) \in \mathbb{R}$. The opinion is one-dimensional, representing a position on a single issue along a continuous axis; the scale is normalized so that the unsigned model evolves on $[0, 1]$ and the signed model on the symmetric interval $[-1, 1]$.

centered at the neutral opinion 0. In bounded-confidence dynamics, an agent is influenced only by those agents whose opinions are sufficiently close to its own.

Building on the classical Hegselmann–Krause (HK) framework [6], we adopt an HK-type fixed-step relaxation in which an agent moves toward the agents in its confidence set with a common interpersonal weight τ . The update of agent i 's opinion follows the protocol

$$x_i(t+1) = x_i(t) + \tau \sum_{v_j \in \mathcal{I}_i(t)} (x_j(t) - x_i(t)), \quad (1)$$

where $\tau \in (0, 1/n]$ is a constant interpersonal influence weight, and $\mathcal{I}_i(t)$ denotes the confidence set of agent v_i , defined by

$$\mathcal{I}_i(t) = \{v_j \in \mathcal{V} : |x_j(t) - x_i(t)| < \varepsilon_i, j \neq i\}. \quad (2)$$

Equation (1) can be rewritten as

$$x_i(t+1) = (1 - \tau|\mathcal{I}_i(t)|)x_i(t) + \tau \sum_{v_j \in \mathcal{I}_i(t)} x_j(t),$$

which shows that each neighbor receives the same weight τ , whereas the self-weight is given by $1 - \tau|\mathcal{I}_i(t)|$. Therefore, (1) does not coincide with the classical HK averaging rule unless $\tau = 1/(1 + |\mathcal{I}_i(t)|)$; instead, it should be regarded as a fixed-step HK-type relaxation. When all agents share the same confidence bound, namely $\varepsilon_i = \varepsilon$ for all v_i , model (1) reduces to a homogeneous bounded-confidence dynamic, which has been widely studied in the literature [9,10,15–17,20,27].

3.2. Edge-Heterogeneous Bounded-Confidence Model

In the node-heterogeneous model, each agent v_i uses a single confidence bound for all other agents. We now generalize this mechanism by introducing a pair-dependent confidence bound ε_{ij} , which specifies the largest opinion difference under which agent v_i is willing to take agent v_j into account. In this way, the confidence bound may depend on social status, familiarity, or other interpersonal attributes.

The edge-heterogeneous bounded-confidence model is defined as

$$x_i(t+1) = x_i(t) + \tau \sum_{v_j \in \tilde{\mathcal{I}}_i(t)} (x_j(t) - x_i(t)), \quad (3)$$

where $\tilde{\mathcal{I}}_i(t)$ denotes the neighbor set of agent v_i , given by

$$\tilde{\mathcal{I}}_i(t) = \{v_j \in \mathcal{V} : |x_j(t) - x_i(t)| < \varepsilon_{ij}, j \neq i\}, \quad (4)$$

and ε_{ij} represents the confidence bound adopted by agent v_i when responding to agent v_j . Compared with (1) and (2), the update structure of (3) and (4) remains unchanged; the only difference is that the agent-dependent confidence bound ε_i is replaced by the pair-dependent confidence bound ε_{ij} . For later use, we define the minimal and maximal confidence bounds as

$$\underline{\varepsilon} = \min\{\varepsilon_{ij} : i, j = 1, 2, \dots, n, i \neq j\}, \quad \bar{\varepsilon} = \max\{\varepsilon_{ij} : i, j = 1, 2, \dots, n, i \neq j\}.$$

In model (3), the interaction network evolves over time because it depends on both the opinion profile and the pair-dependent confidence bounds. For convenience, let $\mathcal{G}[A(t)] = \{\mathcal{V}, \mathcal{E}[A(t)], A(t)\}$ denote the time-varying unsigned digraph describing the interactions at time t , where the adjacency matrix is defined by

$$A_{ij}(t) = \begin{cases} 1, & v_j \in \check{\mathcal{I}}_i(t), \\ 0, & v_j \notin \check{\mathcal{I}}_i(t). \end{cases} \quad (5)$$

Because ε_{ij} and ε_{ji} may be different, the digraph $\mathcal{G}[A(t)]$ is generally asymmetric. In particular, for any pair of agents v_i and v_j , their relationship may fall into one of the following three cases:

- (1) *Two-way neighboring relationship*: Agents v_i and v_j are neighbors of each other if

$$|x_j(t) - x_i(t)| < \min\{\varepsilon_{ij}, \varepsilon_{ji}\},$$

that is, $v_j \in \check{\mathcal{I}}_i(t)$ and $v_i \in \check{\mathcal{I}}_j(t)$.

- (2) *One-way neighboring relationship*: If

$$\varepsilon_{ji} \leq |x_j(t) - x_i(t)| < \varepsilon_{ij},$$

then v_j is a neighbor of v_i , but v_i is not a neighbor of v_j , that is, $v_j \in \check{\mathcal{I}}_i(t)$ and $v_i \notin \check{\mathcal{I}}_j(t)$.

- (3) *Non-neighboring relationship*: If

$$|x_j(t) - x_i(t)| \geq \max\{\varepsilon_{ij}, \varepsilon_{ji}\},$$

then v_i and v_j are not neighbors of each other.

All edges in $\mathcal{G}[A(t)]$ have positive weights. Therefore, model (3) describes purely attractive interactions.

3.3. Edge-Heterogeneous Signed Bounded-Confidence Model

Distrust and antagonism are also prevalent in social networks. To capture such effects, we introduce a signed edge-heterogeneous bounded-confidence model in which opinions within the confidence bound generate attractive influence, whereas opinions outside the confidence bound generate oppositional influence.

For each agent v_i , define the trusted and distrustful neighbor sets as

$$\begin{aligned} \tilde{\mathcal{I}}_i^+(t) &= \{v_j \in \mathcal{V} : |x_j(t) - x_i(t)| < \varepsilon_{ij}, j \neq i\}, \\ \tilde{\mathcal{I}}_i^-(t) &= \{v_j \in \mathcal{V} : |x_j(t) - x_i(t)| \geq \varepsilon_{ij}, j \neq i\}, \end{aligned} \quad (6)$$

where $\tilde{\mathcal{I}}_i^+(t)$ contains the agents trusted by v_i , and $\tilde{\mathcal{I}}_i^-(t)$ contains those distrusted by v_i .

The signed update rule is given by

$$x_i(t+1) = x_i(t) + \tau \sum_{v_j \in \tilde{\mathcal{I}}_i^+(t)} (x_j(t) - x_i(t)) - \tau \sum_{v_j \in \tilde{\mathcal{I}}_i^-(t)} (x_j(t) + x_i(t)). \quad (7)$$

The first summation represents attractive interactions, whereas the second represents oppositional interactions induced by opinions lying outside the confidence bound. In the oppositional term, the quantity $x_j(t) + x_i(t)$ is used deliberately: since $-\tau(x_j(t) + x_i(t)) = \tau((-x_j(t)) - x_i(t))$, a distrusted neighbor drives v_i toward the *antipodal* opinion $-x_j(t)$. This is the structurally consistent (Altafini-type) repulsion [23,24], which is precisely the form that renders the signed dynamics gauge-equivalent to a cooperative averaging model and thereby yields structurally balanced polarization; a plain distancing term $-\tau(x_j(t) - x_i(t))$ would not possess this property. We also note that all updates in (3) and (7) are *synchronous*: at each step, every agent updates simultaneously using the com-

mon profile $x(t)$. For convenience, let $\mathcal{G}[W(t)] = \{\mathcal{V}, \mathcal{E}[W(t)], W(t)\}$ denote the signed interaction digraph at time t , where the adjacency matrix is defined by

$$W_{ij}(t) = \begin{cases} 1, & v_j \in \tilde{\mathcal{I}}_i^+(t), \\ -1, & v_j \in \tilde{\mathcal{I}}_i^-(t), \\ 0, & i = j. \end{cases} \quad (8)$$

Model (7) allows both supportive and oppositional interactions, and therefore provides a natural framework for studying structural balance, polarization, and convergence to neutrality on online social platforms.

4. Main Results

In this section, we establish rigorous sufficient conditions for cluster consensus. The analysis of fully asymmetric switching dynamics is particularly challenging, since the interaction pattern is state-dependent, discontinuous, and time-varying. Therefore, the results are formulated in terms of finite-time interval configurations.

4.1. Convergence of the Edge-Heterogeneous Bounded-Confidence Model

To state the result of the edge-heterogeneous bounded-confidence model, we first identify a finite-time geometric configuration under which the subsequent evolution becomes tractable. The following definition characterizes the situation where the opinions have already separated into narrow groups, and the gaps between different groups are sufficiently large to exclude further inter-group interactions.

Definition 1. Consider model (3) and a time instant T . The opinion is said to be cluster-separated at time T if there exist nonempty subsets $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_q$ forming a partition of $\mathcal{V}$, real numbers $a_1 < a_2 < \dots < a_q$, and a scalar $\delta \in (0, \bar{\varepsilon})$ such that

$$x_i(T) \in [a_r, a_r + \delta], \quad v_i \in \mathcal{V}_r, \quad r = 1, 2, \dots, q, \quad (9)$$

and

$$a_{r+1} - a_r - \delta > \bar{\varepsilon}, \quad r = 1, 2, \dots, q - 1. \quad (10)$$

Definition 1 identifies the geometric configuration under which further inter-group mixing is excluded. In general asymmetric bounded-confidence dynamics, the existence of such a time instant is not automatic. The next lemma gives a compact clustered-leader criterion ensuring finite-time entry into this regime.

Lemma 1. Consider model (3) under a clustered-leader architecture. Suppose that the node set admits a partition $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2 \cup \dots \cup \mathcal{V}_q$, and that each subset $\mathcal{V}_r$ contains a stubborn leader ℓ_r satisfying

$$x_{\ell_r}(t) \equiv s_r, \quad t = 0, 1, 2, \dots$$

Assume that there exists an integer $h \geq 1$ such that, for every r and every $t \geq 0$, the windowed digraph

$$\mathcal{G}_r[t, t+h-1] := \left(\mathcal{V}_r, \bigcup_{k=t}^{t+h-1} (\mathcal{E}[A(k)] \cap (\mathcal{V}_r \times \mathcal{V}_r)) \right)$$

satisfies the following:

1. For every r , every follower $v_i \in \mathcal{V}_r \setminus \{\ell_r\}$, and every $t \geq 0$, one has $\tilde{\mathcal{I}}_i(t) \subseteq \mathcal{V}_r \setminus \{v_i\}$;

2. For every r and every $t \geq 0$, the digraph $\mathcal{G}_r[t, t + h - 1]$ contains a spanning tree rooted at ℓ_r ;
3. After relabeling the subsets if necessary, the leader values satisfy

$$s_1 < s_2 < \dots < s_q, \quad s_{r+1} - s_r > \bar{\varepsilon}, \quad r = 1, 2, \dots, q - 1.$$

Then there exists a finite-time instant T such that the opinion profile is cluster-separated at time T in the sense of Definition 1.

Proof. Fix a subset $\mathcal{V}_r$ and let $n_r = |\mathcal{V}_r|$. By assumption (1), the dynamics of $\mathcal{V}_r$ are decoupled from the other subsets. For every follower $v_i \in \mathcal{V}_r \setminus \{\ell_r\}$, define the deviation

$$z_i(t) = x_i(t) - s_r.$$

Since $x_{\ell_r}(t) \equiv s_r$, the leader satisfies $z_{\ell_r}(t) \equiv 0$. For the followers, the update rule takes the form

$$z_F(t + 1) = Q_r(t)z_F(t),$$

where $z_F(t)$ collects the follower deviations in $\mathcal{V}_r$, and $Q_r(t)$ is a nonnegative matrix. Because each nonzero interpersonal weight equals τ and $0 < \tau \leq 1/n$, every diagonal entry of $Q_r(t)$ is bounded below by

$$\beta := 1 - \tau(n - 1) > 0.$$

Moreover, every row sum of $Q_r(t)$ is at most 1, and any follower directly influenced by the leader at time t has a row sum at most $1 - \tau$.

Now set $H_r = (n_r - 1)h$. By assumption (2), during any block of H_r consecutive steps, the leader influence can propagate from ℓ_r to every follower in $\mathcal{V}_r$ along a time-respecting directed path. Since each active edge contributes at least τ and each waiting step contributes at least β through the self-weight, there exists a constant

$$\mu_r := \tau^{n_r-1} \beta^{H_r-(n_r-1)} \in (0, 1)$$

such that every follower receives total weight of at least μ_r from the leader over each block of length H_r . Consequently,

$$\|Q_r(t + H_r - 1) \cdots Q_r(t)\|_\infty \leq 1 - \mu_r,$$

and therefore

$$\|z_F(t + H_r)\|_\infty \leq (1 - \mu_r)\|z_F(t)\|_\infty, \quad t \geq 0.$$

It follows that $z_F(t) \rightarrow 0$ exponentially; that is, $\lim_{t \rightarrow \infty} x_i(t) = s_r$ for all $v_i \in \mathcal{V}_r$. Since r is arbitrary, every subset converges to its own leader value.

Next choose

$$0 < \eta < \frac{1}{4} \min \left\{ \varepsilon, \min_{1 \leq r \leq q-1} (s_{r+1} - s_r - \bar{\varepsilon}) \right\}.$$

By the above convergence, there exists a finite-time instant T such that $|x_i(T) - s_r| < \eta$ for $v_i \in \mathcal{V}_r$, $r = 1, \dots, q$. Define $a_r = s_r - \eta$ and $\delta = 2\eta$. Then for every $v_i \in \mathcal{V}_r$, $x_i(T) \in [a_r, a_r + \delta]$, and $\delta = 2\eta < \varepsilon$. In addition,

$$a_{r+1} - a_r - \delta = s_{r+1} - s_r - 4\eta > \bar{\varepsilon}, \quad r = 1, 2, \dots, q - 1.$$

Hence, conditions (9) and (10) are satisfied at time T , and the opinion profile is cluster-separated. $\square$

Remark 1. Lemma 1 concerns only the existence of a finite entry time into the cluster-separated regime of Definition 1. Once such a time instant has been reached, the subsequent evolution no longer depends on the leader-driven assumptions used to produce it. Theorem 1 below characterizes the invariant decomposition and the asymptotic consensus behavior generated by any cluster-separated profile.

Theorem 1. Consider model (3) with $0 < \tau \leq \frac{1}{n}$. If the opinion is cluster-separated at some time instant T , then the associated partition remains invariant for all $t \geq T$. Moreover, each subset evolves independently and converges exponentially to an internal consensus. More precisely, if $n_r = |\mathcal{V}_r|$ and

$$\bar{x}_r(T) = \frac{1}{n_r} \sum_{v_i \in \mathcal{V}_r} x_i(T),$$

then

$$\lim_{t \rightarrow \infty} x_i(t) = \bar{x}_r(T), \quad v_i \in \mathcal{V}_r, \quad r = 1, 2, \dots, q.$$

Proof. By Definition 1, for any $v_i, v_j \in \mathcal{V}_r$, condition (9) implies

$$|x_j(T) - x_i(T)| \leq \delta < \varepsilon \leq \varepsilon_{ij},$$

and $v_j \in \check{\mathcal{I}}_i(T)$. Therefore, each subset $\mathcal{V}_r$ induces a complete directed subgraph at time T . On the other hand, if $v_i \in \mathcal{V}_r$ and $v_j \in \mathcal{V}_s$ with $r < s$, then by (9) and (10),

$$x_j(T) - x_i(T) \geq a_s - (a_r + \delta) \geq a_{r+1} - a_r - \delta > \bar{\varepsilon} \geq \varepsilon_{ij},$$

which yields $v_j \notin \check{\mathcal{I}}_i(T)$. Thus, there are no edges between different subsets at time T . Moreover, since the update rule is a convex combination of the opinions within each subset, every $x_i(t)$ with $v_i \in \mathcal{V}_r$ remains in $[a_r, a_r + \delta]$ for all $t \geq T$. Consequently, the partition is invariant.

Fix a subset $\mathcal{V}_r$ and let $n_r = |\mathcal{V}_r|$. Because the partition is invariant and each $\mathcal{V}_r$ induces a complete directed subgraph for all $t \geq T$, every follower $v_i \in \mathcal{V}_r$ has the fixed neighbor set $\check{\mathcal{I}}_i(t) = \mathcal{V}_r \setminus \{v_i\}$. Substituting this into the original update rule (3) and isolating the self-opinion term gives, for $v_i \in \mathcal{V}_r$,

$$\begin{aligned} x_i(t+1) &= x_i(t) + \tau \sum_{v_j \in \mathcal{V}_r \setminus \{v_i\}} (x_j(t) - x_i(t)) \\ &= \underbrace{[1 - \tau(n_r - 1)]}_{\text{self-opinion weight}} x_i(t) + \tau \sum_{v_j \in \mathcal{V}_r \setminus \{v_i\}} x_j(t) = (1 - n_r \tau) x_i(t) + \tau \sum_{v_j \in \mathcal{V}_r} x_j(t), \end{aligned}$$

where the last equality follows by adding and subtracting the self-term $\tau x_i(t)$, so that the self-opinion contribution to the cluster dynamics is exactly the diagonal weight $1 - n_r \tau$. Collecting the opinions of $\mathcal{V}_r$ into the vector $x_{\mathcal{V}_r}(t)$, this reads

$$x_{\mathcal{V}_r}(t+1) = M_r x_{\mathcal{V}_r}(t), \quad M_r = (1 - n_r \tau) I_{n_r} + \tau \mathbf{1}_{n_r} \mathbf{1}_{n_r}^\top.$$

Since $0 < \tau \leq \frac{1}{n} \leq \frac{1}{n_r}$, all entries of M_r are nonnegative; each row sum and each column sum equals $(1 - n_r \tau) + n_r \tau = 1$, so M_r is symmetric and doubly stochastic, and this property is unaffected by the endogenous interaction structure because, after separation, that structure is frozen at the complete intra-cluster graph. Doubly stochasticity gives $\mathbf{1}_{n_r}^\top M_r = \mathbf{1}_{n_r}^\top$; hence, the cluster sum, and therefore the cluster average $\bar{x}_r = \frac{1}{n_r} \mathbf{1}_{n_r}^\top x_{\mathcal{V}_r}$, is the exact invariant of the reduced dynamics

$$\bar{x}_r(t+1) = \frac{1}{n_r} \mathbf{1}_{n_r}^\top M_r x_{\mathcal{V}_r}(t) = \frac{1}{n_r} \mathbf{1}_{n_r}^\top x_{\mathcal{V}_r}(t) = \bar{x}_r(t), \quad t \geq T.$$

Moreover, M_r has the simple eigenvalue 1 with eigenvector $\mathbf{1}_{n_r}$ and the eigenvalue $1 - n_r\tau \in [0, 1)$ with multiplicity $n_r - 1$ on the subspace $\mathbf{1}_{n_r}^\perp$. Consequently, for any $v_i, v_k \in \mathcal{V}_r$,

$$x_i(t+1) - x_k(t+1) = (1 - n_r\tau)(x_i(t) - x_k(t)),$$

so every pairwise within-cluster difference decays geometrically with the rate $1 - n_r\tau$. Combining the preserved average with the vanishing differences yields $\lim_{t \rightarrow \infty} x_i(t) = \bar{x}_r(T)$ for all $v_i \in \mathcal{V}_r$. $\square$

4.2. Convergence of the Edge-Heterogeneous Signed Bounded-Confidence Model

For the signed model, we again focus on a finite-time configuration from which the subsequent dynamics become transparent. The relevant structure is a two-camp profile that is structurally balanced with respect to a prescribed bipartition of the node set and becomes purely attractive after a gauge transformation.

Theorem 2. Consider model (7) with $0 < \tau \leq \frac{1}{n}$. Suppose that there exist a partition $\mathcal{V} = \mathcal{V}_1 \cup \mathcal{V}_2$ into two nonempty subsets, a scalar $a > \bar{\varepsilon}/2$, and a number $\delta \in (0, \varepsilon)$ such that

$$y_i(T) = \begin{cases} x_i(T), & v_i \in \mathcal{V}_1, \\ -x_i(T), & v_i \in \mathcal{V}_2, \end{cases}$$

satisfies $y_i(T) \in [a, a + \delta]$ for all $i = 1, 2, \dots, n$. Then the trust–distrust pattern remains invariant and structurally balanced for all $t \geq T$. Moreover, the network polarizes: there exists $\alpha \in [a, a + \delta]$ such that

$$\lim_{t \rightarrow \infty} x_i(t) = \alpha, \quad v_i \in \mathcal{V}_1, \quad \lim_{t \rightarrow \infty} x_i(t) = -\alpha, \quad v_i \in \mathcal{V}_2.$$

Proof. For any $v_i, v_j \in \mathcal{V}_1$ or $v_i, v_j \in \mathcal{V}_2$,

$$|x_j(T) - x_i(T)| = |y_j(T) - y_i(T)| \leq \delta < \varepsilon \leq \varepsilon_{ij},$$

so the corresponding edge is positive. If $v_i \in \mathcal{V}_1$ and $v_j \in \mathcal{V}_2$, then $x_i(T) \in [a, a + \delta]$ and $x_j(T) \in [-a - \delta, -a]$, so

$$|x_j(T) - x_i(T)| \geq 2a > \bar{\varepsilon} \geq \varepsilon_{ij},$$

and the corresponding edge is negative. The signed graph at time T is therefore complete and structurally balanced with respect to $\mathcal{V}_1 \cup \mathcal{V}_2$.

We now derive the gauge-transformed dynamics explicitly. Introduce the camp signs $\sigma_i = +1$ for $v_i \in \mathcal{V}_1$ and $\sigma_i = -1$ for $v_i \in \mathcal{V}_2$, so that $y_i = \sigma_i x_i$ and $x_i = \sigma_i y_i$ (because $\sigma_i^2 = 1$). At the complete, structurally balanced configuration just established, each v_i trusts exactly its own camp ($\tilde{\mathcal{I}}_i^+ = \{v_j : \sigma_j = \sigma_i, j \neq i\}$) and distrusts exactly the opposite camp ($\tilde{\mathcal{I}}_i^- = \{v_j : \sigma_j = -\sigma_i\}$). Multiplying the signed update (7) by σ_i and using $\sigma_i x_j = \sigma_i \sigma_j y_j$, which equals y_j when $\sigma_j = \sigma_i$ and $-y_j$ when $\sigma_j = -\sigma_i$, we obtain

$$\begin{aligned} y_i(t+1) &= \sigma_i x_i(t+1) = y_i(t) + \tau \sum_{v_j \in \tilde{\mathcal{I}}_i^+} (\sigma_i x_j(t) - \sigma_i x_i(t)) - \tau \sum_{v_j \in \tilde{\mathcal{I}}_i^-} (\sigma_i x_j(t) + \sigma_i x_i(t)) \\ &= y_i(t) + \tau \sum_{\sigma_j = \sigma_i, j \neq i} (y_j(t) - y_i(t)) - \tau \sum_{\sigma_j = -\sigma_i} (-y_j(t) + y_i(t)) \\ &= y_i(t) + \tau \sum_{j \neq i} (y_j(t) - y_i(t)) = (1 - n\tau) y_i(t) + \tau \sum_{j=1}^n y_j(t). \end{aligned}$$

The crucial cancellation occurs in the distrust sum: the term $-(x_j + x_i)$ of (7), which encodes attraction of v_i toward the antipodal opinion $-x_j$ in the structurally consistent (Altafini) sense [23], maps under the gauge to $-(-y_j + y_i) = +(y_j - y_i)$ since $-(x_j + x_i) = (-x_j) - x_i$, i.e., to a purely attractive term among the transformed variables. Replacing $-(x_j + x_i)$ by a plain distancing term $-(x_j - x_i)$ would instead produce $+(y_j + y_i)$, which is not a consensus term and would prevent the bipartite limit below; this confirms that the sign structure in (7) is exactly the one required for structurally balanced polarization. The resulting recursion $y_i(t+1) = (1 - n\tau)y_i(t) + \tau \sum_{j=1}^n y_j(t)$ is a convex combination of $\{y_1(t), \dots, y_n(t)\}$ because $0 < \tau \leq \frac{1}{n}$. Therefore, $y_i(t) \in [a, a + \delta]$ for all $t \geq T$, and the trust–distrust pattern persists. Moreover,

$$y_i(t+1) - y_k(t+1) = (1 - n\tau)(y_i(t) - y_k(t)),$$

so all transformed opinions converge exponentially to

$$\alpha = \frac{1}{n} \sum_{j=1}^n y_j(T) = \frac{1}{n} \left(\sum_{v_j \in \mathcal{V}_1} x_j(T) - \sum_{v_j \in \mathcal{V}_2} x_j(T) \right) \in [a, a + \delta].$$

Returning to the original coordinates yields the polarized limits. $\square$

Remark 2. It is important to clarify the status of structural balance in our setting. Unlike classical signed-graph models, in which the sign pattern is prescribed exogenously, the signed graph $\mathcal{G}[W(t)]$ here is generated endogenously from opinion distances and pair-dependent confidence bounds through (8). Theorem 2 does not assert that an arbitrary endogenous signed graph is structurally balanced. Rather, it identifies a concrete geometric precondition, the two-camp profile with inter-camp separation exceeding $\bar{\epsilon}$ and intra-camp spread below $\underline{\epsilon}$, under which the induced graph is provably complete and structurally balanced at that time, and the proof further shows this property is forward-invariant because the gauge-transformed dynamics keep every $y_i(t)$ inside $[a, a + \delta]$. In particular, the triad/sign-consistency requirement is met by construction: any two agents in the same camp are within their mutual confidence bounds (positive edge), while any cross-camp pair lies outside them (negative edge), so no frustrated edge can arise as long as the configuration persists. The structural-balance claim is therefore a sufficient, configuration-dependent statement, not a generic property of the model; Example 5 below exhibits endogenous signed graphs that are not structurally balanced, for which polarization correctly fails to form.

5. Implications for Online Social Platforms

The theoretical results in Section 4 characterize the long-term behavior of opinions under edge-heterogeneous and signed bounded-confidence dynamics. In this section, we discuss how these results can be interpreted in the context of contemporary online social platforms. We stress at the outset that the rigorous content of this paper is confined to Theorems 1 and 2, which hold under specific finite-time geometric configurations and the step-size condition $\tau \leq 1/n$. The platform-level statements in this section, concerning echo chambers, recommendation systems, and governance levers, are interpretive analogies that connect the model's proven asymptotic behavior to qualitatively similar phenomena reported in the empirical literature; they are not themselves theorems and are not claimed to be derived with the same rigor. We therefore mark clearly, throughout this section, where a statement is a mathematical consequence of our results and where it is a qualitative implication.

5.1. Pair-Dependent Confidence Bounds as Algorithmic and Social Trust

In real online platforms, the confidence bound ε_{ij} that agent v_i assigns to agent v_j can be regarded as the joint outcome of two factors: (i) platform-level algorithmic exposure, including follower relationships, recommendation weights, and ranking priorities, and (ii) user-level social trust, including familiarity, perceived expertise, ideological alignment, or institutional credibility. The proposed edge-heterogeneous formulation therefore provides a tractable abstraction of platform-mediated interpersonal trust: a user is willing to incorporate diverse opinions from a verified expert or close friend (large ε_{ij}), but quickly disengages from unfamiliar or low-credibility accounts (small ε_{ij}).

5.2. Cluster Consensus and Community Formation

Theorem 1 states that once the opinion profile reaches a cluster-separated regime, the induced partition becomes invariant and each subgroup converges to its internal consensus. On online platforms, this corresponds to the empirical observation that, after a certain stage of debate, users self-organize into topic-aligned communities whose internal opinions become tightly clustered while inter-community exchange essentially vanishes. From a platform-design perspective, this offers a theoretical lens to interpret community detection outcomes and to assess when recommendation algorithms inadvertently reinforce the formation of isolated clusters.

5.3. Polarization and Echo Chambers

Theorem 2 shows that in the signed model, once a two-camp configuration with sufficient inter-camp separation emerges, the trust–distrust pattern becomes structurally balanced and the network polarizes. This formal result captures the dynamic essence of echo chambers observed on platforms such as Twitter and Reddit during politically sensitive events: positive engagement is concentrated within each ideological camp, while negative engagement (e.g., quote-retweet rebuttals, downvotes, or hostile replies) dominates across camps. The model therefore provides a mechanism-level explanation for why moderate cross-camp exposure, once distrust has crystallized, fails to bridge the divide.

5.4. Operational Implications

The theoretical conditions also suggest several operational levers for platform governance. We present these as qualitative implications of the model’s demonstrated behavior rather than as formally derived or optimized policies; a rigorous control-theoretic treatment of each is left for future work.

- **Recommendation diversification:** Because Theorem 1 requires inter-group gaps to exceed $\bar{\varepsilon}$ before the partition freezes, injecting a small set of moderate cross-cluster contacts before the cluster-separated regime is reached may, qualitatively, delay or prevent the partition from becoming invariant.
- **Source-quality signaling:** Visible credibility labels can be viewed as enlarging the confidence bound ε_{ij} from credible sources, which in the model mitigates premature distrust and reduces the likelihood of structural balance.
- **Polarization early warning:** Monitoring the empirical gap between intra-camp and inter-camp opinion distances against the bound $\bar{\varepsilon}$ provides a low-cost indicator of imminent polarization, a reading directly suggested by the separation hypothesis of Theorem 2.

Interpreted in this qualitative sense, these implications connect our convergence guarantees to candidate platform policies and motivate the design of healthier online discussion environments.

6. Numerical Examples

In this section, we present five numerical examples: two synthetic experiments illustrating Theorems 1 and 2, a real-world case study on the Slashdot Zoo signed network, a Monte-Carlo sensitivity study, and a set of counterexamples beyond the sufficient conditions.

Example 1 (Edge-Heterogeneous Bounded-Confidence Model). Consider model (3) with $n = 15$ and $\tau = 0.018$. Let

$$\mathcal{V}_1 = \{v_1, \dots, v_5\}, \quad \mathcal{V}_2 = \{v_6, \dots, v_{10}\}, \quad \mathcal{V}_3 = \{v_{11}, \dots, v_{15}\},$$

and choose the initial opinion vector as

$$x(0) = (0.02, 0.07, 0.11, 0.18, 0.23, 0.16, 0.21, 0.27, 0.33, 0.39, 0.34, 0.40, 0.47, 0.55, 0.62)^\top.$$

The pair-dependent confidence bounds are sampled independently once at the beginning and then kept fixed throughout the simulation according to

$$\varepsilon_{ij} \sim \begin{cases} \text{Unif}[0.14, 0.18], & v_i, v_j \text{ in the same subset, } i \neq j, \\ \text{Unif}[0.045, 0.070], & v_i \in \mathcal{V}_r, v_j \in \mathcal{V}_s, |r - s| = 1, \\ \text{Unif}[0.015, 0.025], & v_i \in \mathcal{V}_r, v_j \in \mathcal{V}_s, |r - s| = 2. \end{cases}$$

Figure 1 depicts the initial topology of the 15 agents (left, $t = 0$) and the final reordered topology induced by the opinion evolution (right, $t = 90$). Although the confidence bounds are fixed at the initial time, the communication topology is initially disordered; after evolution, the agents split into three subgroups, within each of which a common opinion is reached. The opinion trajectories in Figure 2 converge to three cluster centers near 0.1244, 0.2749, and 0.4780, in agreement with Theorem 1.

Example 2 (Edge-Heterogeneous Signed Bounded-Confidence Model). Consider model (7) with $n = 20$ and $\tau = 0.010$. Let

$$\mathcal{V}_1 = \{v_1, \dots, v_{10}\}, \quad \mathcal{V}_2 = \{v_{11}, \dots, v_{20}\},$$

and choose the initial opinion vector as

$$x(0) = (-0.03, -0.01, 0.01, 0.03, 0.05, 0.07, 0.09, 0.11, 0.13, 0.15, \\ -0.15, -0.13, -0.11, -0.09, -0.07, -0.05, -0.03, -0.01, 0.01, 0.03)^\top.$$

The confidence bounds are sampled independently once at the beginning according to

$$\varepsilon_{ij} \sim \begin{cases} \text{Unif}[0.19, 0.22], & v_i, v_j \text{ in the same subset, } i \neq j, \\ \text{Unif}[0.05, 0.08], & v_i \in \mathcal{V}_1, v_j \in \mathcal{V}_2 \text{ or vice versa.} \end{cases}$$

Figure 3 shows the evolution of the communication interactions among the clusters. The dynamics reorganize the mixed signed graph into a clear two-camp structure: by $t = 40$, positive reciprocal edges are concentrated within $\mathcal{V}_1$ and $\mathcal{V}_2$. The opinion trajectories in Figure 4 converge to the polarized levels ± 0.0543 . We note that this example is a qualitative illustration that goes beyond the strict sufficient conditions of Theorem 2: the theorem is stated with the uniform bounds $\underline{\varepsilon}$ and $\bar{\varepsilon}$, and here the terminal cross-camp distance $2 \times 0.0543 \approx 0.109$ does not exceed $\bar{\varepsilon} \approx 0.22$. The polarized outcome is nevertheless fully consistent with the mechanism of the theorem, because the separation holds pairwise: the cross-camp distance exceeds every cross-camp bound (all below 0.08), while the intra-camp

spread contracts to zero, far below every intra-camp bound (all above 0.19), so the trust–distrust pattern is locked exactly as in the proof. This also illustrates that the uniform-bound hypotheses are conservative, in line with the counterexamples of Example 5.

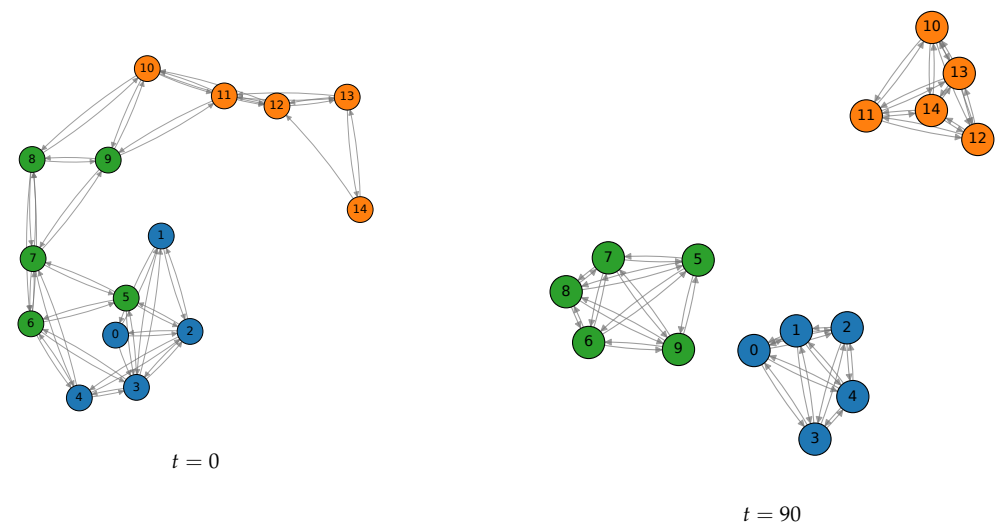


Figure 1. Topology snapshots for Example 1. **Left:** interaction graph at $t = 0$. **Right:** interaction graph at $t = 90$. Node colors indicate cluster membership (blue: $\mathcal{V}_1$, green: $\mathcal{V}_2$, orange: $\mathcal{V}_3$).

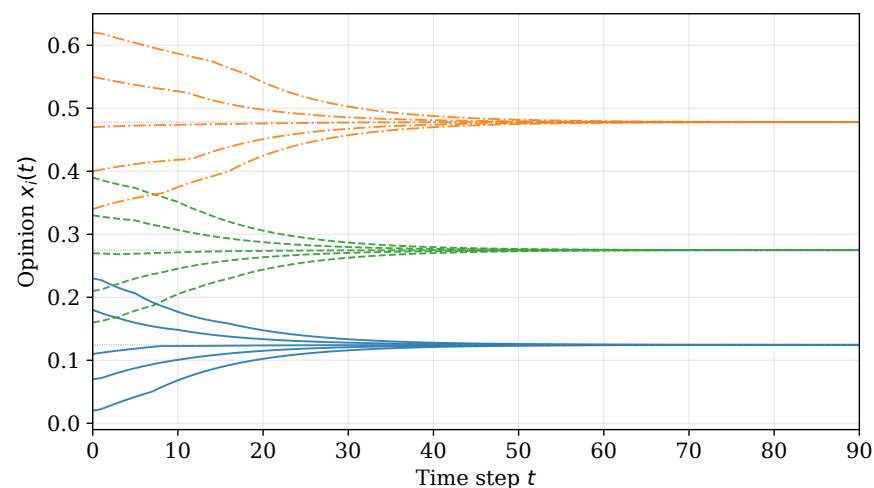


Figure 2. Opinion trajectories for Example 1 under model (3). Solid lines: $\mathcal{V}_1$; dashed lines: $\mathcal{V}_2$; dash-dotted lines: $\mathcal{V}_3$.

Example 3 (Case Study on a Real-World Signed Social Network). *To assess the practical relevance of model (7) on a real online community, we conduct a case study on the Slashdot Zoo signed network [28,29], which records explicit friend/foe relationships among Slashdot users. The full dataset contains 82,140 users and 549,202 signed directed edges. From the largest weakly connected component, we extract a subgraph of $n = 50$ controversial users, ranked by the controversy score $\min(d_i^+, d_i^-)$, where d_i^+ and d_i^- denote the number of friend and foe ties of user v_i ; this score is large only for users who are simultaneously highly trusted and highly distrusted. The induced subgraph contains 266 edges, including 121 positive (friend) and 145 negative (foe) edges. A signed spectral bipartition (the leading eigenvector of the symmetrized adjacency matrix, split at the median) identifies two voting blocs of size 25 each; the partition has 209 structurally aligned edges (intra-bloc positives plus inter-bloc negatives) and 57 frustrated edges, corresponding to a structural balance ratio of approximately 78.6%. The complete extraction and simulation procedure is specified above and in Section 6, so that the case study can be reproduced directly from the publicly available dataset.*

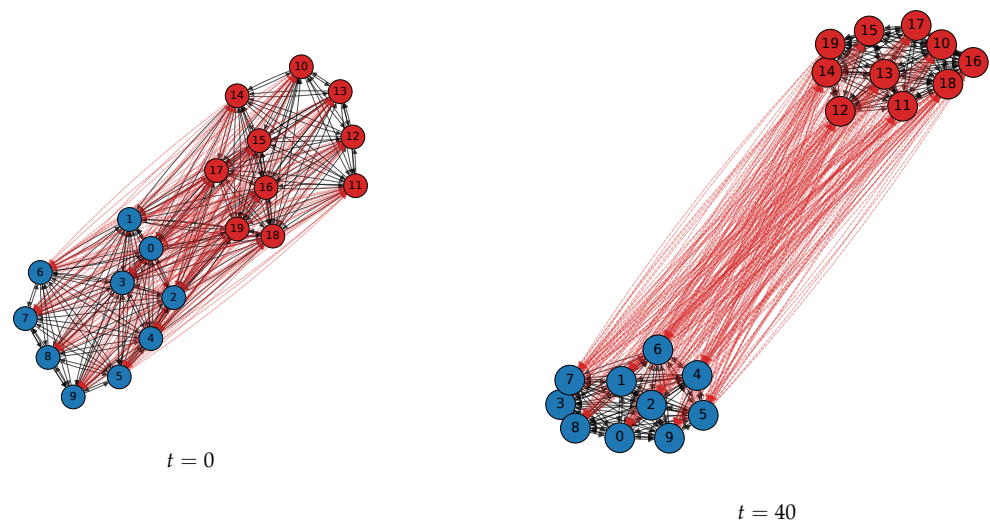


Figure 3. Topology snapshots for Example 2. **Left:** signed interaction graph at $t = 0$. **Right:** signed interaction graph at $t = 40$. Node colors distinguish the two camps (blue: V_1 , red: V_2); black solid edges represent cooperative interactions, whereas red dashed edges represent competitive interactions.

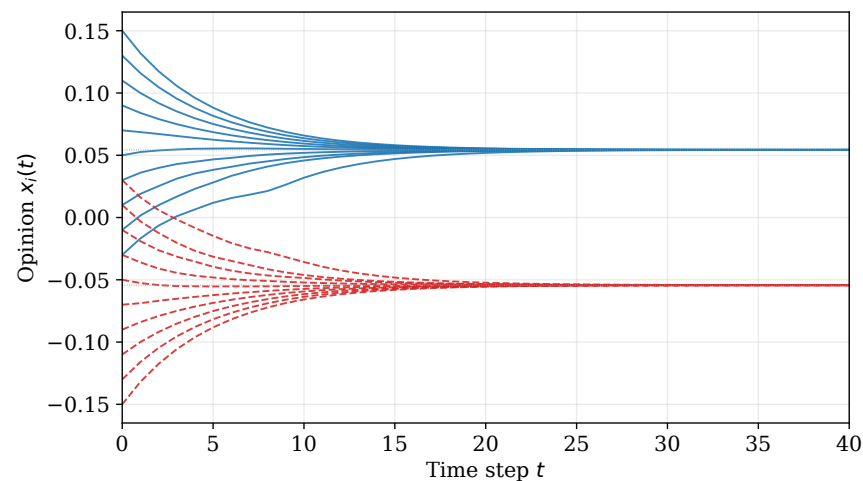


Figure 4. Opinion trajectories for Example 2 under model (7). Solid lines: V_1 ; dashed lines: V_2 .

The pair-dependent confidence bounds are sampled once and held fixed throughout the simulation according to

$$\varepsilon_{ij} \sim \begin{cases} \text{Unif}[0.18, 0.22], & \text{if there is a positive edge } (v_j, v_i), \\ \text{Unif}[0.02, 0.05], & \text{if there is a negative edge } (v_j, v_i), \\ \text{Unif}[0.05, 0.09], & \text{otherwise,} \end{cases}$$

matching the real signed topology with strong trust on friend ties, strong distrust on foe ties, and a neutral prior on unobserved pairs. The initial opinions are set to a polarized configuration consistent with the assumptions of Theorem 2, namely $x_i(0) = +a + \eta_i$ for $v_i \in V_1$ and $x_i(0) = -a + \eta_i$ for $v_i \in V_2$, with $a = 0.15 > \bar{\varepsilon}/2$ and $|\eta_i| \leq 0.070$. The step size is fixed as $\tau = 0.005$.

We emphasize the precise epistemic status of this study. The signed topology is genuinely data-driven, being read off directly from the real friend/foe edges of the Slashdot Zoo network. The pair-dependent confidence bounds ε_{ij} , by contrast, are not observed in the dataset; they are assigned by sampling from the three prescribed ranges according to the observed type of each edge (friend, foe, or absent), so that the bound structure mirrors,

but is not measured from, the real trust relations. The initial opinion profile, by contrast, is deliberately set to satisfy the precondition of Theorem 2. Accordingly, this case study should be read as an illustrative simulation on a real signed topology, demonstrating that, once the Theorem 2 precondition holds on a genuine adversarial network, structurally balanced polarization persists despite a substantial number of frustrated edges, rather than as a fully data-driven inference of the initial opinions from raw data. The latter would require an external opinion measurement that the dataset does not provide.

Figure 5 reports the resulting opinion trajectories of model (7) on two real-world signed social networks. The right panel shows the Slashdot subgraph: despite the presence of 57 structurally frustrated edges and an initial spread that brings the two bloc envelopes close to zero, the dynamics preserve the two-camp configuration throughout the entire horizon and converge to the polarized levels ± 0.1254 . For this run the hypotheses of Theorem 2 can be verified explicitly along the trajectory: with $\varepsilon \approx 0.020$ and $\bar{\varepsilon} \approx 0.220$, the gauge-transformed opinions first satisfy the required configuration at $T = 15$, where they lie in $[a, a + \delta]$ with $a \approx 0.113 > \bar{\varepsilon}/2 = 0.110$ and $\delta \approx 0.014 < \varepsilon$; from that time on, the theorem applies and guarantees the observed invariant polarization. The terminal partition induced by the sign of $x_i(t)$ coincides with the spectral voting-bloc partition on all 50 users (100% agreement), which illustrates that, once the conditions of Theorem 2 are met, the proposed signed edge-heterogeneous model reproduces structurally balanced polarization on a real signed social network. The left panel shows the same model applied to the Wiki-RfA Wikipedia administrator-vote network [30]. Although both networks have a similar global fraction of positive edges (about 77% for Slashdot and 78% for Wiki-RfA), their extracted controversy subgraphs differ sharply in structural balance: the Wiki-RfA subgraph attains a balance ratio of only about 57%, close to the random baseline, so it admits no clean bipartition into two antagonistic camps. As a result, no stable two-camp configuration forms and the dynamics collapse into a single near-zero cluster, consistent with the $q = 1$ case of Theorem 1. This contrast illustrates that the proposed framework distinguishes between a structurally balanced adversarial subgraph (Slashdot), which sustains polarization, and a poorly balanced one (Wiki-RfA), which relaxes to neutral consensus.

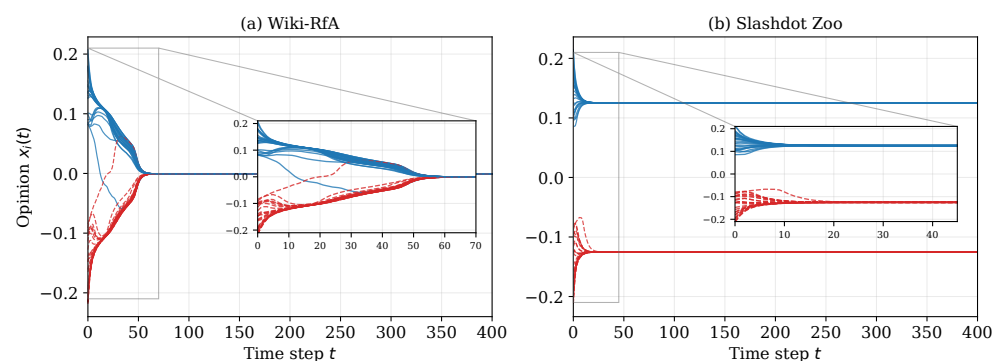


Figure 5. Comparison of the proposed signed model (7) on two real-world signed networks: (a) Wiki-RfA, a cooperative platform where the controversy subgraph collapses to a single cluster; (b) Slashdot Zoo, an adversarial platform where the dynamics sustain a structurally balanced two-camp polarization.

Example 4 (Sensitivity to the Statistics of the Confidence Bounds). *The preceding examples fix one realization of the bounds ε_{ij} . To examine how the statistics of the bounds shape the dynamics, we perform a Monte-Carlo sensitivity study with $n = 15$, mean bound $\bar{\mu} = 0.10$, step size $\tau = 1/n$, random initial opinions on $[0, 1]$, and 80 independent trials per parameter value. The outcome metric is the mean number of terminal opinion clusters (opinions grouped within tolerance 10^{-3}). Figure 6*

reports four scans. Panel (a) varies the half-width w of $\varepsilon_{ij} \sim \text{Unif}[\bar{\mu} - w, \bar{\mu} + w]$: increasing the variance lowers the mean cluster count from about 4.2 to 3.2, because occasional large bounds bridge would-be groups. Panel (b) varies an asymmetry amplitude a by setting $\varepsilon_{ij} = \bar{\mu} + a \xi_{ij}$ with $\xi_{ij} = -\xi_{ji}$; stronger trust asymmetry has a similar coarsening effect, reducing the cluster count to about 3.5, while convergence is preserved in every trial. Panel (c) varies the Pearson correlation $\rho(\varepsilon_{ij}, \varepsilon_{ji})$ between reciprocal bounds, whose effect is mild and non-monotone. Crucially, across all of panels (a)–(c), not a single run failed to converge. Panel (d) isolates the role of the step size under fixed strong asymmetry: oscillatory (non-convergent) runs are entirely absent for $\tau \leq 2/n$ and appear only for $\tau \gtrsim 2.5/n$. This supports the intuition that strong asymmetry combined with a large step size can destabilize the dynamics, while showing that the assumption $\tau \leq 1/n$ used throughout this paper lies safely inside the convergent regime with a comfortable margin.

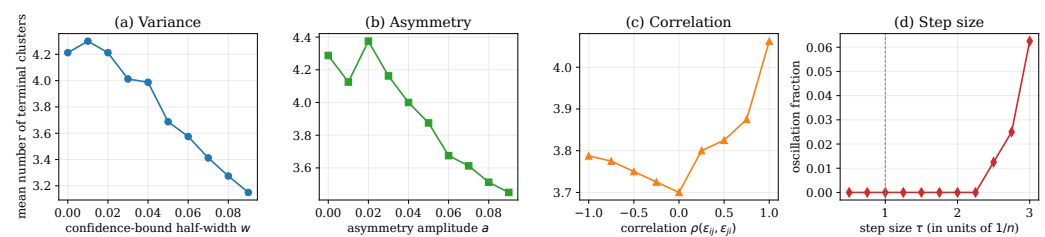


Figure 6. Sensitivity study (Example 4). Mean number of terminal clusters versus (a) confidence-bound variance, (b) trust asymmetry, and (c) reciprocal correlation; and (d) fraction of non-convergent (oscillatory) runs versus step size τ (in units of $1/n$) under fixed strong asymmetry. The dashed line marks $\tau = 1/n$.

Example 5 (Beyond the Sufficient Conditions: Two Counterexamples). *The conditions of Theorems 1 and 2 are sufficient, not necessary. Figure 7 illustrates two regimes in which the hypotheses fail at $t = 0$ yet ordered collective behavior still emerges. In panel (a) ($n = 20$, homogeneous bound $\varepsilon = 0.18$, $\tau = 1/n$), the initial interaction graph is fully connected—a single interacting population, so the cluster-separated precondition of Definition 1 (which forbids inter-group edges) is strongly violated—yet the dynamics still fragment into three well-separated terminal clusters, each in internal consensus. In panel (b) ($n = 20$, heterogeneous $\varepsilon_{ij} \sim \text{Unif}[0.02, 0.05]$, $\tau = 1/n$), the initial signed graph is not structurally balanced (82 of 380 edges are frustrated and no clean bipartition exists), so the precondition of Theorem 2 fails; nonetheless, the opinions still converge to a definite collective state, namely neutral consensus at 0, consistent with the $q = 1$ case of Theorem 1. These examples delineate the gap between the sufficient theoretical conditions and the substantially larger basin of ordered behavior observed in practice, and they indicate that the conditions identify verifiable certificates of clustering/polarization rather than necessary prerequisites.*

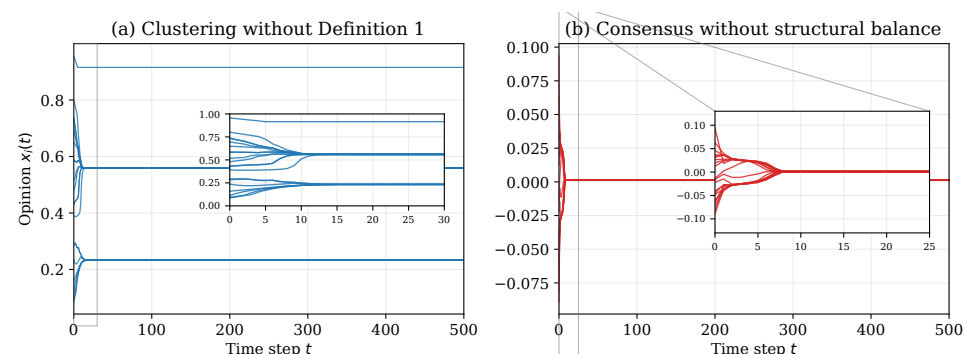


Figure 7. Counterexamples (Example 5). (a) A connected initial graph—violating Definition 1—still fragments into three internally consensual clusters. (b) A structurally unbalanced initial signed graph—violating the hypothesis of Theorem 2—still converges to neutral consensus.

7. Discussion: Limitations and Computational Complexity

In this section we state the scope and limitations of the proposed framework explicitly, and analyze its computational cost.

7.1. Limitations

Like much of the bounded-confidence literature, the proposed framework deliberately favors analytical tractability, and a few resulting limitations are worth noting. Theoretically, our main results give sufficient, not necessary, conditions: they rely on a finite-time geometric configuration (cluster separation in Definition 1, or a separated two-camp profile) together with $\tau \leq 1/n$, and finite-time entry into this regime is guaranteed (Lemma 1) under a clustered-leader architecture. As Example 5 shows, ordered behavior also arises well outside these hypotheses, so the conditions are best read as verifiable certificates of clustering and polarization rather than as a full characterization of when they occur. On the modeling side, the framework adopts a few simplifying assumptions, each a natural direction for future work: the confidence bounds ε_{ij} are fixed and exogenous rather than evolving with reputation or repeated exposure; opinions are scalar rather than multidimensional; the dynamics are synchronous (rather than asynchronous or event-triggered “microscopic” message-passing, which may converge to different asymptotic states), deterministic, and closed, without noise or exogenous shocks such as media events or coordinated campaigns; and links form from opinion proximity alone, without strategic or platform-driven tie formation. Finally, the case study is an illustrative simulation on real signed topologies rather than a fit of measured opinions, since node-level opinion ground truth is unavailable.

7.2. Computational Complexity

Simulating either model is inexpensive. Each synchronous step evaluates the pairwise opinion gaps and the induced (signed) interaction matrix, costing $O(n^2)$ operations and $O(n^2)$ memory for a dense confidence matrix; running T steps therefore costs $O(Tn^2)$. For the sparse, large-scale networks typical of online platforms, restricting interactions to the $O(|\mathcal{E}|)$ existing edges reduces the per-step cost to $O(|\mathcal{E}|)$, so the method scales to networks with millions of edges. Verifying the theorem hypotheses is likewise cheap: cluster separation (Definition 1) is an $O(n \log n)$ sort-and-gap test, and the structural-balance check used in the case study reduces to one leading-eigenvector computation of a sparse symmetric matrix, i.e., $O(|\mathcal{E}|)$ per iteration via the power method.

8. Conclusions

This paper investigated edge-heterogeneous bounded-confidence opinion dynamics with pair-dependent confidence bounds in both attractive and signed cases. For the attractive model, we identified a finite-time cluster-separated regime under which the dynamics split into invariant subsystems, each converging exponentially to an internal consensus; for the signed model, we established a corresponding two-camp regime that is invariant under a gauge transformation and yields polarization. Numerical examples, including a case study on the Slashdot Zoo real-world signed social network, illustrate the theoretical findings, and the framework offers a mechanism-level interpretation of clustering, echo chambers, and polarization on online social platforms; its scope, limitations, and computational cost are discussed in Section 7. Future work will address weaker entry conditions, sharper clustering and polarization criteria, more general switching signed interactions, the extension to multi-dimensional opinions, data-driven estimation of the heterogeneous confidence bounds, and large-scale validation on platform-level traces.

Author Contributions: Conceptualization, Z.H.; methodology, L.S.; validation, L.C.; formal analysis, Z.H.; writing—original draft preparation, Z.M.; writing—review and editing, L.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the China Postdoctoral Science Foundation under Grant 2025T180475.

Data Availability Statement: The original contributions presented in this study are included in the article. Further inquiries can be directed to the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Tan, E.; Stemler, T.; Small, M. Cognitive dissonance and introversion effects on opinion dynamics and echo chamber formation. *Phys. A* **2024**, *654*, 130130. [\[CrossRef\]](#)
2. Cahill, P.; Gottwald, G.A. A modified Hegselmann–Krause model for interacting voters and political parties. *Phys. A* **2025**, *665*, 130490. [\[CrossRef\]](#)
3. Bautista, A. Opinion dynamics in bounded confidence models with manipulative agents: Moving the Overton window. *Phys. A* **2025**, *660*, 130379. [\[CrossRef\]](#)
4. Shang, L.; Zhao, M.; Ai, J.; Su, Z. Opinion evolution in the Sznajd model on interdependent chains. *Phys. A* **2021**, *565*, 125558. [\[CrossRef\]](#)
5. Zhang, J.; Zhang, W. Opinions dynamics for a generalized nonlinear model with state-dependent susceptibility in signed networks on time scales. *J. Frankl. Inst.* **2024**, *361*, 107313. [\[CrossRef\]](#)
6. Hegselmann, R.; Krause, U. Opinion dynamics and bounded confidence: Models, analysis and simulation. *J. Artif. Soc. Soc. Simul.* **2002**, *5*, 1–33.
7. Su, W.; Guo, J.; Chen, X.; Chen, G.; Li, J. Robust fragmentation modeling of Hegselmann–Krause-type dynamics. *J. Frankl. Inst.* **2019**, *356*, 9867–9880. [\[CrossRef\]](#)
8. Cyranka, J.; Mucha, P.B. A polarization opinion model inspired by bounded confidence communications. *J. Frankl. Inst.* **2024**, *361*, 107140. [\[CrossRef\]](#)
9. Mohajer, S.; Touri, B. On convergence rate of scalar Hegselmann–Krause dynamics. In Proceedings of the American Control Conference, Washington, DC, USA, 17–19 June 2013; pp. 206–210.
10. Stamoulas, S.T.; Rathinam, M. Convergence, stability, and robustness of multidimensional opinion dynamics in continuous time. *SIAM J. Control Optim.* **2018**, *56*, 1938–1967. [\[CrossRef\]](#)
11. Hou, J.; Li, W.; Jiang, M. Opinion dynamics in modified expressed and private model with bounded confidence. *Phys. A* **2021**, *574*, 125968. [\[CrossRef\]](#)
12. Hickok, A.; Kureh, Y.; Brooks, H.Z.; Feng, M.; Porter, M.A. A bounded-confidence model of opinion dynamics on hypergraphs. *SIAM J. Appl. Dyn. Syst.* **2022**, *21*, 1–32. [\[CrossRef\]](#)
13. Peng, Y.; Zhao, Y.; Hu, J. On the role of community structure in evolution of opinion formation: A new bounded confidence opinion dynamics. *Inf. Sci.* **2023**, *621*, 672–690. [\[CrossRef\]](#)
14. Kan, U.; Feng, M.; Porter, M.A. An adaptive bounded-confidence model of opinion dynamics on networks. *J. Complex Netw.* **2023**, *11*, 415–444. Baccelli, F.; Chatterjee, A.; Vishwanath, S. Pairwise stochastic bounded confidence opinion dynamics: Heavy tails and stability. *IEEE Trans. Autom. Control* **2017**, *62*, 5678–5693. [\[CrossRef\]](#)
15. Baccelli, F.; Chatterjee, A.; Vishwanath, S. Pairwise stochastic bounded confidence opinion dynamics: Heavy tails and stability. *IEEE Trans. Autom. Control* **2017**, *62*, 5678–5693.
16. Kolarijani, M.A.S.; Proskurnikov, A.V.; Esfahani, P.M. Macroscopic noisy bounded confidence models with distributed radical opinions. *IEEE Trans. Autom. Control* **2021**, *66*, 1174–1189. [\[CrossRef\]](#)
17. Su, W.; Wang, X.; Chen, G.; Yu, Y.; Hadzibeganovic, T. Noise-based synchronization of bounded confidence opinion dynamics in heterogeneous time-varying communication networks. *Inf. Sci.* **2020**, *528*, 219–230. [\[CrossRef\]](#)
18. Shi, L.; Cheng, Y.; Shao, J.; Wang, X.; Sheng, H. Leader–follower opinion dynamics of signed social networks with asynchronous trust/distrust level evolution. *IEEE Trans. Netw. Sci. Eng.* **2022**, *9*, 495–509. [\[CrossRef\]](#)
19. Berenbrink, P.; Hoefer, M.; Kaaser, D.; Lenzner, P.; Rau, M.; Schmand, D. Asynchronous opinion dynamics in social networks. *Distrib. Comput.* **2024**, *37*, 207–224. [\[CrossRef\]](#)
20. Bu, Z.; Li, H.J.; Zhang, C.; Cao, J.; Li, A.; Shi, Y. Graph k -means based on leader identification, dynamic game, and opinion dynamics. *IEEE Trans. Knowl. Data Eng.* **2020**, *32*, 1348–1361. [\[CrossRef\]](#)

21. Jiang, J.; Wang, L. A study of the influence of different opinion leaders on the evolution of followers' views in social e-commerce networks. In Proceedings of the International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES), Anyang, China, 10–12 December 2021; pp. 49–52.
22. Wang, G.; Huang, N.; Zhang, L. Optimal control and dynamic analysis of a new Atangana–Baleanu fractional rumor dissemination model involving media. *Nonlinear Anal. Real World Appl.* **2026**, *87*, 104457. [[CrossRef](#)]
23. Altafini, C.; Ceragioli, F. Signed bounded confidence models for opinion dynamics. *Automatica* **2018**, *93*, 114–125. [[CrossRef](#)]
24. He, G.; Liu, J.; Hu, H.; Fang, J.A. Discrete-time signed bounded confidence model for opinion dynamics. *Neurocomputing* **2021**, *425*, 53–61. [[CrossRef](#)]
25. Brehmer, B. Social judgment theory and the analysis of interpersonal conflict. *Psychol. Bull.* **1976**, *83*, 985–1003. [[CrossRef](#)]
26. Ceragioli, F.; Frasca, P. Continuous and discontinuous opinion dynamics with bounded confidence. *Nonlinear Anal. Real World Appl.* **2012**, *13*, 1239–1251. [[CrossRef](#)]
27. Urbig, D.; Lorenz, J.; Herzberg, H. Opinion dynamics: The effect of the number of peers met at once. *J. Artif. Soc. Soc. Simul.* **2008**, *11*, 4.
28. Kunegis, J.; Lommatzsch, A.; Bauckhage, C. The Slashdot Zoo: Mining a social network with negative edges. In Proceedings of the 18th International Conference on World Wide Web (WWW), Madrid, Spain, 20–24 April 2009; pp. 741–750.
29. Leskovec, J.; Krevl, A. SNAP Datasets: Stanford Large Network Dataset Collection. Available online: <http://snap.stanford.edu/data> (accessed on 1 June 2026).
30. West, R.; Paskov, H.S.; Leskovec, J.; Potts, C. Exploiting social network structure for person-to-person sentiment analysis. *Trans. Assoc. Comput. Linguist.* **2014**, *2*, 297–310. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.