Patterns, Models, and Challenges in Online Social Media: A Survey

NICCOLÒ DI MARCO, Department of Legal, Social, and Educational Sciences, University of Tuscia, Viterbo, Italy
ANITA BONETTI, University of Rome La Sapienza, Rome, Italy
EDOARDO DI MARTINO, University of Rome La Sapienza, Rome, Italy
EDOARDO LORU, University of Rome La Sapienza, Rome, Italy
JACOPO NUDO, University of Rome La Sapienza, Rome, Italy
MARIO EDOARDO PANDOLFO, University of Rome La Sapienza, Rome, Italy
GIULIO PECILE, University of Rome La Sapienza, Rome, Italy
EMANUELE SANGIORGIO, University of Rome La Sapienza, Rome, Italy
IRENE SCALCO, University of Rome La Sapienza, Rome, Italy
SIMON ZOLLO, University of Rome La Sapienza, Rome, Italy
MATTEO CINELLI, University of Rome La Sapienza, Rome, Italy
FABIANA ZOLLO, Environmental Sciences, Informatics and Statistics, Universita Ca' Foscari, Venezia, Italy
WALTER QUATTROCIOCCHI, University of Rome La Sapienza, Rome, Italy

The rise of digital platforms has enabled the large-scale observation of individual and collective behavior through high-resolution interaction data. This development has opened new analytical pathways for investigating how information circulates, how opinions evolve, and how coordination emerges in online environments. Yet despite a growing body of research, the field remains fragmented, marked by methodological heterogeneity, limited model validation, and weak integration across domains. In this survey, we address this

The work is supported by IRIS Infodemic Coalition (UK government, grant no. SCH-00001-3391), SERICS (PE00000014) under the NRRP MUR program funded by the European Union - NextGenerationEU, project CRESP from the Italian Ministry of Health under the program CCM 2022, PON project "Ricerca e Innovazione" 2014-2020 and project SEED n. SP122184858BEDB3.
Authors' Contact Information: Niccolò Di Marco, Department of Legal, Social, and Educational Sciences, University of Tuscia, Viterbo, Italy; e-mail: niccolo.dimarco@unitus.it; Anita Bonetti, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: anita.bonetti@uniroma1.it; Edoardo Di Martino, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: edoardo.dimartino@uniroma1.it; Edoardo Loru, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: edoardo.loru@uniroma1.it; Jacopo Nudo, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: jacopo.nudo@uniroma1.it; Mario Edoardo Pandolfo, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: mario.edoardopandolfo@uniroma1.it; Giulio Pecile, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: giulio.pecile@uniroma1.it; Emanuele Sangiorgio, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: emanuele.sangiorgio@uniroma1.it; Irene Scalco, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: irene.scalco@uniroma1.it; Simon Zollo, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: simon.zollo@uniroma1.it; Matteo Cinelli, University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: matteo.cinelli@uniroma1.it; Fabiana Zollo, Environmental Sciences, Informatics and Statistics, Universita Ca' Foscari, Venezia, VE, Italy; e-mail: fabiana.zollo@unive.it; Walter Quattrociocchi (corresponding author), University of Rome La Sapienza, Rome, Lazio, Italy; e-mail: walter.quattrociocchi@uniroma1.it.

This work is licensed under a Creative Commons Attribution 4.0 International License.

© 2026 Copyright held by the owner/author(s).
ACM 1559-1131/2026/05-ART26
https://doi.org/10.1145/3796548

gap by systematically reviewing the literature on online collective behavior, integrating empirical findings with formal modeling approaches. We examine platform-level regularities, the methodological choices used to identify them, and the extent to which existing modeling frameworks capture the observed dynamics. Rather than aiming for exhaustive coverage of individual subfields, we provide a structural and comparative synthesis of recurring empirical patterns, methodological approaches, and modeling assumptions that span across platforms and domains. The overarching goal is to consolidate a shared empirical baseline and to clarify the structural constraints shaping inference in this area, thereby laying the groundwork for more robust, comparable, and actionable analyses of online social media.

CCS Concepts: • **Applied computing** → *Sociology*; • **Information systems** → World Wide Web; • **Security and privacy** → Human and societal aspects of security and privacy;

Additional Key Words and Phrases: Social media, online behaviour, computational social science, data science

ACM Reference Format:
Niccolò Di Marco, Anita Bonetti, Edoardo Di Martino, Edoardo Loru, Jacopo Nudo, Mario Edoardo Pandolfo, Giulio Pecile, Emanuele Sangiorgio, Irene Scalco, Simon Zollo, Matteo Cinelli, Fabiana Zollo, and Walter Quattrociocchi. 2026. Patterns, Models, and Challenges in Online Social Media: A Survey. *ACM Trans. Web* 20, 2, Article 26 (May 2026), 34 pages. https://doi.org/10.1145/3796548

1 Introduction

The widespread adoption of digital platforms has profoundly reshaped how individuals access information [1, 2], interact with content [3–5], and engage in public discourse [6, 7]. Social media platforms, in particular, mediate these interactions through algorithmic personalization [8], content ranking [9], and engagement-based feedback loops [10]. As a result, they have become not only channels of communication, but also active components in the structure of opinion formation, public debate, and collective behavior [11, 12].

Empirical research over the past decade has identified several recurring behavioral patterns associated with these systems [13]. Users tend to engage more with content that confirms their pre-existing views [1, 14–16], contributing to ideological polarization and reducing exposure to alternative perspectives [17–21]. False or misleading content can achieve wide circulation, often outperforming accurate information in terms of reach and engagement [1, 2]. These dynamics have been linked to downstream effects in diverse domains, including public health [22–27], mental health [25, 28–31], and political communication [12, 32–35].

A key factor enabling observations has been the increasing availability of large-scale digital behavioral data [36, 37]. In fact, logs of user interactions—such as likes, shares, and comments—have provided direct, time-stamped evidence of platform-mediated behavior at scale [36, 38–40].

This has led to a shift in the methodological foundations of social research: from small-scale, self-reported, or experimental studies toward empirical analyses grounded in observed behavior. This transition has catalyzed the growth of **computational social science** (**CSS**) [41–44], an interdisciplinary field that applies methods from network science, machine learning, and statistics to study social phenomena through digital traces.

However, the emergence of behavioral data as the main source of evidence has also introduced new challenges [37]. While it enables the detection of reproducible patterns and statistical regularities, it provides limited access to underlying cognitive processes or motivations [45]. Observational data are also subject to structural biases and rarely allow for causal inference in the absence of experimental variation [46].

Moreover, as research has expanded, the field has become increasingly fragmented [47]. Separate literatures have developed around specific domains—such as polarization, misinformation, attention, algorithmic exposure, and coordination—each with distinct assumptions, datasets, and methodologies [48–51]. This fragmentation has limited the comparability of results across studies and platforms, and has made it difficult to build a cumulative understanding of online social dynamics [52].

Much of the existing literature remains tied to individual platforms, making it difficult to determine whether observed effects are driven by human behavior, content properties, or platform-specific design [40, 53]. As a possible solution, recent works have begun to adopt multiplatform and longitudinal strategies to address this limitation [39, 54, 55]. Interestingly, these approaches provide important insights into which patterns are system-specific and which are more general, offering a stronger basis for explanation, prediction, and intervention.

Despite this abundance of research, limitations suggest the need for a comparative and conceptually integrated approach. The present survey aims to fill this gap by providing a structured synthesis of current research on online behavior from a data science perspective. Our approach seeks to integrate three layers—empirical evidence, modeling frameworks, and system design implications to guide future works and research on online social dynamics. Specifically, we aim to answer the following questions:

RQ1 **How have digital platforms and behavioral data transformed the empirical study of online social dynamics?**
RQ2 **Which structural patterns and mechanisms are driving key online phenomena, such as polarization, misinformation, algorithmic influence, and coordinated behavior?**
RQ3 **What are the strengths and limitations of methodological approaches used to study online social behavior?**

Answering these questions provides both methodological reflections and actionable suggestions, making the work a valuable roadmap for the interdisciplinary community studying online behavior. At the same time, it delivers essential insights for practitioners, such as platform designers and policymakers, by clarifying how dynamics like selective exposure and algorithmic amplification lead to phenomena like polarization and misinformation. Unlike previous surveys that concentrate on single phenomena—such as misinformation [16, 56], polarization [57], or echo chambers [58–60]—we approach the study of online behavior from a more comprehensive view, integrating phenomena into a unified analytical framework. We deliberately do not aim to provide a comprehensive enumeration of all existing works within each domain. Instead, our goal is to identify cross-cutting regularities, methodological trade-offs, and modeling constraints that recur across literatures that are often treated in isolation. At the same time, we provide explicit pointers to existing surveys that discuss each subfield in detail. Ultimately, this work helps build the foundation for new, evidence-based policies capable of addressing these pressing online challenges.

Search strategy. In this paragraph, we define the selection criteria that guided the composition of this survey. Our goal is not to provide an exhaustive bibliometric review but to synthesize the main empirical and modeling approaches that collectively define the field of online social dynamics. The survey primarily focuses on studies published between 2008 and 2025, allowing a coverage of the full period, from when digital trace data and platform analytics first began to be collected to when they became central in social research. Inclusion criteria cover peer-reviewed journal articles and major conference papers that employ empirical, computational, or hybrid methods to analyze online social systems. Similar to previous surveys [57, 60–63], the literature base was constructed using the authors' expertise and a three-stage selection process. In the first stage, we compiled an initial list of candidate studies integrating each topic through keyword searches on Google Scholar. In the second stage, each paper was assessed for its relevance to determine whether it should be included. Finally, in the third stage, we manually reviewed the references of the identified

Table 1. Keywords Used to Identify Relevant Studies Across the Sections of the Survey

Section / Subsection	Keywords
2. The Shift Toward Behavioral Data	behavioral data; CSS; digital trace data; big data.
3.1 Empirical Strategies: Strengths and Structural Constraints	network analysis; field experiments; survey; laboratory experiments; algorithmic audits; browser-based instrumentation; panel-based instrumentation; agent-based models; generative AI.
3.2 Comparative Synthesis and Empirical Trade-offs	causality; cross-platform analysis; validation models; research reproducibility.
4.1 Selective Exposure and Attention Allocation	attention economy; selective exposure; information overload.
4.2 Agenda Setting and Content Visibility	agenda setting; issue salience.
4.3 Algorithmic Amplification and Echo Chambers	echo chambers; filter bubbles; algorithmic amplification.
4.4 (Mis)information Dynamics and Polarization	misinformation; disinformation; fact-checking; prebunking.
4.5 Coordinated Behavior and Collective Signaling	**coordinated inauthentic behavior (CIB)**.
5. Modeling Opinion and Information Dynamics	opinion dynamics; **bounded confidence models (BCMs)**; voter models; Ising models; majority rule dynamics.
6. Comparative Analysis at Scale	cross-platform; multiplatform.

works to uncover additional relevant studies and expand the list of included research. To ensure methodological transparency, the keywords used in these searches are detailed in Table 1.

Paper organization. To help the reader navigate more easily through the content of this work, here we outline the structure of the survey. Specifically, each section addresses a specific dimension of the study on online social dynamics and contributes to the overarching goal of integrating empirical, methodological, and theoretical perspectives.

Section 2 traces the historical and epistemological origins of CSS. It explains the transition from traditional research paradigms—based on surveys, laboratory experiments, and self-reported data—to large-scale analyses grounded in behavioral traces collected from digital platforms. The section highlights how this shift has expanded the empirical reach of social inquiry while introducing new methodological and interpretive challenges.

Section 3 provides a systematic overview of the main empirical strategies used to investigate online behavior. It discusses the assumptions, capabilities, and limitations of diverse methodological frameworks and suggests how their integration enables a more cumulative and comparable understanding of online phenomena.

Section 4 synthesizes major findings from the literature concerning the behavioral and structural dynamics observed on digital platforms. The section is organized around five central domains: selective exposure and attention allocation, agenda-setting and content visibility, algorithmic amplification and echo chambers, misinformation dynamics, and coordinated behavior. Each subsection links empirical findings to the underlying socio-technical mechanisms and platform affordances that generate them.

Section 5 reviews the predominant modeling traditions developed to explain opinion formation and information diffusion. The section presents an overview of the most influential frameworks, emphasizing their evolution, theoretical assumptions, and recent efforts to align them with real-world behavioral data.

Building on these foundations, Section 6 introduces the multiplatform and longitudinal perspective to address the limitations arising from the previous sections. It examines how this analytical approach enables identifying dynamics that persist across online platforms, and it discusses the methodological implications of such approaches for model validation, causal inference, and

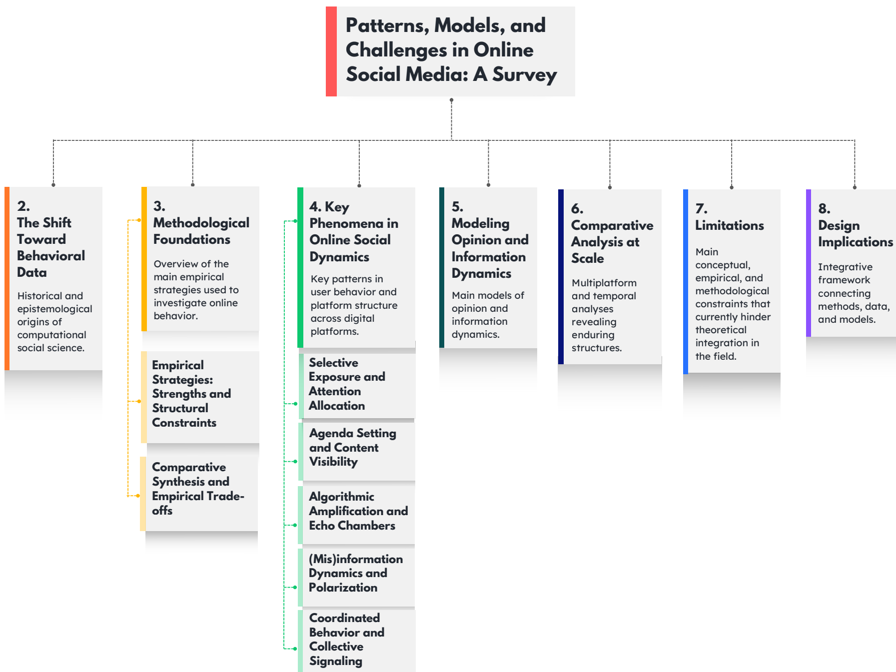

Fig. 1. Visual organization of the survey. Each section covers a different aspect of the research on online behavior.

generalizability. This section positions comparative research as a necessary step toward a cumulative science of digital behavior.

Section 7 completes our discussion by identifying the main conceptual, empirical, and methodological constraints that currently hinder theoretical integration in the field. It highlights the need for better empirical calibration of models, improved causal inference designs, and more consistent use of temporal and cross-platform data. Section 8 complements this analysis by translating empirical findings into actionable design implications for digital platforms, focusing on transparency, exposure diversity, and systemic resilience.

Finally, Section 9 concludes the survey by summarizing its central insights and outlining directions for future research. It emphasizes the need for sustained collaboration across disciplines and for the development of data infrastructures that support reproducible, empirically grounded studies of online social spaces.

To aid the reader, Figure 1 contains a visual map of the organization of the present work. Moreover, Table 2 contains a collection of the most common concepts discussed in the survey, with their operational definition.

2 The Shift Toward Behavioral Data

Social behavior studies have traditionally relied on controlled experiments, small-sample surveys, and self-reported measures [64, 65]. While these methods have produced important theoretical insights, they face well-documented limitations [66]. Experimental findings often struggle to generalize beyond specific contexts [67, 68], and self-reports are affected by recall errors, framing

Table 2. Glossary Table of the Most Important Terms Discussed in the Present Work

Construct	Operational Definition
Echo Chambers	Closed environment within online social media where individuals are exposed predominantly to information, opinions, and beliefs that reinforce their existing views, while contradictory perspectives are filtered out or ignored. This effect may lead to reduced viewpoint diversity and polarized group opinions.
Filter Bubbles	Personalized information environment created by algorithmic filtering, where content is selectively shown to users based on their past behavior, preferences, and interactions.
Exposure Diversity	Breadth of distinct topics, sources, or viewpoints encountered within a defined observation window.
Algorithmic Amplification	Systematic over-representation of content relative to baseline prevalence due to ranking or engagement-optimization.
Coordinated Inhautentic Behavior	Coordinated Inauthentic Behaviour refers to organized activity by groups of accounts on social media that deceptively work together to mislead audiences about their identity, intentions, or actions. Coordination typically involves fake accounts, impersonation, or manipulated content designed to amplify certain narratives, influence public opinion, or disrupt authentic discourse. The key feature is not just false information, but the deceptive and coordinated nature of the behavior.
Virality	Fast spread of a piece of content driven by individuals sharing it widely and quickly, often within a short time frame.
Diffusion	Process through which information, innovations, or behaviors spread across a social network or population over time. It focuses on patterns of adoption and transmission paths, not just the rate of spread
Affective Polarization	Emotional divide between opposing political or social groups, characterized not just by disagreement on issues, but by strong feelings of dislike, distrust, or hostility toward members of the other side.
Attention Inequality	Phenomenon where a small number of sources or individuals capture a disproportionate share of visibility and audience attention, while the majority remain largely unseen. More broadly, it describes the uneven distribution of attention across users, topics, or content within digital platforms.

For each construct, we provide an operational definition, methods of measurement, and common interpretative caveats.

effects, and social desirability biases [69–71]. As a result, key aspects of online behavior—such as attention allocation, content exposure, and interaction dynamics—have long remained difficult to observe directly and with sufficient scale.

The emergence of digital platforms has introduced a structural change in how human behavior can be measured [72]. User actions—such as likes, comments, shares, and clicks–generate large volumes of time-stamped, structured data [73, 74]. These digital traces capture in-situ behavior unfolding in real-world conditions and at population scale [75–77]. This shift has redefined the empirical foundations of social research.

From a methodological perspective, the impact is twofold. First, digital data allow for large-scale, longitudinal observation across millions of users, enabling the detection of macro-level regularities that were previously inaccessible [78, 79]. Second, they support diagnostic analysis of platform-mediated processes, such as information diffusion [80, 81], community formation [82, 83], and feedback amplification [84, 85].

This transition has contributed to the emergence of CSS [41, 42], an interdisciplinary field integrating network analysis, machine learning, and statistical inference to study social dynamics through digital evidence. The core premise is that online platforms generate behavioral signals

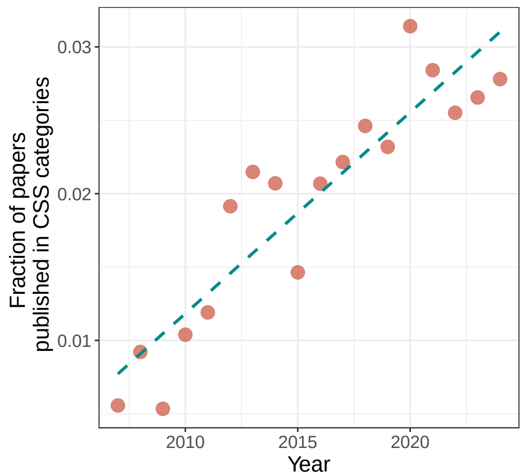

Fig. 2. Fraction of papers published on arXiv in cs.CY, cs.SI, and physics.soc-ph—three key categories in CSS. The notable increase highlights the growing influence of digital behavioral data and computational methods in the empirical study of online social dynamics.

that offer new opportunities for empirical inquiry into collective behavior [86]. However, the quest remains highly fragmented with both challenges and opportunities [37, 87].

To illustrate this shift, Figure 2 presents the fraction of papers published on arXiv between 2007 and 2024 in three relevant categories: Computers and Society (cs.CY), Social and Information Networks (cs.SI), and Physics and Society (physics.soc-ph). A notable increase in submissions is evident, coinciding with the broader availability of web-scale behavioral data. This growth reflects expanded research capacity and an alternative epistemological orientation: from hypothesis-driven studies based on small samples to large-scale analyses grounded in direct observation.

However, this new empirical approach also introduces important challenges. Among them, access to behavioral data is often restricted due to proprietary controls or a lack of standardized APIs [88]. Moreover, the abundance of data does not guarantee explanatory clarity: observational signals require careful methodological treatment to rule out confounding factors and support causal inference. In addition, digital behavior is not neutral: the design of interfaces, algorithmic ranking systems, and platform affordances shapes it [89, 90]. This complicates the task of disentangling endogenous user preferences from exogenous system effects. Finally, we note that datasets often—but not always [91]—contain only a sample of the overall social media traffic, due to collection methods such as keyword-based searches.

Despite its limitations, the increased use of behavioral data marked a significant turning point in research [36, 92]. It opened up new research questions, validation strategies, and applications in system design, content governance, and policy evaluation [93–95]. The following sections explore how this shift unfolds across various methodological strategies and research domains.

3 Methodological Foundations

The study of digital social behavior relies on a diverse set of empirical strategies, each offering access to different aspects of online dynamics. These approaches—ranging from large-scale observation to controlled experimentation and computational modeling—have individually produced valuable

Table 3. Empirical Strategies for Studying Online Behavior: Scope, Structural Limits, and Typical Applications

Method	Scope	Limitations	Examples of applications
Observational network analysis	Large-scale interaction patterns, content flow, and polarization	No causal inference; visibility bias; lacks exposure data	Tracking the diffusion of a misinformation tweet; mapping how anti-vaccine Facebook groups interconnect; identifying which Instagram creators act as superspreaders for political content.
Field experiments	Causal effects of ranking changes, labels, friction, and exposure modulation	Requires platform collaboration; limited replicability	Testing impact of algorithmic interventions; randomly varying the prominence of news posts in users' feeds to measure effects on political leaning.
Surveys and lab experiments	Attitudes, belief shifts, and perceived credibility under control	Low external validity; sample skew; disconnect from real behavior	Showing participants true vs. false headlines to measure belief accuracy; testing how exposure to prebunking videos reduces susceptibility to deepfakes; running lab tasks to measure confirmation-bias in content selection.
Algorithmic audits	Personalization dynamics, recommendation logic, and content visibility	Synthetic behavior; platform ToS limits; hard to generalize	Creating controlled YouTube accounts to test how watch history affects recommended political videos; simulating different user profiles on TikTok to see if extremist content is surfaced; measuring how search results change for identical queries across users.
Browser/panel instrumentation	Real exposure traces, behavior across demographics	Opt-in bias; poor mobile coverage; privacy attrition	Reconstructing media diets and exposure-belief links; tracking exposure to political ads across mobile and desktop.
Agent-based / simulation models	Mechanisms of influence, polarization, and opinion formation	Stylized assumptions; often unvalidated; limited predictive scope	Exploring counterfactuals and systemic dynamics; simulating how rumor-spreading accelerates under different network topologies; modeling how small changes in recommender systems can increase polarization.
Usage of generative AI	Simulation of user behavior, large-scale automated annotation, algorithmic auditing through synthetic data generation	Model bias and hallucination, limited interpretability, risk of synthetic artifacts	Simulating communication dynamics, generating synthetic datasets, auditing recommender systems.

insights. However, their integration remains limited [96, 97]. As a result, the field is marked by fragmentation: different methods often lead to non-comparable results, and cumulative knowledge remains difficult to achieve [47, 48].

To move beyond isolated findings, a consolidated methodological framework is needed: one that makes explicit the assumptions, capabilities, and structural limitations of each approach. In this section, we review the principal empirical strategies used to investigate phenomena such as polarization, algorithmic exposure, misinformation, and coordinated behavior, phenomena that will be discussed in detail in Section 4. Rather than aiming for taxonomic completeness, our goal is conceptual clarity, i.e., to highlight what each method reveals, what it obscures, and how triangulation across methods can mitigate blind spots.

3.1 Empirical Strategies: Strengths and Structural Constraints

Research on platform-mediated behavior typically draws on seven broad methodological families, which are summarized in Table 3. The following section discusses the merits and limitations of each approach.

Observational social network analyses use digital traces—such as follower graphs [98, 99], sharing cascades [1, 40, 100], or hyperlink co-occurence structures [101, 102]. These methods are fundamental for detecting emergent regularities (see Section 4.1), but are limited to correlational inference. They often suffer from visibility bias [103] and provide little information about actual exposure or user-level interpretation.

Field experiments, including randomized interventions on algorithms [104–107], friction elements, or content labels, allow for causal identification of treatment effects in real-world settings [104–107]. These studies offer high external validity and practical relevance (e.g., for agenda-setting dynamics and Section 4.2), but typically depend on privileged access through research partnerships with platforms [108]. This restricts both transparency and reproducibility.

Survey and laboratory experiments enable precise measurement of beliefs, attitudes, and perception under controlled conditions [109, 110]. These methods are strong in internal validity and useful for theory testing. However, these studies often rely on small samples, and their experiments are typically conducted in simulated—and therefore simplified—scenarios. As a result, their relevance to real platform behavior remains limited, especially in contexts involving collective behaviour.

Algorithmic audits simulate user behavior via bots or shadow accounts to examine the functioning of content recommendation systems [111–114]. These methods are valuable for identifying structural biases in exposure (see Section 4.3), but face limits in extrapolating from synthetic agents to real users. In some cases, platform terms of service constrain their implementation or reproducibility.

Browser-based and panel instrumentation, often based on opt-in tracking or user panels, provides fine-grained data on actual exposure and media consumption [115–118]. These methods are particularly relevant for studying the informational environment surrounding phenomena such as polarization and misinformation (Section 4.4). However, they are vulnerable to selection bias, underrepresentation of mobile usage, and privacy-related dropout.

Agent-based and simulation models formalize assumptions about cognition, influence, and feedback loops to explore systemic outcomes and counterfactuals [119–122]. These models provide mechanistic insight and aid in conceptualizing dynamics under uncertainty. Yet, their explanatory value depends heavily on empirical calibration—an aspect often neglected when real-world data are unavailable or siloed.

Generative AI is increasingly used as an alternative way to simulate user behavior, analyze large text datasets, and prototype interventions in online environments [123–127]. However, its usage introduces important methodological and ethical challenges as models may inherit and amplify biases embedded in their training data [128–130]. At the same time, synthetic simulations risk overfitting to models' tendencies rather than replicating authentic social processes [131–134].

3.2 Comparative Synthesis and Empirical Trade-offs

Despite their heterogeneity, the previous strategies face a common set of challenges that hinder their effectiveness and applicability.

First, there is a trade-off between *scale and causality*. Large-scale observational studies are well-suited to detect aggregate patterns [39, 101] but typically cannot support causal claims [135]. Conversely, methods that enable causal inference—such as randomized field experiments—often operate on limited populations, within narrow contexts, or under artificial conditions.

Second, *access asymmetry* remains a major obstacle. The most informative interventions, especially those involving the internal mechanisms of platforms (e.g., feed ranking and interface design), require deep integration with proprietary infrastructures. This restricts both academic independence and reproducibility [136–138].

Third, the field is affected by a persistent *validation gap* between theoretical models and empirical observations. Simulation-based approaches offer plausible mechanisms, but are often disconnected from data. Without a systematic grounding in behavioral evidence, models risk internal coherence without real-world relevance [139–141].

Recognizing these structural boundaries is crucial for the interpretation of empirical findings, the design of methodologically sound studies, and the advancement of a more precise understanding of online social behavior [41, 42, 142].

4 Key Phenomena in Online Social Dynamics

Social media platforms have redefined the architecture of public discourse [39, 143, 144], altering how attention is allocated [145–147] and how content spreads [1, 22, 148, 149].

In the sections that follow, we adopt a dual perspective. We review substantive findings across key domains: attention and exposure (Section 4.1), agenda-setting (Section 4.2), algorithmic influence (Section 4.3), misinformation (Section 4.4), and coordination (Section 4.5)—while keeping the methodological foundations in view.

Together, these processes form the backbone that shapes how information is created, shared, and spread on social media. Beyond reviewing the existing literature, our goal is to show how those insights depend on specific research setups and to identify where future progress is needed. However, given the impossibility of discussing each phenomenon in all its aspects, we also refer the interested reader to specific surveys that allow for a deeper exploration of each topic.

4.1 Selective Exposure and Attention Allocation

The attention economy plays a central role in structuring online content consumption [150–153]. In a context of chronic information abundance [154–156], visibility becomes a limited and contested resource. Rather than emerging from user-driven search and intentional discovery, exposure is increasingly mediated by feed architectures, push-based notification systems [156–158]. These affordances prioritize immediacy and reactivity, delivering a continuous stream of stimuli that compete for cognitive bandwidth in real time [159, 160].

Empirical analyses of user interactions on platforms such as Facebook, YouTube, and X (formerly known as Twitter) consistently show a strong concentration of attention: a small fraction of content receives a disproportionately large share of engagement. The distribution of such engagement typically follows heavy-tailed distributions, remarkably consistent across platforms, content types, and time windows [100, 161, 162].

This skew is not primarily determined by the size of the audience or the institutional reputation of the content producer. Some studies show that even pages with relatively few followers can outperform established outlets when their content aligns with emotionally salient or identity-relevant narratives [3, 163, 164].

This concentration becomes even more pronounced during major events, where content that triggers emotional reactions or reinforces group identity tends to attract more engagement than content selected for accuracy or relevance [165–168].

These patterns may not be driven by user behavior alone. They could be reinforced by platform algorithms that tend to rank content based on engagement metrics rather than on indicators of accuracy or credibility [169, 170]. Therefore, if engagement functions as the main proxy for relevance, a feedback loop may emerge: content that attracts interaction becomes more visible, which, in turn, increases its chances of spreading [171–173].

This mechanism may systematically favor posts that are emotional, polarizing, or identity-driven, selecting for content optimized for immediate reactions rather than thoughtful analysis [83, 174, 175].

Research has focused on the task of detecting bursty patterns and patterns of information diffusion, proposing numerous and heterogeneous techniques [176–181]. Nonetheless, temporal analyses of viral content show that visibility is not just concentrated, but also highly volatile [182–186]. Spikes in attention usually last only a few hours or days, yet they can have lasting effects on how content is ranked and how users perceive what is important [187].

The implications may be structural. The way attention is allocated appears increasingly influenced by optimization processes that prioritize behavioral signals over informational value [185, 188–191].

While in this section we have highlighted core aspects of selective exposure and attention allocation, interested readers may consult specialized reviews offering integrative perspectives on the history and causes of information overload [156, 192] and attention economy [193] and its implications for information processing.

Rather than distorting exposure in isolated instances, these dynamics could be reshaping the default conditions under which content gains visibility, with potential consequences for pluralism, deliberation, and the quality of public information.

4.2 Agenda Setting and Content Visibility

Agenda setting refers to the process by which certain topics, issues, or narratives gain prominence in public discourse, shaping what people perceive as salient or worthy of attention [194]. In classical media theory, this process was conceptualized as largely top-down: editors, broadcasters, and political institutions selected and framed content for mass audiences.

In today's platform-mediated environments, agenda setting has become decentralized, continuous, and personalized [195, 196]. Content exposure is no longer orchestrated by a limited set of gatekeepers, but emerges from the interaction between algorithmic curation, user engagement patterns, and platform-specific design features [197–199]. This shift alters not only who can influence the public agenda, but also how visibility itself is structured and distributed [200–202].

Empirical studies suggest that visibility on digital platforms is no longer limited to legacy media or high-follower accounts. Individual users and niche communities—especially when their content aligns with platform engagement dynamics—can reach exposure levels comparable to those of institutional actors [163]. Algorithmic ranking systems contribute to this shift by prioritizing interaction metrics, which may weaken the link between visibility and traditional indicators such as source credibility or editorial oversight [203].

As a result, the salience of a topic is increasingly determined not by its institutional relevance or civic importance, but by its capacity to trigger engagement at scale [83, 204–206].

Network-based analyses add an additional layer of complexity [207]. Information diffusion often occurs within tightly clustered communities, driven by homophily and shared identity markers [18, 83, 208, 209]. In such contexts, content salience may be amplified not through broad dissemination but through localized resonance. In [210], the authors show that topics rise to prominence not necessarily because they are imposed from above, but because they activate affective or identity-based responses. In this context, literature reviews on agenda-setting theory have emphasized the need to bridge methodological and disciplinary divides [211, 212]. Integrating insights from network science, CSS, and critical algorithm studies could help the Agenda-Setting framework evolve in step with the ever-changing media ecosystems it seeks to explain.

Taken together, these findings raise questions about the continued relevance of hierarchical media models for explaining how content gains visibility online. Agenda setting now appears to operate as a distributed process shaped by feedback loops, algorithmic filtering, and platform-specific design choices. Explaining which topics rise to prominence and under what conditions calls for analytical approaches that account for the interplay between user behavior, network structure, and system-level constraints.

4.3 Algorithmic Amplification and Echo Chambers

The mechanisms of agenda setting discussed above are structurally intertwined with algorithmic curation—the process by which digital platforms personalize content exposure based on predicted user engagement [213]. These curation systems operate through continuous optimization of ranking algorithms, leveraging behavioral signals such as clicks, shares, watch time, and interaction rate to maximize retention and activity [214]. In doing so, they establish feedback loops in which user behavior influences content visibility, which, in turn, shapes subsequent behavior [170, 215, 216]. Over time, this dynamic fosters homophily-based reinforcement and the emergence of ideologically homogeneous environments, commonly referred to as echo chambers [18, 217–220].

These structures are not explicitly designed but arise endogenously [18] from the cumulative effects of personalization strategies. As platforms optimize content exposure to maximize short-term engagement, users are increasingly shown material that aligns with their prior preferences, social identity, or emotional disposition [1, 206]. This narrowing of informational inputs reduces exposure diversity and inhibits cross-cutting interaction [17, 221], particularly in politically or emotionally charged contexts [51, 222].

Sentiment analyses further indicate that echo chambers are not only informationally homogeneous but also emotionally unstable [83, 223, 224]. Moreover, analyses show that content expressing out-group animosity, particularly from political actors, consistently drives higher engagement across platforms, reinforcing the emotional polarization of online discussions [225]. As homogeneity increases, conversations within these environments tend to become more negative, reactive, and polarized [220]. This affective decay is reinforced by the reward structure of engagement-driven ranking: content that elicits outrage, fear, or in-group affirmation is more likely to be promoted, accelerating the circulation of emotionally charged narratives [83, 165, 226–228].

In this landscape, recent literature reviews have mapped the diverse conceptualizations, measurement approaches and possible solutions of echo chambers [58], showing how methodological choices have an impact on findings [59, 60].

The amplification processes generated by algorithmic recommendation systems can function as intervening mechanisms in the formation and persistence of echo chambers. By reinforcing homophilous exposure and limiting cross-cutting information flows, these dynamics contribute to the reduction of informational diversity and, consequently, to the distortion of opinion formation processes. A systematic understanding of these mechanisms is therefore crucial for interpreting empirical manifestations of polarization and for developing interventions that target their structural underpinnings.

4.4 (Mis)information Dynamics and Polarization

The dynamics of algorithmic amplification described above may contribute directly to the persistence and reach of misinformation in digital ecosystems. In environments optimized for engagement, the diffusion of content is shaped less by its factual accuracy than by its emotional salience, identity alignment, and resonance with existing beliefs [17, 227, 229].

The tendency to favor ideologically-aligned content, as discussed in the previous section, creates favorable conditions for the spread of misinformation [230, 231]. False or misleading claims can outperform accurate information [2, 232] in terms of engagement because they tend to be simpler, more emotionally charged, and more aligned with in-group narratives [1].

Attempts to mitigate misinformation through fact-checking [233], crowd-sourced systems [234], warning labels [235, 236], or inoculation campaigns have produced mixed outcomes [237]. A large-scale analysis of over 50,000 debunking posts on Facebook revealed that corrective content mostly circulates within communities already predisposed to accept the correction, with negligible penetration into conspiratorial or misinformed segments [238]. Several works highlight that debunking often fails to reach those most exposed to or influenced by misinformation [239–241]. However, we also note how a series of fact-checking experiments conducted simultaneously across several countries have shown the opposite results. [242].

Further research suggests that cognitive alignment often takes precedence over informational intent [243]. Users with strong ideological or conspiratorial predispositions may interpret corrections, satire, or disclaimers not as challenges to their beliefs, but as further confirmation [242, 244, 245].

The result is a form of epistemic fragmentation, where even well-designed informational interventions may have limited impact due to structural and contextual constraints [246]. As discussed earlier, algorithms typically prioritize content that generates engagement rather than content selected for accuracy or reliability. At the same time, social dynamics can reinforce belief persistence through mechanisms of group identity and emotional resonance [247]. In this context, misinformation can be understood as a systemic outcome of digital communication processes, arising from platform design, content circulation dynamics, and patterns of user interaction.

A growing body of research explores an alternative intervention strategy: prebunking [248, 249]. Unlike debunking, which seeks to correct false beliefs after exposure, prebunking operates preemptively by exposing users to weakened or generic versions of common manipulative tactics before they encounter actual misinformation [250]. This approach leverages principles from inoculation theory and has been shown to reduce susceptibility to false claims across domains and populations [251]. Unlike post-hoc corrections, which often fail due to motivational resistance or selective exposure, prebunking can act upstream—buffering cognitive vulnerabilities before belief consolidation occurs [252]. However, its effectiveness depends on timely delivery, contextual relevance, and integration into platform-level design, highlighting the need for scalable and adaptive deployment strategies.

Complementary to these results, recent reviews on misinformation reveal a growing gap between public narratives and empirical evidence, showing that exposure to false content is often limited and concentrated among small, highly motivated groups [16]. They also highlight inconsistencies in how misinformation is defined and measured, particularly in science and health communication [56]. Moreover, recent surveys on polarization caution that much of the evidence remains fragmented and largely U.S.-centric, limiting the generalizability of current findings [57]. They highlight that polarization is driven not only by elite and partisan cues, but also by emotional and group-based dynamics at the base of the information processing [12]. These insights underscore the need for interventions like prebunking to be context-sensitive and informed by cross-disciplinary evidence.

4.5 Coordinated Behavior and Collective Signaling

One of the most impactful dynamics in digital ecosystems is the emergence of coordinated behavior [253–255], often involving non-human actors or strategically organized groups [256, 257]. Automated accounts (bots) [258, 259], sockpuppets, and troll networks [260] are frequently used to simulate authentic activity, manipulate engagement metrics, and distort perceptions of consensus [261–264]. These practices—collectively referred to as CIB—can amplify targeted narratives, suppress alternative viewpoints, or create the illusion of spontaneous grassroots support (astroturfing) [265–269]. By generating high-volume, synchronized signals across multiple accounts, such campaigns can hijack visibility algorithms, flood discourse spaces, and influence user perception of legitimacy or popularity [262, 270, 271].

Alongside these artificial interventions, platforms also exhibit forms of spontaneous coordination, where large numbers of users engage in similar behaviors without centralized planning [272–274]. Trends, memes, and hashtag cascades can create synchronized patterns of attention and signaling, often driven by affective contagion or identity salience rather than strategic intent [253, 275]. Understanding the spectrum from scripted manipulation to emergent synchronization is key to interpreting how influence and visibility are shaped in digital environments.

Following advances in generative **Artificial Intelligence** (**AI**), research has also begun to explore how **Large Language Models** (**LLMs**) can simulate collective human behavior. Experimental results show that social conventions, such as collective biases, may spontaneously emerge between interacting LLM agents [276] and that group size drives the observed dynamics [277]. Particularly, some highlight the risks associated with malicious AI swarms when used in influence campaigns, which may accelerate the weakening of democratic information ecosystems [278]. For instance, recent work found that autonomous LLM agents can reproduce coordination strategies observed in real Information Operations, even without human guidance [279].

Detecting these behaviors requires methodological precision [259, 280, 281]. Recent studies have employed similarity metrics, temporal correlation, content redundancy, and network structural features to identify coordinated entities across platforms [254, 282–285]. In the context of political or toxic content, coordinated accounts tend to exhibit bursty activation patterns during high-salience events—behaving in synchrony, targeting the same content, and frequently escaping detection by moderation systems [286–288]. While these operations are often short-lived, they can exert outsized influence on narrative framing, agenda saturation, and emotional tone.

Coordination represents a key point of convergence between individual behavior and systemic structure. Alongside attention dynamics, agenda setting, algorithmic amplification, and misinformation diffusion, it constitutes a foundational dimension in the study of online social systems. A cumulative science of digital behavior must account for both engineered manipulation and emergent synchronization, recognizing them as distinct but interacting drivers of visibility, perception, and influence. In this section, we have summarized the main results of research on coordinated behavior, with a focus on the inauthentic kind. The readers interested in a more comprehensive overview of the field may refer to dedicated surveys such as [62].

5 Modeling Opinion and Information Dynamics

Modeling opinion dynamics is a long-standing interdisciplinary effort spanning statistical physics, computational theory, and the social sciences. These models seek to formalize how individuals form, adjust, or reinforce their beliefs through social interaction, typically under uncertainty, bounded information, and network constraints [63, 289–291].

Before the advent of digital platforms and large-scale behavioral datasets, these models operated largely in an abstract space. Their strength resided in analytical tractability and conceptual clarity, not in empirical correspondence [292, 293]. The increasing availability of high-resolution digital trace data has transformed this landscape [294]. Social media platforms now enable direct observation of belief-related behaviors—expressed through likes, shares, comments, and content exposure—across massive user populations and diverse contexts. This shift has introduced new methodological opportunities: recent work has developed procedures for learning model parameters from behavioral traces [295, 296], and systematic efforts are underway to bridge the gap between formal models and real-world data [297]. As a result, empirical adequacy has become a central criterion: models must not only be theoretically coherent, but also capture observable regularities in opinion and information dynamics.

Historically, the first wave of formal models drew heavily from analogies with physical systems. Weidlich's early formulation and subsequent adaptations of the Ising model conceptualized social influence as a process of local alignment: agents tended to adopt the most common opinion in their neighborhood [298–300]. These approaches revealed how global order could emerge from simple local rules, but they relied on strong assumptions—such as symmetric influence, spatial homogeneity, and memoryless dynamics—that limit their applicability in the context of digital communication. Recent reviews have highlighted these limitations and called for models that better reflect the heterogeneity and complexity of real-world settings [291, 301, 302]. Several works have integrated features such as algorithmic filtering [303], personalization mechanisms [304], and cognitive constraints [305, 306]. Others have introduced nonlinear update rules or convergence schemes that depart from naive averaging, offering a closer match to observed opinion trajectories [307].

A second and highly influential class of models is represented by the voter model family [308, 309]. In this framework, agents hold binary opinions and update their state by copying the opinion of a randomly selected neighbor. Classical extensions have explored the impact of network topology, stochastic perturbations, and bounded-confidence constraints [310–316], yielding key insights into metrics like consensus time, fixation probability, and noise-driven transitions. However, their explanatory reach is limited in modern digital ecosystems, where exposure is filtered by algorithms, influence is asymmetric, and identity-based signaling matters. Recent models address these gaps by incorporating algorithmic bias into update rules and network interactions [303, 317]. Other work extends the voter framework by embedding personalization effects and attention decay, illustrating how polarization regimes and convergence rates shift dramatically under such mechanisms [304]. Empirical-oriented models advance even further: for instance, micro-foundational frameworks use weighted-median updating that align better with observed shifts in opinion formation [302, 305]. These developments mark a significant move toward bridging theory with behavioral data.

A third influential class is the majority-rule model: in each round, a randomly selected group of agents updates the majority opinion in that group. Classic results characterize convergence rates, fixation probability, and scaling behavior as functions of group size and network dimension [318, 319]. However, these models often rely on simplifying assumptions—static group structure, uniform mixing, and no external bias—that are misaligned with digitally mediated social systems. Recent extensions introduce greater realism: majority-rule on hypergraphs reveals how modular or heterogeneous group connections alter convergence and prevent universal consensus [320]. Other models incorporate directed or asymmetric influence—analogous to algorithmic recommendation or propaganda—to demonstrate how slight external bias can sustain long-term polarization [321, 322]. Monte-Carlo results from finite-size majority-rule systems further illustrate noise-driven transitions and the emergence of persistent coexistence regimes under realistic filtering conditions [323].

Along the same path, BCMs have been introduced to capture a key empirical feature of online opinion dynamics: selective exposure. In these models, agents hold continuous opinions and interact only with others whose views lie within a defined similarity threshold [290, 324–329]. The BCM framework has proven flexible in reproducing polarization, fragmentation, and stable pluralism. It accommodates the empirical observation that people tend to avoid opinion exchanges perceived as too distant or hostile [330].

However, most implementations rely on numerical simulations and rarely incorporate empirically derived parameters. While BCMs succeed in capturing the form of polarization, they often fail to explain the persistence and asymmetry observed in real systems—especially when echo chambers, identity-driven engagement, or asymmetric exposure are present. Recent extensions have begun addressing these issues: models that integrate heterogeneous sociability show increased fragmentation and slower convergence [331], while adaptive confidence bounds have been proposed to reflect individual variation in openness [332]. Other works embed algorithmic influence or recommender systems into the interaction structure, amplifying polarization even under nominally moderate parameters [333]. The inclusion of repulsive dynamics, where agents actively avoid disagreeing peers, has been shown to generate persistent disagreement and opinion divergence [334]. In parallel, models that explore the emergence of consensus in networked bounded-confidence systems have highlighted the importance of structural constraints and local connectivity [335]. Finally, multilayer models that integrate media competition and interpersonal influence—such as the framework by [199]—demonstrate how fragmented equilibria can emerge from the interplay between media messaging and social filtering in complex networks.

Despite their conceptual richness, most models in this field suffer from a structural limitation: the lack of systematic empirical validation. Often developed in synthetic environments and parameterized without reference to observed behavior, these models provide insight into possible mechanisms, but rarely offer testable predictions. This limitation reflects a deeper epistemological divide between abstract theorizing and data-driven science [294]. Without empirical grounding, models risk becoming internally consistent yet externally irrelevant, disconnected from the behavioral dynamics that characterize real social systems. As a result, their ability to inform platform design, forecast polarization, or guide policy interventions remains limited [295].

Recent works have begun to address this limitation [296, 336] by empirically validating models of polarization using behavioral data from web-based social platforms. This approach embeds empirical measures of exposure and engagement directly into the model, enabling a systematic comparison between simulated dynamics and real-world opinion trajectories. In addition, the studies benchmark their performance against alternative modeling frameworks [337, 338], highlighting how different assumptions affect the ability to reproduce observed distributions. This kind of comparative, data-anchored modeling marks a necessary shift—from metaphorical abstraction toward empirically grounded explanation. The most promising direction for the field lies in models that retain analytical clarity while being structurally constrained by empirical data, reproducing the dynamics observed across platforms, populations, and time periods.

6 Comparative Analysis at Scale

The proliferation of cross-platform and longitudinal research has transformed the empirical baseline for modeling online social behavior. Rather than treating each platform as an isolated case, recent studies trace behavioral patterns that persist across platforms and time, providing a foundation for validating computational models. For example, in [39, 54] the authors document phenomena such as increasing toxicity in longer thread conversations and systematic simplification of user language across platforms and time. These results support a critical insight: algorithmic may expose and amplify human behavior, that is, consistently observed across platforms. This insight has important implications for model builders: theories of opinion dynamics should be tested not only for algorithmic effects but also for their capacity to capture structural regularities intrinsic to large-scale mediated interaction.

Crucially, the multiplatform strategy enables researchers to distinguish between system-specific artifacts and structural regularities [339, 340]. It allows for the formulation of more robust hypotheses about causal mechanisms, enabling inference that transcends individual platform biases [341, 342]. The integration of large-scale trace data, computational modeling, and comparative analysis opens the door to a new class of hybrid models: empirically informed, mechanistically explicit, and cross-context validated [39].

What emerges from this literature is a clear imperative: no model, however, elegant, parsimonious, or theoretically compelling, can be considered explanatory unless it is validated against observed behavioral patterns [40]. The need is especially pressing in algorithmically infused societies, where structural dynamics must be rigorously measured based on empirical data. Recent cross-platform analyses uncover invariant conversational and linguistic patterns—such as persistent toxicity dynamics and systematic simplification of language—that transcend platform-specific affordances [39, 54].

Future models must internalize these dimensions—either by embedding them structurally or by calibrating parameters on real-world data. Without such alignment, models risk remaining metaphorical abstractions, disconnected from the dynamics they aim to explain.

In sum, the study of opinion dynamics stands at a critical juncture. Decades of formal modeling have yielded elegant abstractions, but abstraction alone no longer suffices. The rise of digital platforms has introduced a structural discontinuity: opinion formation now unfolds within socio-technical systems that generate massive, traceable, and algorithmically shaped behavioral data. This transformation demands more than theoretical refinement—it calls for empirical accountability. Only by confronting this dual imperative—preserving conceptual clarity while anchoring models in observed regularities—can the field evolve from metaphor to mechanism. Without this convergence, models will remain speculative blueprints. With it, they can become diagnostic instruments for decoding and ultimately counteracting the systemic pathologies of online collective behavior.

7 Limitations

Over the past decade, research on online behavior has revealed a growing number of empirical regularities. These findings, often replicated across platforms and contexts, represent an important step toward a more systematic understanding of the dynamics underlying digital interactions. Yet despite this empirical progress, the field still lacks a shared analytical framework to explain how such patterns emerge, under what conditions they persist, and how they interact with platform design.

The first open challenge concerns the development of models that capture behavioral dynamics beyond idealized scenarios. Many current models remain loosely connected to empirical observables and rely on assumptions that are difficult to validate or calibrate with real-world data. More systematic efforts are needed to link model parameters to measurable quantities.

A second limitation lies in isolating the role of platform architecture. While similar behavioral phenomena appear across diverse systems, the contribution of algorithmic curation, moderation policies, and interface design remains difficult to disentangle from endogenous dynamics. Comparative studies, leveraging natural variation across platforms or temporal policy shifts, may offer a viable strategy but require coordinated efforts in data collection and methodological design.

Causal inference poses a third conceptual and practical challenge. Randomized experiments, though powerful, are rare and typically limited to specific partnerships or restricted user groups. Alternative strategies—such as quasi-experimental designs, instrumental variables, or simulation-informed counterfactuals—remain underused in this context. Their systematic integration into the study of digital systems could substantially improve our ability to test hypotheses about mechanism and effect.

Another crucial area is the identification of observables that capture system-level dynamics over time. Much of the current literature relies on static metrics, which offer limited insight into how fragmentation, synchronization, or volatility evolve. Temporal measures—such as entropy in interaction networks, burst alignment, or shifts in exposure diversity—could offer more sensitive indicators of systemic transitions or emerging risks.

Finally, most behavioral models treat exposure as exogenous, neglecting the recursive interplay between user activity and algorithmic filtering. This simplification limits our ability to understand feedback effects, such as amplification or convergence, that may be critical in shaping outcomes. Even minimal models that incorporate adaptive exposure mechanisms could provide valuable insight into how interventions propagate through user—system coupling. However, we emphasize that this goal is generally difficult to achieve, as algorithmic filtering is hard to study without specific partnerships.

These limitations do not undermine the progress made so far, but they highlight the need for a more integrated approach, one that connects empirical findings, theoretical models, and system-level inference within a coherent analytical structure. Rather than proposing a unified agenda, we suggest that addressing these issues represents a necessary step toward a cumulative science of online behavior, capable of moving from descriptive mapping to explanatory understanding. Clarifying these foundational aspects may also help inform future system design and policy evaluation, grounding interventions in empirically supported mechanisms.

iewWe conclude by noting that, by design, this survey prioritizes structural synthesis and a field-level perspective over exhaustive coverage. While this choice necessarily limits the level of detail devoted to individual subdomains, it enables us to surface invariant patterns that often remain invisible in narrowly focused surveys. At the same time, we provide direct references to surveys dedicated to each subdomain for readers seeking more detailed information.

8 Design Implications

While the preceding sections have examined distinct empirical and theoretical dimensions of online behavior, their interrelation deserves explicit articulation. We propose an integrative framework

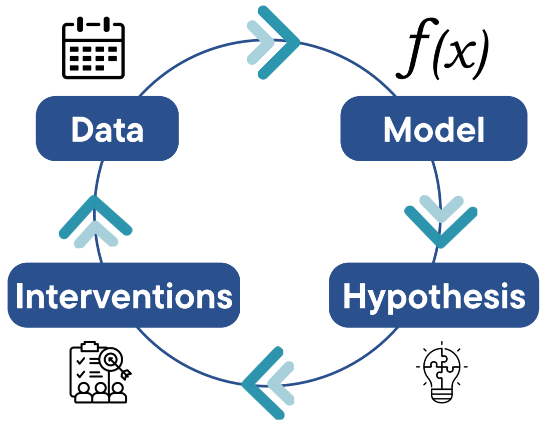

Fig. 3. Visual representation of the interplay between data, models, hypotheses, and interventions.

that links the methodological, empirical, and modeling perspectives outlined in this survey, offering a conceptual map to guide progress in the field. In particular, Figure 3 shows a visual representation of our proposed approach.

At its core, the framework rests on three complementary dimensions: observation, modeling, and intervention. Observation encompasses the empirical analysis of behavioral traces. It provides the factual basis upon which theories of online interaction are grounded. Modeling translates these empirical regularities into formal mechanisms, enabling the identification of causal relationships and the simulation of counterfactual scenarios. Intervention involves the application of empirically informed and model-based insights to platform design, governance, and policy evaluation.

These dimensions are connected through an iterative logic of validation and feedback. Empirical analyses inform model calibration; models, in turn, generate testable hypotheses that guide new data collection; and both contribute to the design of interventions that can be empirically monitored and refined. This recursive structure ensures that advances in one domain systematically reinforce the others, promoting theoretical coherence and methodological convergence across subfields such as polarization, misinformation, algorithmic curation, and coordination.

By framing research on online social dynamics within this triadic structure, we aim to move toward a unified paradigm that combines data-driven observation, formal abstraction, and applied design. The proposed framework does not prescribe a single method or theory; rather, it aims to delineate the epistemological relationships among them, offering a shared foundation upon which a cumulative and comparative science of online behavior can be built.

To conclude the section, we now propose how these scientific findings can serve as a foundation for identifying key implications and informing future developments. In this view, we outline below a set of design-relevant implications that follow from the empirical and modeling literature:

— **Real-time indicators of systemic change.** Because platforms often restrict data access, most current research depends on aggregated engagement metrics, such as like or view counts, which offer only a static snapshot of user behavior. However, providing researchers with access to longitudinal data and measures could enable earlier detection of patterns such as fragmentation, synchronization, or volatility.
— **Introduce calibrated exposure diversity.** Algorithmic curation focused on short-term engagement can amplify polarizing content and reinforce users' existing preferences. Adding signals like diversity, novelty, or reputation—and designing for controlled variation—can maintain relevance while broadening the range of viewpoints users encounter.

– **Enhance interpretability of content ranking.** Users rarely understand why specific content appears in their feed. Simple, human-readable explanations of ranking decisions could improve transparency and support external auditing efforts, without requiring full algorithmic disclosure.

These implications highlight structural levers through which visibility, diversity, and stability can be modulated. Moreover, it is important to emphasize that these changes, when paired with appropriate data access, will enable researchers to better understand the mechanisms underlying online behavior.

9 Conclusions

The emergence of digital platforms has redefined the structural conditions under which information is produced, disseminated, and internalized. This transformation is not transient; it marks a long-term reconfiguration of the informational landscape, with implications for how social behavior is observed, modeled, and interpreted. The convergence of algorithmic mediation, user-generated content, and large-scale behavioral traces has enabled a shift in the study of online dynamics, from anecdotal or small-scale analyses to systematic empirical observation on a large scale.

This survey has synthesized findings across key domains. In each, computational approaches have uncovered common patterns that challenge long-standing assumptions and suggest the existence of dynamics that transcend specific platforms or contexts.

Yet despite substantial progress, the field continues to face persistent challenges, as highlighted in Section 7.

Addressing them is key to developing a more coherent understanding of online social dynamics. It calls for a joint effort between researchers and platforms to establish validated pipelines and public infrastructures for data access. Most critically, it requires a shared commitment to methodological rigor and empirical grounding across the research community.

The promise of CSS lies in its ability to render complex behavioral systems observable, interpretable, ultimately, and actionable. However, CSS has often functioned as an umbrella category, that is, broad in scope but weak in+ methodological cohesion. Recent works in the field show the benefits of a more integrated paradigm: one relying on the analytical tools of data science and the availability of large-scale social data, but also grounded on the formal frameworks of complex systems science. This convergence can mark a transition to a discipline, that is, capable of explanatory depth and predictive precision.

Future work should proceed along four converging directions. First, the development of empirically validated models that integrate cognitive constraints, algorithmic mediation, and observed behavioral patterns across platforms. Second, the construction of multiplatform observatories that support synchronized, longitudinal data collection under consistent ethical and technical protocols. Third, the advancement of causal inference techniques—such as counterfactual simulations, synthetic controls, and large-scale interventions—that can disentangle behavioral dynamics from infrastructural effects. Second, the institutionalization of open, independent infrastructures for behavioral data access, designed to ensure reproducibility, transparency, and long-term availability of high-quality digital trace data. These directions define the structural foundations for a robust and cumulative science of online social behavior. Finally, this effort requires a clear separation between narrative plausibility and empirical validity in interpreting large-scale behavioral traces—what recent work has termed *epistemia* [343]

References

[1] Michela Del Vicario, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli, H Eugene Stanley, and Walter Quattrociocchi. 2016. The spreading of misinformation online. *Proceedings of the National Academy of Sciences* 113, 3 (2016), 554–559.

[2] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *science* 359, 6380 (2018), 1146–1151.

[3] Ana Lucía Schmidt, Fabiana Zollo, Michela Del Vicario, Alessandro Bessi, Antonio Scala, Guido Caldarelli, H. Eugene Stanley, and Walter Quattrociocchi. 2017. Anatomy of news consumption on Facebook. *Proceedings of the National Academy of Sciences* 114, 12 (2017), 3035–3039.

[4] Anna Sophie Kümpel, Veronika Karnowski, and Till Keyling. 2015. News sharing in social media: A review of current research on news sharing users, content, and networks. *Social Media+ Society* 1, 2 (2015), 2056305115610141.

[5] Manos Tsagkias, Maarten De Rijke, and Wouter Weerkamp. 2011. Linking online news and social media. In *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*. 565–574.

[6] Damon Centola. 2010. The spread of behavior in an online social network experiment. *Science* 329, 5996 (2010), 1194–1197.

[7] Neil F. Johnson, Rhys Leahy, N. Johnson Restrepo, Nicholas Velásquez, Minzhang Zheng, Pedro Manrique, Prajwal Devkota, and Stefan Wuchty. 2019. Hidden resilience and adaptive dynamics of the global online hate ecology. *Nature* 573, 7773 (2019), 261–265.

[8] Ido Guy, Naama Zwerdling, Inbal Ronen, David Carmel, and Erel Uziel. 2010. Social media recommendation based on people and tags. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*. 194–201.

[9] Elaheh Momeni, Claire Cardie, and Nicholas Diakopoulos. 2015. A survey on assessment and ranking methodologies for user-generated content on the web. *ACM Computing Surveys (CSUR)* 48, 3 (2015), 1–49.

[10] Tarleton Gillespie. 2018. *Custodians of the Internet: Platforms, content moderation, and the hidden decisions that shape social media*. Yale University Press.

[11] Christopher A. Bail, Lisa P. Argyle, Taylor W. Brown, John P. Bumpus, Haohan Chen, M. B. Fallin Hunzaker, Jaemin Lee, Marcus Mann, Friedolin Merhout, and Alexander Volfovsky. 2018. Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences* 115, 37 (Aug. 2018), 9216–9221. DOI: http://doi.org/10.1073/pnas.1804840115

[12] Joshua A. Tucker, Andrew Guess, Pablo Barberá, Cristian Vaccari, Alexandra Siegel, Sergey Sanovich, Denis Stukal, and Brendan Nyhan. 2018. Social media, political polarization, and political disinformation: A review of the scientific literature. *Political polarization, and political disinformation: A review of the scientific literature (March 19, 2018)*. DOI: https://doi.org/10.2139/ssrn.3144139

[13] Mohammad Noor Injadat, Fadi Salo, and Ali Bou Nassif. 2016. Data mining techniques in social media: A survey. *Neurocomputing* 214 (2016), 654–670. DOI: https://doi.org/10.1016/j.neucom.2016.06.045

[14] Seth Flaxman, Sharad Goel, and Justin M. Rao. 2016. Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly* 80, S1 (2016), 298–320.

[15] Andrew Guess, Brendan Nyhan, and Jason Reifler. 2018. Selective exposure to misinformation: Evidence from the consumption of fake news during the 2016 US presidential campaign. *European Research Council* 9, 3 (2018), 4.

[16] Ceren Budak, Brendan Nyhan, David M. Rothschild, Emily Thorson, and Duncan J. Watts. 2024. Misunderstanding the harms of online misinformation. *Nature* 630, 8015 (2024), 45–53.

[17] Alessandro Bessi, Mauro Coletto, George Alexandru Davidescu, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Science vs conspiracy: Collective narratives in the age of misinformation. *PLOS One* 10, 2 (2015), e0118093.

[18] Matteo Cinelli, Gianmarco De Francisci Morales, Alessandro Galeazzi, Walter Quattrociocchi, and Michele Starnini. 2021. The echo chamber effect on social media. *Proceedings of the National Academy of Sciences* 118, 9 (2021), e2023301118.

[19] Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, and Michael Mathioudakis. 2018. Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship. In *Proceedings of the 2018 world wide web conference*. 913–922.

[20] Pablo Barberá, John T. Jost, Jonathan Nagler, Joshua A. Tucker, and Richard Bonneau. 2015. Tweeting from left to right: Is online political communication more than an echo chamber? *Psychological science* 26, 10 (2015), 1531–1542.

[21] Michela Del Vicario, Fabiana Zollo, Guido Caldarelli, Antonio Scala, and Walter Quattrociocchi. 2017. Mapping social dynamics on Facebook: The Brexit debate. *Social Networks* 50 (2017), 6–16. DOI: https://doi.org/10.1016/j.socnet.2017.02.002

[22] Matteo Cinelli, Walter Quattrociocchi, Alessandro Galeazzi, Carlo Michele Valensise, Emanuele Brugnoli, Ana Lucia Schmidt, Paola Zola, Fabiana Zollo, and Antonio Scala. 2020. The COVID-19 social media infodemic. *Scientific Reports* 10, 1 (2020), 16598.

[23] Israel Junior Borges Do Nascimento, Ana Beatriz Pizarro, Jussara M. Almeida, Natasha Azzopardi-Muscat, Marcos André Gonçalves, Maria Björklund, and David Novillo-Ortiz. 2022. Infodemics and health misinformation: A systematic review of reviews. *Bulletin of the World Health Organization* 100, 9 (2022), 544.

[24] Javier Alvarez-Galvez, Victor Suarez-Lledo, and Antonio Rojas-Garcia. 2021. Determinants of infodemics during disease outbreaks: a systematic review. *Frontiers in Public Health* 9 (2021), 603603. DOI: https://doi.org/10.3389/fpubh.2021.603603

[25] Hilal Bashir and Shabir Ahmad Bhat. 2017. Effects of social media on mental health: A review. *International Journal of Indian Psychology* 4, 3 (2017), 125–131.

[26] Francesco Pierri, Brea L. Perry, Matthew R. DeVerna, Kai-Cheng Yang, Alessandro Flammini, Filippo Menczer, and John Bryden. 2022. Online misinformation is linked to early COVID-19 vaccination hesitancy and refusal. *Scientific Reports* 12, 1 (2022), 5966.

[27] Talha Burki. 2019. Vaccine misinformation and social media. *The Lancet Digital Health* 1, 6 (2019), e258–e259.

[28] Amy Orben, Andrew K. Przybylski, Sarah-Jayne Blakemore, and Rogier A. Kievit. 2022. Windows of developmental sensitivity to social media. *Nature Communications* 13, 1 (March 2022). DOI: http://doi.org/10.1038/s41467-022-29296-3

[29] J. A. Naslund, K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels. 2016. The future of mental health care: Peer-to-peer support and social media. *Epidemiology and Psychiatric Sciences* 25, 2 (Jan. 2016), 113–122. DOI: http://doi.org/10.1017/s2045796015001067

[30] Paul E. Weigle and Reem M. A. Shafi. 2024. Social media and youth mental health. *Current Psychiatry Reports* 26, 1 (2024), 1–8.

[31] Christopher J. Ferguson, Linda K. Kaye, Dawn Branley-Bell, and Patrick Markey. 2025. There is no evidence that time spent on social media is correlated with adolescent mental health problems: Findings from a meta-analysis. *Professional Psychology: Research and Practice* 56, 1 (2025), 73–83. DOI: https://doi.org/10.1037/pro0000589

[32] Giulio Pecile, Niccolò Di Marco, Matteo Cinelli, and Walter Quattrociocchi. 2025. Mapping the global election landscape on social media in 2024. *PLOS One* 20, 2 (2025), e0316271.

[33] Ekaterina Zhuravskaya, Maria Petrova, and Ruben Enikolopov. 2020. Political effects of the internet and social media. *Annual Review of Economics* 12, 1 (2020), 415–438.

[34] Panagiotis T. Metaxas and Eni Mustafaraj. 2012. Social media and the elections. *Science* 338, 6106 (2012), 472–473.

[35] Sinan Aral and Dean Eckles. 2019. Protecting Elections from Social Media Manipulation. *Science* 365, 6456 (2019), 858–861.

[36] Derek Ruths and Jürgen Pfeffer. 2014. Social media for large studies of behavior. *Science* 346, 6213 (2014), 1063–1064.

[37] David M. J. Lazer, Alex Pentland, Duncan J. Watts, Sinan Aral, Susan Athey, Noshir Contractor, Deen Freelon, Sandra Gonzalez-Bailon, Gary King, Helen Margetts, et al. 2020. Computational social science: Obstacles and opportunities. *Science* 369, 6507 (2020), 1060–1062.

[38] Sandra González-Bailón, Javier Borge-Holthoefer, Alejandro Rivero, and Yamir Moreno. 2011. The dynamics of protest recruitment through an online network. *Scientific Reports* 1, 1 (2011), 1–7.

[39] Michele Avalle, Niccoló Di Marco, Gabriele Etta, Emanuele Sangiorgio, Shayan Alipour, Anita Bonetti, Lorenzo Alvisi, Antonio Scala, Andrea Baronchelli, Matteo Cinelli, et al. 2024. Persistent interaction patterns across social media platforms and over time. *Nature* 628, 8008 (March 2024), 582–589. DOI: http://doi.org/10.1038/s41586-024-07229-y

[40] Claudia Wagner, Markus Strohmaier, Alexandra Olteanu, Emre Kıcıman, Noshir Contractor, and Tina Eliassi-Rad. 2021. Measuring algorithmically infused societies. *Nature* 595, 7866 (2021), 197–204.

[41] David Lazer, Alex Pentland, Lada Adamic, Sinan Aral, Albert-László Barabási, Devon Brewer, Nicholas Christakis, Noshir Contractor, James Fowler, Myron Gutmann, et al. 2009. Computational social science. *Science* 323, 5915 (2009), 721–723.

[42] R. Conte, N. Gilbert, G. Bonelli, C. Cioffi-Revilla, G. Deffuant, J. Kertesz, V. Loreto, S. Moat, J. P. Nadal, A. Sanchez, et al. 2012. Manifesto of computational social science. *The European Physical Journal Special Topics* 214, 1 (Nov. 2012), 325–346. DOI: http://doi.org/10.1140/epjst/e2012-01697-8

[43] Duncan J. Watts. 2007. A twenty-first century science. *Nature* 445, 7127 (2007), 489–489.

[44] Sandra González-Bailón and Ning Wang. 2013. Networked discontent: The anatomy of protest campaigns in social media. *Social networks* 44 (2013), 95–104. DOI: https://doi.org/10.2139/ssrn.2268165

[45] Mohsen Mosleh, Gordon Pennycook, and David G. Rand. 2020. Self-reported willingness to share political news articles in online surveys correlates with actual sharing on Twitter. *PLOS One* 15, 2 (2020), e0228882.

[46] Zeynep Tufekci. 2014. Big questions for social media big data: Representativeness, validity and other methodological pitfalls. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 8. 505–514.

[47] Xiaohui Wang, Yunya Song, and Youzhen Su. 2023. Less fragmented but highly centralized: A bibliometric analysis of research in computational social science. *Social Science Computer Review* 41, 3 (2023), 946–966.

[48] Juho Pääkkönen, Matti Nelimarkka, and Samuli Reijula. 2025. The emergence of computational social science: Intellectual integration or persistent fragmentation? Center for Open Science. DOI: https://doi.org/10.31235/osf.io/g6m34_v2

[49] Barberá Pablo. 2014. How social media reduces mass political polarization. Evidence from Germany, Spain, and the US. *Job Market Paper* (2014).

[50] Yphtach Lelkes, Gaurav Sood, and Shanto Iyengar. 2017. The hostile audience: The effect of access to broadband internet on partisan affect. *American Journal of Political Science* 61, 1 (2017), 5–20.

[51] Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, and Fabiana Zollo. 2019. Polarization and fake news: Early warning of potential misinformation targets. *ACM Transactions on the Web (TWEB)* 13, 2 (2019), 1–22.

[52] Deen Freelon. 2018. Computational research in the post-API age. *Political Communication* 35, 4 (2018), 665–668.

[53] Andrew M. Guess and Benjamin A. Lyons. 2020. Misinformation, disinformation, and online propaganda. *Social Media and Democracy: The State of the Field, Prospects for Reform* 10 (2020), 10–33. DOI: https://doi.org/10.1017/9781108890960.003

[54] Niccolò Di Marco, Edoardo Loru, Anita Bonetti, Alessandra Olga Grazia Serra, Matteo Cinelli, and Walter Quattrociocchi. 2024. Patterns of linguistic simplification on social media platforms over time. *Proceedings of the National Academy of Sciences* 121, 50 (Dec. 2024). DOI: http://doi.org/10.1073/pnas.2412105121

[55] Edoardo Di Martino, Alessandro Galeazzi, Michele Starnini, Walter Quattrociocchi, and Matteo Cinelli. 2025. Ideological fragmentation of the social media ecosystem: From echo chambers to echo platforms. *PNAS Nexus* (Aug. 2025). DOI: http://doi.org/10.1093/pnasnexus/pgaf262

[56] Michael A. Cacciatore. 2021. Misinformation and public opinion of science and health: Approaches, findings, and future directions. *Proceedings of the National Academy of Sciences* 118, 15 (April 2021). DOI: http://doi.org/10.1073/pnas.1912437117

[57] Max Falkenberg, Matteo Cinelli, Alessandro Galeazzi, Christopher A. Bail, Rosa M. Benito, Axel Bruns, Anatoliy Gruzd, David Lazer, Jae K. Lee, Jennifer McCoy, et al. 2025. Towards global equity in political polarization research. (2025). DOI: http://doi.org/10.48550/ARXIV.2504.11090

[58] Faisal Alatawi, Lu Cheng, Anique Tahir, Mansooreh Karami, Bohan Jiang, Tyler Black, and Huan Liu. 2021. A survey on echo chambers on social media: Description, detection and mitigation. arXiv:2112.05084. Retrieved from https://arxiv.org/abs/2112.05084 (2021).

[59] David Hartmann, Sonja Mei Wang, Lena Pohlmann, and Bettina Berendt. 2025. A systematic review of echo chamber research: Comparative analysis of conceptualizations, operationalizations, and varying outcomes. *Journal of Computational Social Science* 8, 2 (2025), 52.

[60] Ludovic Terren Ludovic Terren and Rosa Borge-Bravo Rosa Borge-Bravo. 2021. Echo chambers on social media: A systematic review of the literature. *Review of Communication Research* 9 (March 2021). DOI: http://doi.org/10.12840/issn.2255-4165.028

[61] Long Jin, Yang Chen, Tianyi Wang, Pan Hui, and Athanasios V. Vasilakos. 2013. Understanding user behavior in online social networks: a survey. *IEEE Communications Magazine* 51, 9 (Sept. 2013), 144–150. DOI: http://doi.org/10.1109/mcom.2013.6588663

[62] Lorenzo Mannocci, Michele Mazza, Anna Monreale, Maurizio Tesconi, and Stefano Cresci. 2024. Detection and characterization of coordinated online behavior: A survey. (2024). DOI: http://doi.org/10.48550/ARXIV.2408.01257

[63] Guido Caldarelli, Oriol Artime, Giulia Fischetti, Stefano Guarino, Andrzej Nowak, Fabio Saracco, Petter Holme, and Manlio de Domenico. 2025. The physics of news, rumors, and opinions. (2025). DOI: http://doi.org/10.48550/ARXIV.2510.15053

[64] William J Reid, Bonnie Davis Kenaley, and Julanne Colvin. 2004. Do some interventions work better than others? A review of comparative social work experiments. *Social Work Research* 28, 2 (2004), 71–81.

[65] Ming-Sum Tsui. 1997. Empirical research on social work supervision: The state of the art (1970-1995). *Journal of Social Service Research* 23, 2 (1997), 39–54.

[66] Stephen H. Bell and Laura R. Peck. 2012. Obstacles to and limitations of social experiments: 15 false alarms. *Abt Thought Leadership Paper, Abt Associates* (2012).

[67] Katherine S. Button, John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson, and Marcus R. Munafò. 2013. Power failure: Why small sample size undermines the reliability of neuroscience. *Nature Reviews Neuroscience* 14, 5 (April 2013), 365–376. DOI: http://doi.org/10.1038/nrn3475

[68] Matthew C. Makel, Jonathan A. Plucker, and Boyd Hegarty. 2012. Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science* 7, 6 (Nov. 2012), 537–542. DOI: http://doi.org/10.1177/1745691612460688

[69] Adrian Furnham. 1986. Response bias, social desirability and dissimulation. *Personality and Individual Differences* 7, 3 (Jan. 1986), 385–400. DOI: http://doi.org/10.1016/0191-8869(86)90014-0

[70] Robert Rosenman, Vidhura Tennekoon, and Laura G. Hill. 2011. Measuring bias in self-reported data. *International Journal of Behavioural and Healthcare Research* 2, 4 (2011), 320. DOI: http://doi.org/10.1504/ijbhr.2011.043414

[71] Katharina Koller, Paulina K. Pankowska, and Cameron Brick. 2023. Identifying bias in self-reported pro-environmental behavior. *Current Research in Ecological and Social Psychology* 4 (2023), 100087. DOI: http://doi.org/10.1016/j.cresp.2022.100087

[72] Angela Xiao Wu and Harsh Taneja. 2021. Platform enclosure of human behavior and its measurement: Using behavioral trace data against platform episteme. *New Media & Society* 23, 9 (2021), 2650–2667.

[73] Cheonsoo Kim and Sung-Un Yang. 2017. Like, comment, and share on Facebook: How each behavior differs from the other. *Public relations review* 43, 2 (2017), 441–449.

[74] Melanie Schreiner, Thomas Fischer, and Rene Riedl. 2019. Impact of content characteristics and emotion on behavioral engagement in social media: literature review and research agenda. *Electronic Commerce Research* 21, 2 (2019), 329–345. DOI: https://doi.org/10.1007/s10660-019-09353-8

[75] Gordon Bell, Tony Hey, and Alex Szalay. 2009. Beyond the data deluge. *Science* 323, 5919 (March 2009), 1297–1298. DOI: http://doi.org/10.1126/science.1170411

[76] Douglas A. Parry, Brittany I. Davidson, Craig J. R. Sewall, Jacob T. Fisher, Hannah Mieczkowski, and Daniel S. Quintana. 2021. A systematic review and meta-analysis of discrepancies between logged and self-reported digital media use. *Nature Human Behaviour* 5, 11 (May 2021), 1535–1547. DOI: http://doi.org/10.1038/s41562-021-01117-5

[77] Mubashir Sultan, Christin Scholz, and Wouter van den Bos. 2023. Leaving traces behind: Using social media digital trace data to study adolescent wellbeing. *Computers in Human Behavior Reports* 10 (2023), 100281. DOI: https://doi.org/10.1016/j.chbr.2023.100281

[78] Shamanth Kumar, Reza Zafarani, and Huan Liu. 2011. Understanding user migration patterns in social media. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 25. 1204–1209.

[79] Mariam Adedoyin-Olowe, Mohamed Medhat Gaber, and Frederic Stahl. 2014. A survey of data mining techniques for social media analysis. *Journal of Data Mining & Digital Humanities* 2014 (2014).

[80] Mei Li, Xiang Wang, Kai Gao, and Shanshan Zhang. 2017. A survey on information diffusion in online social networks: Models and methods. *Information* 8, 4 (2017), 118.

[81] Adrien Guille, Hakim Hacid, Cecile Favre, and Djamel A Zighed. 2013. Information diffusion in online social networks: A survey. *ACM Sigmod Record* 42, 2 (2013), 17–28.

[82] Lars Backstrom, Dan Huttenlocher, Jon Kleinberg, and Xiangyang Lan. 2006. Group formation in large social networks: membership, growth, and evolution. In *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. 44–54.

[83] Michela Del Vicario, Gianna Vivaldo, Alessandro Bessi, Fabiana Zollo, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2016. Echo chambers: Emotional contagion and group polarization on facebook. *Scientific Reports* 6, 1 (2016), 37825.

[84] Soo Ling Lim and Peter J. Bentley. 2022. Opinion amplification causes extreme polarization in social networks. *Scientific Reports* 12, 1 (2022), 18131.

[85] Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. 2022. Algorithmic amplification of politics on Twitter. *Proceedings of the National Academy of Sciences* 119, 1 (2022), e2025334119.

[86] Ray M. Chang, Robert J. Kauffman, and YoungOk Kwon. 2014. Understanding the paradigm shift to computational social science in the presence of big data. *Decision Support Systems* 63 (July 2014), 67–80. DOI: http://doi.org/10.1016/j.dss.2013.08.008

[87] Deen Freelon. 2014. On the interpretation of digital trace data in communication and social computing research. *Journal of Broadcasting; Electronic Media* 58, 1 (Jan. 2014), 59–75. DOI: http://doi.org/10.1080/08838151.2013.875018

[88] Amelia Acker and Adam Kreisberg. 2020. Social media data archives in an API-driven world. *Archival Science* 20, 2 (2020), 105–123.

[89] Kyra Milan Abrams and Madelyn Rose Sanfilippo. 2025. Neutrality or Contextuality. *Social Informatics* (2025).

[90] Blake Hallinan, Rebecca Scharlach, and Limor Shifman. 2022. Beyond neutrality: Conceptualizing platform values. *Communication Theory* 32, 2 (2022), 201–222.

[91] Juergen Pfeffer, Daniel Matter, Kokil Jaidka, Onur Varol, Afra Mashhadi, Jana Lasser, Dennis Assenmacher, Siqi Wu, Diyi Yang, Cornelia Brantner, et al. 2023. Just another day on Twitter: A complete 24 hours of Twitter data. (2023). DOI: http://doi.org/10.48550/ARXIV.2301.11429

[92] Reza Zafarani and Huan Liu. 2014. Behavior analysis in social media. *IEEE Intelligent systems* 29, 4 (2014), 1–4.

[93] Gema Bello-Orgaz, Jason J. Jung, and David Camacho. 2016. Social big data: Recent achievements and new challenges. *Information Fusion* 28 (March 2016), 45–59. DOI: http://doi.org/10.1016/j.inffus.2015.08.005

[94] Ibrahim Abaker Targio Hashem, Ibrar Yaqoob, Nor Badrul Anuar, Salimah Mokhtar, Abdullah Gani, and Samee Ullah Khan. 2015. The rise of "big data" on cloud computing: Review and open research issues. *Information Systems* 47 (Jan. 2015), 98–115. DOI: http://doi.org/10.1016/j.is.2014.07.006

[95] Min Chen, Shiwen Mao, and Yunhao Liu. 2014. Big data: A survey. *Mobile Networks and Applications* 19, 2 (Jan. 2014), 171–209. DOI: http://doi.org/10.1007/s11036-013-0489-0

[96] Raviv Cohen and Derek Ruths. 2013. Classifying political orientation on Twitter: It's not easy!. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 7. 91–99.

[97] David Jurgens, Tyler Finethy, James McCorriston, Yi Xu, and Derek Ruths. 2015. Geolocation prediction in Twitter using social networks: A critical analysis and review of current practice. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 9. 188–197.

[98] Alexandre Bovet and Hernán A. Makse. 2019. Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications* 10, 1 (2019), 7.

[99] James Flamino, Alessandro Galeazzi, Stuart Feldman, Michael W. Macy, Brendan Cross, Zhenkun Zhou, Matteo Serafino, Alexandre Bovet, Hernán A. Makse, and Boleslaw K. Szymanski. 2023. Political polarization of news media and influencers on Twitter in the 2016 and 2020 US presidential elections. *Nature Human Behaviour* 7, 6 (2023), 904–916.

[100] Lilian Weng, Filippo Menczer, and Yong-Yeol Ahn. 2013. Virality prediction and community structure in social networks. *Scientific Reports* 3, 1 (2013), 2522.

[101] Neil F. Johnson, Nicolas Velásquez, Nicholas Johnson Restrepo, Rhys Leahy, Nicholas Gabriel, Sara El Oud, Minzhang Zheng, Pedro Manrique, Stefan Wuchty, and Yonatan Lupu. 2020. The online competition between pro-and anti-vaccination views. *Nature* 582, 7811 (2020), 230–233.

[102] Han Woo Park and Mike Thelwall. 2003. Hyperlink analyses of the World Wide Web: A review. *Journal of computer-mediated communication* 8, 4 (2003), JCMC843.

[103] Bing Han, David Hirshleifer, and Johan Walden. 2023. Visibility bias in the Transmission of Consumption Beliefs and Undersaving. *The Journal of Finance* 78, 3 (April 2023), 1647–1704. DOI: http://doi.org/10.1111/jofi.13223

[104] Jack Bandy and Nicholas Diakopoulos. 2023. Facebook's news feed algorithm and the 2020 us election. *Social Media+ Society* 9, 3 (2023), 20563051231196898.

[105] Anne Oeldorf-Hirsch, Mike Schmierbach, Alyssa Appelman, and Michael P Boyle. 2020. The ineffectiveness of fact-checking labels on news memes and articles. *Mass Communication and Society* 23, 5 (2020), 682–704.

[106] Cameron Martel and David G. Rand. 2024. Fact-checker warning labels are effective even for those who distrust fact-checkers. *Nature Human Behaviour* 8, 10 (2024), 1957–1967.

[107] David Lazer. 2015. The rise of the social algorithm. *Science* 348, 6239 (2015), 1090–1091.

[108] Meta Platforms, Inc. 2020. 2020 Election Research. Retrieved November 12, 2025 from https://research.facebook.com/2020-election-research/. (2020).

[109] Stephan Lewandowsky and Sander Van Der Linden. 2021. Countering misinformation and fake news through inoculation and prebunking. *European Review of Social Psychology* 32, 2 (2021), 348–384.

[110] Mohsen Mosleh, Gordon Pennycook, and David G. Rand. 2022. Field experiments on social media. *Current Directions in Psychological Science* 31, 1 (2022), 69–75.

[111] Nathan Bartley, Andres Abeliuk, Emilio Ferrara, and Kristina Lerman. 2021. Auditing algorithmic bias on Twitter. In *Proceedings of the 13th ACM Web Science Conference 2021*. 65–73.

[112] Hong Shen, Alicia DeVos, Motahhare Eslami, and Kenneth Holstein. 2021. Everyday algorithm auditing: Understanding the power of everyday users in surfacing harmful algorithmic behaviors. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.

[113] Jack Bandy. 2021. Problematic machine behavior: A systematic literature review of algorithm audits. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–34.

[114] Peter R. Scott and J. Mike Jacka. 2011. *Auditing Social Media: A Governance and Risk Guide*. John Wiley & Sons.

[115] Sharad Goel, Jake Hofman, and M. Sirer. 2012. Who does what on the web: A large-scale study of browsing behavior. In *Proceedings of the International AAAI Conference on web and Social Media*, Vol. 6. 130–137.

[116] Ravi Kumar and Andrew Tomkins. 2010. A characterization of online browsing behavior. In *Proceedings of the 19th International Conference on World Wide Web*. 561–570.

[117] Randolph E. Bucklin and Catarina Sismeiro. 2003. A model of web site browsing behavior estimated on clickstream data. *Journal of Marketing Research* 40, 3 (2003), 249–267.

[118] Miriam J. Metzger, Christo Wilson, and Ben Y. Zhao. 2018. Benefits of browsing? The prevalence, nature, and effects of profile consumption behavior in social network sites. *Journal of Computer-Mediated Communication* 23, 2 (2018), 72–89.

[119] William Rand, Jeffrey Herrmann, Brandon Schein, and Neža Vodopivec. 2015. An agent-based model of urgent diffusion in social media. *Journal of Artificial Societies and Social Simulation* 18, 2 (2015), 1.

[120] Fujio Toriumi, Hitoshi Yamamoto, and Isamu Okada. 2012. Why do people use social media? agent-based simulation and population dynamics analysis of the evolution of cooperation in social media. In *Proceedings of the 2012 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology*, Vol. 2. IEEE, 43–50.

[121] Darius Plikynas, Aistis Raudys, and Sarunas Raudys. 2015. Agent-based modelling of excitation propagation in social media groups. *Journal of Experimental & Theoretical Artificial Intelligence* 27, 4 (2015), 373–388.

[122] Chaitanya Kaligotla, Enver Yücesan, and Stephen E. Chick. 2015. An agent based model of spread of competing rumors through online interactions on social media. In *Proceedings of the 2015 Winter Simulation Conference (WSC)*. IEEE, 3985–3996.

[123] Erica Cau, Valentina Pansanella, Dino Pedreschi, and Giulio Rossetti. 2025. Selective agreement, not sycophancy: investigating opinion dynamics in LLM interactions. *EPJ Data Science* 14, 1 (Aug. 2025). DOI: http://doi.org/10.1140/epjds/s13688-025-00579-1

[124] Erica Cau, Andrea Failla, and Giulio Rossetti. 2025. *Bots of a Feather: Mixing Biases in LLMs' Opinion Dynamics*. Springer Nature Switzerland, 166–176. DOI: http://doi.org/10.1007/978-3-031-82439-5_14

[125] Marcel Binz, Stephan Alaniz, Adina Roskies, Balazs Aczel, Carl T. Bergstrom, Colin Allen, Daniel Schad, Dirk Wulff, Jevin D. West, Qiong Zhang, et al. 2025. How should the advancement of large language models affect the practice of science? *Proceedings of the National Academy of Sciences* 122, 5 (Jan. 2025). DOI: http://doi.org/10.1073/pnas.2401227121

[126] Lisa Messeri and M. J. Crockett. 2024. Artificial intelligence and illusions of understanding in scientific research. *Nature* 627, 8002 (March 2024), 49–58. DOI: http://doi.org/10.1038/s41586-024-07146-0

[127] Igor Grossmann, Matthew Feinberg, Dawn C. Parker, Nicholas A. Christakis, Philip E. Tetlock, and William A. Cunningham. 2023. AI and the transformation of social science research. *Science* 380, 6650 (June 2023), 1108–1109. DOI: http://doi.org/10.1126/science.adi1778

[128] Yao Qu and Jue Wang. 2024. Performance and biases of Large Language Models in public opinion simulation. *Humanities and Social Sciences Communications* 11, 1 (Aug. 2024). DOI: http://doi.org/10.1057/s41599-024-03609-x

[129] Rickard Stureborg, Dimitris Alikaniotis, and Yoshi Suhara. 2024. Large Language Models are Inconsistent and Biased Evaluators. (2024). DOI: http://doi.org/10.48550/ARXIV.2405.01724

[130] Tiancheng Hu, Yara Kyrychenko, Steve Rathje, Nigel Collier, Sander van der Linden, and Jon Roozenbeek. 2024. Generative language models exhibit social identity biases. *Nature Computational Science* 5, 1 (Dec. 2024), 65–75. DOI: http://doi.org/10.1038/s43588-024-00741-1

[131] Nicolas Yax, Hernán Anlló, and Stefano Palminteri. 2024. Studying and improving reasoning in humans and machines. *Communications Psychology* 2, 1 (June 2024). DOI: http://doi.org/10.1038/s44271-024-00091-8

[132] James W. A. Strachan, Dalila Albergo, Giulia Borghini, Oriana Pansardi, Eugenio Scaliti, Saurabh Gupta, Krati Saxena, Alessandro Rufo, Stefano Panzeri, Guido Manzi, et al. 2024. Testing theory of mind in large language models and humans. *Nature Human Behaviour* 8, 7 (May 2024), 1285–1295. DOI: http://doi.org/10.1038/s41562-024-01882-z

[133] Emily M. Bender, Timnit Gebru, Angelina McMillan-Major, and Shmargaret Shmitchell. 2021. On the dangers of stochastic parrots: Can language models be too big?. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. ACM, 610–623. DOI: http://doi.org/10.1145/3442188.3445922

[134] Melanie Mitchell and David C. Krakauer. 2023. The debate over understanding in AI's large language models. *Proceedings of the National Academy of Sciences* 120, 13 (March 2023). DOI: http://doi.org/10.1073/pnas.2215907120

[135] Fabiana Zollo, Andrea Baronchelli, Cornelia Betsch, Marco Delmastro, and Walter Quattrociocchi. 2024. Understanding the complex links between social media and health behaviour. *bmj* 385 (2024), e075645. DOI: https://doi.org/10.1136/bmj-2023-075645

[136] Michael W. Wagner. 2023. Independence by permission. *Science* 381, 6656 (2023), 388–391.

[137] Inder M. Verma. 2014. Editorial expression of concern: Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences of the United States of America* 111, 29 (2014), 10779–10779.

[138] Asunción Esteve. 2017. The business of personal data: Google, Facebook, and privacy issues in the EU and the USA. *International Data Privacy Law* 7, 1 (2017), 36–47.

[139] Giorgio Fagiolo, Paul Windrum, and Alessio Moneta. 2006. *Empirical Validation of Agent-Based Models: A Critical Survey*. Technical Report. LEM Working Paper Series.

[140] Gianluca Manzo and Toby Matthews. 2014. Potentialities and limitations of agent-based simulations. *Revue française de sociologie* 55, 4 (2014), 653–688.

[141] Paul Windrum, Giorgio Fagiolo, and Alessio Moneta. 2007. Empirical validation of agent-based models: Alternatives and prospects. *Journal of Artificial Societies and Social Simulation* 10, 2 (2007), 8.

[142] Norjihan Abdul Ghani, Suraya Hamid, Ibrahim Abaker Targio Hashem, and Ejaz Ahmed. 2019. Social media big data analytics: A survey. *Computers in Human Behavior* 101 (2019), 417–428. DOI: https://doi.org/10.1016/j.chb.2018.08.039

[143] Maksim Kitsak, Lazaros K. Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H. Eugene Stanley, and Hernán A. Makse. 2010. Identification of influential spreaders in complex networks. *Nature Physics* 6, 11 (2010), 888–893.

[144] Zoe Corbyn. 2012. Facebook experiment boosts US voter turnout. *Nature* 12 (2012).

[145] Filippo Menczer and Thomas Hills. 2020. The attention economy. *Scientific American* 323, 6 (2020), 54–61.

[146] Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2015. The production of information in the attention economy. *Scientific Reports* 5, 1 (2015), 9452.

[147] Michael H. Goldhaber. 1997. The attention economy and the net. *First Monday* (1997). DOI: https://doi.org/10.5210/fm.v2i4.519

[148] Sen Pei, Lev Muchnik, José S. Andrade, Jr, Zhiming Zheng, and Hernán A. Makse. 2014. Searching for superspreaders of information in real-world social media. *Scientific Reports* 4, 1 (2014), 5547.

[149] Ahmad Zareie and Rizos Sakellariou. 2021. Minimizing the spread of misinformation in online social networks: A survey. *Journal of Network and Computer Applications* 186 (2021), 103094. DOI: https://doi.org/10.1016/j.jnca.2021.103094
[150] Jisun An, Daniele Quercia, and Jon Crowcroft. 2013. Fragmented social media: A look into selective exposure to political news. In *Proceedings of the 22nd International Conference on World Wide Web*. 51–52.
[151] Thomas H. Davenport and John C. Beck. 2001. The attention economy. *Ubiquity* 2001, May (May 2001), 1. DOI: http://doi.org/10.1145/376625.376626
[152] Josef Falkinger. 2008. Limited attention as a scarce resource in information-rich economies. *The Economic Journal* 118, 532 (Sept. 2008), 1596–1620. DOI: http://doi.org/10.1111/j.1468-0297.2008.02182.x
[153] Dominic Spohr. 2017. Fake news and ideological polarization: Filter bubbles and selective exposure on social media. *Business Information Review* 34, 3 (2017), 150–160.
[154] Simon P. Anderson and André De Palma. 2012. Competition for attention in the information (overload) age. *The RAND Journal of Economics* 43, 1 (2012), 1–25.
[155] Eszter Hargittai, W. Russell Neuman, and Olivia Curry. 2012. Taming the information tide: Perceptions of information overload in the american home. *The Information Society* 28, 3 (May 2012), 161–173. DOI: http://doi.org/10.1080/01972243.2012.669450
[156] David Bawden and Lyn Robinson. 2020. Information overload: An introduction. In *Oxford Research Encyclopedia of Politics*. Oxford University Press. DOI: http://doi.org/10.1093/acrefore/9780190228637.013.1360
[157] Matteo Cinelli, Emanuele Brugnoli, Ana Lucia Schmidt, Fabiana Zollo, Walter Quattrociocchi, and Antonio Scala. 2020. Selective exposure shapes the Facebook news diet. *PLOS one* 15, 3 (2020), e0229129.
[158] Andrew M. Guess, Neil Malhotra, Jennifer Pan, Pablo Barberá, Hunt Allcott, Taylor Brown, Adriana Crespo-Tenorio, Drew Dimmery, Deen Freelon, Matthew Gentzkow, et al. 2023. How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science* 381, 6656 (July 2023), 398–404. DOI: http://doi.org/10.1126/science.abp9364
[159] Zeynep Tufekci. 2013. "Not this one": Social movements, the attention economy, and microcelebrity networked activism. *American Behavioral Scientist* 57, 7 (March 2013), 848–870. DOI: http://doi.org/10.1177/0002764213479369
[160] Vikram R. Bhargava and Manuel Velasquez. 2020. Ethics of the attention economy: The problem of social media addiction. *Business Ethics Quarterly* 31, 3 (Oct. 2020), 321–359. DOI: http://doi.org/10.1017/beq.2020.32
[161] Alessandro Bessi, Fabiana Zollo, Michela Del Vicario, Michelangelo Puliga, Antonio Scala, Guido Caldarelli, Brian Uzzi, and Walter Quattrociocchi. 2016. Users polarization on Facebook and Youtube. *PLOS One* 11, 8 (2016), e0159641.
[162] Xu Cheng, Cameron Dale, and Jiangchuan Liu. 2008. Statistics and social network of youtube videos. In *Proceedings of the 2008 16th Interntional Workshop on Quality of Service*. IEEE, 229–238.
[163] Emanuele Sangiorgio, Matteo Cinelli, Roy Cerqueti, and Walter Quattrociocchi. 2024. Followers do not dictate the virality of news outlets on social media. *PNAS Nexus* 3, 7 (June 2024). DOI: http://doi.org/10.1093/pnasnexus/pgae257
[164] Kristina Lerman. 2007. Social information processing in news aggregation. *IEEE Internet Computing* 11, 6 (2007), 16–28.
[165] Gabriele Etta, Emanuele Sangiorgio, Niccolò Di Marco, Michele Avalle, Antonio Scala, Matteo Cinelli, and Walter Quattrociocchi. 2023. Characterizing engagement dynamics across topics on Facebook. *PLOS One* 18, 6 (June 2023), e0286150. DOI: http://doi.org/10.1371/journal.pone.0286150
[166] Claire E. Robertson, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J. Van Bavel, and Stefan Feuerriegel. 2023. Negativity drives online news consumption. *Nature Human Behaviour* 7, 5 (March 2023), 812–822. DOI: http://doi.org/10.1038/s41562-023-01538-4
[167] Maddalena Torricelli, Max Falkenberg, Alessandro Galeazzi, Fabiana Zollo, Walter Quattrociocchi, and Andrea Baronchelli. 2023. How does extreme weather impact the climate change discourse? Insights from the Twitter discussion on hurricanes. *PLOS Climate* 2, 11 (2023), e0000277.
[168] Riccardo Gallotti, Francesco Valle, Nicola Castaldo, Pierluigi Sacco, and Manlio De Domenico. 2020. Assessing the risks of 'infodemics' in response to COVID-19 epidemics. *Nature Human Behaviour* 4, 12 (2020), 1285–1293.
[169] Matteo Cinelli, Gabriele Etta, Michele Avalle, Alessandro Quattrociocchi, Niccolò Di Marco, Carlo Valensise, Alessandro Galeazzi, and Walter Quattrociocchi. 2022. Conspiracy theories and social media platforms. *Current Opinion in Psychology* 47 (Oct. 2022), 101407. DOI: http://doi.org/10.1016/j.copsyc.2022.101407
[170] Dino Pedreschi, Luca Pappalardo, Emanuele Ferragina, Ricardo Baeza-Yates, Albert-László Barabási, Frank Dignum, Virginia Dignum, Tina Eliassi-Rad, Fosca Giannotti, János Kertész, et al. 2025. Human-AI coevolution. *Artificial Intelligence* 339 (Feb. 2025), 104244. DOI: http://doi.org/10.1016/j.artint.2024.104244
[171] Alessandro Bessi, Fabio Petroni, Michela Del Vicario, Fabiana Zollo, Aris Anagnostopoulos, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Viral misinformation: The role of homophily and polarization. In *Proceedings of the 24th International Conference on World Wide Web*. 355–356.
[172] Joseph B. Bak-Coleman, Ian Kennedy, Morgan Wack, Andrew Beers, Joseph S. Schafer, Emma S. Spiro, Kate Starbird, and Jevin D. West. 2022. Combining interventions to reduce the spread of viral misinformation. *Nature Human Behaviour* 6, 10 (2022), 1372–1380.

[173] Ji Won Kim. 2018. They liked and shared: Effects of social media virality metrics on perceptions of message influence and behavioral intentions. *Computers in Human Behavior* 84 (2018), 153–161. DOI: https://doi.org/10.1016/j.chb.2018.01.030
[174] Maryam Mousavi, Hasan Davulcu, Mohsen Ahmadi, Robert Axelrod, Richard Davis, and Scott Atran. 2022. Effective messaging on social media: What makes online content go viral?. In *Proceedings of the ACM Web Conference 2022*. 2957–2966.
[175] Angeliki Nikolinakou and Karen Whitehill King. 2018. Viral video ads: Emotional triggers and social media virality. *Psychology & Marketing* 35, 10 (2018), 715–726.
[176] Jon Kleinberg. 2003. Bursty and hierarchical structure in streams. *Data Mining and Knowledge Discovery* 7, 4 (Oct. 2003), 373–397. DOI: http://doi.org/10.1023/a:1024940629314
[177] Riley Crane and Didier Sornette. 2008. Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences* 105, 41 (Oct. 2008), 15649–15653. DOI: http://doi.org/10.1073/pnas.0803685105
[178] Yasuko Matsubara, Yasushi Sakurai, B. Aditya Prakash, Lei Li, and Christos Faloutsos. 2012. Rise and fall patterns of information diffusion: Model and implications. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '12)*. ACM, 6–14. DOI: http://doi.org/10.1145/2339530.2339537
[179] Pedro Olmo S. Vaz de Melo, Christos Faloutsos, Renato Assunção, and Antonio Loureiro. 2013. The self-feeding process: A unifying model for communication dynamics in the web. In *Proceedings of the 22nd International Conference on World Wide Web (WWW '13)*. ACM, 1319–1330. DOI: http://doi.org/10.1145/2488388.2488503
[180] Rodrigo Alves, Antoine Ledent, Renato Assunção, Pedro Vaz-De-Melo, and Marius Kloft. 2024. Unraveling the Dynamics of Stable and Curious Audiences in Web Systems. In *Proceedings of the ACM Web Conference 2024 (WWW '24)*. ACM, 2464–2475. DOI: http://doi.org/10.1145/3589334.3645473
[181] Haoming Zhang, Yiping Yao, Wenjie Tang, Jiefan Zhu, and Yonghua Zhang. 2023. Opinion-aware information diffusion model based on multivariate marked Hawkes process. *Knowledge-Based Systems* 279 (Nov. 2023), 110883. DOI: http://doi.org/10.1016/j.knosys.2023.110883
[182] Miriam Fernandez, Alejandro Bellogín, and Iván Cantador. 2024. Analysing the effect of recommendation algorithms on the spread of misinformation. In *Proceedings of the 16th ACM Web Science Conference*. 159–169.
[183] Marco Guerini, Carlo Strapparava, and Gozde Ozbal. 2011. Exploring text virality in social networks. In *proceedings of the International AAAI Conference on Web and Social Media*, Vol. 5. 506–509.
[184] Yue Han, Theodoros Lappas, and Gaurav Sabnis. 2020. The importance of interactions between content characteristics and creator characteristics for studying virality in social media. *Information Systems Research* 31, 2 (2020), 576–588.
[185] Niccolò Di Marco, Matteo Cinelli, Shayan Alipour, and Walter Quattrociocchi. 2024. Users volatility on reddit and voat. *IEEE Transactions on Computational Social Systems* 11, 5 (2024), 5871–5879. DOI: https://doi.org/10.1109/tcss.2024.3379318
[186] Giulia Cencetti, Federico Battiston, Bruno Lepri, and Márton Karsai. 2021. Temporal properties of higher-order interactions in social networks. *Scientific Reports* 11, 1 (March 2021). DOI: http://doi.org/10.1038/s41598-021-86469-8
[187] Emanuele Sangiorgio, Niccolò Di Marco, Gabriele Etta, Matteo Cinelli, Roy Cerqueti, and Walter Quattrociocchi. 2025. Evaluating the effect of viral posts on social media engagement. *Scientific Reports* 15, 1 (2025), 639.
[188] Ullrich K. H. Ecker, Stephan Lewandowsky, John Cook, Philipp Schmid, Lisa K. Fazio, Nadia Brashier, Panayiota Kendeou, Emily K. Vraga, and Michelle A. Amazeen. 2022. The psychological drivers of misinformation belief and its resistance to correction. *Nature Reviews Psychology* 1, 1 (2022), 13–29.
[189] Sylvie C. Briand, Matteo Cinelli, Tim Nguyen, Rosamund Lewis, Dimitri Prybylski, Carlo M. Valensise, Vittoria Colizza, Alberto Eugenio Tozzi, Nicola Perra, Andrea Baronchelli, et al. 2021. Infodemics: A new challenge for public health. *Cell* 184, 25 (2021), 6010–6014.
[190] Stefan Stieglitz and Linh Dang-Xuan. 2013. Emotions and information diffusion in social media–sentiment of microblogs and sharing behavior. *Journal of Management Information Systems* 29, 4 (2013), 217–248.
[191] Arash Amini, Yigit Ege Bayiz, Eun-Ju Lee, Zeynep Somer-Topcu, Radu Marculescu, and Ufuk Topcu. 2025. How media competition fuels the spread of misinformation. *Science Advances* 11, 25 (June 2025). DOI: http://doi.org/10.1126/sciadv.adu7743
[192] Miriam Arnold, Mascha Goldschmitt, and Thomas Rigotti. 2023. Dealing with information overload: A comprehensive review. *Frontiers in Psychology* 14 (June 2023). DOI: http://doi.org/10.3389/fpsyg.2023.1122200
[193] George Loewenstein and Zachary Wojtowicz. 2025. The economics of attention. *Journal of Economic Literature* 63, 3 (Sept. 2025), 1038–1089. DOI: http://doi.org/10.1257/jel.20241665
[194] Maxwell E. McCombs and Donald L. Shaw. 1972. The agenda-setting function of mass media. *Public Opinion Quarterly* 36, 2 (1972), 176–187.
[195] Fabrizio Gilardi, Theresa Gessler, Maël Kubli, and Stefan Müller. 2022. Social media and political agenda setting. *Political Communication* 39, 1 (2022), 39–60.

[196] Xinxin Yang, Bo-Chiuan Chen, Mrinmoy Maity, and Emilio Ferrara. 2016. Social politics: Agenda setting and political communication on social media. In *Proceedings of the Social Informatics: 8th International Conference, SocInfo 2016, Bellevue, WA, USA, November 11-14, 2016, Proceedings, Part I 8*. Springer, 330–344.
[197] Henry Jenkins. 2004. The cultural logic of media convergence. *International Journal of Cultural Studies* 7, 1 (March 2004), 33–43. DOI: http://doi.org/10.1177/1367877904040603
[198] Victor Bekkers, Henri Beunders, Arthur Edwards, and Rebecca Moody. 2011. New media, micromobilization, and political agenda setting: Crossover effects in political mobilization and media usage. *The Information Society* 27, 4 (July 2011), 209–219. DOI: http://doi.org/10.1080/01972243.2011.583812
[199] Walter Quattrociocchi, Guido Caldarelli, and Antonio Scala. 2014. Opinion dynamics on interacting networks: Media competition and social influence. *Scientific Reports* 4, 1 (2014), 4938.
[200] Rachel M. Kim and Ashton Anderson. 2025. The agenda-setting function of social media. In *Proceedings of the ACM on Web Conference 2025*. 601–613.
[201] Keyan Ding, Ronggang Wang, and Shiqi Wang. 2019. Social media popularity prediction: A multiple feature fusion approach with deep neural networks. In *Proceedings of the 27th ACM International Conference on Multimedia*. 2682–2686.
[202] Qingchao Kong, Wenji Mao, Guandan Chen, and Daniel Zeng. 2018. Exploring trends and patterns of popularity stage evolution in social media. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* 50, 10 (2018), 3817–3827.
[203] Andrea Zaccaria, Michela Del Vicario, Walter Quattrociocchi, Antonio Scala, and Luciano Pietronero. 2019. PopRank: Ranking pages' impact and users' engagement on Facebook. *PLOS One* 14, 1 (2019), e0211038.
[204] Mariapina Trunfio and Simona Rossi. 2021. Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing* 2021, 3 (2021), 267–292.
[205] Alex Moehring, Avinash Collis, Kiran Garimella, M Amin Rahimian, Sinan Aral, and Dean Eckles. 2023. Providing normative information increases intentions to accept a COVID-19 vaccine. *Nature Communications* 14, 1 (2023), 126.
[206] Eytan Bakshy, Solomon Messing, and Lada A. Adamic. 2015. Exposure to ideologically diverse news and opinion on Facebook. *Science* 348, 6239 (2015), 1130–1132.
[207] Lilian Weng, Márton Karsai, Nicola Perra, Filippo Menczer, and Alessandro Flammini. 2018. Attention on weak ties in social and communication networks. *Complex Spreading Phenomena in Social Systems: Influence and Contagion in Real-World Social Networks* (2018), 213–228. DOI: https://doi.org/10.1007/978-3-319-77332-2_12
[208] Bruno Gonçalves, Nicola Perra, and Alessandro Vespignani. 2011. Modeling users' activity on Twitter networks: Validation of dunbar's number. *PLOS One* 6, 8 (2011), e22656.
[209] Max Falkenberg, Alessandro Galeazzi, Maddalena Torricelli, Niccolò Di Marco, Francesca Larosa, Madalina Sas, Amin Mekacher, Warren Pearce, Fabiana Zollo, Walter Quattrociocchi, et al. 2022. Growing polarization around climate change on social media. *Nature Climate Change* 12, 12 (2022), 1114–1121.
[210] Edoardo Loru, Alessandro Galeazzi, Anita Bonetti, Emanuele Sangiorgio, Niccolò Di Marco, Matteo Cinelli, Max Falkenberg, Andrea Baronchelli, and Walter Quattrociocchi. 2025. Ideology and polarization set the agenda on social media. *Scientific Reports* 15, 1 (Oct. 2025). DOI: http://doi.org/10.1038/s41598-025-19776-z
[211] Maxwell McCombs. 2005. A look at agenda-setting: Past, present and future. *Journalism Studies* 6, 4 (Nov. 2005), 543–557. DOI: http://doi.org/10.1080/14616700500250438
[212] Safran Safar Almakaty. 2025. Agenda setting theory in the digital media age: A comprehensive and critical literature review. *Future Technology* 4, 2 (May 2025), 51–60. DOI: http://doi.org/10.55670/fpll.futech.4.2.6
[213] Paul Covington, Jay Adams, and Emre Sargin. 2016. Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*. 191–198.
[214] Asia J. Biega, Krishna P. Gummadi, and Gerhard Weikum. 2018. Equity of attention: Amortizing individual fairness in rankings. In *Proceedings of the 41st international ACM Sigir Conference on Research & Development in Information Retrieval*. 405–414.
[215] Eytan Bakshy, Dean Eckles, and Michael S. Bernstein. 2014. Designing and deploying online field experiments. In *Proceedings of the 23rd International Conference on World wide web*. 283–292.
[216] Walter Quattrociocchi. 2017. Inside the echo chamber. *Scientific American* 316, 4 (2017), 60–63.
[217] Amin Mahmoudi, Dariusz Jemielniak, and Leon Ciechanowski. 2024. Echo chambers in online social networks: A systematic literature review. *IEEE Access* 12 (2024), 9594–9620. DOI: https://doi.org/10.1109/access.2024.3353054
[218] Miriam Amendola, Danilo Cavaliere, Carmen De Maio, Giuseppe Fenza, and Vincenzo Loia. 2024. Towards echo chamber assessment by employing aspect-based sentiment analysis and gdm consensus metrics. *Online Social Networks and Media* 39–40 (2024), 100276. DOI: https://doi.org/10.1016/j.osnem.2024.100276
[219] Hywel T. P. Williams, James R. McMurray, Tim Kurz, and F. Hugo Lambert. 2015. Network analysis reveals open forums and echo chambers in social media discussions of climate change. *Global Environmental Change* 32 (2015), 126–138. DOI: https://doi.org/10.1016/j.gloenvcha.2015.03.006
[220] Fabiana Zollo, Petra Kralj Novak, Michela Del Vicario, Alessandro Bessi, Igor Mozetič, Antonio Scala, Guido Caldarelli, and Walter Quattrociocchi. 2015. Emotional dynamics in the age of misinformation. *PLOS One* 10, 9 (Sept. 2015), e0138740. DOI: http://doi.org/10.1371/journal.pone.0138740

[221] Elizabeth Dubois and Grant Blank. 2018. The echo chamber is overstated: the moderating effect of political interest and diverse media. *Information, Communication & Society* 21, 5 (2018), 729–745.

[222] Max Falkenberg, Fabiana Zollo, Walter Quattrociocchi, Jürgen Pfeffer, and Andrea Baronchelli. 2024. Patterns of partisan toxicity and engagement reveal the common structure of online political communication across countries. *Nature Communications* 15, 1 (2024), 9560.

[223] Emanuele Brugnoli, Matteo Cinelli, Walter Quattrociocchi, and Antonio Scala. 2019. Recursive patterns in online echo chambers. *Scientific Reports* 9, 1 (2019), 20118.

[224] Kazutoshi Sasahara, Wen Chen, Hao Peng, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2021. Social influence and unfollowing accelerate the emergence of echo chambers. *Journal of Computational Social Science* 4, 1 (2021), 381–402.

[225] Steve Rathje, Jay J. Van Bavel, and Sander van der Linden. 2021. Out-group animosity drives engagement on social media. *Proceedings of the National Academy of Sciences* 118, 26 (June 2021). DOI: http://doi.org/10.1073/pnas.2024292118

[226] Daejin Choi, Selin Chun, Hyunchul Oh, Jinyoung Han, and Ted "Taekyoung" Kwon. 2020. Rumor propagation is amplified by echo chambers in social media. *Scientific Reports* 10, 1 (2020), 310.

[227] Killian L. McLoughlin, William J. Brady, Aden Goolsbee, Ben Kaiser, Kate Klonick, and M. J. Crockett. 2024. Misinformation exploits outrage to spread online. *Science* 386, 6725 (Nov. 2024), 991–996. DOI: http://doi.org/10.1126/science.adl2829

[228] Claire E. Robertson, Nicolas Pröllochs, Kaoru Schwarzenegger, Philip Pärnamets, Jay J. Van Bavel, and Stefan Feuerriegel. 2023. Negativity drives online news consumption. *Nature Human Behaviour* 7, 5 (2023), 812–822.

[229] Andrew M. Guess, Brendan Nyhan, and Jason Reifler. 2020. Exposure to untrustworthy websites in the 2016 US election. *Nature Human Behaviour* 4, 5 (2020), 472–480.

[230] Giovanni Luca Ciampaglia, Alexios Mantzarlis, Gregory Maus, and Filippo Menczer. 2018. Research challenges of digital misinformation: Toward a trustworthy web. *AI Magazine* 39, 1 (2018), 65–74.

[231] David M. J. Lazer, Matthew A. Baum, Yochai Benkler, Adam J. Berinsky, Kelly M. Greenhill, Filippo Menczer, Miriam J. Metzger, Brendan Nyhan, Gordon Pennycook, David Rothschild, et al. 2018. The science of fake news. *Science* 359, 6380 (2018), 1094–1096.

[232] Jonas L. Juul and Johan Ugander. 2021. Comparing information diffusion mechanisms by matching on cascade size. *Proceedings of the National Academy of Sciences* 118, 46 (2021), e2100786118.

[233] Cameron Martel, Mohsen Mosleh, and David G Rand. 2021. You're definitely wrong, maybe: Correction style has minimal effect on corrections of misinformation online. *Media and Communication* 9, 1 (2021), 120–133.

[234] Isaac Slaughter, Axel Peytavin, Johan Ugander, and Martin Saveski. 2025. Community notes reduce engagement with and diffusion of false information online. *Proceedings of the National Academy of Sciences* 122, 38 (Sept. 2025). DOI: http://doi.org/10.1073/pnas.2503413122

[235] Cameron Martel and David G. Rand. 2023. Misinformation warning labels are widely effective: A review of warning effects and their moderating features. *Current Opinion in Psychology* 54 (2023), 101710. DOI: https://doi.org/10.31234/osf.io/48p2a

[236] Chenyan Jia, Alexander Boltz, Angie Zhang, Anqing Chen, and Min Kyung Lee. 2022. Understanding effects of algorithmic vs. community label on perceived accuracy of hyper-partisan misinformation. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW2 (2022), 1–27.

[237] Rakoen Maertens, Jon Roozenbeek, Melisa Basol, and Sander van der Linden. 2021. Long-term effectiveness of inoculation against misinformation: Three longitudinal experiments. *Journal of Experimental Psychology: Applied* 27, 1 (2021), 1.

[238] Fabiana Zollo, Alessandro Bessi, Michela Del Vicario, Antonio Scala, Guido Caldarelli, Louis Shekhtman, Shlomo Havlin, and Walter Quattrociocchi. 2017. Debunking in a world of tribes. *PLOS One* 12, 7 (July 2017), e0181821. DOI: http://doi.org/10.1371/journal.pone.0181821

[239] Stephan Lewandowsky, Ullrich K. H. Ecker, Colleen M. Seifert, Norbert Schwarz, and John Cook. 2012. Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest* 13, 3 (2012), 106–131.

[240] Ryusuke Iizuka, Fujio Toriumi, Mao Nishiguchi, Masanori Takano, and Mitsuo Yoshida. 2022. Impact of correcting misinformation on social disruption. *PLOS One* 17, 4 (2022), e0265734.

[241] Man-pui Sally Chan and Dolores Albarracín. 2023. A meta-analysis of correction effects in science-relevant misinformation. *Nature Human Behaviour* 7, 9 (2023), 1514–1525.

[242] Ethan Porter and Thomas J. Wood. 2021. The global effectiveness of fact-checking: Evidence from simultaneous experiments in Argentina, Nigeria, South Africa, and the United Kingdom. *Proceedings of the National Academy of Sciences* 118, 37 (Sept. 2021). DOI: http://doi.org/10.1073/pnas.2104235118

[243] Sean J. Taylor, Lev Muchnik, Madhav Kumar, and Sinan Aral. 2023. Identity effects in social media. *Nature Human Behaviour* 7, 1 (2023), 27–37.

[244] Nathan Walter, Jonathan Cohen, R. Lance Holbert, and Yasmin Morag. 2019. Fact-Checking: A meta-analysis of what works and for whom. *Political Communication* 37, 3 (Oct. 2019), 350–375. DOI: http://doi.org/10.1080/10584609.2019.1668894

[245] Briony Swire-Thompson, Joseph DeGutis, and David Lazer. 2020. Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition* 9, 3 (Sept. 2020), 286–299. DOI: http://doi.org/10.1016/j.jarmac.2020.06.006

[246] Briony Swire-Thompson, Joseph DeGutis, and David Lazer. 2020. Searching for the backfire effect: Measurement and design considerations. *Journal of Applied Research in Memory and Cognition* 9, 3 (2020), 286–299.

[247] Brendan Nyhan. 2021. Why the backfire effect does not explain the durability of political misperceptions. *Proceedings of the National Academy of Sciences* 118, 15 (2021), e1912440117.

[248] Sander Van Der Linden. 2022. Misinformation: Susceptibility, spread, and interventions to immunize the public. *Nature Medicine* 28, 3 (2022), 460–467.

[249] Melisa Basol, Jon Roozenbeek, Manon Berriche, Fatih Uenal, William P. McClanahan, and Sander van der Linden. 2021. Towards psychological herd immunity: Cross-cultural evidence for two prebunking interventions against COVID-19 misinformation. *Big Data & Society* 8, 1 (2021), 20539517211013868.

[250] Sander van der Linden and Jon Roozenbeek. 2024. "Inoculation" to resist misinformation. *Jama* 331, 22 (2024), 1961–1962.

[251] Cecilie S. Traberg, Jon Roozenbeek, and Sander van der Linden. 2022. Psychological inoculation against misinformation: Current evidence and future directions. *The ANNALS of the American Academy of Political and Social Science* 700, 1 (2022), 136–151.

[252] Sander Van der Linden and Jon Roozenbeek. 2020. Psychological inoculation against fake news. *The Psychology Of Fake News: Accepting, Sharing, and Correcting Misinformation* (2020), 147–169.

[253] Diogo Pacheco, Pik-Mai Hui, Christopher Torres-Lugo, Bao Tran Truong, Alessandro Flammini, and Filippo Menczer. 2021. Uncovering coordinated networks on social media: methods and case studies. In *Proceedings of the international AAAI conference on web and social media*, Vol. 15. 455–466.

[254] Serena Tardelli, Leonardo Nizzoli, Maurizio Tesconi, Mauro Conti, Preslav Nakov, Giovanni Da San Martino, and Stefano Cresci. 2024. Temporal dynamics of coordinated online behavior: Stability, archetypes, and influence. *Proceedings of the National Academy of Sciences* 121, 20 (May 2024). DOI: http://doi.org/10.1073/pnas.2307038121

[255] Matteo Cinelli, Stefano Cresci, Walter Quattrociocchi, Maurizio Tesconi, and Paola Zola. 2022. Coordinated inauthentic behavior and information spreading on Twitter. *Decision Support Systems* 160 (Sept. 2022), 113819. DOI: http://doi.org/10.1016/j.dss.2022.113819

[256] Kristina Hristakieva, Stefano Cresci, Giovanni Da San Martino, Mauro Conti, and Preslav Nakov. 2022. The spread of propaganda by coordinated communities on social media. In *Proceedings of the 14th ACM web Science Conference 2022*. 191–201.

[257] Derek Weber and Frank Neumann. 2021. Amplifying influence through coordinated behaviour in social networks. *Social Network Analysis and Mining* 11, 1 (2021). DOI: https://doi.org/10.1007/s13278-021-00815-2

[258] Emilio Ferrara, Onur Varol, Clayton Davis, Filippo Menczer, and Alessandro Flammini. 2016. The rise of social bots. *Communications of the ACM* 59, 7 (2016), 96–104.

[259] Stefano Cresci. 2020. A decade of social bot detection. *Communications of the ACM* 63, 10 (Sept. 2020), 72–83. DOI: http://doi.org/10.1145/3409116

[260] Justin Cheng, Michael Bernstein, Cristian Danescu-Niculescu-Mizil, and Jure Leskovec. 2017. Anyone can become a troll: Causes of trolling behavior in online discussions. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing (CSCW '17)*. ACM, 1217–1230. DOI: http://doi.org/10.1145/2998181.2998213

[261] Leonardo Nizzoli, Serena Tardelli, Marco Avvenuti, Stefano Cresci, and Maurizio Tesconi. 2021. Coordinated behavior on social media in 2019 UK general election. *Proceedings of the International AAAI Conference on Web and Social Media* 15 (May 2021), 443–454. DOI: http://doi.org/10.1609/icwsm.v15i1.18074

[262] Adam Badawy, Emilio Ferrara, and Kristina Lerman. 2018. Analyzing the digital traces of political manipulation: The 2016 Russian interference Twitter campaign. In *Proceedings of the 2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*. IEEE, 258–265.

[263] Kai Shu, Suhang Wang, Dongwon Lee, and Huan Liu. 2020. *Mining Disinformation and Fake News: Concepts, Methods, and Recent Advancements*. Springer International Publishing, 1–19. DOI: http://doi.org/10.1007/978-3-030-42699-6_1

[264] Kate Starbird, Ahmer Arif, and Tom Wilson. 2019. Disinformation as collaborative work: Surfacing the participatory nature of strategic information operations. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (Nov. 2019), 1–26. DOI: http://doi.org/10.1145/3359229

[265] Massimo Stella, Emilio Ferrara, and Manlio De Domenico. 2018. Bots increase exposure to negative and inflammatory content in online social systems. *Proceedings of the National Academy of Sciences* 115, 49 (2018), 12435–12440.

[266] Luca Luceri, Ashok Deb, Adam Badawy, and Emilio Ferrara. 2019. Red bots do it better: Comparative analysis of social bot partisan behavior. In *Companion Proceedings of the 2019 World Wide Web Conference*. 1007–1012.

[267] Franziska B. Keller, David Schoch, Sebastian Stier, and JungHwan Yang. 2020. Political astroturfing on Twitter: How to coordinate a disinformation campaign. *Political Communication* 37, 2 (2020), 256–280.

[268] David Schoch, Franziska B. Keller, Sebastian Stier, and JungHwan Yang. 2022. Coordination patterns reveal online political astroturfing across the world. *Scientific Reports* 12, 1 (2022). DOI: https://doi.org/10.1038/s41598-022-08404-9

[269] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Goncalves, Alessandro Flammini, and Filippo Menczer. 2021. Detecting and tracking political abuse in social media. *Proceedings of the International AAAI Conference on Web and Social Media* 5, 1 (Aug. 2021), 297–304. DOI: http://doi.org/10.1609/icwsm.v5i1.14127

[270] Jacob Ratkiewicz, Michael Conover, Mark Meiss, Bruno Gonçalves, Snehal Patil, Alessandro Flammini, and Filippo Menczer. 2011. Truthy: Mapping the spread of astroturf in microblog streams. In *Proceedings of the 20th International Conference Companion on World Wide Web*. 249–252.

[271] M. Cinelli, M. Conti, L. Finos, F. Grisolia, P. Kralj Novak, A. Peruzzi, M. Tesconi, F. Zollo, and W. Quattrociocchi. 2019. (Mis)Information Operations: An integrated perspective. *Journal of Information Warfare* 18, 3 (2019), 83–98. Retrieved from https://www.jstor.org/stable/26894683

[272] W. Lance Bennett and Alexandra Segerberg. 2012. The logic of connective action: Digital media and the personalization of contentious politics. *Information, Communication & Society* 15, 5 (2012), 739–768.

[273] Katrin Weller, Axel Bruns, Jean Elizabeth Burgess, Merja Mahrt, and Cornelius Puschmann. 2013. *Twitter and Society*. Peter Lang New York.

[274] Damon Centola, Joshua Becker, Devon Brackbill, and Andrea Baronchelli. 2018. Experimental evidence for tipping points in social convention. *Science* 360, 6393 (June 2018), 1116–1119. DOI: http://doi.org/10.1126/science.aas8827

[275] Carlos Castillo, Mohammed El-Haddad, Jürgen Pfeffer, and Matt Stempeck. 2014. Characterizing the life cycle of online news stories using social media reactions. In *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*. 211–223.

[276] Ariel Flint Ashery, Luca Maria Aiello, and Andrea Baronchelli. 2025. Emergent social conventions and collective bias in LLM populations. *Science Advances* 11, 20 (May 2025). DOI: http://doi.org/10.1126/sciadv.adu9368

[277] Ariel Flint, Luca Maria Aiello, Romualdo Pastor-Satorras, and Andrea Baronchelli. 2025. Group size effects and collective misalignment in LLM multi-agent systems. (2025). DOI: http://doi.org/10.48550/ARXIV.2510.22422

[278] Daniel Thilo Schroeder, Meeyoung Cha, Andrea Baronchelli, Nick Bostrom, Nicholas A. Christakis, David Garcia, Amit Goldenberg, Yara Kyrychenko, Kevin Leyton-Brown, Nina Lutz, et al. 2026. How malicious AI swarms can threaten democracy. *Science* 391, 6783 (2026), 354–357. DOI: https://doi.org/10.1126/science.adz1697

[279] Gian Marco Orlando, Jinyi Ye, Valerio La Gatta, Mahdi Saeedi, Vincenzo Moscato, Emilio Ferrara, and Luca Luceri. 2025. Emergent coordinated behaviors in networked LLM agents: Modeling the strategic dynamics of information operations. (2025). DOI: http://doi.org/10.48550/ARXIV.2510.25003

[280] Eiman Alothali, Nazar Zaki, Elfadil A. Mohamed, and Hany Alashwal. 2018. Detecting social bots on Twitter: A literature review. In *Proceedings of the 2018 International Conference on Innovations in Information Technology (IIT)*. IEEE, 175–180.

[281] Valeria Pantè, David Axelrod, Alessandro Flammini, Filippo Menczer, Emilio Ferrara, and Luca Luceri. 2025. Beyond interaction patterns: Assessing claims of coordinated inter-state information operations on Twitter/X. In *Companion Proceedings of the ACM on Web Conference 2025 (WWW '25)*. ACM, 1234–1238. DOI: http://doi.org/10.1145/3701716.3715575

[282] Niccolò Di Marco, Sara Brunetti, Matteo Cinelli, and Walter Quattrociocchi. 2025. Post-hoc evaluation of nodes influence in information cascades: The case of coordinated accounts. *ACM Transactions on the Web* 19, 2 (May 2025), 1–19. DOI: http://doi.org/10.1145/3700644

[283] Marco Minici, Luca Luceri, Francesco Fabbri, and Emilio Ferrara. 2025. IOHunter: Graph foundation model to uncover online information operations. *Proceedings of the AAAI Conference on Artificial Intelligence* 39, 27 (April 2025), 28258–28266. DOI: http://doi.org/10.1609/aaai.v39i27.35046

[284] Alexander C. Nwala, Alessandro Flammini, and Filippo Menczer. 2023. A language framework for modeling social media account behavior. *EPJ Data Science* 12, 1 (Aug. 2023). DOI: http://doi.org/10.1140/epjds/s13688-023-00410-9

[285] Federico Cinus, Marco Minici, Luca Luceri, and Emilio Ferrara. 2025. Exposing cross-platform coordinated inauthentic activity in the run-up to the 2024 U.S. Election. In *Proceedings of the ACM on Web Conference 2025 (WWW '25)*. ACM, 541–559. DOI: http://doi.org/10.1145/3696410.3714698

[286] Matteo Cinelli, Andraž Pelicon, Igor Mozetič, Walter Quattrociocchi, Petra Kralj Novak, and Fabiana Zollo. 2021. Dynamics of online hate and misinformation. *Scientific Reports* 11, 1 (Nov. 2021). DOI: http://doi.org/10.1038/s41598-021-01487-w

[287] Edoardo Loru, Matteo Cinelli, Maurizio Tesconi, and Walter Quattrociocchi. 2024. The influence of coordinated behavior on toxicity. *Online Social Networks and Media* 43–44 (Nov. 2024), 100289. DOI: http://doi.org/10.1016/j.osnem.2024.100289

[288] Stefano Cresci, Marinella Petrocchi, Angelo Spognardi, and Stefano Tognazzi. 2019. On the capability of evolved spambots to evade detection via genetic engineering. *Online Social Networks and Media* 9 (Jan. 2019), 1–16. DOI: http://doi.org/10.1016/j.osnem.2018.10.005

[289] Claudio Castellano, Santo Fortunato, and Vittorio Loreto. 2009. Statistical physics of social dynamics. *Reviews of Modern Physics* 81, 2 (May 2009), 591–646. DOI: http://doi.org/10.1103/revmodphys.81.591

[290] Hegselmann Rainer and Ulrich Krause. 2002. Opinion dynamics and bounded confidence: Models, analysis and simulation. *Journal of Artificial Societies and Social Simulation (JASSS'02)* 5, 3 (2002). http://jasss.soc.surrey.ac.uk/5/3/2.html

[291] Michele Starnini, Fabian Baumann, Tobias Galla, David Garcia, Gerardo Iñiguez, Márton Karsai, Jan Lorenz, and Katarzyna Sznajd-Weron. 2025. Opinion dynamics: Statistical physics and beyond. (2025). DOI: http://doi.org/10.48550/ARXIV.2507.11521

[292] Andreas Flache, Michael Mäs, Thomas Feliciani, Edmund Chattoe-Brown, Guillaume Deffuant, Sylvie Huet, and Jan Lorenz. 2017. Models of social influence: Towards the next frontiers. *Jasss-The journal of artificial societies and social simulation* 20, 4 (2017), 2.

[293] Abhimanyu Das, Sreenivas Gollapudi, and Kamesh Munagala. 2014. Modeling opinion dynamics in social networks. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*. 403–412.

[294] Idris Adjerid and Ken Kelley. 2018. Big data in psychology: A framework for research advancement. *American Psychologist* 73, 7 (2018), 899.

[295] Corrado Monti, Gianmarco De Francisci Morales, and Francesco Bonchi. 2020. Learning opinion dynamics from social traces. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. 764–773.

[296] Carlo M. Valensise, Matteo Cinelli, and Walter Quattrociocchi. 2023. The drivers of online polarization: Fitting models to data. *Information Sciences* 642 (Sept. 2023), 119152. DOI: http://doi.org/10.1016/j.ins.2023.119152

[297] Antonio F. Peralta, János Kertész, and Gerardo Iñiguez. 2022. Opinion dynamics in social networks: From models to data. arXiv:2201.01322. Retrieved from https://arxiv.org/abs/2201.01322 (2022).

[298] W. Weidlich. 1971. The statistical description of polarization phenomena in society†. *British Journal of Mathematical and Statistical Psychology* 24, 2 (Nov. 1971), 251–266. DOI: http://doi.org/10.1111/j.2044-8317.1971.tb00470.x

[299] Serge Galam, Yuval Gefen (Feigenblat), and Yonathan Shapir. 1982. Sociophysics: A new approach of sociological collective behaviour. I. mean-behaviour description of a strike. *The Journal of Mathematical Sociology* 9, 1 (Nov. 1982), 1–13. DOI: http://doi.org/10.1080/0022250x.1982.9989929

[300] Serge Galam and Serge Moscovici. 1991. Towards a theory of collective phenomena: Consensus and attitude changes in groups. *European Journal of Social Psychology* 21, 1 (Jan. 1991), 49–74. DOI: http://doi.org/10.1002/ejsp.2420210105

[301] Hossein Noorazar, Kevin R. Vixie, Arghavan Talebanpour, and Yunfeng Hu. 2020. From classical to modern opinion dynamics. *International Journal of Modern Physics C* 31, 07 (2020), 2050101.

[302] Dino Carpentras. 2023. Why we are failing at connecting opinion dynamics to the empirical world. *Review of Artificial Societies and Social Simulations* (2023).

[303] Janos Kertesz, Alina Sirbu, Fosca Gianotti, and Dino Pedreschi. 2019. Algorithmic bias amplifies opinion polarization: A bounded confidence model. In *Proceedings of the StatPhys 27 Main Conference*.

[304] Nicola Perra and Luis E. C. Rocha. 2019. Modelling opinion dynamics in the age of algorithmic personalisation. *Scientific reports* 9, 1 (2019), 7261.

[305] Wenjun Mei, Francesco Bullo, Ge Chen, Julien M. Hendrickx, and Florian Dörfler. 2022. Micro-foundation of opinion dynamics: Rich consequences of the weighted-median mechanism. *Physical Review Research* 4, 2 (2022), 023213.

[306] Wenjun Mei, Julien M Hendrickx, Ge Chen, Francesco Bullo, and Florian Dörfler. 2024. Convergence, consensus and dissensus in the weighted-median opinion dynamics. *IEEE Transactions on Automatic Control* 69, 10 (2024), 6700–6714. DOI: https://doi.org/10.1109/tac.2024.3376752

[307] Anastasia Bizyaeva, Alessio Franci, and Naomi Ehrich Leonard. 2022. Nonlinear opinion dynamics with tunable sensitivity. *IEEE Transactions on Automatic Control* 68, 3 (2022), 1415–1430.

[308] PETER CLIFFORD and AIDAN SUDBURY. 1973. A model for spatial conflict. *Biometrika* 60, 3 (1973), 581–588. DOI: http://doi.org/10.1093/biomet/60.3.581

[309] Richard A. Holley and Thomas M. Liggett. 1975. Ergodic theorems for weakly interacting infinite systems and the voter model. *The Annals of Probability* 3, 4 (Aug. 1975). DOI: http://doi.org/10.1214/aop/1176996306

[310] Sidney Redner. 2001. *A Guide to First-Passage Processes*. Cambridge University Press. DOI: http://doi.org/10.1017/cbo9780511606014

[311] Thomas M. Liggett. 1985. *Interacting Particle Systems*. Springer New York. DOI: http://doi.org/10.1007/978-1-4613-8542-4

[312] Luca Dall'Asta and Claudio Castellano. 2007. Effective surface-tension in the noise-reduced voter model. *Europhysics Letters (EPL)* 77, 6 (March 2007), 60005. DOI: http://doi.org/10.1209/0295-5075/77/60005

[313] Boris L. Granovsky and Neal Madras. 1995. The noisy voter model. *Stochastic Processes and their Applications* 55, 1 (Jan. 1995), 23–43. DOI: http://doi.org/10.1016/0304-4149(94)00035-r

[314] Tommaso Zillio, Igor Volkov, Jayanth R. Banavar, Stephen P. Hubbell, and Amos Maritan. 2005. Spatial scaling in model plant communities. *Physical Review Letters* 95, 9 (Aug. 2005). DOI: http://doi.org/10.1103/physrevlett.95.098101

[315] T. Antal, S. Redner, and V. Sood. 2006. Evolutionary dynamics on degree-heterogeneous graphs. *Physical Review Letters* 96, 18 (May 2006). DOI: http://doi.org/10.1103/physrevlett.96.188104

[316] Beáta Oborny. 2023. Lost in translation? – Caveat to the application of the voter model in ecology and evolutionary biology. *Science Progress* 106, 2 (April 2023). DOI: http://doi.org/10.1177/00368504231175324

[317] Antonio F. Peralta, János Kertész, and Gerardo Iñiguez. 2021. Opinion formation on social networks with algorithmic bias: dynamics and bias imbalance. *Journal of Physics: Complexity* 2, 4 (2021), 045009.

[318] Paul L. Krapivsky and Sidney Redner. 2003. Dynamics of majority rule in two-state interacting spin systems. *Physical Review Letters* 90, 23 (2003), 238701.

[319] P. L. Krapivsky and S. Redner. 2021. Divergence and consensus in majority rule. *Physical Review E* 103, 6 (2021), L060301.

[320] James Noonan and Renaud Lambiotte. 2021. Dynamics of majority rule on hypergraphs. *Physical Review E* 104, 2 (2021), 024316.

[321] Fabricio L. Forgerini, Nuno Crokidakis, and Márcio AV Carvalho. 2024. Directed propaganda in the majority-rule model. *International Journal of Modern Physics C* 35, 07 (2024), 2450082.

[322] Ashwin Sah and Mehtaab Sawhney. 2024. Majority dynamics: The power of one. *Israel Journal of Mathematics* 267, 1 (2024), 85–133. DOI: https://doi.org/10.1007/s11856-024-2690-1

[323] Marco A. Montes de Oca, Eliseo Ferrante, Alexander Scheidler, Carlo Pinciroli, Mauro Birattari, and Marco Dorigo. 2011. Majority-rule opinion dynamics with differential latency: A mechanism for self-organized collective decision-making. *Swarm Intelligence* 5, 3–4 (2011), 305–327. DOI: https://doi.org/10.1007/s11721-011-0062-z

[324] S. Chatterjee and E. Seneta. 1977. Towards consensus: Some convergence theorems on repeated averaging. *Journal of Applied Probability* 14, 1 (March 1977), 89–97. DOI: http://doi.org/10.2307/3213262

[325] M. Stone. 1961. The opinion pool. *The Annals of Mathematical Statistics* 32, 4 (Dec. 1961), 1339–1342. DOI: http://doi.org/10.1214/aoms/1177704873

[326] Guillaume Deffuant, David Neau, Frederic Amblard, and Gérard Weisbuch. 2000. Mixing beliefs among interacting agents. *Advances in Complex Systems* 03, 01n04 (Jan. 2000), 87–98. DOI: http://doi.org/10.1142/s0219525900000078

[327] SANTO FORTUNATO. 2004. Universality of the Threshold for complete consensus for the opinion dynamics of deffuant et al. *International Journal of Modern Physics C* 15, 09 (Nov. 2004), 1301–1307. DOI: http://doi.org/10.1142/s0129183104006728

[328] JAVIER GÓMEZ-SERRANO, CARL GRAHAM, and JEAN-YVES LE BOUDEC. 2012. The bounded confidence model of opinion dynamics. *Mathematical Models and Methods in Applied Sciences* 22, 02 (Feb. 2012). DOI: http://doi.org/10.1142/s0218202511500072

[329] Benedetto Piccoli and Francesco Rossi. 2021. Generalized solutions to bounded-confidence models. *Mathematical Models and Methods in Applied Sciences* 31, 06 (April 2021), 1237–1276. DOI: http://doi.org/10.1142/s0218202521400054

[330] Carmela Bernardo, Claudio Altafini, Anton Proskurnikov, and Francesco Vasca. 2024. Bounded confidence opinion dynamics: A survey. *Automatica* 159 (Jan. 2024), 111302. DOI: http://doi.org/10.1016/j.automatica.2023.111302

[331] Grace J. Li, Jiajie Luo, and Mason A. Porter. 2025. Bounded-confidence models of opinion dynamics with adaptive confidence bounds. *SIAM Journal on Applied Dynamical Systems* 24, 2 (2025), 994–1041.

[332] Grace Jingying Li. 2024. *Some Generalizations of Bounded-Confidence Models of Opinion Dynamics.* University of California, Los Angeles.

[333] Kush R Varshney. 2014. Bounded confidence opinion dynamics in a social network of Bayesian decision makers. *IEEE Journal of Selected Topics in Signal Processing* 8, 4 (2014), 576–585.

[334] Jesús Giráldez-Cru, Carmen Zarco, and Oscar Cordón. 2022. Analyzing the extremization of opinions in a general framework of bounded confidence and repulsion. *Information Sciences* 609 (2022), 1256–1270. DOI: https://doi.org/10.1016/j.ins.2022.07.164

[335] Marco Alberto Javarone. 2014. Social influences in opinion dynamics: The role of conformity. *Physica A: Statistical Mechanics and its Applications* 414 (2014), 19–30. DOI: https://doi.org/10.1016/j.physa.2014.07.018

[336] Henrique Ferraz de Arruda, Kleber Andrade Oliveira, and Yamir Moreno. 2024. Echo chamber formation sharpened by priority users. *iScience* 27, 11 (Nov. 2024), 111098. DOI: http://doi.org/10.1016/j.isci.2024.111098

[337] Fabian Baumann, Philipp Lorenz-Spreen, Igor M. Sokolov, and Michele Starnini. 2020. Modeling echo chambers and polarization dynamics in social networks. *Physical Review Letters* 124, 4 (Jan. 2020). DOI: http://doi.org/10.1103/physrevlett.124.048301

[338] Henrique Ferraz de Arruda, Felipe Maciel Cardoso, Guilherme Ferraz de Arruda, Alexis R. Hernández, Luciano da Fontoura Costa, and Yamir Moreno. 2022. Modelling how social network algorithms can influence opinion polarization. *Information Sciences* 588 (April 2022), 265–278. DOI: http://doi.org/10.1016/j.ins.2021.12.069

[339] Ashwini Kumar Singh, Vahid Ghafouri, Jose Such, and Guillermo Suarez-Tangil. 2024. Differences in the toxic language of cross-platform communities. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 1463–1476.

[340] Anees Baqir, Alessandro Galeazzi, and Fabiana Zollo. 2024. News and misinformation consumption: A temporal comparison across European countries. *PLOS One* 19, 5 (2024), e0302473.

[341] Erik Saberski, Tom Lorimer, Delia Carpenter, Ethan Deyle, Ewa Merz, Joseph Park, Gerald M. Pao, and George Sugihara. 2024. The impact of data resolution on dynamic causal inference in multiscale ecological networks. *Communications Biology* 7, 1 (2024), 1442.

[342] Isabel E. Murdock. 2024. *Information Diffusion Over Diverse Social Media Platforms and the Simulated Cross-Platform Impacts of Interventions.* Ph.D. Dissertation. Carnegie Mellon University.

[343] Edoardo Loru, Jacopo Nudo, Niccolò Di Marco, Alessandro Santirocchi, Roberto Atzeni, Matteo Cinelli, Vincenzo Cestari, Clelia Rossi-Arnaud, and Walter Quattrociocchi. 2025. The simulation of judgment in LLMs. *Proceedings of the National Academy of Sciences* 122, 42 (2025), e2518443122.

Received 15 July 2025; revised 12 January 2026; accepted 14 January 2026