# Opinion Dynamics in Social Media: Models, Analysis, and Tools for Understanding Digital Polarization in Crisis Contexts

May 30, 2025

**Abstract**

The proliferation of social media platforms has fundamentally transformed how opinions form, spread, and polarize in modern society. This comprehensive survey examines the state-of-the-art research on opinion dynamics in social media, with particular emphasis on models that incorporate external factors such as epidemics, emergencies, and other critical events. We systematically review computational and mathematical frameworks used to model opinion formation, including bounded confidence models, epidemic-like approaches, agent-based simulations, and hybrid methodologies. Our analysis reveals that dynamic network structures predominate over static configurations, with multilayer and platform-specific networks playing increasingly important roles. External factors—including media influence, algorithmic filtering, fake news propagation, and epidemic contexts—are explicitly incorporated across numerous studies, demonstrating the field's evolution toward more realistic and contextually aware models. Key parameters identified across the literature include network topology characteristics, confidence thresholds, transmission and convergence rates, influence weights, and susceptibility measures. Empirical validation approaches range from simulation experiments to sophisticated statistical fits using real-world data from platforms such as Twitter, Facebook, and Reddit. We conclude by proposing a novel data-informed model that integrates political community structures with external crisis fields, specifically designed to capture polarization dynamics during events like the COVID-19 pandemic. This work provides a foundation for developing parametric, open-source software implementations that can advance both theoretical understanding and practical applications in the domain of digital opinion dynamics.

## 1 Introduction

The digital revolution has fundamentally altered the landscape of human communication and opinion formation. Social media platforms have emerged as primary venues for information exchange, debate, and the formation of collective beliefs, creating unprecedented opportunities to observe and model opinion dynamics at scale [36]. Unlike traditional media environments, these digital ecosystems exhibit complex network structures, algorithmic mediation, and real-time interaction patterns that challenge conventional models of opinion formation and social influence.

The importance of understanding opinion dynamics in social media has been amplified by recent global events, including political polarization, the spread of misinformation, and the COVID-19 pandemic [9]. These phenomena have demonstrated how digital platforms can both facilitate rapid information dissemination and contribute to the formation of echo chambers, conspiracy theories, and deeply polarized communities [4]. The consequences extend far beyond the digital realm, influencing public health responses, democratic processes, and social cohesion.

Traditional models of opinion dynamics, such as the voter model, the Ising model, and bounded confidence approaches, were primarily developed for understanding face-to-face interactions or simple network structures [12]. While these foundational frameworks provide important theoretical insights, they often fail to capture the complexity of modern digital communication environments. Social media platforms introduce several unique characteristics that require specialized modeling approaches: algorithmic content curation that creates filter bubbles [34], the presence of automated agents (bots) that can artificially amplify certain viewpoints [8], the multi-layered nature of platform interactions, and the rapid temporal dynamics of viral content spread.

Furthermore, the integration of external factors—such as major news events, natural disasters, pandemics, or political crises—into opinion dynamics models represents a critical frontier in the field [14]. These external shocks can dramatically alter the parameters of opinion formation, potentially breaking down established community structures, accelerating polarization processes, or catalyzing consensus formation around shared concerns. The COVID-19 pandemic, in particular, has provided a natural experiment for observing how external crises interact with digital opinion dynamics, influencing everything from vaccine hesitancy to political trust [16].

The development of computational models that can accurately capture these complex dynamics is not merely an academic exercise. Such models have practical applications in understanding and potentially mitigating the spread of misinformation, designing more effective public health communication strategies, predicting social unrest, and developing policy interventions that promote healthier democratic discourse. However, the diversity of modeling approaches, parameter choices, and validation methodologies in the current literature makes it challenging to synthesize findings and develop robust, generalizable tools.

This survey addresses these challenges by providing a comprehensive review of the current state-of-the-art in opinion dynamics modeling for social media contexts. We systematically examine the various modeling paradigms employed in the literature, analyze the key parameters and network structures that shape opinion formation processes, and evaluate the empirical validation approaches used to test these models against real-world data. Our analysis reveals both the significant progress made in the field and the remaining gaps that future research should address.

A particular focus of this review is on models that incorporate external factors, recognizing that opinion dynamics in social media rarely occur in isolation from broader social, political, and environmental contexts [13]. We examine how researchers have modeled the influence of media campaigns, algorithmic interventions, epidemic contexts, and other external perturbations on opinion formation processes. This analysis provides insights into the mechanisms through which external events can reshape digital discourse and offers lessons for understanding future crisis communications.

The survey is structured to serve multiple audiences: researchers seeking to understand the current landscape of opinion dynamics modeling, practitioners interested in applying these models to real-world problems, and software developers working to create tools that implement these theoretical frameworks. To support the latter goal, we pay particular attention to the specific parameters, computational requirements, and implementation considerations that emerge from the literature review.

Our contribution extends beyond synthesis to propose a novel modeling framework that addresses several limitations in current approaches. We present a data-informed model that integrates political community structures with external crisis fields, specifically designed to capture the polarization dynamics observed during events like the COVID-19 pandemic. This model builds on insights from the reviewed literature while addressing the need for approaches that can be empirically calibrated using real social media data and validated against observed crisis responses.

The remainder of this survey is organized as follows. Section 2 provides background on the fundamental concepts and challenges in opinion dynamics modeling. Section 3 presents our systematic analysis of current modeling approaches, categorizing them by methodology and examining their key characteristics. Section 4 focuses specifically on how external factors are incorporated into these models. Section 5 analyzes the network structures and interaction mechanisms employed across the literature. Section 6 examines validation methodologies and empirical findings. Section 7 synthesizes the key insights and identifies patterns across studies. Finally, Section 8 presents our proposed framework for modeling opinion dynamics with external crisis fields, and Section 9 concludes with directions for future research.

# 2 Background and Conceptual Framework

## 2.1 Fundamental Concepts in Opinion Dynamics

Opinion dynamics, as a field of study, concerns itself with understanding how individual beliefs, preferences, and attitudes evolve through social interactions and external influences. In the context of social media, this evolution occurs within complex digital ecosystems that mediate, amplify, and sometimes distort the natural processes of social influence. To understand the current state of research, it is essential to establish the key concepts that underpin modern opinion dynamics modeling.

At its core, an opinion can be conceptualized as a quantitative or qualitative representation of an individual's stance on a particular issue. Early models often simplified this to a scalar value on a continuous scale (e.g., -1 to +1 representing opposing viewpoints) or discrete states (e.g., binary pro/con positions) [46]. However, social media contexts have revealed the limitations of such simplifications, leading to more sophisticated representations that can capture multi-dimensional opinion spaces, uncertainty, and temporal variations in belief strength.

Social influence, the mechanism through which individual opinions change through interaction with others, operates through several psychological and social processes. Homophily—the tendency for similar individuals to interact more frequently—creates clustering patterns that can lead to echo chambers and polarization [18]. Conversely, heterophilous interactions, while less common, can promote consensus or at least moderate extreme positions. The strength of social influence is often modeled as a function of factors such as social distance, trust, expertise, and frequency of interaction.

Network structure plays a crucial role in determining how influence propagates through a population. Traditional models often assumed complete mixing (everyone interacts with everyone) or simple regular networks. Social media platforms, however, exhibit complex topologies characterized by power-law degree distributions, community structure, and temporal dynamics [42]. These structural features can dramatically alter opinion dynamics, creating bottlenecks that slow consensus formation or bridges that enable rapid opinion spread.

The concept of bounded confidence, introduced by [28], has proven particularly relevant to social media contexts. This principle suggests that individuals are only influenced by others whose opinions fall within a certain range of their own beliefs. When confidence bounds are narrow, populations tend to fragment into polarized clusters; when they are wide, consensus becomes more likely. Social media algorithms, by curating content based on user preferences, effectively modify these confidence bounds in complex ways.

## 2.2 Unique Characteristics of Social Media Environments

Social media platforms introduce several characteristics that distinguish them from traditional social influence contexts and necessitate specialized modeling approaches. Understanding these characteristics is crucial for developing realistic models of digital opinion dynamics.

Algorithmic mediation represents perhaps the most significant departure from traditional social influence contexts. Platforms like Facebook, Twitter, and YouTube use sophisticated algorithms to determine which content users see, fundamentally altering the natural flow of information and social influence [40]. These algorithms typically optimize for engagement, which can inadvertently promote controversial or emotionally charged content, potentially accelerating polarization processes. The feedback loops between user behavior and algorithmic content curation create complex dynamics that are difficult to predict and model.

The scale and speed of social media interactions create opportunities for rapid opinion cascades and viral phenomena that would be impossible in face-to-face contexts [29]. A single post can reach millions of users within hours, potentially triggering widespread opinion shifts or solidifying existing divisions. This temporal compression of social influence processes challenges models that assume gradual, local propagation of opinions.

The presence of automated agents—social bots—introduces artificial influence sources that can manipulate opinion dynamics in systematic ways [8]. These agents can amplify certain viewpoints, create the illusion of grassroots support for particular positions, or inject misinformation into ongoing debates. Their behavior is often coordinated and strategic, rather than following the psychological principles that govern human opinion change.

Anonymity and reduced social cues in online interactions can lead to different influence patterns compared to face-to-face communication. The lack of visual and social context can make users more susceptible to certain types of influence while reducing the effectiveness of others. Additionally, the ability to interact with distant others creates influence networks that span geographic and cultural boundaries in unprecedented ways.

Multi-platform dynamics add another layer of complexity, as users often maintain presences across multiple social media platforms simultaneously. Opinion formation may occur differently across platforms due to varying algorithms, user bases, and interaction patterns, yet these processes are interconnected through users who bridge platforms.

### 2.3 External Factors and Crisis Contexts

One of the most important developments in opinion dynamics research has been the recognition that social influence processes rarely occur in isolation from broader social, political, and environmental contexts [45]. External events—ranging from natural disasters to political scandals to global pandemics—can dramatically alter the parameters of opinion formation, sometimes overriding established patterns of social influence.

Crisis contexts, in particular, have been shown to accelerate opinion dynamics and potentially reshape long-standing community structures. During the COVID-19 pandemic, for example, public health information, political responses, and individual risk perceptions created a complex field of external influences that interacted with existing social networks and opinion patterns [38]. Understanding these interactions is crucial for predicting how societies will respond to future crises and for designing effective communication strategies during emergencies.

The concept of an "external field"—borrowed from physics models of phase transitions—provides a useful framework for thinking about how crises influence opinion dynamics. Just as magnetic fields can align spins in physical systems, external events can create pressures that push opinions in particular directions, potentially overcoming the local influence patterns that would otherwise dominate. The strength and direction of these external fields can vary over time, creating complex temporal dynamics that challenge simple models of opinion evolution.

## 3 Modeling Approaches and Frameworks

The diversity of modeling approaches employed in social media opinion dynamics research reflects both the complexity of the phenomena under study and the multidisciplinary nature of the field. Researchers have drawn from physics, computer science, sociology, psychology, and applied mathematics to develop frameworks capable of capturing different aspects of digital opinion formation. This section provides a comprehensive analysis of the major modeling paradigms identified in our literature review.

### 3.1 Mathematical and Analytical Models

Mathematical models provide the theoretical foundation for understanding opinion dynamics, offering analytical insights that complement computational approaches. These models typically represent opinions as continuous or discrete variables and describe their evolution through differential equations, Markov processes, or other mathematical frameworks.

The Friedkin-Johnsen model has emerged as one of the most influential frameworks in this category, appearing in five of the reviewed studies [1, 5, 17, 31, 32]. This model captures the persistence of disagreement in social networks by incorporating individual stubbornness parameters that resist social influence. In the context of social media, these parameters can represent users' commitment to their initial beliefs or their susceptibility to algorithmic filter bubbles. The model's mathematical tractability allows for analytical results about consensus formation and the effects of network structure on opinion outcomes.

The DeGroot model, used in three studies [1, 17, 46], represents a simpler but still powerful framework where individuals update their opinions as weighted averages of their neighbors' opinions. While the standard DeGroot model predicts eventual consensus under mild conditions, extensions that incorporate stubborn agents or time-varying networks can produce more realistic dynamics, including persistent polarization. The model's linear structure makes it particularly suitable for optimization problems, such as determining optimal intervention strategies.

Bounded confidence models, including the Deffuant and Hegselmann-Krause variants, capture the psychological principle that individuals are only influenced by others with sufficiently similar opinions [25, 40]. These models naturally produce polarization when confidence thresholds are below critical values, making them particularly relevant for understanding filter bubble effects in social media. The models can be analyzed using techniques from dynamical systems theory, revealing phase transitions between consensus and polarization regimes.

Kinetic approaches, exemplified by [2], bridge microscopic interaction rules with macroscopic population dynamics. These models describe the evolution of opinion distributions using partial differential equations derived from underlying interaction kernels. The kinetic framework is particularly powerful for analyzing large-scale systems where individual-level tracking becomes computationally prohibitive, while still maintaining theoretical rigor.

Mean-field approximations appear in several studies [19], providing analytical tractability for complex systems by replacing individual interactions with average field effects. These approaches are particularly useful for systems with large numbers of agents and complex interaction patterns, allowing researchers to derive analytical predictions about phase transitions and long-term behavior.

### 3.2 Agent-Based Models

Agent-based models (ABMs) represent the most popular approach in the reviewed literature, appearing in 21 studies. These computational frameworks simulate individual agents that interact according to specified rules, allowing for the emergence of complex collective behaviors from simple individual-level processes. The flexibility of ABMs makes them particularly well-suited for capturing the heterogeneity and complexity of social media environments.

The appeal of agent-based approaches lies in their ability to incorporate multiple factors simultaneously: individual psychology, network structure, algorithmic influences, and external events can all be represented explicitly. For example, [35] develops an agent-based model that incorporates both opinion dynamics and network topology evolution, allowing the study of co-evolutionary processes where opinions influence connection patterns and vice versa.

Several studies use agent-based frameworks to explore the effects of algorithmic filtering and personalization [15, 34]. These models typically implement recommendation algorithms that determine which content agents see, creating feedback loops between opinion evolution and information exposure. The computational nature of ABMs allows researchers to experiment with different algorithmic designs and study their effects on polarization and consensus formation.

Hybrid approaches that combine agent-based simulation with analytical elements are particularly common in epidemic-related studies [13, 16]. These models often implement compartmental structures (susceptible-infected-recovered type models) at the individual level while tracking population-level dynamics. The combination allows for detailed analysis of individual behavior changes while maintaining computational efficiency for large populations.

The validation of agent-based models presents unique challenges, as their complexity makes analytical analysis difficult. Most studies rely on computational experiments that explore parameter spaces and compare outcomes with empirical data. Some recent work has focused on developing systematic approaches to ABM calibration using machine learning techniques and statistical inference methods.

## 3.3 Epidemic-Like and Complex Contagion Models

The analogy between opinion spread and epidemic processes has proven particularly fruitful, leading to a class of models that adapt epidemiological frameworks to social influence contexts. These approaches are especially relevant for understanding how information, misinformation, and behavioral changes propagate through social networks.

Simple contagion models, where exposure to a single "infected" individual can cause opinion change, provide baseline frameworks for understanding rapid information spread [13]. However, social media contexts often exhibit complex contagion patterns where multiple exposures or reinforcement from multiple sources are required for opinion change. This has led to the development of threshold models where agents change opinions only after receiving influence from a minimum number of neighbors or after cumulative influence exceeds a threshold.

The integration of actual epidemic dynamics with opinion formation represents a particularly sophisticated development in the field [23, 38]. These models recognize that during health crises like COVID-19, epidemiological states (susceptible, infected, recovered) interact with opinion states (pro-vaccine, anti-vaccine, undecided) in complex ways. For example, infection experience might influence vaccine opinions, while vaccine opinions influence preventive behaviors that affect transmission rates.

Complex contagion models have been extended to capture the role of social reinforcement in opinion adoption. Unlike simple contagion, where single exposure can lead to adoption, complex contagion requires multiple exposures or confirmation from multiple sources [14]. This framework is particularly relevant for understanding how controversial or counterintuitive ideas spread, as they typically require more social validation than simple factual information.

Multilayer contagion models address the reality that users participate in multiple social networks simultaneously, and that contagion processes can occur across different platforms or social contexts. These models track opinion states across multiple network layers, allowing for cross-platform influence and the study of how opinion formation differs across different social media environments.

### 3.4 Hybrid and Multiscale Approaches

The complexity of social media opinion dynamics has led many researchers to develop hybrid approaches that combine elements from multiple modeling paradigms. These frameworks aim to capture both the microscopic details of individual interactions and the macroscopic patterns of population-level dynamics.

Ten studies in our review employ hybrid methodologies [13,14,16,24,26,30,36,41,46]. These approaches typically combine agent-based simulation for individual-level processes with analytical or statistical methods for population-level analysis. For example, [36] develops a model that uses agent-based simulation for social influence within communities while employing game-theoretic analysis for media competition between communities.

Multiscale approaches address the challenge of computational complexity in large-scale systems by using different modeling techniques at different scales. Individual-level interactions might be modeled using detailed psychological rules, while population-level dynamics are captured using mean-field approximations or kinetic equations. This approach allows researchers to maintain computational efficiency while preserving important microscopic details.

Machine learning integration represents an emerging trend in hybrid modeling, where traditional opinion dynamics models are combined with data-driven approaches. [30] develops a framework that learns opinion dynamics patterns from social media traces, combining generative models with traditional influence mechanisms. This approach allows for more accurate calibration to real-world data while maintaining interpretability.

Probabilistic frameworks, used in three studies [10, 11, 21], incorporate uncertainty and stochasticity into opinion dynamics models. These approaches are particularly important for capturing the inherent randomness in social media interactions and the uncertainty associated with opinion measurements from text data.

### 3.5 Optimization and Control Models

A specialized but important category of models focuses on optimization and control problems in opinion dynamics. These frameworks address questions about how external actors—whether platform designers, policymakers, or malicious agents—can influence opinion formation processes to achieve specific outcomes.

[17] develops models for understanding adversarial perturbations of opinion dynamics, analyzing how small, strategic interventions can lead to large changes in population-level outcomes. This work is particularly relevant for understanding how social media manipulation campaigns might operate and how they can be detected or countered.

Intervention optimization models, such as those developed by [20] and [31], address the problem of how to optimally allocate limited resources to influence opinion outcomes. These models can inform the design of public health campaigns, political interventions, or platform policy changes. The optimization framework allows for principled comparison of different intervention strategies and the identification of optimal targeting approaches.

Control-theoretic approaches treat opinion dynamics as dynamical systems that can be steered through external inputs [6]. These models are particularly useful for understanding how sustained interventions, such as media campaigns or algorithmic changes, can be designed to achieve desired long-term outcomes while accounting for system dynamics and feedback effects.

The integration of game-theoretic elements into opinion dynamics models addresses strategic interactions between multiple influence agents. These frameworks are relevant for understanding competition between different information sources, the strategic behavior of political actors, or the interaction between platform algorithms and user behavior.

# 4 Integration of External Factors

The recognition that opinion dynamics in social media occur within broader social, political, and environmental contexts has led to sophisticated approaches for incorporating external factors into computational models. This integration represents one of the most important developments in the field, as it bridges the gap between theoretical models of social influence and the complex realities of digital discourse during crises, emergencies, and major social events.

## 4.1 Media Influence and Competition

Media influence represents one of the most extensively studied external factors, appearing in seven of the reviewed studies. Traditional models of opinion dynamics often assume that influence occurs primarily through peer-to-peer interactions, but social media environments are characterized by the simultaneous presence of professional media content, user-generated content, and algorithmically curated information streams.

[36] develops a pioneering framework for modeling media competition in multilayer networks, where different media sources compete for attention and influence within and across communities. Their model demonstrates how media competition can lead to stable polarized cultures, even when individual users are not inherently partisan. The mathematical framework treats media sources as external influence agents that can shift opinion distributions, while communities respond strategically to maintain coherence and identity.

The concept of media pressure as an external field is explored by [18], who show how external media campaigns can overcome homophily-driven segregation and create temporary consensus around specific issues. Their agent-based model incorporates media influence as a global field that affects all agents simultaneously, but with varying effectiveness depending on individual susceptibility parameters. This approach reveals how media campaigns must reach critical intensity thresholds to overcome existing opinion clustering.

[19] presents a sophisticated mathematical framework for analyzing influencer and media strategies using partial differential equations and mean-field approaches. Their model treats external media influence as a control parameter that can be optimized to achieve specific opinion outcomes. The analytical nature of their approach allows for the derivation of optimal media strategies under different network conditions and population characteristics.

The integration of traditional media with social media dynamics is addressed by [33], who develop a bounded confidence model that incorporates both peer influence and mass media exposure. Their framework demonstrates how mass media can serve as a consensus-promoting force by providing common reference points that transcend local network structures, but can also amplify polarization when media sources themselves are polarized.

Recent work has begun to address the complex interplay between algorithmic content curation and traditional media influence. These hybrid approaches recognize that social media algorithms often prioritize content from traditional media sources, creating complex feedback loops between professional journalism, algorithmic amplification, and user engagement patterns.

## 4.2 Algorithmic Filtering and Personalization

The role of algorithmic content curation in shaping opinion dynamics has received increasing attention, with three studies explicitly modeling these effects. Social media algorithms, designed to maximize user engagement, can inadvertently create filter bubbles and echo chambers that accelerate polarization processes.

[34] develops a mechanistic model of algorithmic personalization that captures how recommendation systems modify the natural flow of social influence. Their framework treats algorithmic filtering as a dynamic process that evolves based on user behavior, creating feedback loops between opinion formation and content exposure. The model demonstrates how seemingly neutral algorithms that optimize for relevance can lead to increased polarization by reducing exposure to diverse viewpoints.

The concept of algorithmic bias is explored by [40], who extend bounded confidence models to incorporate systematic biases in content exposure. Their mathematical framework shows how small biases in algorithmic content selection can be amplified over time, leading to significant changes in population-level opinion distributions. The model provides analytical conditions for when algorithmic bias will lead to increased fragmentation versus consensus formation.

[15] develops an agent-based model that explicitly simulates social media algorithms as they interact with user behavior and opinion formation. Their approach allows for detailed analysis of how different algorithmic designs affect polarization outcomes, providing insights for platform designers seeking to promote healthier discourse environments.

Recent work has begun to explore how users adapt to algorithmic systems, potentially gaming or manipulating these systems to achieve desired information environments. These co-evolutionary approaches recognize that the relationship between algorithms and users is dynamic, with both sides adapting to each other over time.

## 4.3 Epidemic and Health Crisis Contexts

The COVID-19 pandemic has provided a natural experiment for understanding how health crises interact with social media opinion dynamics, leading to several innovative modeling approaches that integrate epidemiological and social influence processes.

[13] develops a hybrid model that combines SEIR (Susceptible-Exposed-Infected-Recovered) epidemic dynamics with opinion formation about health behaviors. Their framework demonstrates how social media can accelerate consensus formation around health behaviors, but also shows how misinformation can reduce risk awareness and undermine public health responses. The model incorporates feedback loops between epidemic states and opinion states, where infection experience influences health opinions, which in turn affect behaviors that influence transmission rates.

The role of awareness and attention during health crises is explored by [38], who develop a multiplex network model that tracks both epidemic spread and attention dynamics across multiple social media platforms. Their mathematical framework treats collective attention as a dynamical variable that influences both information processing and behavioral responses to health threats.

[16] presents a sophisticated kinetic approach to modeling COVID-19 vaccination hesitancy in Italy, incorporating both social influence mechanisms and external media effects. Their model demonstrates how fake news and misinformation can create persistent bimodal opinion distributions, even in the presence of authoritative health information. The framework allows for the analysis of different intervention strategies and their effectiveness under various network conditions.

Vaccination dynamics represent a particularly important application area, as vaccine adoption involves both individual risk assessment and social influence processes. [24] develops an ensemble approach that combines multiple opinion dynamics models to understand vaccine hesitancy, showing how different mechanisms (social pressure, information exposure, personal experience) contribute to vaccination decisions under different conditions.

The temporal dynamics of health crises create unique challenges for opinion dynamics modeling, as the salience and urgency of health issues change over time. Recent work has begun to explore how opinion dynamics models can incorporate time-varying external fields that represent the evolving nature of health threats and policy responses.

### 4.4 Misinformation and Fake News

The spread of misinformation represents one of the most pressing challenges in social media environments, leading to specialized modeling approaches that capture the unique characteristics of false information propagation.

[39] develops a minimalistic model that incorporates bias, polarization, and misinformation in a unified framework. Their approach demonstrates how misinformation can exploit existing biases and accelerate polarization processes, creating self-reinforcing cycles where false information becomes increasingly accepted within polarized communities.

The interaction between misinformation and legitimate information is explored by [25], who show how opinion amplification mechanisms can cause extreme polarization when false information is present. Their agent-based model demonstrates that misinformation can be particularly effective when it confirms existing beliefs or biases, leading to asymmetric acceptance patterns across different communities.

[3] provides a theoretical framework for understanding how fake news propagates through social networks and interacts with existing polarization patterns. Their model shows how false information can be strategically deployed to exacerbate existing divisions or create new ones, with effectiveness depending on network structure and initial opinion distributions.

Recent work has begun to explore how fact-checking and content moderation efforts can be incorporated into opinion dynamics models. These approaches recognize that platforms actively intervene in information flows through labeling, removal, or algorithmic down-ranking of false content, creating complex dynamics between misinformation producers, platform responses, and user behavior.

### 4.5 Social Bots and Artificial Agents

The presence of automated agents in social media environments introduces artificial influence sources that can systematically manipulate opinion dynamics. [8] develops a model of how social bots interfere with public opinion formation, showing how coordinated bot networks can create the illusion of grassroots support for particular viewpoints.

Bot influence operates through several mechanisms: amplification of certain content, creation of artificial social proof, and manipulation of trending topics and visibility metrics. These artificial agents can be strategically deployed to influence opinion formation at critical moments, such as during elections or public health crises.

The detection and mitigation of bot influence represents an active area of research, with models being developed to understand how bot networks operate and how their effects can be countered. These approaches often incorporate game-theoretic elements, as bot operators and platform defenders engage in strategic interactions.

### 4.6 Policy Interventions and Social Norms

The modeling of policy interventions represents an important application area for opinion dynamics research, as policymakers seek evidence-based approaches to promoting healthy discourse and countering harmful information spread.

[6] develops optimization frameworks for understanding how centralized interventions can influence opinion formation in communities with complex structures. Their mathematical approach provides analytical insights into the effectiveness of different intervention strategies and the conditions under which consensus or diversity can be promoted.

The role of social norms in opinion dynamics is explored through models that incorporate normative pressure as an external influence factor. These approaches recognize that opinion formation occurs within broader cultural and institutional contexts that shape what opinions are considered acceptable or socially desirable.

Recent work has begun to explore how platform design changes—such as modifications to user interface elements, recommendation algorithms, or social feedback mechanisms—can be modeled as external interventions that influence opinion dynamics. These approaches provide insights for evidence-based platform governance and the design of systems that promote constructive discourse.

## 5 Network Structures and Interaction Mechanisms

The structure of social networks and the mechanisms through which agents interact within these networks fundamentally determine how opinions form, spread, and evolve in social media environments. Our analysis reveals that network structure effects represent one of the most sophisticated aspects of current opinion dynamics research, with the field having moved far beyond simple assumptions of random mixing or regular lattices to embrace the complex, dynamic, and multilayered nature of real social media networks.

### 5.1 Dynamic versus Static Network Structures

A striking finding from our literature review is the predominance of dynamic network structures, which appear in 27 studies compared to only 7 that employ static configurations. This shift reflects the recognition that social media networks are inherently dynamic, with users continuously forming new connections, breaking existing ones, and modifying the strength of their relationships based on ongoing interactions and opinion alignment.

Dynamic networks introduce several important phenomena that cannot be captured by static models. Homophily-driven link formation, where agents preferentially connect to others with similar opinions, can lead to increasing segregation over time even when initial conditions are relatively mixed [18]. Conversely, heterophily-driven processes, while less common, can promote opinion diversity and consensus formation by maintaining bridges between different communities.

[26] presents a sophisticated analysis of opinion polarization in coevolving networks, where both opinions and network structure evolve simultaneously according to coupled dynamics. Their model demonstrates that network coevolution can lead to stable bipolarization patterns that persist even when external perturbations attempt to promote consensus. The mathematical framework reveals phase transitions between consensus and polarization regimes that depend on the relative timescales of opinion updates and network rewiring.

The temporal dynamics of network evolution introduce memory effects that can significantly influence opinion outcomes. [5] develops a framework for analyzing local edge dynamics, showing how the history of previous connections and interactions influences current opinion formation processes. Their approach reveals that recommendation systems and confirmation bias can create feedback loops that amplify small initial differences into large-scale polarization patterns.

Activity-driven networks represent a particularly important class of dynamic structures that capture the bursty nature of social media interactions [4]. These models recognize that users are not continuously active but rather engage in intermittent bursts of activity that can dramatically alter local network structure and influence patterns. The integration of activity dynamics with opinion formation reveals how temporal patterns of engagement can influence polarization outcomes.

### 5.2 Multilayer and Platform-Specific Networks

The recognition that users participate in multiple social networks simultaneously has led to the development of multilayer network models that capture cross-platform dynamics and the interaction between different types of social relationships. Five studies in our review explicitly employ multilayer frameworks, reflecting the growing sophistication of network modeling approaches in capturing the complexity of modern digital communication environments.

[36] pioneers the use of multilayer networks in opinion dynamics by modeling media competition across different communication channels. Their framework distinguishes between a media layer, represented as a complete graph where different media sources compete for attention, and a social layer with scale-free or small-world topology where users interact directly. The interaction between these layers creates complex dynamics where media influence and peer influence compete and interact in non-trivial ways.

The concept of multiplex social networks, where the same set of users maintain different types of relationships across multiple platforms, is explored by [38]. Their model incorporates weighted connections that represent different interaction modalities (likes, shares, comments, direct messages) and shows how the relative importance of these different interaction types can influence the speed and direction of opinion changes.

Platform-specific modeling has emerged as users increasingly specialize their behavior across different social media environments. Twitter networks, characterized by their asymmetric following relationships and real-time information sharing, exhibit different opinion dynamics compared to Facebook networks with their emphasis on reciprocal friendship connections and curated content sharing [47]. These platform differences necessitate specialized modeling approaches that capture the unique affordances and constraints of each environment.

Reddit represents a particularly interesting case for multilayer modeling due to its subreddit structure, where users participate in multiple topical communities with different norms, cultures, and discussion patterns [27]. The modeling of opinion dynamics in such hierarchically structured environments requires frameworks that can capture both local community effects and global platform-wide phenomena.

### 5.3 Community Structure and Polarization Dynamics

Community structure—the tendency for networks to organize into densely connected clusters with sparse connections between clusters—plays a crucial role in opinion dynamics by creating natural boundaries for opinion propagation and potential sources of polarization.

[6] develops sophisticated mathematical frameworks for analyzing opinion manipulation in networks with community structure. Their approach shows how community boundaries can be exploited by external actors seeking to influence opinion formation, with different intervention strategies being optimal for different community structures. The mathematical analysis reveals that communities with strong internal cohesion but weak inter-community connections are particularly vulnerable to targeted manipulation campaigns.

The emergence of echo chambers—communities where members are primarily exposed to information that confirms their existing beliefs—represents one of the most studied phenomena in social media opinion dynamics [7]. Echo chamber formation can occur through multiple mechanisms: algorithmic filtering that reduces exposure to diverse content, homophily-driven network formation that brings similar users together, and confirmation bias that leads users to selectively attend to agreeable information.

[4] presents a comprehensive analysis of echo chamber formation and polarization dynamics, showing how the interplay between network structure and opinion updating rules can lead to different types of polarization patterns. Their model distinguishes between opinion polarization (divergence of average opinions between groups) and network polarization (structural separation between groups), showing that these two types of polarization can occur independently and interact in complex ways.

The role of bridge nodes—users who maintain connections across different communities—has received particular attention as these individuals can either facilitate consensus formation or serve as conduits for conflict escalation [45]. The strategic importance of bridge nodes makes them natural targets for intervention strategies, but also potential points of vulnerability for adversarial attacks.

## 5.4 Hub Structures and Influence Hierarchies

The heterogeneous nature of social media networks means that some users have disproportionate influence due to their large numbers of followers, high engagement rates, or strategic network positions. Understanding how these influential users affect opinion dynamics is crucial for predicting and potentially controlling opinion formation processes.

Hub structures in social media networks often follow power-law distributions, where a small number of users have extremely large numbers of connections while most users have relatively few connections. This heterogeneity can dramatically accelerate opinion propagation when influential users adopt and promote particular viewpoints [44]. The mathematical analysis of opinion dynamics on power-law networks reveals phase transitions where small changes in hub behavior can lead to dramatic shifts in population-level opinion distributions.

The temporal dynamics of influence hierarchies add another layer of complexity, as user influence can change rapidly based on the virality of their content, their position in trending topics, or their involvement in current events. [19] develops frameworks for analyzing how influencers and media figures can strategically time their interventions to maximize their impact on opinion formation processes.

The interaction between organic influence (based on genuine social connections and trust) and artificial influence (created through purchased followers, bot networks, or algorithmic manipulation) creates complex dynamics that challenge traditional models of social influence. Recent work has begun to explore how these different types of influence interact and how authentic influence networks can be distinguished from artificial ones.

## 5.5 Interaction Mechanisms and Social Influence Rules

Beyond network structure, the specific rules governing how agents interact and influence each other play crucial roles in determining opinion dynamics outcomes. The field has developed increasingly sophisticated models of social influence that go beyond simple averaging or majority rules.

Confidence-based influence mechanisms, where agents are only influenced by others whose opinions fall within certain bounds of their own beliefs, have proven particularly relevant for social media contexts [28]. These mechanisms naturally produce clustering and polarization effects, as agents become increasingly isolated from those with dissimilar opinions. The mathematical analysis of bounded confidence models reveals critical threshold effects where small changes in confidence parameters can lead to qualitatively different outcomes.

Social proof mechanisms, where agents are influenced by the perceived popularity or social validation of particular opinions, capture important aspects of social media dynamics where likes, shares, and comments serve as signals of social approval [25]. These mechanisms can create herding effects where individuals adopt popular opinions regardless of their personal preferences or private information.

The integration of emotional and affective factors into influence mechanisms represents an important development in capturing the psychological realism of social media interactions [43]. Emotional contagion—the tendency for emotions to spread through social networks—can accelerate opinion changes and make them more resistant to rational counterarguments.

Reinforcement learning approaches, where agents adapt their opinion updating strategies based on the outcomes of previous interactions, capture the adaptive nature of social media users who learn to navigate complex information environments [43]. These approaches reveal how users can develop sophisticated strategies for information processing and social influence that go beyond simple heuristics.

### 5.6 Temporal Dynamics and Memory Effects

The temporal dimension of network interactions introduces memory effects and path dependence that can significantly influence opinion outcomes. Unlike static analyses that assume instantaneous equilibration, realistic models must account for the fact that social influence processes unfold over time and are influenced by the history of past interactions.

Temporal networks, where connections appear and disappear over time, create opportunities for opinion influence that depend on the precise timing of interactions [42]. The analysis of temporal network effects reveals that the order and timing of interactions can be as important as their existence, with early interactions having disproportionate influence on final outcomes.

Memory effects in opinion dynamics can occur at both individual and network levels. Individual memory affects how agents weight recent versus distant interactions, while network memory captures how past interaction patterns influence current network structure. The mathematical modeling of these memory effects requires sophisticated approaches that can capture both short-term fluctuations and long-term trends.

Burst dynamics, where periods of intense activity alternate with periods of relative quiescence, are characteristic of many social media environments and can dramatically influence opinion formation processes. The modeling of bursty activity patterns requires approaches that can capture the statistical properties of human activity while maintaining computational tractability for large-scale simulations.

## 6 Empirical Validation and Data-Driven Approaches

The validation of opinion dynamics models against real-world data represents one of the most challenging aspects of the field, requiring sophisticated methodological approaches that can bridge the gap between theoretical predictions and empirical observations. Our analysis reveals a diverse landscape of validation strategies, from controlled simulation experiments to sophisticated statistical fits using large-scale social media datasets.

### 6.1 Data Sources and Collection Methodologies

The choice of data sources fundamentally shapes both the types of models that can be validated and the conclusions that can be drawn about opinion dynamics in social media. Twitter emerges as the most frequently used platform in our reviewed studies, appearing in five validation studies

[1, 2, 4, 21, 47]. This preference reflects Twitter's relatively open API policies, the public nature of most tweets, and the platform's role as a primary venue for real-time public discourse.

[2] exemplifies sophisticated data-driven modeling by developing a kinetic framework that is directly calibrated using Twitter follower networks and sentiment analysis of tweet content. Their approach demonstrates how modern natural language processing techniques can be integrated with traditional opinion dynamics models to create data-informed frameworks that maintain theoretical rigor while achieving empirical accuracy.

Facebook data, despite being more difficult to obtain due to privacy restrictions, appears in three studies [7, 24, 26]. The Facebook studies often focus on specific controversial topics where public posts and comments provide insights into opinion formation processes. [7] conducts a particularly thorough analysis of echo chamber formation in Facebook discussions, using network analysis techniques to identify communities and track opinion evolution within and between these communities.

Reddit represents an increasingly important data source due to its hierarchical community structure and the availability of comprehensive historical data [27, 30]. The subreddit structure provides natural experiments for studying how opinions form within different community contexts, while the voting system provides explicit measures of opinion popularity and social validation.

The integration of multiple data sources has become increasingly common as researchers recognize that opinion formation processes often span multiple platforms. Cross-platform studies face additional methodological challenges in linking user identities across platforms and accounting for the different interaction patterns and user bases of each platform.

## 6.2 Statistical Validation Approaches

The statistical validation of opinion dynamics models requires sophisticated approaches that can account for the complex, high-dimensional nature of social media data while providing meaningful tests of theoretical predictions.

Goodness-of-fit measures represent the most straightforward approach to model validation, comparing predicted opinion distributions or network statistics with observed data. [42] demonstrates how analytical models can improve explanatory power for news sentiment data, using $R^2$ improvements and empirical network analysis to validate their theoretical predictions. Their approach shows how incorporating noise and topology effects into traditional models can significantly improve empirical fit.

Time series validation approaches compare predicted temporal dynamics with observed opinion evolution over time. [47] develops empirical comparison methodologies that track opinion formation processes in real-time, showing how their agent-based model reproduces key features of Twitter opinion evolution during major events.

Statistical fit approaches using advanced optimization techniques have become increasingly sophisticated. [2] employs nonlinear least-squares fitting combined with sentiment analysis to calibrate their kinetic model parameters using Twitter data. This approach demonstrates how modern computational tools can enable precise parameter estimation even for complex mathematical models.

Correlation analysis between model predictions and real-world outcomes provides another validation approach. [20] tests their stubborn agent optimization model by comparing predicted opinion shifts with actual Twitter sentiment changes during political events, using correlation measures to assess model accuracy.

The development of null model approaches for detecting significant cascade events represents an important methodological advance. These approaches establish statistical baselines for opinion spread that account for network structure and user activity patterns, allowing researchers to identify opinion cascades that are statistically significant beyond what would be expected from random processes.

### 6.3 Experimental and Simulation-Based Validation

While real-world data provides the ultimate test of model validity, controlled experiments and sophisticated simulation studies play important roles in understanding model behavior and testing theoretical predictions under controlled conditions.

Large-scale simulation studies allow researchers to explore parameter spaces and test model robustness under conditions that may be difficult to observe in real data. [34] conducts extensive simulation experiments to test their algorithmic personalization model, exploring how different parameter combinations affect polarization outcomes and identifying critical threshold effects.

Comparative simulation studies, where multiple models are tested against the same empirical phenomena, provide insights into the relative strengths and weaknesses of different theoretical approaches. [24] develops an ensemble approach that combines multiple opinion dynamics models and tests their collective performance against vaccination debate data from Facebook.

Controlled perturbation experiments, where models are tested against synthetic data with known ground truth, allow for precise assessment of model accuracy and bias. These approaches are particularly important for validating complex models where analytical analysis is difficult and empirical data may be noisy or incomplete.

The use of agent-based models as "virtual laboratories" for testing hypotheses about opinion dynamics has become increasingly sophisticated. These simulations can explore counterfactual scenarios (what would happen if network structure were different?) and test intervention strategies before implementing them in real systems.

### 6.4 Machine Learning and Data-Driven Modeling

The integration of machine learning techniques with traditional opinion dynamics modeling represents one of the most promising developments in the field, offering approaches that can learn complex patterns from data while maintaining interpretability and theoretical grounding.

[30] develops a hybrid framework that learns opinion dynamics patterns from social media traces, combining generative models with traditional influence mechanisms. Their approach demonstrates how modern machine learning techniques can be used to identify patterns in social media data that inform more accurate opinion dynamics models.

Deep learning approaches for opinion extraction from text data have revolutionized the field by enabling automated analysis of large-scale datasets. Sentiment analysis, topic modeling, and other NLP techniques allow researchers to extract opinion measurements from social media content at unprecedented scale and accuracy.

Reinforcement learning frameworks capture the adaptive nature of social media users who learn to navigate complex information environments. [43] incorporates reinforcement learning into their agent-based model, showing how users can develop sophisticated strategies for information processing and social influence.

Network embedding techniques provide new approaches for representing complex social media networks in mathematical models. These approaches can capture high-dimensional network structures in lower-dimensional spaces while preserving important topological properties that influence opinion dynamics.

### 6.5 Methodological Challenges and Limitations

Despite significant advances in validation methodologies, several important challenges remain in empirically validating opinion dynamics models.

The measurement of opinions from social media data introduces significant methodological challenges. Text-based sentiment analysis can be noisy and may not capture the nuanced, multi-dimensional nature of real opinions. The distinction between expressed opinions (what users post) and private beliefs (what users actually think) remains difficult to address systematically.

Temporal alignment between models and data presents another challenge, as theoretical models often assume continuous-time dynamics while empirical data is typically sampled at discrete intervals. The choice of temporal resolution can significantly affect validation results, and optimal sampling strategies remain an active area of research.

Selection bias in social media data represents a fundamental challenge for model validation. Users who are active on social media may not be representative of the broader population, and the types of opinions that are expressed publicly may systematically differ from private beliefs. This bias can lead to models that accurately predict social media behavior but fail to generalize to broader social influence processes.

The privacy and ethical constraints surrounding social media data collection create limitations on the types of validation studies that can be conducted. Researchers must balance the need for comprehensive data with respect for user privacy and platform terms of service.

Causal inference remains difficult in observational social media studies. While models may accurately predict correlations in the data, establishing that theoretical mechanisms actually cause observed outcomes requires sophisticated methodological approaches that account for confounding factors and alternative explanations.

## 7 Synthesis of Key Insights and Patterns

Our comprehensive analysis of the opinion dynamics literature reveals several fundamental patterns and insights that transcend individual studies and point toward underlying principles governing opinion formation in social media environments. This synthesis identifies convergent findings across different modeling approaches, highlights robust phenomena that appear consistently across studies, and reveals important gaps that future research should address.

## 7.1 Convergent Findings Across Modeling Approaches

Despite the diversity of modeling frameworks employed in the literature, several key findings emerge consistently across different theoretical approaches and empirical contexts. These convergent results suggest fundamental principles that govern opinion dynamics in social media environments.

The relationship between network structure and polarization outcomes represents one of the most robust findings in the literature. Across agent-based models, mathematical frameworks, and empirical studies, homophily-driven network formation consistently leads to increased polarization and echo chamber formation [4, 7, 18]. This finding holds regardless of the specific influence mechanism employed, suggesting that structural effects may be more important than detailed psychological assumptions in determining polarization outcomes.

The critical role of confidence thresholds or interaction boundaries appears consistently across bounded confidence models, epidemic-like frameworks, and hybrid approaches. Studies repeatedly demonstrate that small changes in these threshold parameters can lead to qualitatively different outcomes, with phase transitions between consensus and polarization regimes occurring at critical values [28, 40]. This finding has important implications for understanding how algorithmic filtering and content curation can influence population-level opinion dynamics.

The amplifying effects of algorithmic mediation on natural social influence processes represent another convergent finding. Whether modeled through explicit algorithmic filtering [34], recommendation system effects [5], or bias parameters in mathematical models [40], studies consistently show that algorithmic intervention can significantly alter opinion dynamics outcomes even when the intervention appears neutral or minimal.

External perturbations, whether modeled as media campaigns, crisis events, or adversarial attacks, consistently demonstrate the ability to override local influence patterns and create system-wide opinion changes [13, 17, 19]. This finding suggests that opinion dynamics systems, while often exhibiting stable local patterns, remain vulnerable to external shocks that can reshape the entire system.

## 7.2 Robust Phenomena and Universal Patterns

Several phenomena appear so consistently across studies that they can be considered robust features of social media opinion dynamics, representing universal patterns that transcend specific platforms, topics, or user populations.

Echo chamber formation emerges as perhaps the most universal phenomenon, appearing in eight separate studies across different modeling frameworks and empirical contexts [5, 7, 15, 25, 26, 34, 36, 40]. The consistency of this finding across different theoretical approaches suggests that echo chamber formation may be an emergent property of social media systems rather than the result of specific design choices or user behaviors.

Polarization amplification through opinion-based network rewiring represents another robust phenomenon. Studies consistently show that when users can modify their social connections based on opinion similarity, the resulting dynamics lead to increased segregation and more extreme opinion distributions [5, 26]. This finding has important implications for platform design and suggests that features enabling easy connection modification may inadvertently promote polarization.

The asymmetric effects of positive versus negative information appear consistently across studies of misinformation and fake news spread. Multiple studies demonstrate that false or negative information tends to spread faster and have more persistent effects than positive or correcting information [16, 25, 39]. This asymmetry appears to be a fundamental feature of human psychology that is amplified in social media environments.

Critical mass effects, where small initial advantages can be amplified into dominant positions through social influence processes, appear across epidemic-like models, cascade studies, and empirical analyses [14, 44]. These effects suggest that early intervention or influence may be disproportionately effective compared to later attempts to change established opinion patterns.

## 7.3 Parameter Sensitivity and Model Robustness

The analysis of parameter sensitivity across studies reveals important insights about which aspects of opinion dynamics models are most critical for producing realistic outcomes and which parameters may be less important for practical applications.

Network topology parameters consistently emerge as having strong effects on opinion dynamics outcomes. Studies across different modeling frameworks show that measures of network density, clustering, and degree heterogeneity significantly influence polarization and consensus formation [28, 42]. This finding suggests that accurate network representation is crucial for realistic opinion dynamics modeling.

Influence strength parameters, while important for quantitative predictions, often have less impact on qualitative outcomes than might be expected. Many studies find that opinion dynamics models are relatively robust to changes in influence strength parameters, with outcomes being determined more by network structure and threshold effects than by the precise magnitude of social influence [22, 46].

Temporal parameters, including update frequencies and memory effects, show complex and sometimes counterintuitive effects on opinion dynamics outcomes. Some studies find that slower update rates can actually accelerate consensus formation by reducing the impact of noise and temporary fluctuations [42], while others show that memory effects can either stabilize or destabilize opinion dynamics depending on network structure.

External influence parameters demonstrate high sensitivity across multiple studies, with small changes in media influence strength or algorithmic bias parameters leading to large changes in population-level outcomes [19, 33]. This sensitivity suggests that external intervention strategies may be highly effective but also that their effects may be difficult to predict precisely.

## 7.4 Scale and Generalizability Considerations

The question of how findings from computational models and small-scale empirical studies generalize to large-scale real-world opinion dynamics represents one of the most important challenges facing the field.

Population size effects appear in several studies, with different models showing varying sensitivity to the number of agents or users included in simulations. Some phenomena, such as critical mass effects and phase transitions, appear to be more pronounced in larger populations, while others, such as local influence patterns, may be relatively scale-invariant [2].

Platform-specific effects suggest that findings from one social media environment may not directly generalize to others. Studies comparing Twitter, Facebook, and Reddit dynamics reveal significant differences in opinion formation patterns that appear to be related to platform affordances and user behavior norms [7, 27, 47].

Cultural and temporal generalizability remains largely unexplored in the current literature. Most studies focus on specific events, topics, or time periods, making it difficult to assess whether findings represent universal principles or context-specific phenomena. The few studies that examine multiple time periods or cultural contexts suggest that significant variation may exist [37].

Cross-topic generalizability represents another important limitation. Studies often focus on particular controversial topics (politics, health, social issues) that may not be representative of opinion dynamics for less controversial or more technical topics. The mechanisms governing opinion formation about entertainment preferences, consumer products, or professional topics may differ significantly from those observed in political contexts.

### 7.5 Implications for Theory and Practice

The synthesis of findings across the reviewed literature suggests several important implications for both theoretical development and practical applications of opinion dynamics research.

For theoretical development, the convergent findings suggest that future models should prioritize network structure effects and external influence mechanisms over detailed individual-level psychological assumptions. The robustness of structural effects across different modeling approaches suggests that simplified models that accurately capture network dynamics may be more useful than complex models with sophisticated individual behavior but simplified network representations.

The consistent finding that small parameter changes can lead to qualitatively different outcomes suggests that opinion dynamics systems may be inherently difficult to predict precisely, even when general patterns are well understood. This has important implications for policy interventions and platform design changes, suggesting that pilot testing and careful monitoring may be essential for successful interventions.

For practical applications, the asymmetric effects of negative versus positive information suggest that correcting misinformation may require different strategies than promoting accurate information. The finding that false information tends to have persistent effects suggests that prevention may be more effective than correction, with important implications for content moderation and fact-checking strategies.

The critical role of network structure in determining opinion outcomes suggests that network-based interventions (connecting users across communities, promoting diverse information sources) may be more effective than content-based interventions (fact-checking, content labeling) for promoting healthy discourse environments.

# 8 A Data-informed model for Opinion Dynamics evolution during crises

Building on the existing literature and with the aim of addressing the critical gaps that we identified in current modeling approaches, we present a novel framework to model the complex dynamics of opinion formation and evolution on social media platforms during times of crisis. The proposed framework departs significantly from previous approaches in that it explicitly integrates political community structures with dynamically evolving crisis signals. The model parameters will be continuously calibrated against large-scale social media data streams.

## 8.1 Theoretical Foundation and Motivation

Our framework addresses several limitations found in previous modeling efforts. In particular, we incorporate the dynamical effects within well-established political groups, which are often treated as static entities rather than fluid networks capable of evolving under pressure. We also capture the inherently dynamic nature of crisis scenarios, which fluctuate in response to unfolding events, shifting media coverage, and evolving policy responses. Finally, we provide empirical validation to improve real-world applicability.

We adopt an inherently interdisciplinary approach that synthesizes insights from multiple fields of study. From physics, we borrow the concept of external fields that can align local interaction patterns, much like magnetic fields influence the behavior of spin systems in condensed matter physics. This concept proves particularly powerful when applied to social systems, where external crises can be viewed as forces that disrupt the clustering and interaction patterns within social networks. From sociology, we incorporate the crucial understanding of the role that political identity and community structure play in shaping how individuals process and respond to new information. Finally, from computer science, we leverage advanced real-time social media data processing capabilities that enable continuous calibration and validation of the model.

The COVID-19 pandemic provides an ideal natural laboratory for testing our framework. As a truly global crisis with measurable severity indicators, constantly shifting policy landscapes, clear politicization across different demographic and political groups, and an unprecedented volume of social media discourse, the pandemic offers rich empirical data for both model development and validation.

## 8.2 Model Architecture: A Comprehensive Framework

We model society through adaptive agents embedded within evolving networks. Each agent possesses multiple attributes that change over time in response to both local interactions and external, global crisis conditions.

Each agent in our model, described by a multidimensional state vector $\mathbf{s}_i(t)$, is characterized by four key attributes that evolve throughout the simulation period:

$$\mathbf{s}_i(t) = [o_i(t), c_i(t), a_i(t), h_i(t)]$$

First, we track their opinion stance $o_i(t) \in [-1, 1]$ regarding the specific crisis issue at hand—for instance, their level of support for public health measures during a pandemic. This opinion is not treated as a binary position but rather as a continuous variable that can shift gradually or dramatically depending on various influences. Second, we model community identity, which captures an individual's membership in political or social groups. This identity $c_i(t) \in \{1, 2, ..., K\}$ influences how they interpret information and which sources they find credible, but importantly, our framework allows for identity shifts during periods of crisis-induced social upheaval. Third, we monitor activity levels on social media platforms through $a_i(t) \in [0, 1]$, recognizing that more active users often have disproportionate influence on network dynamics and information flow. Finally, we track each individual's susceptibility to homophily via $h_i(t) \in [0, 1]$; this is their tendency to form connections with like-minded users, which can either reinforce existing beliefs or, when disrupted by crisis conditions, lead to unexpected alliance formations.

Opinion changes in our model emerge from the complex interplay of three distinct influence mechanisms:

$$\frac{do_i}{dt} = \alpha_{\text{social}} \sum_{j \in N_i(t)} w_{ij}(t)[o_j(t) - o_i(t)] + \alpha_{\text{community}} \mathbf{F}_{\text{comm}}(c_i, t) + \alpha_{\text{external}} \mathbf{F}_{\text{ext}}(t) \tag{1}$$

The first term in (1) measures peer interactions within an individual's evolving ego-network, which provide the most direct source of influence, as users encounter and respond to the opinions expressed by those in their immediate social circle. In the expression, $N_i(t)$ represents individual $i$'s neighbors at time $t$, and $w_{ij}(t)$ represents the influence weight between individuals $i$ and $j$. The community influence term $\mathbf{F}_{\text{comm}}(c_i, t)$ encodes the idea that individuals experience pressure from the community's prevailing narrative, which may either reinforce or conflict with the messages they receive from immediate contacts. Finally, the overall crisis environment creates a dynamic backdrop that can amplify certain types of messages while suppressing others, controlled by the external field term $\mathbf{F}_{\text{ext}}(t)$.

The network structure in our framework is not static but evolves continuously through processes that reflect real social media behavior. Connections form most rapidly among highly active users who frequently engage with crisis-related content, creating dense clusters of information exchange around the most vocal participants in the discourse. More gradually, connections also develop among users who share similar opinions, reflecting the natural tendency toward homophily in social networks. This idea is quantified by an adaptive network model that takes the form

$$P(\text{link formation})_{ij} = \lambda_{\text{activity}} a_i(t) a_j(t) + \lambda_{\text{homophily}} h_i(t) h_j(t) \exp\left(-\frac{|o_i(t) - o_j(t)|^2}{2\theta^2_{\text{tolerance}}}\right)$$

Crucially, our model recognizes that an intense crisis can disrupt established social patterns. Long-standing community ties that were stable during normal times may weaken under crisis pressure, prompting some individuals to switch group allegiances or adopt more fluid, less committed group identities. This capability to model community transformation is what distinguishes our approach more clearly from previous frameworks that treat political affiliations as fixed characteristics. At the first stage, the dynamics of community membership are given by:

$$P(c_i(t+\Delta t)=k|c_i(t)=j)=\begin{cases}1-\beta_{\text{switch}}|\mathbf{F}_{\text{ext}}(t)| & \text{if } k=j\\ \frac{\beta_{\text{switch}}|\mathbf{F}_{\text{ext}}(t)|}{K-1} & \text{if } k\neq j\end{cases} \tag{2}$$

Equation (2), incorporates crisis-induced community switching, which depends on the intensity of the external field $\mathbf{F}\text{ext}(t)$ and is modulated by the parameter $\beta$switch. In our initial formulation, we assume that an agent is equally likely to switch to any other community; however, a plausible alternative is to assign higher switching probabilities toward like-minded communities.

Central to our framework is a representation of the external crisis signal that combines four measurable components, each carefully normalized to ensure comparability across different types of crises and time periods:

$$\mathbf{F}_{\text{ext}}(t)=\gamma_1 I(t)+\gamma_2 M(t)+\gamma_3 P(t)+\gamma_4 U(t)$$

We use objective "severity measures" to quantify aspects of the crisis itself: $I(t)$ represents case counts or hospitalization rates during a pandemic, economic indicators during a financial crisis, or casualty figures during a conflict. We also recognise that public perception of crisis severity is often as important as the above objective crisis measures: the media attention term $M(t)$ reflects the volume and intensity of news coverage, including traditional media articles and social media trending topics. We use policy intensity measures $P(t)$ to quantify the stringency and scope of official responses, including government announcements, regulatory changes, and institutional actions taken in response to the crisis. Finally, information uncertainty $U(t)$ quantifies the degree of conflicting expert opinions, policy reversals, and contradictory messaging that often characterizes crisis periods. This composite signal evolves as the crisis unfolds, influencing individual opinion formation while also modulating the strength of peer and community relations.

## 8.3 Data Integration and Empirical Calibration

Our framework relies on sophisticated real-time processing of social media streams to extract multiple types of information simultaneously.

For the opinion extraction, we employ state-of-the-art transformer-based models, specifically trained and validated for sentiment analysis and stance detection in crisis contexts. These models undergo continuous validation against survey data and human annotations to ensure accuracy across different crisis types and temporal contexts. Community identification occurs through advanced network clustering algorithms that analyze patterns of sharing, following, and engagement, with particular attention to ensuring temporal stability of community assignments while still allowing for crisis-induced changes in group membership. Network reconstruction captures both the rapid formation of new connections around crisis-related content and the slower evolution of underlying social structures through comprehensive analysis of follower relationships, reply patterns, retweets, and mentions. Activity measurements reflect normalized metrics of posting frequency and engagement levels across different platforms, which allow for a fair comparison between users with different baseline activity patterns and across platforms with different engagement norms.

For parameter estimation, we employ a hierarchical Bayesian approach that explicitly quantifies uncertainty at both population and individual levels. This approach recognizes that while some behavioral patterns may be consistent across large populations, individual users may exhibit significant deviations from average behavior that need to be captured and modeled appropriately.

We implement variational inference techniques combined with rigorous cross-validation procedures to ensure that our model scales effectively to large datasets while guarding against overfitting that can occur given the high-dimensional nature of social media data. Our parameter estimation carefully distinguishes between network effects driven by high activity levels and those arising from homophily-based connection formation, as these different mechanisms have distinct implications for information flow and opinion change.

Perhaps most importantly, our validation framework explicitly relates observed shifts in opinion distributions to the evolving crisis signal, accounting for both time lags between crisis events and opinion changes, and the nonlinear relationships that often characterize human responses to crisis conditions.

Our initial implementation focuses on Twitter data spanning from January 2020 through December 2022, and covering multiple countries to capture diverse political and cultural contexts in pandemic response. We concentrate particularly on attitudes toward public health measures and vaccination, as these topics generated substantial debate and clear opinion evolution throughout the pandemic period.

Community detection algorithms applied to pre-pandemic data provide a baseline measurement, allowing us to track how crisis-related content and discourse gradually reshaped existing political and social communities. Our crisis signal components draw from authoritative sources, including official health authorities for objective severity measures, comprehensive news aggregators for media attention metrics, policy tracking databases for governmental response intensity, and expert forecasting platforms to quantify information uncertainty levels.

## 8.4 Model Predictions

Our framework generates specific, testable predictions about how polarization evolves throughout different phases of a crisis. During the early crisis period, we hypothesize that high levels of objective threat tend to create temporary opinion convergence as the crisis signal becomes strong enough to overwhelm existing community divisions. This initial unity reflects a fundamental human tendency to coordinate responses when facing clear external threats.

As crises progress into their middle phases, however, we predict that opinions will increasingly polarize along pre-existing community lines. In this phase, political narratives become more established and communities develop distinct interpretive frameworks for interpreting crisis-related information. As a consequence, opinion positions can diverge significantly from what might be predicted from objective evidence, and community identity becomes a stronger predictor of individual opinions than factual information alone.

In the late crisis phase, we predict that fluctuations in crisis intensity, combined with policy reversals and changing information landscapes, will drive cyclical patterns of consensus and polarization. During this period, we also expect to observe realignment effects as some individuals, having experienced the limitations of their initial community responses, reconsider their group affiliations and potentially shift to different political or social communities.

These predictions offer insights into the temporal dynamics of crisis-driven opinion change that can inform both academic understanding and practical crisis communication strategies.

# 9 Discussion and Future Directions

Our comprehensive survey of opinion dynamics in social media reveals a field that has rapidly evolved from simple theoretical models to sophisticated frameworks capable of capturing the complex realities of digital discourse. The synthesis of findings across diverse modeling approaches demonstrates both the significant progress made and the important challenges that remain in understanding and predicting opinion formation in digital environments.

## 9.1 Theoretical Advances and Methodological Innovations

The field has made substantial theoretical advances in several key areas. The recognition that network structure and external influences often dominate individual-level psychological factors represents a fundamental shift from early opinion dynamics models that focused primarily on individual behavior. This insight has important implications for both theoretical development and practical applications, suggesting that interventions targeting network structure or information environments may be more effective than those attempting to change individual attitudes directly.

The development of multilayer and temporal network models represents another significant advance, acknowledging the complex, dynamic nature of social media environments. These frameworks have revealed phenomena such as cross-platform opinion contagion, temporal memory effects, and the importance of activity-driven network formation that would be invisible in simpler modeling approaches.

The integration of external factors—particularly crisis events, media influence, and algorithmic mediation—has moved the field toward more realistic and policy-relevant models. The COVID-19 pandemic has provided a natural experiment that has accelerated development in this area, leading to sophisticated frameworks that can capture the interaction between global crises and local social influence processes.

Methodological innovations in empirical validation have strengthened the field's connection to real-world phenomena. The development of sophisticated data processing pipelines, statistical validation approaches, and machine learning integration has enabled more rigorous testing of theoretical predictions against large-scale empirical data.

## 9.2 Persistent Challenges and Limitations

Despite significant progress, several fundamental challenges continue to limit the field's development and practical impact. The measurement of opinions from social media data remains problematic, with ongoing questions about the relationship between expressed opinions and private beliefs, the representativeness of social media users, and the accuracy of automated opinion extraction techniques.

The scalability and generalizability of current models represent another significant challenge. Most studies focus on specific platforms, time periods, or topics, making it difficult to assess whether findings represent universal principles or context-specific phenomena. The development of models that can generalize across different social media environments and cultural contexts remains an important goal.

The causal interpretation of observational social media data continues to be problematic. While models may accurately predict correlations in the data, establishing that theoretical mechanisms actually cause observed outcomes requires sophisticated methodological approaches that are often difficult to implement with available data.

The dynamic nature of social media platforms themselves creates ongoing challenges for model validation and application. Platform algorithm changes, user interface modifications, and evolving user behavior patterns mean that models calibrated on historical data may not accurately predict future behavior.

## 9.3 Implications for Platform Design and Policy

The findings from our literature review have important implications for social media platform design and public policy. The consistent finding that algorithmic filtering and content curation can significantly influence opinion dynamics suggests that platforms bear substantial responsibility for the health of democratic discourse.

The research demonstrates that seemingly neutral algorithmic choices—such as optimizing for user engagement or personalizing content recommendations—can have profound effects on opinion polarization and echo chamber formation. This suggests that platform design decisions should be evaluated not only for their immediate user experience effects but also for their long-term impacts on social cohesion and democratic discourse.

The importance of network structure in determining opinion outcomes suggests that interventions promoting diverse social connections may be more effective than content-based approaches for improving discourse quality. Features that facilitate cross-community interaction, reduce algorithmic filtering, or promote exposure to diverse viewpoints may help counteract natural tendencies toward polarization and segregation.

The research on external crisis effects highlights the special responsibilities that platforms face during emergencies and major social events. The ability of external crises to rapidly reshape opinion landscapes suggests that platform policies and algorithmic behaviors during these periods may have disproportionate and lasting effects on social dynamics.

## 9.4 Research Priorities and Future Directions

Several priority areas emerge from our analysis as particularly important for future research development. The development of more sophisticated approaches to causal inference in social media data represents a critical need, as policy applications require understanding not just correlations but actual causal mechanisms.

The extension of current models to capture cross-platform dynamics and multi-modal information environments represents another important direction. As users increasingly consume information across multiple platforms and media types, models that can capture these complex information ecosystems will become increasingly important.

The integration of psychological and cognitive factors with network-based models represents a promising direction for future theoretical development. While current research has established the importance of structural effects, understanding how these interact with individual psychological processes may enable more accurate predictions and more effective interventions.

The development of real-time monitoring and intervention systems represents an important practical application area. The ability to detect emerging polarization, misinformation campaigns, or social unrest in real-time could enable more effective responses to social media-related social problems.

International and cross-cultural validation of opinion dynamics models represents a critical gap in current research. Most studies focus on Western, English-speaking populations, limiting our understanding of how cultural factors influence digital opinion formation processes.

### 9.5 Ethical Considerations and Responsible Research

The growing sophistication of opinion dynamics models raises important ethical questions about their potential applications and the responsibilities of researchers working in this area. The ability to predict and potentially manipulate public opinion through social media carries significant risks that must be carefully considered.

The dual-use nature of opinion dynamics research—where techniques developed for understanding social phenomena could also be used for manipulation or control—requires careful attention to research ethics and responsible disclosure practices. Researchers must balance the benefits of open scientific inquiry with the risks of enabling harmful applications.

The privacy implications of large-scale social media data analysis require ongoing attention, particularly as models become more sophisticated and potentially capable of inferring sensitive personal information from public social media behavior.

The potential for research findings to be misinterpreted or misapplied by policymakers or platform designers creates responsibilities for clear communication about model limitations, uncertainty ranges, and appropriate applications.

## 10 Conclusion

This comprehensive survey of opinion dynamics in social media has revealed a rapidly evolving field that has made significant theoretical and methodological advances while grappling with fundamental challenges posed by the complexity of digital communication environments. The analysis of 40 high-quality studies spanning multiple modeling paradigms and empirical contexts demonstrates both the substantial progress made and the important work that remains.

The convergent findings across different theoretical approaches—particularly the critical importance of network structure, the amplifying effects of algorithmic mediation, and the system-reshaping potential of external crises—suggest that fundamental principles governing digital opinion formation are beginning to emerge. These insights provide a foundation for developing more effective approaches to promoting healthy democratic discourse in digital environments.

The proposed framework for modeling opinion dynamics with external crisis fields represents a concrete step toward addressing several limitations in current approaches. By integrating political community structures with time-varying external influences and enabling empirical calibration using real social media data, this framework offers a path toward more realistic and policy-relevant opinion dynamics models.

The field's evolution toward more sophisticated, data-driven, and empirically validated approaches suggests that opinion dynamics research is maturing into a discipline capable of providing actionable insights for addressing some of the most pressing challenges facing democratic societies in the digital age. The development of open-source, parametric software tools will be crucial for realizing this potential by enabling broader access to sophisticated modeling capabilities and facilitating collaborative research efforts.

As social media platforms continue to evolve and new digital communication technologies emerge, the need for sophisticated understanding of opinion dynamics will only increase. The foundation provided by current research, combined with continued methodological innovation and empirical validation, offers hope that the field will be able to meet these challenges and contribute to the development of digital communication environments that support healthy democratic discourse and social cohesion.

The integration of crisis contexts into opinion dynamics modeling represents a particularly important development, as global challenges such as climate change, pandemics, and technological disruption require coordinated social responses that depend critically on effective digital communication. The ability to understand and predict how opinions form and evolve during crisis periods may be essential for society's ability to respond effectively to future challenges.

Ultimately, the research reviewed in this survey demonstrates that opinion dynamics in social media are neither random nor entirely predictable, but follow patterns that can be understood, modeled, and potentially influenced through thoughtful intervention. The continued development of this understanding represents both a scientific challenge and a social imperative for maintaining democratic discourse in an increasingly digital world.

# References


[1] Rediet Abebe, Jon Kleinberg, David Parkes, and Charalampos E. Tsourakakis. Opinion Dynamics with Varying Susceptibility to Persuasion. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1089–1098. ACM, jul 19 2018.

[2] Giacomo Albi, Elisa Calzola, and Giacomo Dimarco. A data-driven kinetic model for opinion dynamics with social network contacts. *arXiv*, 2023.

[3] Marina Azzimonti and Marcos Fernandes. Social Media Networks, Fake News, and Polarization. Technical report, National Bureau of Economic Research, 3 2018.

[4] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. Modeling Echo Chambers and Polarization Dynamics in Social Networks. *Physical Review Letters*, 124(4), jan 27 2020.

[5] Nikita Bhalla, Adam Lechowicz, and Cameron Musco. Local Edge Dynamics and Opinion Polarization. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 6–14. ACM, feb 27 2023.

[6] Paolo Bolzern, Alessandro Colombo, and Carlo Piccardi. Manipulating Opinions in Social Networks With Community Structure. *IEEE Transactions on Network Science and Engineering*, 11(1):185–196, 1 2024.

[7] Emanuele Brugnoli, Matteo Cinelli, Walter Quattrociocchi, and Antonio Scala. Recursive patterns in online echo chambers. *Scientific Reports*, 9(1), dec 27 2019.

[8] Chun Cheng, Yun Luo, and Changbin Yu. Dynamic mechanism of social bots interfering with public opinion in network. *Physica A: Statistical Mechanics and its Applications*, 551:124163, 8 2020.

[9] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. The COVID-19 social media infodemic. *Scientific Reports*, 10(1), oct 6 2020.

[10] A. De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya, and M. Gomez-Rodriguez. Learning and Forecasting Opinion Dynamics in Social Networks. *Neural Information Processing Systems*, 2015.

[11] A. De, Isabel Valera, Niloy Ganguly, Sourangshu Bhattacharya, and M. Gomez-Rodriguez. Modeling Opinion Dynamics in Diffusion Networks. *arXiv.org*, 2015.

[12] Michela Del Vicario, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. Modeling confirmation bias and polarization. *Scientific Reports*, 7(1), jan 11 2017.

[13] Erhu Du, Eddie Chen, Ji Liu, and Chunmiao Zheng. How do social media and individual behaviors affect epidemic transmission and control? *Science of The Total Environment*, 761:144114, 3 2021.

[14] Clara Eminente, Oriol Artime, and Manlio De Domenico. Interplay between exogenous triggers and endogenous behavioral changes in contagion processes on social networks. *Chaos, Solitons & Fractals*, 165:112759, 12 2022.

[15] Henrique Ferraz de Arruda, Felipe Maciel Cardoso, Guilherme Ferraz de Arruda, Alexis R. Hernández, Luciano da Fontoura Costa, and Yamir Moreno. Modelling how social network algorithms can influence opinion polarization. *Information Sciences*, 588:265–278, 4 2022.

[16] Jonathan Franceschi, Lorenzo Pareschi, Elena Bellodi, Marco Gavanelli, and Marco Bresadola. Modeling opinion polarization on social media: Application to Covid-19 vaccination hesitancy in Italy. *PLOS ONE*, 18(10):e0291993, oct 2 2023.

[17] Jason Gaitonde, Jon Kleinberg, and Éva Tardos. Adversarial Perturbations of Opinion Dynamics in Networks. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472. ACM, jul 13 2020.

[18] Floriana Gargiulo and Yerali Gandica. The Role of Homophily in the Emergence of Opinion Controversies. *Journal of Artificial Societies and Social Simulation*, 20(3), 2017.

[19] Luzie Helfmann, Nataša Djurdjevac Conrad, Philipp Lorenz-Spreen, and Christof Schütte. Modelling opinion dynamics under the impact of influencer and media strategies. *Scientific Reports*, 13(1), nov 8 2023.

[20] David Scott Hunter and Tauhid Zaman. Optimizing Opinions with Stubborn Agents. *Operations Research*, 70(4):2119–2137, 7 2022.

[21] Minkyoung Kim, David Newth, and Peter Christen. Modeling Dynamics of Diffusion Across Heterogeneous Social Networks: News Diffusion in Social Media. *Entropy*, 15(10):4215–4242, oct 8 2013.

[22] Ivan V. Kozitsin. Opinion dynamics of online social network users: a micro-level analysis. *The Journal of Mathematical Sociology*, 47(1):1–41, aug 4 2021.

[23] Viney Kumar and Samit Bhattacharyya. Nonlinear effect of sentiments and opinion sharing on vaccination decision in face of an outbreak: A multiplex network approach. *Chaos, Solitons & Fractals*, 175:114014, 10 2023.

[24] Jacopo Lenti and Giancarlo Ruffo. Ensemble of opinion dynamics models to understand the role of the undecided about vaccines. *Journal of Complex Networks*, 10(3), apr 25 2022.

[25] Soo Ling Lim and Peter J. Bentley. Opinion amplification causes extreme polarization in social networks. *Scientific Reports*, 12(1), oct 28 2022.

[26] Jiazhen Liu, Shengda Huang, Nathaniel M. Aden, Neil F. Johnson, and Chaoming Song. Emergence of Polarization in Coevolving Networks. *Physical Review Letters*, 130(3), jan 18 2023.

[27] A. Mancini, A. Desiderio, Riccardo Di Clemente, and G. Cimini. Self-induced emergence of consensus in social networks: Reddit and the GameStop short squeeze. *arXiv.org*, 2021.

[28] X. Flora Meng, Robert A. Van Gorder, and Mason A. Porter. Opinion formation and distribution in a bounded-confidence model on various networks. *Physical Review E*, 97(2), feb 22 2018.

[29] Bjarke Mønsted, Piotr Sapieżyński, Emilio Ferrara, and Sune Lehmann. Evidence of complex contagion of information in social media: An experiment using Twitter bots. *PLOS ONE*, 12(9):e0184148, sep 22 2017.

[30] Corrado Monti, Gianmarco De Francisci Morales, and Francesco Bonchi. Learning Opinion Dynamics From Social Traces. In *Proceedings of the 26th ACM SIGKDD International

*Conference on Knowledge Discovery & Data Mining*, pages 764–773. ACM, aug 20 2020.

[31] Cameron Musco, Christopher Musco, and Charalampos E. Tsourakakis. Minimizing Polarization and Disagreement in Social Networks. In *Proceedings of the 2018 World Wide Web Conference on World Wide Web - WWW '18*, pages 369–378. ACM Press, 2018.

[32] Charlotte Out, Sijing Tu, Stefan Neumann, and Ahad N. Zehmakan. The Impact of External Sources on the Friedkin–Johnsen Model. In *Proceedings of the 33rd ACM International Conference on Information and Knowledge Management*, pages 1815–1824. ACM, oct 21 2024.

[33] Valentina Pansanella, Alina Sîrbu, Janos Kertesz, and Giulio Rossetti. Mass media impact on opinion evolution in biased digital environments: a bounded confidence model. *Scientific Reports*, 13(1), sep 5 2023.

[34] Nicola Perra and Luis E. C. Rocha. Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific Reports*, 9(1), may 13 2019.

[35] Hafizh A. Prasetya and Tsuyoshi Murata. A model of opinion and propagation structure polarization in social media. *Computational Social Networks*, 7(1), jan 9 2020.

[36] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. Opinion dynamics on interacting networks: media competition and social influence. *Scientific Reports*, 4(1), may 27 2014.

[37] Armin A. Rad, Mohammad S. Jalali, and Hazhir Rahmandad. How exposure to different opinions impacts the life cycle of social media. *Annals of Operations Research*, 268(1-2):63–91, jun 23 2017.

[38] Marialisa Scata, Barbara Attanasio, Grazia Veronica Aiosa, and Aurelio La Corte. The Dynamical Interplay of Collective Attention, Awareness and Epidemics Spreading in the Multiplex Social Networks During COVID-19. *IEEE Access*, 8:189203–189223, 2020.

[39] Orowa Sikder, Robert E. Smith, Pierpaolo Vivo, and Giacomo Livan. A minimalistic model of bias, polarization and misinformation in social networks. *Scientific Reports*, 10(1), mar 26 2020.

[40] Alina Sîrbu, Dino Pedreschi, Fosca Giannotti, and János Kertész. Algorithmic bias amplifies opinion fragmentation and polarization: A bounded confidence model. *PLOS ONE*, 14(3):e0213246, mar 5 2019.

[41] Pawel Sobkowicz, Michael Kaschesky, and Guillaume Bouchard. *Opinion Formation in the Social Web: Agent-Based Simulations of Opinion Convergence and Divergence*, pages 288–303. Springer Berlin Heidelberg, 2012.

[42] Samuel Stern and Giacomo Livan. The impact of noise and topology on opinion dynamics in social networks. *Royal Society Open Science*, 8(4), 4 2021.

[43] Petter Törnberg, Claes Andersson, Kristian Lindgren, and Sven Banisch. Modeling the emergence of affective polarization in the social media society. *PLOS ONE*, 16(10):e0258259, oct 11 2021.

[44] Sijing Tu and Stefan Neumann. A Viral Marketing-Based Model For Opinion Dynamics in Online Social Networks. In *Proceedings of the ACM Web Conference 2022*, pages 1570–1578. ACM, apr 25 2022.

[45] Xin Wang, Antonio D. Sirianni, Shaoting Tang, Zhiming Zheng, and Feng Fu. Public Discourse and Social Network Echo Chambers Driven by Socio-Cognitive Biases. *Physical Review X*, 10(4), dec 1 2020.

[46] Zhibin Wu, Qinyue Zhou, Yucheng Dong, Jiuping Xu, Abdulrahman H. Altalhi, and Francisco Herrera. Mixed Opinion Dynamics Based on DeGroot Model and Hegselmann–Krause Model in Social Networks. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(1):296–308, 1 2023.

[47] Fei Xiong and Yun Liu. Opinion formation on social media: An empirical approach. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 24(1), mar 1 2014.