Text Accumulation: Conceptualizing and Measuring Collective Attributes on Social Media

Discussion on social media appears different characteristics from the offline public environment, which initiates various research. We argue, that classical political communication topics involving polarization, uncivil behaviors, and online radicalization on social media all demonstrate a collective attribute. This attribute can be conceptualized by Text Accumulation and be measured by the scale and homogeneity of text. This paper unravels the comparability between online opinion dynamics and offline crowd behavior and reimposes that self-awareness is important to understanding the collective attribute on social media from a perspective of social role. Social role, maintaining consistency based on social cues and feedback, differentiates from social identity or group norms. We proposed a self-integration mechanism and an algorithmic impact mechanism to build an interpretation model of Text Accumulation and use Agent-Based Modeling to verify their impact on the fluctuations of texts. This general model of online behavior can help us to understand the difference between opinion leaders and normal participation online and interpret why we observe different impacts of social media towards polarization and uncivil behavior in different discussion areas.

Keywords: text accumulation, self-integration, social media, affective polarization, online uncivil behavior

Since social media entered the mainstream of life, new problems have arisen. Topics on different social media often cause large-scale discussions, social issues, serious scientific reports, and even screenshots from common individuals. Comments below can be irrational and aggressive, often shaping a sense of oppression and segregation. Sometimes, posting memes and meaningless words can be the best way to express attitudes. Several academic branches have already studied these phenomena: online polarization, digital activism, online uncivil behavior, and so on. They often focus on how social media influences human behavior and what's the result.

In the early stages of studies of computer-mediated communication(CMC), scholars focused on the differences between CMC and face-to-face communication(FTF) and proposed that online communication has fewer social cues than offline interaction. They related their social psychology interpretation with polarization and generated a common model of online individuals. For example, using the social identity approach(Tanis & Postmes, 2008), or online disinhibition(Suler, 2004) to grasp the inner dimension of human actors and explain why they became affective online.

However, answers about whether we are more uncivil when we are using social media vary(K. Chang & Park, 2021; Mooijman, Hoover, Lin, et al., 2018), depending on specific topics and areas. In a more general social theory framework, scholars proposed the fundamental nature of the Internet such as connectivity(Haythornthwaite, 2005) , hyperconnectivity(Brubaker, 2020) , and simulation theory for digital platforms(Turkle, 2009) . At the core of these theories is the idea that digital platforms offer highly customized spaces that reflect users' offline preferences while simultaneously eroding clear contextual boundaries(Gil-López, Shen, Benefield, et al., 2018). This shift underpins online interactions and behaviors, where the design of algorithms emulates offline interactions within fundamentally different spatial and temporal frameworks. Therefore, text that is left on social media by users can be both communication and action, and the consequences of text distribution can be taken as polarization and social movements at the same time.

Based on this attribute, we argue that Text Accumulation is the general outcome of the changing nature of action on social media, where the action leaves its digital traces and often does not have steady, contextualized feedback from others. The impact of mediated social cues can be demonstrated by the self-integration process, which relates to social roles in classic sociology theory. This paper will give a mechanism explanation and its formal representation verified by agent-based modeling.

Why Text Accumulation?

Shirky (2009) explores the potential of the internet to enable uncoordinated public action in Here Comes Everybody. Online, many large-scale public opinion movements or politically significant actions do not follow the structured organization typical of traditional collective action but can still have considerable public impact. In the book's title, EVERYBODY is taken as a whole, symbolizing the generalized clustering effect that emerges from such online actions. These events spread information rapidly, collapsing the usual social processes required for collective action across different cyberspaces. As large numbers of people participate simultaneously through mobile devices, individual contributions blend, creating a form of distributed responsibility. Although Shirky argues that the internet provides the power of organizing without organizations, the constant emergence of extreme rhetoric and anti-social behavior online has led scholars to link it back to Le Bon (2017) and the characteristics showed in The Crowd. and his studies on crowd psychology, which emphasized emotional intensity. Before the internet, collective gatherings relied on physical presence to transmit emotions and foster an affective entity, a dynamic also explored in Collins's Interaction Ritual Chains(Collins, 2014) and Durkheim's studies of religious rituals(Durkheim, 2016).

On social media, the action follows different distributed laws and emerges as a collectivity. Previous studies dealing with these online collective properties can be identified as three focuses: the clustering of text in online discussions, the group nature of participants signaled by diminished individual rationality, and the collective force of actions influenced by heightened social impact. This paper tries to use Text

Accumulation to conceptualize these properties and comprehend former discussions.

Take Text Accumulation as Information and Its Discontent

Network science partially analyses the unique spatio-temporal conditions of online actions from the perspective of information distribution patterns. In the digital realm, actions translate into information, which follows a power-law distribution; in other words, the more attention a piece of information gains, the more likely it is to continue attracting attention, reaching a wider audience(Barabási, 2013). On social media, large-scale discussion happens all the time, and the clustering of opinions forms an important object of study in communication.

Different from offline discussions on public issues, discourse online seems to show more populist and polarized attitudes(Gidron, Adams, & Horne, 2019; Hameleers, Schmuck, Bos, et al., 2021), indicating that social media might make individuals more irrational. The most important research branch of it is studies of polarization, broadly defined as the process by which public opinion splits into opposing extremes, often leading to entrenched group divides. Polarization is studied across multiple dimensions, including emotional polarization(Bakker & Lelkes, 2024; Iyengar, Lelkes, Levendusky, Malhotra, & Westwood, 2019), cultural polarization(Flache & Macy, 2014; Mouw & Sobel, 2001), and political polarization(Callander & Carbajal, 2022; Van Bavel & Pereira, 2018), with some research focusing on the interactions between these dimensions(Lelkes, 2018; Röllicke, 2023). These studies show, that despite increased opportunities for online dialogue, the digital environment often fosters division rather than consensus and compromise.

Various classical theories have emerged to explain the evolution of opinions on social media. Pariser (2011) filter bubble theory posits that algorithms restrict individuals to a narrow range of information, reinforcing preexisting beliefs and closing off alternative viewpoints over time. Similarly, theories like the “echo chamber effect”(Cinelli, De Francisci Morales, Galeazzi, et al., 2021) and the “spiral of silence”(Noelle-Neumann, 1974) suggest that people are less likely to express opposite views, resulting in clusters of like-minded content. These theories aim to explain why the internet often fails to provide a conducive environment for democratic discourse and how polarized views evolve by analyzing algorithmic features. These studies show that online opinion clustering is not a direct reflection of people's cognition, but an outcome of interaction within a mediated environment since we can feel the overall inclination that influences us before we send a comment. Studies on "false polarization"(Blatz & Mercier, 2018; Fernbach & Van Boven, 2022) support this, showing that there remains a distance between individuals' online expressions and their stable political beliefs. Some studies now aim to explore the relationships between them(Monin & Norton, 2003; Vissers & Stolle, 2014).

Therefore, the evolution of online views reflects the nature of interaction, in which people's views may be reinforced by conflicts, call-to-action group symbols, and other immediate stimuli. Although recent studies can be controversies with the classic theories that are mentioned above, for example, suggest that exposure to divergent views can fail to foster openness and may instead reinforce defensive attitudes(Bail, Argyle, Brown, et al., 2018), the interaction attribute remains interpretative.

The interaction attributes of online discourse have been reflected in many studies, such as examining the tribal characteristics of online communities of interest(W. J. Chang & Park, 2019), analyzing that expressing opinions online can trigger reactions that reinforce group identity and encourage divisive discourse(van der Maas, Dalege, & Waldorp, 2020; Wu, Hauert, Kremen, & Zhao, 2022). Contemporary studies also explore how low-cost communication and anonymity contribute to polarized and extremist communities(Koehler, 2014) . While polarization focuses on the evolution of people's views, these research topics deal with the characteristics and styles within online discourse, including trolling(Cheng, Bernstein, Danescu-Niculescu-Mizil, et al., 2017; Phillips, 2015), memes(Paz, Mayagoitia-Soria, & González-Aguilar, 2021; Shifman, 2012) and other aggressive communication styles(Jane, 2015; Lea, O'Shea, Fung, et al., 1992). In many cases, the behavior on social media displays anti-social tendencies, which forms fragmented research apart from the topic of polarization, but in parallel.

In conclusion, research on the evolution of online text accumulation largely centers on cultural or political opinions, analyzing factors such as algorithms, social identity, and communicative interactions within controversial contexts. On social media, opinions often have an action or interaction nature because they will influence others.

Take Text Accumulation as Action and Its Discontent

In the last section, we suggest that online discourse is not a direct reflection of people's ideas but has an interactive nature. Although there remain controversies about whether can we take online posts, comments, and likings as action or interaction, scholars who focus on the impact of these behaviors already produced abundant literature to analyze online text from the perspective of social action. This section will go through these studies, which can be divided into two categories, digital activism and online behavior study.

Digital activism is a kind of social activism mediated through digital technologies, and its heart lies in social movements(George & Leidner, 2019). In many studies, digital media acts as an amplifier of political ideas, giving people more ways to act and be visible. Digital activism specializes in the study of Internet-based political action or actors engaging in political action via the Internet, which includes both the use of mobile devices by people to engage in political campaigns and the combination of online and offline campaigns(Joyce, 2010; Kaun & Uldam, 2018), and again includes purely textual actions like information distribution, such as Yuce, Agarwal, Wigand, et al. (2014) focused on studying the process of formation of online collective action by analyzing the diffusion of hashtags. Hashtags play the role of gathering people's attention and triggering public discussion, and some scholars have studied hashtag activism, whose classic studies include hashtags related to Black Lives Matter and the women's movement, etc(George & Leidner, 2019).

In the latest trend of digital activism, online discourse can be seen as a participation in social movements. In these movements, online collective actions are usually low-threshold compared to offline participation, which some researchers also consider as digital sub-activism, where many of the political actions are everyday actions(Bakardjieva, 2012). A key characteristic of these online political actions is demonstrated by digital sub-activism, which implies daily posts can be politically important. In a traditional collective action, participants are always aligned with the purpose of the organization. However, participants on social media engage with varying levels of commitment(Bennett & Segerberg, 2013; Selander & Jarvenpaa, 2016; Vaast, Safadi, Lapointe, & Negoita, 2017). Bennett and Segerberg (2013) handle this difference in The Logic of Connective Action, proposing that Connective Action, which argues that in large-scale action networks, digital media have altered the core dynamics of action from the logic of collective action driven by organizational resources and identities. In these networks, individual expression often replaces clear political purpose as the fundamental motivation for textual action.

Under the fact that people are less purposeful in online action, some studies propose slacktivism to argue that online participation is not facilitative but rather a form of laziness(Christensen, 2011), potentially demobilizing groups aiming to achieve collective purposes(Hassid, 2012; MacKinnon, 2008; Schumann & Klein, 2015). This characteristic is consistent with some of the qualities presented in the previous studies of online polarization, in which social media seems to present the phenomenon of the so-called "affective public"(Papacharissi, 2016). This concept refers to the fact that clusters generated on the web are often dominated by emotional factors and have less stable consequent trends. Studies on digital activism take connective emotion as a primary variable of action(George & Leidner, 2019) and show that both the dynamics of opinions and political participation on social media are easily influenced by emotion.

This inclination is also shown in another research path, online behavior study. Different from digital activism, which puts online actions in a broad context, online behavior studies underline the characteristics of online behaviors themselves. Social media seems to present a large amount of uncivil behavior and extreme speech and scholars want to analyze whether social media harms human sentiment and why(K. Chang & Park, 2021; Mooijman et al., 2018; Müller & Schwarz, 2021). Their research objects overlap with text accumulation because many studies focus on the overall presentation of individuals, for example, online flaming(Jane, 2015), digital vigilantism(Trottier, 2017), and online fandom(W. J. Chang & Park, 2019). These uncivil behaviors showed up as an irrational crowd within changes in individuals' behavior patterns. Despite social media also providing a positive influence on specific aspects(Faizi, El Afia, & Chiheb, 2013; Korda & Itani, 2013), there is much research showing that the usage of social media can increase uncivil human behavior.

Overall, studies on online action or behavior focus not only on the dynamics of text but also the inherent meaning and social impact it involves. Scholars take text accumulation as a new form or supplement of action in a social context, which can facilitate goals. The concepts of action and behavior are not the same, but in the online environment, their differences are kind of diminished, and online behavior studies can encompass a broad range of phenomena, for example, some scholars take online political engagement as online collective behavior(Qiu, Lin, Chiu, et al., 2015). This attribute also shows a need for a more specific demonstration of online text.

Mediated Actors Within These Studies

The research above mentions two important characteristics of the landscape in nowadays social media. The first is the public deviant emotion, the second is the different commitment levels towards the post. Usually, those studies who take text accumulation as dynamics of information or opinions may find the distribution pattern of the affective public is obvious, and those studies who take text accumulation as action will focus on the process of how actors participate in an event and post a message. However, they both need a fundamental assumption on individuals.

This question has long been discussed in CMC, scholars want to analyze why CMC shows more hostility than FTF, and soon some of them abandon the assumption that we behave more uncivil online. Researchers went through a lot of psychological explanatory frameworks.

Deindividuation theory is a foundational approach to understanding unregulated behavior, originating with Le Bon's early empirical observations. Le Bon's ideas were subsequently adopted by experimental social psychology(Festinger, Pepitone, & Newcomb, 1952) and further developed over the following decades(Diener, 1980; Prentice-Dunn & Rogers, 1989; Zimbardo, 1969) . This theory has been reinterpreted in the context of online actions, where anonymity is seen as a central factor explaining the irrational aspects of behavior. Some scholars argue that the anonymity afforded by the Internet makes it more likely that people will ignore the consequences of their actions. (Suler, 2004) systematized these arguments by identifying six interacting factors affecting individuals' online disinhibition: dissociative anonymity, invisibility, asynchronicity, solipsistic introjection, dissociative imagination, and minimization of authority. In brief, the lack of social cues in online environments can alter the personalities of those engaging in online interaction, creating differences from face-to-face socialization.

The difference between face-to-face and online interaction was a central topic of early CMC research, with an important focus on the group dynamics resulting from reduced social clues(Walther, Van Der Heide, Ramirez Jr, et al., 2015; Yao & Ling, 2020), Although the theories are more complex in their composition, they all start from the point of view of explaining the irrational characteristics of actors in CMC. However, these theories, which emphasize the suppression of individual rationality and the production of anti-sociality, have not convinced all researchers. They propose alternative explanations such as depersonalization theory(Voci, 2006) and social identity approach(Turner, 1991; Turner, Hogg, Oakes, Reicher, & Wetherell, 1987). It would be argued that the depersonalization of individual embodiment is due to the integration of the individual by the collective, in which the self-identity is replaced by a group identity, and therefore the collective does not necessarily produce more extreme expressions, but may also produce more prudent outcomes. The anonymity of individuals instead makes people more social, and there is evidence that depersonalized online collaboration is associated with stronger identification, social identity salience, and group cohesion(Lea, Spears, & De Groot, 2001; Lea, Spears, & Watt, 2007).

Several studies have found that depersonalization is associated with a greater tendency to view the collaborative group or dyad as an entity or “one”(Postmes & Brunsting, 2002; Sassenberg & Postmes, 2002; Tanis & Postmes, 2008)

Theories such as SIDE suggest that CMC reinforces group identity through the inhibition of certain individual attributes(Spears & Postmes, 2015), promoting both polarization and certain pro-social collective actions and deliberative collective dynamics. This theory has criticized group psychology since Le Bon, argued that the social psychology of social identity can be an alternative to this theory. This line of thinking can explain a wide range of social phenomena, such as the clustering of emotions of online behavior using Intergroup Emotions Theory(Marchal, 2022), which inherits the perspective of social identity theory that emphasizes the social nature of the self(Smith & Mackie, 2015). Group emotions differ from individual emotions in that they depend on the degree of group identity(Smith, Seger, & Mackie, 2007).

However, Group identity is an extrinsic concept, and how an individual’s social identity relates to the feelings of the self has complex internal processes. Moreover, it can not explain the internal process of different commitment levels. This level is different from the salient of identity but relates to the stability of emotions and goals. Despite recent studies applying social network position and interaction between individuals(Yarchi, Baden, & Kligler-Vilenchik, 2021) to grasp more details about online actors in the polarization formation, they didn’t completely go through an inner dimension and give this varying commitment level an explanation. Facing one outer group’s opinion can cause different reactions, and these differences are important to understanding the true impact of social media on collective identity.

To sum up, the dynamics of online text are not only the information nor action, the formation of the dynamics is the evolution of textual actions which leave traces. In other words, in the context of social media, polarization and uncivil behavior share the same process and they all embody the accumulation of text, in which previous text can gain visibility and thus invoke group identity. But this group identity can not explain all the actors because actors online have different self-awareness, some are more resolute while some are just having fun. We will propose a concrete explanation framework in this paper.

Concepts and Mechanisms

As I mentioned in the last section, in many research, the impact of online actions or opinions is shown as a whole. The text on social media seems to have an inner inclination toward collective. There remains other research trying to analyze the collective attributes of online text. Some continental theoretical discussions emphasize the irreducibility of online groups. Le Bon's influence also goes beyond social psychology, as his term crowd has been borrowed to describe the ontological qualities of online collective behavior(Stage, 2013) , crowd implies an emotionally synchronized, and therefore non-individualized, gathering, presenting the state of a "spiritual unity"(Le Bon, 2017) , is irreducible. online actions are emitted by assemblages of human and nonhuman actors, with some scholars describing the emitters of online actions as swarm or network(Wiedemann, 2014) , referring to the collectivity in understanding online action through an approach that is not based on shared identity. The final condition of individual textual actions is influenced by several factors with complex evolution processes and finally emerges as an irreducible outcome.

Like the segregation process(Schelling, 1969) which relates to the group dynamic, the explanation of text accumulation requires an analytical model to elaborate precisely. In this part, we will propose two key concepts for the model and introduce the mechanisms involved.

Text Accumulation and Self-Integration

Text Accumulation. This concept is used to feature the collective attributes and outcomes of textual action on social media.

Textual actions refer to actions that can leave visible traces of others on social media, which include comments, likes, retweets, and chats. Text accumulation is the process and pattern of evolution of public or semi-public textual actions on the Internet. The concept of text accumulation consists of two dimensions, one is the number of texts and the other is the degree of text similarity, which is portrayed as high or low, and the typical area of the Internet with the highest degree of text accumulation is the popular comment section filled with copy-and-paste copywriting. It is also able to deal with actual disagreements and contradictions within different viewpoints, avoiding, as in political action, the view of individuals as actors with fixed ideas and biases, and ignoring the dynamics of users reinforcing or weakening one another's viewpoints in their interactions.

Text accumulation is a more general concept that tries to portray the core of online textual action. It takes online opinions and actions as digital traces and shows their technical characteristics in the digital space.

Self-Integration. This concept portrays the state of self-awareness of the actor's actions. We use this concept to grasp the consistency of users and measure the different commitment levels of actions.

This concept was developed to deal with the tension between the traditional social theory's use of the "self", which receives constant social feedback from the outside world(Mead, 2023) , and the textual actions, which face mediated feedback. In the early studies of CMC, they also suggest there is a loss of self-awareness in online behavior. This is one-sided, but enlightening and can not be replaced by a social identity approach. The distortion of the online self has long been discussed. Sherry TurkleTurkle (1995) studies fragmentation of self-identity due to the different interfaces of the Internet and how interactions are carried out on the Internet. In many social theories, the self has a social nature and the identity is shaped by the self(Mead, 2023). The change of social cues may not necessarily cause a low level of self-monitoring since we can monitor ourselves by quantified information(Lupton, 2016), but the relation between action and self is mediated.

Compared with the political action perspective, online actions are less purposeful. However, this is not the core reason for slacktivism. Giddens deals with actions in everyday life that do not differentiate between purposeful social actions and other, more natural, actions in which reflexive surveillance allows the person to stabilize the self in retrospect of the behavior(Giddens, 1984). In other words, the time-spatial conditions provide actors with the stabilizing process of self-identity. Marya SchechtmanSchechtman (1996) suggests four key dimensions of self in common sense: survival, moral responsibility, self-interested concern, and compensation. These dimensions delineate why one can be seen as one person different from others. This is to put action into a continuous flow of life and make one's action be as themselves. This consistency of self and action is composed in the process of self-narrative. Through these narratives, self-identity can be built(Ricoeur, 1992).

The narrative self requires self-monitoring and retrospection. Self-integration is this process of retrospection carried out by the person, i.e., self-retrospection. On social media, others indexing behavior such as comments can make actors recall their previous text. Therefore, the self retrospection and others' indexing behavior are two important factors of self-integration. The degree of integration is reflected in the consistency and stability of his or her person, which in the Internet is manifested in the consistency of speech. People with a low degree of self-integration have lower consistency of speech on the Internet. The degree of self-integration is influenced by the actor's attention to his or her actions, which in offline actions is reflected in the constant reflexive monitoring of one's actions, whereas in the action environment provided by the Internet, the replies or likes of others can provide feedback.

This concept can explore the inner dimension of actors, measuring the different commitment levels of participants and give a possible method to deal with the false polarization.

Mechanisms

This section provides a unified explanation for the micro-level interactions among individuals on the internet and the macro-level state of text accumulation, and it describes the overall characteristics of text actions on the internet. Text accumulation is a typical result of text actions, which includes both the accumulation of text quantity within a topic due to discussions on the network and the simplification of information and increase in text similarity caused by internet dissemination forms, such as the reduction in the complexity of viewpoints, the emergence of extreme viewpoints, and typical internet phenomena like meme creation and repetition. Therefore, the focus is on two attributes of text: one is its quantity, and the other is the similarity of its content. The process of text accumulation can be simplified as the interaction between the following variables.

Effect of Speech Tendency Mechanism on Text Accumulation. Speech tendency describes the average level of activity of individuals in online topics. The stronger the speech tendency, the more active the actor. A stronger speech tendency leads to a higher degree of text accumulation. (This approach decomposes the role of identity considered in polarization models into prudence in speech (i.e., speech tendency) and self-integration degree (consistency with past statements) discussed below.)

Effect of Positive and Negative Indexing Mechanism on Text Accumulation. Indexing behavior includes actions such as replying to, liking, disliking, or sharing others' viewpoints. Positive indexing behavior refers to indexing actions on texts with similar viewpoints. Positive indexing tendency indicates the propensity to perform indexing actions when encountering similar viewpoints, while negative indexing tendency indicates the propensity to perform indexing actions when encountering differing viewpoints.

Effect of Self-Retrospection Mechanism on Text Accumulation. Self-retrospection refers to the actor's attention to others' indexing actions on their own texts. Its effect lies in influencing the strength of the self-integration mechanism.

Self-retrospection is often for self-monitoring and regulation. Self-monitoring refers to the fact that some people monitor their performance behaviors out of concern for the context of their self-representation and regulate their self-representation accordingly to achieve a desirable public image(Gangestad & Snyder, 2000). This process requires the person to pay attention to his or her actions and to regulate them. Actors can have more retrospection online than offline action since they may become opinion leaders.

Effect of Self-Integration Mechanism on Text Accumulation. The self-integration mechanism is influenced by two aspects. Firstly, the core of self-integration lies in self-retrospection, which is affected by others' indexing actions—the more indexing by others, the more self-retrospection. Secondly, the degree of self-integration is influenced by the overall consistency of the texts the actor produces—the higher the consistency, the higher the degree of self-integration.

The effect of the self-integration mechanism reflects the strength of the impact of the actor's degree of self-integration on their textual actions (i.e., the behavior of making statements).

Effect of Platform Curation Mechanism on Text Accumulation. This mechanism includes the algorithmic settings for text comments (encouraging or discouraging comments) in general research. Curation behavior truncates or expands individuals' exposure to information, shaping the interpersonal influence environment, and ultimately influencing text accumulation through this environment. We reflect this in the "social influence network," i.e., the network formed by interpersonal influence relationships, used to describe interpersonal influence.

As mentioned earlier, there is a significant difference between prudent and active speakers, and the role of feedback from others and algorithmic curation (Curation) further influences an individual's self-retrospective process. Curation influences the information that actors have access to and the visibility of their textual actions, such as whether or not to set "visible only to self" or "visible only to friends" is a common form of curation.

The term "curation" was originally used to describe how museums and art galleries organize and exhibit content. In communications, the term is used to describe the process by which platforms filter, organize, and present information through algorithms or human intervention. Media curation determines the content that users are exposed to on a platform. For example, High Likes topping is the curation of the comments section, and Big Data Push is the curation of information on the homepage.

Through curation, platforms can interact with users more effectively and influence the dissemination of information. Common forms include personalized recommendations, news pushes, and hot search terms, all of which influence public discussion. Filter bubbles theory is an example of information curation(Pariser, 2011).

Models

Entities and Variables

Firstly, to effectively represent complex semantic entities, we introduce the concept of high-dimensional viewpoints. This concept can be expressed as $y_i(t) = (y_{i1}(t)...y_{iN}(t))$, representing all the opinions held by individual i at time t. Each component $y_{ik}(t)$ can have various interpretations. For example, on the Weibo platform, each component can represent an individual i's attitude towards a specific topic or hashtag (from 1 to N). Generally, $y_{ik}(t)$ can be understood as an individual i's opinion on issue k within the digital domain.

Secondly, to characterize the interaction behavior between individuals and their intrinsic psychological processes, a series of variables related to the psychological states of actors need to be considered:

Referencing Behavior $R_i(t)$. :

$R_i(t)$ is used to indicate whether individual i has been referenced or responded to by other individuals at time t. Specifically, the referencing behavior of individual i, it can be represented as $R_i(t) = (R_{i1}(t)...R_{in}(t))$, where $R_{ij}(t)$ indicates whether the individual j referenced the individual i's content at time t through actions such as retweets, likes, or comments. If referenced, $R_{ij}(t) = 1$; if not, $R_{ij}(t) = 0$.

Self-Reflection Degree $T_i(t)$. :

This variable measures the extent to which individual i reflects on their own digital traces, i.e., their previous textual actions at the time t. $T_i(t)$ ranges between $[0, 1]$, where higher values indicate a deeper level of reflection on past behavior.

Self-Integration Degree $I_i(t)$. :

Self-integration describes the extent to which an individual's behavior aligns with their cognitive role expectations. This represents a depiction of the concept of self-integration mentioned earlier. Like $T_i(t)$, the value of $I_i(t)$ ranges between $[0, 1]$, with higher values indicating a greater alignment between actions and role expectations, while lower values reflect a higher degree of separation between actions and self-awareness, as previously discussed.

Role R_i. :

This describes the role of active agents or cautious actors as mentioned earlier. $R_i = 1$ indicates an active agent, and $R_i = 0$ represents a cautious actor. Note that this variable relates not only to the objective tendency of the actor's speech but also to the actor's perception of their own speaking tendencies. Therefore, "role" is used to express this concept.

Lastly, the core micro-level phenomenon depicted by this model—namely, "textual action"—needs to be considered. To this end, this study integrates the viewpoints and speech behaviors of individuals across various issue domains to construct a depiction of textual actions. Firstly, individual i's opinion on region k is represented by $y_i(t)$, where k denotes any region between 1 and N. Secondly, whether individual i expresses an opinion is represented by $M_i(t)$, denoted as $M_i(t) = (m_{i1}(t)...m_{iN}(t))$, where $m_{ik}(t) = 1$ indicates that the individual expressed an opinion in the region k, and $m_{ik}(t) = 0$ means no opinion was expressed. By performing an element-wise multiplication (Hadamard product) of $y_i(t)$ and $M_i(t)$, i.e., $x_i(t) = (y_{i1}(t) * m_{i1}(t)...y_{iN}(t) * m_{iN}(t))$, the textual action of the individual i at time t can be represented.

Specifically, if an individual i expresses an opinion in a region k, i.e., $m_{ik}(t) = 1$, the textual action in region k, $x_{ik}(t) = y_{ik}(t)$, represents that the individual's viewpoint is expressed in the text and leaves a trace. On the other hand, if $m_{ik}(t) = 0$, then $x_{ik}(t) = 0$, meaning no textual action took place. Consequently, by observing the changes in $x_i(t)$ for all individuals in the system, the overall phenomenon of textual accumulation can be described, which shows how individuals' textual behaviors in different regions evolve and accumulate over time.

By introducing these key variables and structures, this model captures the complex behaviors of individuals in social media environments, including opinion formation and evolution, individual interaction and feedback, self-awareness and integration, as well as the process of Text accumulation.

Evolutionary Processes

Interaction: dynamics of viewpoint content and interpersonal indexing. After defining and describing the entity variables in the previous section, we next turn our attention to the evolutionary process of these variables, in particular the evolution of the textual action $y_i(t)$ and the interpersonal index $R_i(t)$. This process relies heavily on interpersonal interactions between individuals and viewpoint interaction effects, combining the classical viewpoint dynamics model with randomization factors in psychosocial mechanisms. To explain these evolutionary mechanisms more clearly, we first discuss several sub-processes in an individual's textual action and elicit the relevant settings of the model accordingly. First, when an individual i thinks about and chooses his or her opinion on the topic k, he or she is usually influenced by two main sources: one is the individual's opinion history, which represents the individual's judgments based on past opinions, and the other is the other people that the individual has access to, in particular the neighboring individuals in his or her social network (i.e., "the opinions of others in the social neighborhood "). Therefore, when an individual i updates her opinion on a topic k, the individual will combine her own past opinions with the opinions of other individuals in her social network based on specific weights. It can be described as individual i learning about individual j. The rate of this learning can be denoted as w_{ij}, and similarly, the rate of learning about one's own established opinions can be denoted as w_{ii}. Thus, the evolution of individual i's opinions can be expressed as follows:

$$y_i(t+1) = \sum w_{ij} y_j(t)$$

This formula portrays how an individual i's viewpoint changes, with the individual i's viewpoint $y_i(t+1)$ at the next time (i.e., $t+1$) being based on the weighted sum of the viewpoints of the other individuals in his or her social network. This formula preserves interpersonal influences and portrays the social properties of textual actions, rather than simply treating leave-behind texts as stable positions of individuals. Second, during the process of an individual i's expression of her viewpoint, i.e., from t to $t+1$, due to the existence of standardized interactions on the platform such as likes, comments, retweets, etc., other individuals j may index the individual i's textual action, i.e., give feedback on individual i's viewpoint, so that individual i receives the attitudes of others, which can be differentiated as either agreeing or rejecting in simple terms, and reflects the "social influence theory" of social influence. This reflects the "interpersonal evaluation" dimension of social influence theory. In this process, when the views of the individual j are different from those of the individual j, the individual j may show rejection of i's viewpoint, while if the views of the two individuals are similar, they may show approval. In social platforms, this attitude is generally objectively expressed through behaviors such as liking, commenting, retweeting, etc., which are also known as the indexing behaviors of individual j to individual i. When the indexing behavior occurs, it is written as $R_{ij}(t) = 1$. The probability that the individual j indexes individual i depends on the similarities and differences of the two viewpoints. This probability can be expressed as the following equation:

$$P(R_{ij}(t) = 1) = c \cdot sim\_cos(y_i(t), y_j(t)) + d \cdot (\frac{1}{Sim\_cos(y_i(t), y_j(t))})$$

where $sim_cos(y_i(t), y_j(t))$denotes the viewpoint similarity between individuals i and j. c and d are unknown parameters. c is used to measure the sensitivity of the individual j to viewpoint similarity, and d is used to measure its sensitivity to viewpoint difference, and both affect the probability that j engages in indexing behavior.

Actors: the dynamics of self-retrospection and self-integration. After discussing the evolution of interactional behavior, we turn to an exploration of the evolutionary process of the individual's intrinsic psychological variables, the actor's intrinsic process by the degree of self-retrospection $T_i(t)$, and the degree of self-integration $I_i(t)$. These variables reflect the individual's self-perception and behavioral adjustment process in social networks. First, individuals are triggered by certain external factors or events when they engage in self-retrospection. Especially in the social network environment, the indexing behaviors (e.g. likes, comments, retweets, etc.) of others towards an individual are important factors that motivate the individual to retrace. When an individual i enters a social platform and views his or her previously posted content, he or she often receives comments and feedback (either positive or negative) from other users. This feedback will make the individual aware of his or her previous statements or behaviors again, thus triggering self-retrospection. The process is modeled and represented as follows:

$$T_i(t+1) = \beta \cdot \sum R_{ij}(t)$$

where is an unknown parameter that measures the degree of sensitivity of individual i to the indexing of others in the retrospective behavior. However, as long as the sensitivity degree is not 0, the more others index, the more frequent the individual's backtracking behavior. Regarding the degree of self-integration $I_i(t)$, its evolution is closely related to the consistency of the individual's historical behavior. During the backtracking process, individuals refer to their previously published opinions or text traces. If these show a high degree of logical or emotional consistency, the individual's degree of self-integration will subsequently increase. In other words, the greater the coherence of historical textual actions, the greater the individual's likelihood of self-integration. The degree of self-integration can be expressed by the following formula:

$$I_i(t+1) = T_i(t+1) + \gamma \cdot \sum sim\_cos(x_{ik}(t), x_{il}(t))$$

The summation component portrays the consistency of an individual's textual traces or opinions under different moments. is an unknown parameter that measures the effect of an individual's historical consistency on self-integration. Overall, when individuals are more consistent in their historical views, they are more self-integrated.

Dynamics of Individuals Issuing Textual Actions. After discussing the process of interpersonal interaction and self-integration, we finally focus on the evolution of text actions $x_i(t)$, which is one of the core variables of this paper. Textual actions are jointly composed of an individual's opinion content $y_i(t)$ and opinion-posting behavior $M_i(t)$. While the previous section focused on the processes of interpersonal indexing as well as self-retrospection that $y_i(t)$ is affected by, this part explores the evolution of $M_i(t)$, i.e., the evolution of whether or not an individual i posts his or her opinion in a region k. Here, we need to draw out the previous distinction between prudent and active actors, and it is clear that active individuals are more likely to post opinions than prudent individuals under the same region. It is important to note that the role R_i is related to the degree of self-integration, in other words, the process of whether an individual perceives himself as "active" or "prudent" and whether his behavior conforms to the expectations of this role is influenced by the degree of self-integration of the individual. The higher the level of self-integration, the more the individual tends to maintain behavior consistent with the role. Empirically, an actor who is a frequent speaker may underestimate his or her level of activism due to a lack of self-retrospective processes. In addition, the viewpoints presented in the region can influence an individual's decision to issue an action, as discussed previously concerning the tendency of actors to avoid conflict, whereas individuals are more likely to engage in speaking if they perceive their viewpoints to be similar to those of other users in the region. Actors naturally make consequence expectations when confronted with statements with different degrees of viewpoint similarity, and the psychological burden of expressing views similar to those of others is low, which makes the herd effect widespread in social networks. Of course, the exclusion effect caused by the difference in the degree of viewpoints also has an impact on the issuance of textual actions, and this difference can be portrayed by different values of the parameter. Based on the above two factors, the probability of an individual posting an opinion in region k can be expressed as:

$$P(m_{ij}(t) = 1) = \alpha \cdot R_i \cdot I_i(t+1) + (1-\alpha) \cdot (\frac{1}{K} \cdot \sum sim(y_i(t), y))$$

The $\alpha * R_i * I_i(t+1)$ of this formula indicates the effect of an individual's role perception and self-integration on the speaking behavior, while the second term indicates the effect of similarity or difference of opinions in the region on the speaking, and $\frac{1}{K}$ is used to measure the overall probability of an individual's speaking.

Simulation and Calibration

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Morbi ac iaculis neque. Sed posuere ac augue vel bibendum. Pellentesque in commodo nibh. Maecenas mattis finibus maximus.

Agent-Based Modeling

This paper portrays the phenomenon of text clustering in digital society from multiple perspectives, focusing on individuals' viewpoint evolution, indexing behavior, self-retrospection, and self-integration on social platforms, and constructs an agent-based modeling (ABM) framework through these processes. Next, the rules based on the above modeling are simulated and the macroscopic results presented in the graphs are illustrated.

Evolution of viewpoint distribution.

– Figure 1 here –

This three-dimensional diagram shows the process of viewpoint evolution, where the three axes represent "Time", "Opinion" and "Prob_Density ". This is a dynamic simulation of an individual's or group's changing views on a topic. The time axis from 0 to 30 indicates the time step of the simulation, where the individual's viewpoints keep changing as time advances. The viewpoint axis from 0 to 1 indicates the viewpoint of an individual or group on a particular topic. The probability density, or vertical axis, indicates how often different viewpoints appear at a given point in time. A high-density region (close to 0.3) means that more individuals or larger groups are clustered on that viewpoint, while a low-density region (close to 0) indicates that fewer individuals hold that viewpoint. As shown in the figure, the clustering of views can be observed over time. In terms of the range of distribution of viewpoints, the distribution of viewpoints is more centralized and less dense in the early stage, more complex and diverse in the middle stage, and back to a centralized state in the late stage, indicating convergence of viewpoints and clustering of texts. In the second half of the graph, higher probability density regions gradually appear on certain viewpoint values. As the discussion progresses, certain viewpoints gain support or consensus from more individuals. This figure shows that the evolution of viewpoints is not linear, but higher-density viewpoint intervals are gradually formed through continuous interaction and feedback, and the textual viewpoints presented in the same region have experienced a process from dispersion to polarization.

The influence of viewpoint similarity on Text Accumulation. As mentioned before, the similarity of viewpoints within a region enables actors to lower their prediction of the cost of actions and thus issue more text actions. It is important to note that clicking activities such as likes and quick spins also fall into the category of text actions with viewpoints. The following figure simulates the herd effect of the Internet to present the strength of the model's explanation of the polarization phenomenon. This figure shows the cumulative change in the number of texts in the five regions over time, with different colored curves representing the amount of text clustering in each of the five regions. The vertical axis of the graph represents the amount of text and the horizontal axis represents time. By observing the growth trend of these curves, the dynamics of text agglomeration in each region can be analyzed, as well as the impact of viewpoint polarization on text agglomeration.

– Figure 2 here –

Among them, the two curves in blue and orange represent the regions with a higher degree of polarization. It can be seen that they accumulate the number of texts significantly faster than other regions, and the final total number of texts is much higher than in other regions. The curves in other colors represent regions with a lower degree of polarization. Compared to the blue and orange curves, these regions have a slower rate of text accumulation and a lower final text total. Early in time, the growth in the number of texts in each region is more similar. This suggests that in the early stages, polarization has not yet manifested itself or had a significant impact on text clustering, and there is not much difference in the amount of text generated in each region. In the time period 5-10, the number of texts in different regions starts to diverge significantly. In the more polarized regions, the rate of text clustering shows a first growth and then a slowdown, but overall the number of texts is much higher than that of other regions.

Positive and Negative Indexing Behavior and Level of Role Activity. The previous section illustrated how indexing behavior affects self-referential behavior, which in turn affects the degree of text clustering. The following heatmap shows the relationship between roles with different levels of activity and positive and negative indexing. The horizontal axis represents the average of the roles (Role), with Ri taking the value of 0 or 1, while the horizontal axis is taken as the average of all people, reflecting the relationship between the average level of activity in the region on text clustering. The horizontal axis is taken from 0 to 12, with 6 being a cut-off point where individuals with values less than 6 are skewed towards prudence and those with values greater than 6 are skewed towards activity. The vertical axis indicates the degree of indexing, with higher values meaning more indexed. The color indicates the amount of text, blue means a high amount of text, and red means a low amount of text.

– Figure 3 here –

As shown above, the vertical axis indicates Positive Reference, where a higher value means that the character is positively referenced by others more often. For the area with more prudence, which is near the left, too much Positive Reference decreases the number of texts, but the right amount of Positive Reference increases the overall number of texts. For the regions with more activists, the number of texts is much larger than that of the regions with more prudence, and the effect of positive indexing on the number of texts is not significant. In the charts with negative indexing (Negative Reference) behavior, it is also the prudes that are more affected. In the left-hand part of the chart, the overall gradual warming of the color represents a decrease in the number of texts in the presence of an increase in Negative Indexing. Many activists make decisions based on the expected consequences of their actions, choosing to remain silent if they anticipate that they may face backlash from others, and this chart shows the same effect. For activists, on the other hand, negative indexing triggers their motivation to respond, and a certain degree of negative indexing in the middle part of the chart's horizontal axis results in a higher number of texts. But when the region has a high percentage of activists, the number of texts is high regardless of whether they are negatively indexed or not.

– Figure 4 here –

The effect of self-integration and speaking tendencies on text Accumulation. The previous section provided a brief explanation of the relationship between the degree of self-integration and roles, and this section interprets the simulation results for this. The following figure shows the effect of the degree of self-integration as well as roles on the number of texts, where the horizontal axis represents the percentage of roles, with higher values representing a higher percentage of activists. Where 6 is a cut-off point, where a fraction less than 6 indicates a higher percentage of prudent actors. The vertical axis represents the degree of self-integration, where higher values indicate a higher degree of self-integration. Where the shade of color represents the number of texts, with blue indicating the highest number of texts. This graph shows that when there is a higher percentage of activists, the higher the degree of self-integration, the more speaking behaviors there are. That is, when most individuals in a given platform tend to express opinions, the more individuals are aligned with their positions and opinions, the more likely they are to engage in textual actions. For more prudent speakers, on the other hand, the higher the degree of self-integration, the more these prudent individuals tend to speak less, and instead, individuals with a lower degree of self-integration will engage in speaking. In the theoretical model, self-integration reflects an individual's consistency in actions and role expectations, and individuals with high levels of integration tend to be more assertive in expressing themselves, as directly reflected by the increase in the number of texts in the figure with increasing levels of integration.

– Figure 5 here –

Case from Weibo

The cases in this study are derived from the microblogging platform, which is a social media platform characterized by real-time content posting and interaction. For the purpose of model calibration and verification, this study uses data from three major hot search phrases under the hotly debated Fat Cat event in 2024. Fat Cat was a game trainer who committed suicide by jumping into a river on April 11, 2024 and was believed by Fat Cat's sister and netizens to be related to his ex-girlfriend's financial scam because he sent her the message "We're done" before his death. The incident initially attracted widespread attention, with many netizens empathizing with Fat Cat, who was regarded as a "warrior of pure love," sending tweets of condolence and accusing his ex, Tan Zhu, of being a "womanizer. As it struck a nerve with the male-female dichotomy on the Internet today, the discussion of the incident once again polarized netizens and triggered a large number of text actions. The three phrases crawled in this paper are "fat cat case details released", "Tan Zhu" and "police report fat cat and Tan's financial dealings", and the criteria for selection are the heat of discussion and the number of participants. The criteria for selection were the heat of discussion and the number of participants. Among the entries discussing the fat cat incident, these three entries have a high level of discussion activity. This study crawls the public user interaction data on Weibo, and the data crawling process is carried out by legal technical means, following the platform's privacy policy and data use norms, to ensure that the information obtained is limited to the public content, and does not involve any user privacy data. Specifically, the data crawled in this study mainly includes the following information: first, the usernames of the actors, which are used to label different actors; second, the textual content of the tweets sent by the actors, the length and content of which can illustrate the intensity of the actors' attitudes and opinions; third, the number of interactions the tweets have received, i.e., the number of likes, retweets, and replies, which are important indicators of the degree of text clustering; and finally, the timestamps of text publication, which are used to perform evolutionary analysis of text actions. After normalizing the data according to the analysis requirements, the final dataset includes a total of 949 tweets over a 45-day period from May 4th to June 14th.

To ensure the validity and realism of the ABM model in this paper, the crawled microblogging data are used in this study for the calibration and verification process of the model. Calibration is the process of adjusting the parameters in the model to make its output as consistent as possible with the observations in reality; while calibration is the process of verifying the predictive ability and accuracy of the model under certain conditions. In the figure below, the horizontal axis represents time and the vertical axis represents the number of texts in the region. represents a randomly selected text region, where different lines represent the evolution process obtained through the model data, which is consistent with the real data. The evolution process of the real data is one of the lines, and it can be seen that the different lines in this figure reflect the same evolution process, with only the difference in the number of texts. This figure illustrates that multiple calibrations show that this model has accuracy under certain conditions.

– Figure 6 here –

Discussion

Research in disciplines like digital sociology and political communication has long emphasized the significance of digital action. Within the realm of online collective action, scholars have predominantly engaged with two overarching perspectives. One approach views digital action as politically significant, focusing on identity change and the forms of collective action on the internet. Conversely, another viewpoint emphasizes the pervasive dissemination of opinions and emotions on social media platforms, drawing parallels to Le Bon's concept of Crowd Behavior. Both approaches tend to simplify the role of the actor within specific contexts. Meanwhile theoretical strand within digital sociology emphasizes the role of the internet in creating fragmented subjects. This paper aims to synthesize the intrinsic dimensions of the subject with analytical sociological models of collective action, thereby uncovering a more universally applicable framework for roles and interactions within digital action contexts. Through this integrative approach, a deeper comprehension of the patterns and mechanisms underlying digital activism and engagement can be achieved.

This paper posits Text Accumulation as an explanatory mechanism for online collective action. Text Accumulation phenomenologically characterizes the concentration of actor engagement within specific areas of the Internet. It refers to the textual scale, consistency, and emotional intensity of text production within a specific online domain. It serves to characterize the dissemination of information across public internet platforms and its influence on actors' actions. To formalize this concept, the paper presents a detailed model outlining its integrative nature and the underlying laws of Text Accumulation. These laws outline the patterns governing the accumulation of textual content and its impact on online collective behavior.

At the model level, this paper delineates the internal mechanisms governing the fluctuation of Text Accumulation, contingent upon dimensions such as media dissemination, technological regulations, and self-cognitive processing. Central to these mechanisms is the concept of "self-integration," which hinges on two primary factors. Firstly, it underscores individuals' introspection regarding their own behaviors, manifested through micro-actions like liking or commenting. Secondly, it relies on the coherence of meanings within individuals' social trajectories, indicating the extent to which multiple micro-actions are interconnected. The manifestation of individuals' micro-actions within media spaces varies significantly based on their perceived roles and levels of self-integration. For instance, individuals who identify as "prudent staters" and exhibit high levels of self-integration are less inclined to conform to crowd behavior or engage in random commentary. Collectively, these intrinsic mechanisms furnish the microfoundations essential for comprehending and modeling digital actions and the phenomenon of Text Accumulation.

At the modeling level, drawing from the outlined patterns and mechanisms, we integrate adaptive networks, stochastic games, and neighborhood mimicry strategies to formulate a comprehensive stochastic dynamics model—a type of Agent-Based Model (ABM). This model is designed to capture the numerical expression of inter-individual interactions (such as engagement within specific online areas), the evolution of viewpoints (reflecting attitudes towards particular topics), and the phenomenon of "indexing" behavior (public assessments of others). Such a model serves as a powerful tool for describing the phenomenon of Text Accumulation. Furthermore, leveraging concepts from heterogeneous mean-field theory and master equations in statistical mechanics, we derive approximate kinetic equations governing the evolution of key variables, including the degree of accumulation and media attitudes. Employing integral transformation and other analytical methods, we obtain corresponding solutions, thus facilitating an in-depth exploration of the fundamental conditions influencing text accumulation dynamics. Moreover, employing Bayesian computation techniques alongside other methodologies, we simulate and calibrate the model using real-world social media data. Through this rigorous process, we attain fitting results that validate the robustness and reliability of the model.

At the level of laws, we obtain a series of conclusions based on formal proofs, model simulation, and data validation: (1)Micro-mechanisms such as preference attachment, reputational recommendation (with high liking prevalence), information integration, and role reproduction/self-awareness significantly impact Text Accumulation levels. (2). Higher levels of viewpoint polarization contribute to increased Text Accumulation in highly public online venues. (3). Text Accumulation tends to rise when self-integration levels are low. (4). Diverse character types and the prioritization of high-reputation texts for attention resources inhibit the upward trend of Text Accumulation. These findings underscore the intertwined roles of cognitive processing, interpersonal interaction, and media regulations.

In summary, our development of a self-integration-role mediation framework offers a comprehensive explanation for the degree of Text Accumulation across various digital actions. This framework not only unveils the underlying mechanisms and stochastic dynamics models, enabling explanation and prediction but also elucidates the conditions and fundamental laws influencing this phenomenon. Through this work, we aim to offer insights and avenues for digital sociology research. This paper also helps us to consider the relation between online antagonism and polarization, which is important for understanding affecive polarization.

References

Bail, C. A., Argyle, L. P., Brown, T. W., et al. (2018). Exposure to opposing views on social media can increase political polarisation. *Proceedings of the National Academy of Sciences*, *115*(37), 9216–9221.

Bakardjieva, M. (2012). Subactivism: Lifeworld and politics in the age of the internet. In *(re) inventing the internet* (pp. 85–108). Brill.

Bakker, B. N., & Lelkes, Y. (2024). Putting the affect into affective polarization. *Cognition and Emotion*, *38*(4), 418–436.

Barabási, A.-L. (2013). Network science. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *371*(1987), 20120375.

Bennett, W. L., & Segerberg, A. (2013). *The logic of connective action: Digital media and the personalization of contentious politics*. Cambridge University Press.

Blatz, C. W., & Mercier, B. (2018). False polarisation and false moderation: political opponents overestimate the extremity of each other's ideologies but underestimate each other's certainty. *Social Psychological and Personality Science*, *9*(5), 521–529.

Brubaker, R. (2020). Digital hyperconnectivity and the self. *Theory and Society*, *49*, 771–801.

Callander, S., & Carbajal, J. C. (2022). Cause and effect in political polarization: a dynamic analysis. *Journal of Political Economy*, *130*(4), 825–880.

Chang, K., & Park, J. (2021). Social media use and participation in dueling protests: The case of the 2016–2017 presidential corruption scandal in south korea. *The International Journal of Press/Politics*, *26*(3), 547–567.

Chang, W. J., & Park, S. E. (2019). The fandom of hallyu, a tribe in the digital network era: The case of army of bts. *Kritika Kultura*.

Cheng, J., Bernstein, M., Danescu-Niculescu-Mizil, C., et al. (2017). Anyone can become a troll: Causes of trolling behaviour in online discussions. In *Proceedings of the 2017 acm conference on computer supported cooperative work and social computing* (pp. 1217–1230).

Christensen, H. S. (2011). Political activities on the internet: Slacktivism or political participation by other means? *First Monday*.

Cinelli, M., De Francisci Morales, G., Galeazzi, A., et al. (2021). The echo chamber effect on social media. *Proceedings of the National Academy of Sciences*, *118*(9), e2023301118.

Collins, R. (2014). Interaction ritual chains and collective effervescence. In *Collective emotions* (pp. 299–311).

Diener, E. (1980). Deindividuation: The absence of self-awareness and self-regulation in group members. In P. Paulus (Ed.), *The psychology of group influence.* Hillsdale, NJ: Erlbaum.

Durkheim, É. (2016). The elementary forms of religious life. In *Social theory re-wired* (pp. 52–67). Routledge.

Faizi, R., El Afia, A., & Chiheb, R. (2013). Exploring the potential benefits of using social media in education. *International Journal of Engineering Pedagogy (iJEP)*, *3*(4), 50–53.

Fernbach, P. M., & Van Boven, L. (2022). False polarization: Cognitive mechanisms and potential solutions. *Current Opinion in Psychology*, *43*, 1–6.

Festinger, L., Pepitone, A., & Newcomb, T. (1952). Some consequences of de-individuation in a group. *Journal of Abnormal and Social Psychology*, *47*, 382–389.

Flache, A., & Macy, M. W. (2014). Small worlds and cultural polarisation. In *Micro-macro links and microfoundations in sociology* (pp. 146–176). Routledge.

Gangestad, S. W., & Snyder, M. (2000). Self-monitoring: Appraisal and reappraisal. *Psychological Bulletin*, *126*(4), 530.

George, J. J., & Leidner, D. E. (2019). From clicktivism to hacktivism: understanding digital activism. *Information and Organization*, *29*(3), 100249.

Giddens, A. (1984). *The constitution of society: Outline of the theory of structuration.* Polity.

Gidron, N., Adams, J., & Horne, W. (2019). Toward a comparative research agenda on affective polarisation in mass publics. *APSA Comp Polit Newsl*, *29*, 30–36.

Gil-López, T., Shen, C.-f., Benefield, G. A., et al. (2018). One size fits all: Context collapse, self-presentation strategies and language styles on facebook. *Journal of Computer-Mediated Communication*, *23*(3), 127–145.

Hameleers, M., Schmuck, D., Bos, L., et al. (2021). Interacting with the ordinary people: how populist messages and styles communicated by politicians trigger users' behaviour on social media in a comparative context. *European journal of communication*, *36*(3), 238–253.

Hassid, J. (2012). Safety valve or pressure cooker? blogs in chinese political life. *Journal of Communication*, *62*(2), 212–230.

Haythornthwaite, C. (2005). Social networks and internet connectivity effects. *Information, Community Society*, *8*(2), 125–147.

Iyengar, S., Lelkes, Y., Levendusky, M. S., Malhotra, N., & Westwood, S. J. (2019). The origins and consequences of affective polarization in the united states. *Annual Review of Political Science*, *22*, 129–146.

Jane, E. A. (2015). Flaming? what flaming? the pitfalls and potentials of researching online hostility. *Ethics and Information Technology*, *17*(1), 65–87.

Joyce, M. C. (2010). *Digital activism decoded: The new mechanics of change*. Idea.

Kaun, A., & Uldam, J. (2018). Digital activism: after the hype. *New Media & Society*, *20*(6), 2099–2106.

Koehler, D. (2014). The radical online: individual radicalisation processes and the role of the internet. *Journal for Deradicalisation*(1), 116–134.

Korda, H., & Itani, Z. (2013). Harnessing social media for health promotion and behavior change. *Health Promotion Practice*, *14*(1), 15–23.

Lea, M., O'Shea, T., Fung, P., et al. (1992). *'flaming' in computer-mediated communication: observations, explanations, implications*. Harvester Wheatsheaf.

Lea, M., Spears, R., & De Groot, D. (2001). Knowing me, knowing you: Effects of visual anonymity on self-categorization, stereotyping and attraction in computer-mediated groups. *Personality and Social Psychology Bulletin*, *27*, 526–537.

Lea, M., Spears, R., & Watt, S. E. (2007). Visibility and anonymity effects on attraction and group cohesiveness. *European Journal of Social Psychology*, *37*, 761–773.

Le Bon, G. (2017). *The crowd: a contradiction in terms?* Routledge.

Lelkes, Y. (2018). Affective polarization and ideological sorting: a reciprocal, albeit weak, relationship. *Forum*, *16*(1), 67–79.

Lupton, D. (2016). The diverse domains of quantified selves: Self-tracking modes and dataveillance. *Economy and Society*, *45*(1), 101–122.

MacKinnon, R. (2008). Flatter world and thicker walls? blogs, censorship and civic discourse in china. *Public Choice*, *134*(1–2), 31–46.

Marchal, N. (2022). "be nice or leave me alone": An intergroup perspective on affective polarization in online political discussions. *Communication Research*, *49*(3), 376–398.

Mead, G. H. (2023). Self. In *Social theory re-wired* (pp. 425–437). Routledge.

Monin, B., & Norton, M. I. (2003). Perceptions of a fluid consensus: Uniqueness bias, false consensus, false polarization, and pluralistic ignorance in a water conservation crisis. *Personality and Social Psychology Bulletin*, *29*(5), 559–567.

Mooijman, M., Hoover, J., Lin, Y., et al. (2018). Moralization in social networks and the emergence of violence during protests. *Nature Human Behaviour*, *2*(6), 389–396.

Mouw, T., & Sobel, M. E. (2001). Culture wars and opinion polarization: the case of abortion. *American Journal of Sociology*, *106*(4), 913–943.

Müller, K., & Schwarz, C. (2021). Fanning the flames of hate: Social media and hate crime.

Journal of the European Economic Association, *19*(4), 2131–2167.

Noelle-Neumann, E. (1974). The spiral of silence a theory of public opinion. *Journal of Communication*, *24*(2), 43–51.

Papacharissi, Z. (2016). Affective publics and structures of storytelling: sentiment, events and mediality. *Information, Communication & Society*, *19*(3), 307–324.

Pariser, E. (2011). *The filter bubble: What the internet is hiding from you.* Penguin UK.

Paz, M. A., Mayagoitia-Soria, A., & González-Aguilar, J. M. (2021). From polarisation to hate: Portrait of the spanish political meme. *Social Media + Society*, *7*(4), 20563051211062920.

Phillips, W. (2015). *This is why we can't have nice things: mapping the relationship between online trolling and mainstream culture.* MIT Press.

Postmes, T., & Brunsting, S. (2002). Collective action in the age of the internet: mass communication and online mobilisation. *Social science computer review*, *20*(3), 290–301.

Prentice-Dunn, S., & Rogers, R. W. (1989). Deindividuation and the self-regulation of behavior. In P. Paulus (Ed.), *The psychology of group influence* (2nd ed., pp. 86–109). Hillsdale, NJ: Lawrence Erlbaum.

Qiu, L., Lin, H., Chiu, C.-y., et al. (2015). Online collective behaviors in china: dimensions and motivations. *Analyses of Social Issues and Public Policy*, *15*(1), 44–68.

Ricoeur, P. (1992). *Oneself as another.* University of Chicago Press.

Röllicke, L. (2023). Polarisation, identity and affect—conceptualising affective polarisation in multi-party systems. *Electoral Studies*, *85*, 102655.

Sassenberg, K., & Postmes, T. (2002). Cognitive and strategic processes in small groups: effects of anonymity of the self and anonymity of the group on social influence. *British Journal of Social Psychology*, *41*(3), 463–480.

Schechtman, M. (1996). *The constitution of selves.* Ithaca, NY: Cornell University Press.

Schelling, T. C. (1969). Models of segregation. *The American Economic Review*, *59*(2), 488–493.

Schumann, S., & Klein, O. (2015). Substitute or stepping stone? assessing the impact of low-threshold online collective actions on offline participation. *European Journal of Social Psychology*, *45*(3), 308–322.

Selander, L., & Jarvenpaa, S. L. (2016). Digital action repertoires and transforming a social movement organization. *MIS Quarterly*, *40*(2), 331–352.

Shifman, L. (2012). An anatomy of a youtube meme. *New Media & Society*, *14*(2), 187–203.

Shirky, C. (2009). *Here comes everybody: how change happens when people come together.* Penguin UK.

Smith, E. R., & Mackie, D. M. (2015). Dynamics of group-based emotions: Insights from intergroup emotions theory. *Emotion Review*, *7*(4), 349–354.

Smith, E. R., Seger, C. R., & Mackie, D. M. (2007). Can emotions be truly group level? evidence regarding four conceptual criteria. *Journal of Personality and Social Psychology*, *93*(3), 431.

Spears, R., & Postmes, T. (2015). Group identity, social influence, and collective action online. extensions and applications of the side model. *The handbook of the psychology of communication technology*, 23–46.

Stage, C. (2013). The online crowd: a contradiction in terms? on the potentials of gustave le bon's crowd psychology in an analysis of affective blogging. *Distinktion: Scandinavian Journal of Social Theory*, *14*(2), 211–226.

Suler, J. (2004). The online disinhibition effect. *Cyberpsychology & Behavior*, *7*(3), 321–326.

Tanis, M., & Postmes, T. (2008). Cues to identity in online dyads: Effects of interpersonal versus intragroup perceptions on performance. *Group Dynamics-Theory Research and Practice*, *12*(2), 96–111.

Trottier, D. (2017). Digital vigilantism as weaponisation of visibility. *Philosophy & Technology*, *30*, 55–72.

Turkle, S. (1995). *Life on the screen: Identity in the age of the internet.* New York: Simon and Schuster.

Turkle, S. (2009). *Simulation and its discontents.* MIT press.

Turner, J. C. (1991). *Social influence.* Milton Keynes, UK: Open University Press.

Turner, J. C., Hogg, M. A., Oakes, P. J., Reicher, S., & Wetherell, M. S. (1987). *Rediscovering the social group: a self-categorisation theory.* Oxford, UK: Basil Blackwell.

Vaast, E., Safadi, H., Lapointe, L., & Negoita, B. (2017). Social media affordances for connective action - an examination of microblogging use during the gulf of mexico oil spill. *MIS Quarterly*, *41*(4), 1179–1205.

Van Bavel, J. J., & Pereira, A. (2018). The partisan brain: an identity-based model of political belief. *Trends in Cognitive Sciences*, *22*(3), 213–224.

van der Maas, H. L. J., Dalege, J., & Waldorp, L. J. (2020). The polarization within and across individuals: the hierarchical ising opinion model. *Journal of Complex Networks*, *8*(2), cnaa010.

Vissers, S., & Stolle, D. (2014). Spill-over effects between facebook and on/offline political evidence from a two-wave panel study. *Journal of Information Technology & Politics*, *11*(3), 259–275.

Voci, A. (2006). Relevance of social categories, depersonalization and group processes: two field tests of self-categorisation theory. *European Journal of Social Psychology*, *36*(1), 73–90.

Walther, J. B., Van Der Heide, B., Ramirez Jr, A., et al. (2015). Interpersonal and hyperpersonal dimensions of computer-mediated communication. In *The handbook of the psychology of communication technology* (pp. 1–22).

Wiedemann, C. (2014). Between swarm, network, and multitude: Anonymous and the infrastructures of the common. *Distinktion: Scandinavian Journal of Social Theory*, *15*(3), 309–326.

Wu, J. S., Hauert, C., Kremen, C., & Zhao, J. (2022). A framework on polarization, cognitive inflexibility, and rigid cognitive specialization. *Frontiers in Psychology*, *13*, 776891.

Yao, M. Z., & Ling, R. (2020). "what is computer-mediated communication?"—an introduction to the special issue. *Journal of Computer-Mediated Communication*, *25*(1), 4–8.

Yarchi, M., Baden, C., & Kligler-Vilenchik, N. (2021). Political polarization on the digital sphere: A cross-platform, over-time analysis of interactional, positional, and affective polarization on social media. *Political Communication*, *38*(1-2), 98–139.

Yuce, S., Agarwal, N., Wigand, R. T., et al. (2014). Studying the evolution of online collective action: Saudi arabian women's 'oct26driving' twitter campaign. In *Social computing, behavioral-cultural modeling and prediction: 7th international conference, sbp 2014, washington, dc, usa, april 1-4, 2014. proceedings 7* (pp. 413–420). Springer International Publishing.

Zimbardo, P. G. (1969). The human choice: Individuation, reason, and order vs. deindividuation, impulse and chaos. In W. J. Arnold & D. Levine (Eds.), *Nebraska symposium on motivation* (Vol. 17, pp. 237–307). Lincoln, NE: University of Nebraska Press.

Figure 1 Viewpoint Distribution

Figure 2 The Effect of Viewpoint Similarity on Text Accumulation

Figure 3 The Effect of Positive Index and Role Activity on Text Accumulation

Figure 4 The Effect of Negative Index and Role Activity on Text Accumulation

Figure 5 The Effects of Self-Integration and Role Activity on Text Accumulation

Figure 6 Multiple Simulation Tests

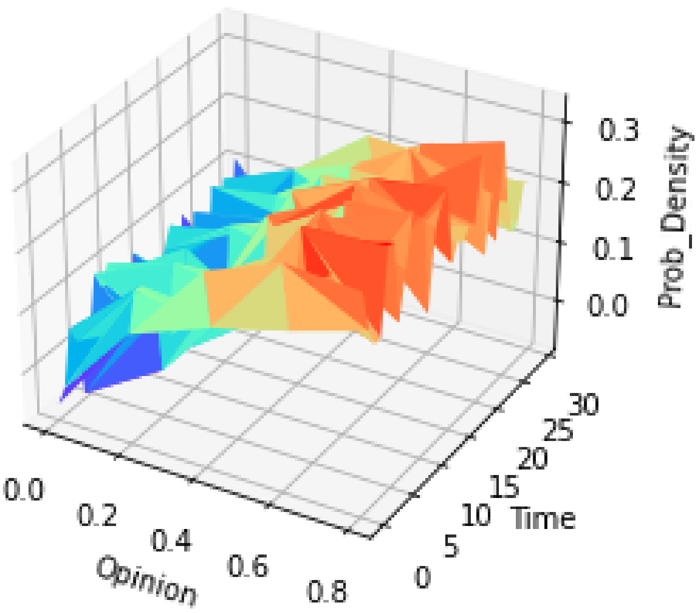

Figure 1. Viewpoint Distribution Evolution

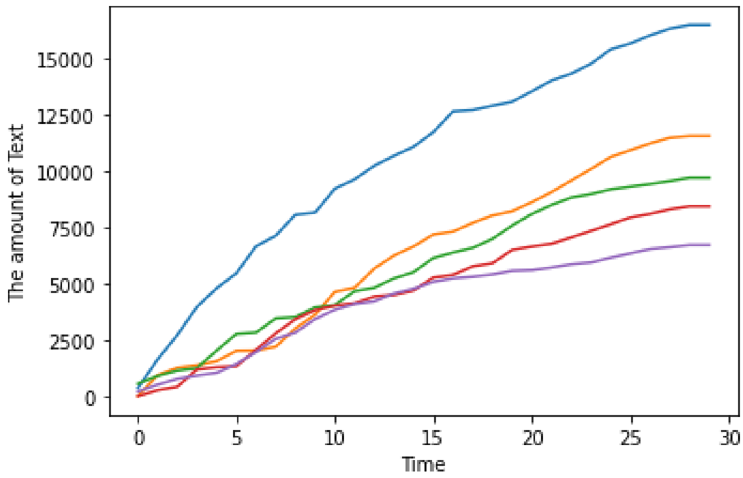

Figure 2. The effect of viewpoint similarity on text Accumulation

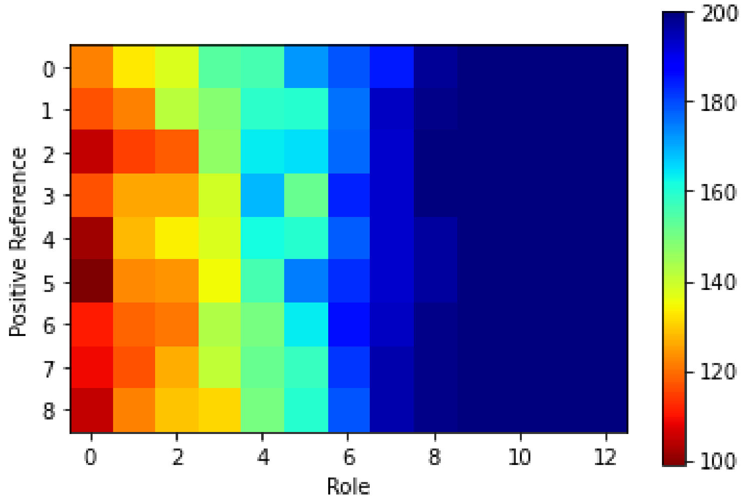

Figure 3. The Effect of Positive Index and Role Activity on Text Accumulation

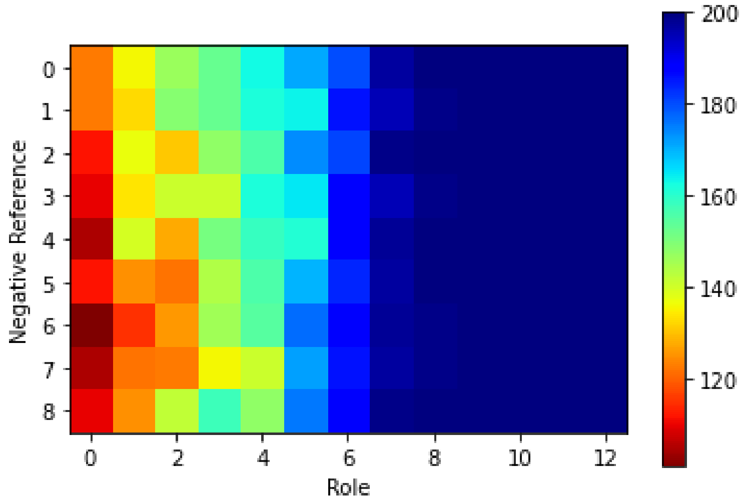

Figure 4. The Effect of Negative Index and Role Activity on Text Accumulation

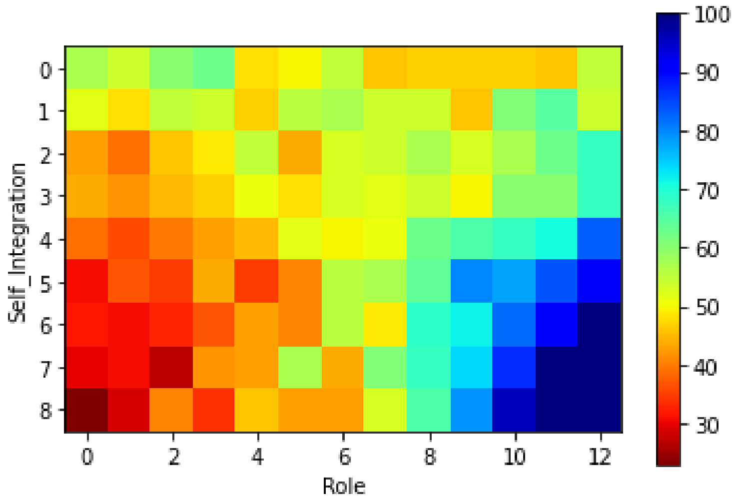

Figure 5. The Effects of Self-Integration and Speaking Tendencies on Textual Accumulation

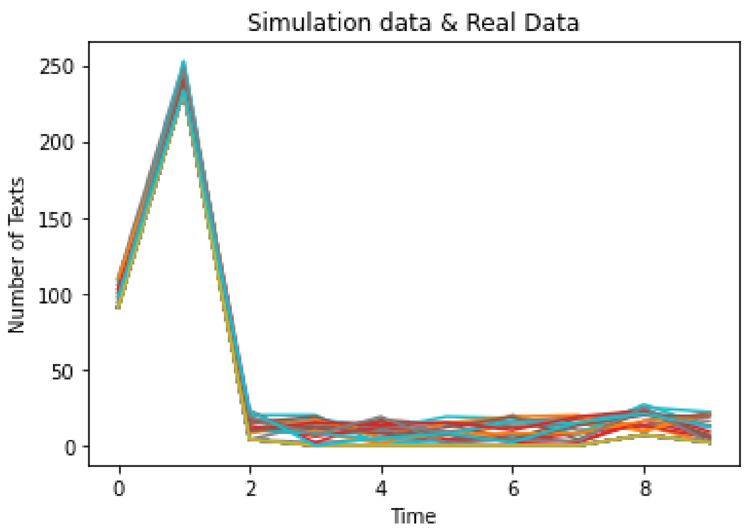

Figure 6. Multiple simulation tests