

Whose Voice in the COVID-19 Pandemic Discussion: a network actor analysis based on Weibo-COV corpus

Keywords: COVID-19, Actor-Network-Theory, Weibo-COV, Network Amplification, Agenda Setting

1 Research Background

As the COVID-19 pandemic spreads rapidly around the world, people are requested to maintain “social distance” and “stay at home.” Extensive social interactions subsequently moved to on-line space. People are using social media posts to share information, express opinions and seek help during the coronavirus pandemic[Hu et al., 2020]. During the duration of this public health emergency, people used social network to understand, disseminate and produce a large amount of epidemic-related information. On the one hand, it helped to make the epidemic information known to the public and to jointly protest. On the other hand, a large amount of false information confuses and misleads the public. The phenomenon of complicated information about the COVID-19 epidemic is called an “infodemic” by the World Health Organization[?]. The infodemic refers to the excessive amount of information online and offline. The message is usually mixed with false and false elements (misinformation, disinformation and mal-information) in the information flow. It can be quickly achieved with the help of information technologies such as the Internet, mobile phones and social media, causing public confusion, panic and distrust.

As the dominant channel for people’s social interaction during the COVID-19, social media has played an important role in the dissemination and diffusion of relevant information. In the empirical research on this infodemic, researchers either focus on the diffused information content and discuss the characteristics and impact of false information[Wang and Chang, 2023]; or focus on the production subjects of the information epidemic and analyze the impact of multiple subjects on the evolution of public opinion[Liu, 2020]. The basis of multiple actors in the social media field is the “new news ecosystem” composed of existing professional media, institutional media, self-media and platform media. In this system, multiple actors participate together and diverse news practices coexist. Liu [2020] pointed out that during the epidemic, the voices of ordinary users converged, creating an open information flow, breaking the structured news field controlled by institutional media outlets, and forming an information environment where “the whole world is talking” and multiple voices collide. Wang and Wu [2023] also pointed out that internet celebrity intellectuals, as the “middle level”, play an important role in helping nationalism ferment on social media.

In the context of diverse public information dissemination discourse subjects, intertwined rationality and emotion in discourse texts, and complex topics, it is worth exploring the analysis of diverse actors represented by different types of media, social media opinion leaders, and official government accounts in the infodemic. Existing examinations of this media information dissemination still remain at the theoretical level of agenda setting, indicating the connection between party media, market-oriented media, social media opinion leaders and the public agenda[Han et al., 2021], but for more detailed issue competition relationships and actors Influence logic and causal relationship inference are relatively lacking in detailed analysis and in-depth analysis[Wang et al., 2023].

Therefore, the research question of this article is: How do multiple actors on social media (e.g., media, opinion leaders, official government accounts) use their network structure to influence the evolution of public discussion networks during the COVID-19 epidemic? This study attempts to use actor network theory to treat the social media public communication network shaped by the connections and interactions of multiple actors as a “black box” for analysis. Drawing on the “network amplification” and agenda-setting perspectives proposed by Yini Zhang, we use computational social science methods and machine learning and other technologies to analyze the role of different actors in this process of social shaping and network generation.

2 Literature Review

2.1 Actor Network Theory

“Actor-Network-Theory” (ANT) originated from the Science and Technology Studies (STS) tradition in the 1980s. Its basic starting point is any given social “reality” is all products of the interaction of a series of complex relationships (ANT calls them “networks”). In these complex relationship networks, there are both human actors in the traditional sociological sense and all nonhuman actors (such as technology and artifacts). When actors in a network (including human actors and non-human actors) form an “association”, ANT believes that a given social field can be configured [Latour, 1987]. Understanding the social shaping of technology is to examine the dynamic relationships between technology and its networks of practices. Looking further, the starting point of ANT is how technological objects themselves construct social links: First, there is no a priori “essential determination” for technological objects (“non-human beings”). The understanding of technology must be placed in the “network” in which the technical object itself is located, that is, observing its practical construction process in the network; secondly, the prerequisite for a network to be constructed is the existence of “coalition” composed of various actors (including both human actors and non-human actors, such as “technical objects”). When the will between actors is stabilized through transformation, ANT believes that a network becomes a “black-box”; again, the power of each actor in the network is realized through “translation”. What is accomplished is to create differences, and the wishes of other actors are presented through “transformation” rather than “representation”; finally, the task of social science research is to open the “black box” and show the connections in the established network of actors and deployment, that is, the “action trajectories” of different actors [Dai, 2019].

The researcher first opened the door for ANT to enter journalism research in the study of news innovation. Subsequent research went beyond the aspect of innovation and applied it more generally to the study of news production practice. Viewing news production as a network of actors has powerful epistemological consequences, a perspective that goes beyond the specific scope of journalistic innovation to discuss the interactions among multiple actors in everyday journalistic practices more generally. Existing research includes both overall research on the newsroom level and examination of some more specific production practices, which has formed a research hotspot focusing on “news networks” [Hongyi and Shiyu, 2021].

“Translation” is the core concept in ANT. Latour believes that any actor is a translator, and any information and conditions will be transformed through actors. The translation process includes four steps: problem definition, interest stimulation, member recruitment, and mobilization. Translation establishes connections between actors, who continuously switch issues and interests to gain mutual understanding. Problem definition mainly involves identifying core actors and common goals. Core actors are followers of network construction. By discussing the

obstacles and interests of each actor in this process, ANT believes that reasonable translation can connect actors together and build a relatively stable network.

Some scholars have used actor networks as a theoretical perspective and linked them with agenda setting to explain the risks and governance of public opinion networks in the context of the heterogeneous demands of multiple actors[Li and Shao, 2023]. The use of actor network theory to explain the evolution and generation of discussion networks in emergencies is considered to have good applicability. First of all, from the perspective of actors, the “human” factor mainly includes the following types of heterogeneous actors: ordinary netizens (individuals/communities), opinion leaders or Internet influencers, platform media and mainstream media, government and policy formulation; From the perspective of “non-human” factors, emotional contagion, technical tools, institutional disciplines, etc. can all be regarded as heterogeneous actors participating in the evolution of the network. ANT abandons preconceived judgments about subject participation in the action process, leaving ample space for the fluid and ever-changing identification of communication practitioners. Secondly, the translation process such as problem presentation, benefit conferment, recruitment and mobilization also presents the operating logic of heterogeneous actors in the actor network, providing an open map for further understanding of public discussion networks under emergencies[Sun et al., 2020].

2.2 Network Amplification

This study will use the perspectives of network amplification and agenda setting to understand the actions and influence logic of multiple actors in the COVID-19 discussion network.

In the field of communication, amplification can generally be defined as the process of increasing the exposure of a specific idea, object, or person and directing public attention. Mass media have long practiced amplification activities, including agenda setting or media hype[McCombs and Shaw, 1972, Vasterman, 2005]. Network amplification on social media is both a mechanism for attracting public attention and a process of information dissemination shaped by online social networks. Network amplification is defined as the sharing of similar or common information by like-minded elites, highlighting the interrelationships and interactions between elite information. Yini Zhang views social media amplification as a mechanism that affects public attention and a diffusion process that forms the structure of online social networks[?]. Elites on social media can drive information dissemination and direct public attention because they have large and loyal networks of followers. This layered and homogeneous network structure creates a network that can influence elite messaging and its effects as an environment.

In addition to being a means of gaining public attention, amplifying an idea, object, or person on social media is a huge cascade of information. The social network structure on social media is a semi-permanent conduit for the transmission of information from one node to another, which shapes the diffusion/amplification process and creates an online communication environment. A large body of literature suggests that both online social network structures and the information diffusion they support exhibit homogeneity and peer influence-driven homogeneous clustering patterns. On public social media such as Twitter, for example, the network of followers has a clear group structure, with like-minded nodes clustering together; information on social media typically spreads in echo chambers: ‘most of the links between successive sharing users are homogeneous’. Furthermore, social network structure and information diffusion are hierarchical, as evidenced by the different levels of connectivity of nodes and their different roles in information diffusion. While it is exciting that ordinary users drive the flow of information on social media in a multilevel diffusion chain, empirical evidence suggests that

most diffusion events, such as game sharing or news sharing, follow a shallow first-level diffusion process. Notably, information tends to flow directly from elites on social media (including media outlets, celebrities, organizations, and bloggers) to different opinion leaders and the general public. Evidence from empirical studies further suggests that the central node attracts most of the attention in the echo chamber of social media. More specifically, the information and communication cocoon of social media is always present, despite different elites advancing competing narratives. When elites in homogenous networks on social media amplify messages that are similar to each other, this reinforcement and repetition can create a "truth effect" that increases the familiarity, credibility and share ability of the message and further amplifies its impact.

Network amplification provides a new lens to scrutinize online actors on social media. Currently, research in this area focuses on content features such as themes and frames, moral and emotional language, group identity cues, and visual features in the social media posts of politicians and the media. Network amplification draws attention to the fact that homogenous and hierarchical social network structures on social media can contribute to what elites see and how they interact, and further influence what their followers are exposed to and how they are influenced through this exposure. In this regard, network amplification may better characterize and explain elite communication patterns and influence on social media than, for example, agenda-setting and framing theories.

It is noteworthy that classical agenda-setting theory has been more widely applied to empirical research. The Internet and social media have created new "hybrid media systems" that have expanded the number and types of actors that can influence political discourse and agendas. The rise of digital platforms therefore raises the question: who sets the agenda in this environment? [Gilardi et al., 2022] With the application of theories such as reverse agenda-setting and multi-level agenda-setting, a large number of empirical studies have discussed topics such as the agenda relationship and time lag between social bots, the media, and the public[?] , the association between the media, the public's attention, and the effectiveness of prevention and control of major public health emergencies , the construction of risk issues of rumors , and agenda-setting of cross-media disinformation[?Ge, 2023, Chen, 2023]. However, the agenda-setting relationships formed by multiple actors on Chinese social media platforms have not been scrutinized enough.

3 Research Design

3.1 Dataset

During the outbreak of New Crown Pneumonia, an infectious disease caused by a coronavirus discovered in December 2019, people's physical contact and interpersonal communication outside of their homes was greatly limited, relying primarily on digital devices to communicate and interact. This made social media platforms like Weibo an even more important source of information during the outbreak. Sina Weibo, a Chinese social media platform founded in 2009, has gained widespread popularity in terms of information dissemination and public discussion, especially as it allows for open, multi-party communication between strangers, thus enabling multi-level dissemination of information.

A researcher constructed a high-quality pool of active Weibo users containing 20 million active users from more than 250 million Weibo users, and then collected posts from all active users during the outbreak and filtered posts related to the Xinguan outbreak with 179 representative keywords, which were named Weibo-COV. This dataset (V2) covers the period from

December 1, 2019, to 2020 December 30 and filtered 65,175,112 tweets from 2,615,185,101 original tweets by keywords. The fields of posts in this dataset are fine-grained, including post level information, interaction information, location information, and retweet networks (see Table I)[Hu et al., 2020].

Table 1: Weibo-COV Data Structure

Field	Description
_id	Unique identifier for the tweet
crawl_time	The time the tweet was crawled, indicating the time researchers retrieved the specific Weibo post (Time Zone: GMT+8)
created_at	The posting time of the tweet
like_num	Number of likes for the tweet at the time of crawling
repost_num	Number of reposts for the tweet at the time of crawling
comment_num	Number of comments for the tweet at the time of crawling
content	Content of the tweet
origin_weibo	The _id of the original Weibo; this field is not empty only when the tweet is a repost
geo_info	Geographic latitude and longitude information; this field is not empty only when the tweet includes geographic information

3.2 Research Methods

This study will be divided into two parts.

3.2.1 Part I: Description and Inference of Actor Network Amplification

Influential user in the dataset were first sampled and different elite networks were identified. Then, their tweets related to the New Crown outbreak were collected and analyzed with descriptive and inferential statistics.

(1) Defining network elites The study uses Vintage Sparse Principal component analysis (VSP) to identify the different actor networks and especially the elite networks in the sample based on the followers. VSP is a spectral clustering technique that estimates the underlying factors in multivariate data. The basic version of the algorithm consists of two steps, low-rank singular value decomposition (SVD) and maximum variance rotation on singular vectors. Through specific quantitative summaries and qualitative judgments, each network generated corresponds to an elite type based on shared followers.

(2) For modeling the topics discussed

Using a list of keywords, Weibo-COV collected tens of millions of tweets about COVID-19. In order to answer the theme about elite amplification, the study will apply automated text techniques to process the data. In the preprocessing stage, the researchers will remove spaces, punctuation (except # and @), generic words and URLs. Next, after discarding all words appearing in less than 0.02% or more than 90% of the tweets and removing tweets containing less than two words, the plan is to construct a document term matrix for topic modeling. The VSP was then applied on the document term matrix to find 50 potential topics (after iterations to find more granular and clear topics). To validate the results of the VSP, latent Delicacy Allocation (LDA) topic modeling can be applied to the same document term matrix, where the model parameters are fitted by Gibbs sampling.

Regarding message similarity within each elite network, the researcher first quantifies the topic distribution of each member based on topic assignments (via VSP or LDA) and tests the correlation of the topic distributions of all possible account pairs in the network using the Pearson correlation coefficient. This study will use the Benjamini-Hochberg (BH) procedure for multiplicity correction. Finally, for each network, the percentage of account pairs that are significantly correlated ($p_{.05}$) in topic distribution is calculated. To assess reciprocal retweets,

two-sample t-tests were used to compare the frequency of in-network retweets for conservative media and influencer network accounts to the frequency of in-network retweets for liberal media and influencer accounts. All tweets posted by network members were also categorized into six types for descriptive analysis, including original tweets, self-retweets, retweets from other accounts within the network, retweets from other networks within the same category, retweets from different categories of networks in the sample, and retweets from accounts outside the sample. To analyze message similarity and mutual forwarding between networks, the researcher applies the same procedures as for message similarity and forwarding within networks.

3.2.2 Part II: Associations and Impacts between Different Types of Actors' Agendas

This study will introduce a vector autoregressive (VAR) model to analyze the relationship between different types of actor agendas. Vector autoregression constructs the model by taking each endogenous variable in the system as a function of the lagged values of all endogenous variables in the system, thus generalizing the univariate regression model to a "vector" autoregression model composed of multivariate time-series variables. The VAR model is often used to predict the interconnectedness of time-series and to analyze the dynamic impacts of stochastic perturbations on a system of variables. Dynamic shocks. Since the agenda setting of different types of actors in a social network is not independent, but rather a dynamic process of interaction, the VAR model can better explain the interactions between the three actors in the time dimension.

Granger causality tests whether variable Y is affected by the lagged variable of variable X. If variable Y is affected by the lagged variable of variable X, they are said to have granger causality. This method has been widely used in papers related to agenda-setting research and has a high degree of confidence in explaining the statistically significant causal relationship between two time series variables.

In addition, this study also constructs a timeline dataset of the COVID-19 epidemic event with the help of archived data, and through Granger causality test, considers the time point of the event as a potential predictor, and examines its predictive ability for other variables in the VAR model[Wikipedia, 2023].

4 Significance and shortcomings of the study

The reflection and re-examination of the New Crown Epidemic as a "critical moment in human history" needs to be discussed in detail in related studies[Chomsky and Žižek, 2020]. As an information dissemination event, the COVID-19 epidemic has greatly increased the public's reliance on relevant information, and this study utilizes Weibo-COV, a large-scale text mining data with a scale of 10 million, to analyze the actor network in the epidemic discussion, which overcomes the problems of under-representation of small samples and measurement bias of self-reporting questionnaires[Institute of Sociology, Chinese Academy of Social Sciences. Center for Social Psychology Research, 2020]. At the same time, the longitudinal modeling of the text data with the help of computational social science methods, combined with network amplification, agenda-setting and other perspectives to explore the influence relationship between different actor networks, providing the possibility of linking actor networks with classical communication theories and emerging computational technologies, enriching the research perspectives of information outbreaks, and the study has certain academic value. At the same

time, this study tries to give some realistic references to the information dissemination path and disinformation governance in public health events.

The shortcomings of this study include the following: First, social robots, as an influential force that cannot be ignored, have a significant impact on both issue construction and selective network amplification, which makes it difficult to effectively define the concept and scope of actors in this study[Duan et al., 2022, Shi and Chen, 2020]. Second, the role of emotional contagion as a unit of meaning in the evolution of influence networks can be further investigated.

References

- Y.-N. K. Chen. Agenda setting and covid-19: Misinformation from social network sites to traditional media. *Communication and Society*, 63:101–133, 2023.
- Noam Chomsky and Slavoj Žižek. Chomsky: In self-isolation, what kind of world do we want to live in? Online, 2020. URL https://www.thepaper.cn/newsDetail_forward_769405. *ThePaper*.
- Yuchen Dai. The old acquaintances with a reunion: Actor-networktheory and the future of media(tization) study. *Chinese Journal of Journalism Communication*, 41(4):68–88, 2019.
- Zening Duan, Jianing Li, Josephine Lukito, Kai-Cheng Yang, Fan Chen, Dhavan V Shah, and Sijia Yang. Algorithmic agents in the hybrid media system: Social bots, selective amplification, and partisan news about covid-19. *Human Communication Research*, 48(3):516–542, 2022.
- Mingsi Ge. Risk issue construction, amplification, and governance of rumors in public health events: Based on the analysis of early covid-19 online rumors. *Journalism University*, (8): 44–58+118, 2023.
- Fabrizio Gilardi, Theresa Gessler, Maël Kubli, and Stefan Müller. Social media and political agenda setting. *Political Communication*, 39(1):39–60, 2022.
- Xiaoning Han, Weiqi Guo, and Yaling Ba. A study of the multilevel agenda setting effect by multiple discourse subjects on epidemic issues. *Contemporary Communication*, (02):41–47, 2021.
- Bai Hongyi and Cao Shiyu. Remaking journalism theory? actor-network theory and sts turn in journalism studies. *Journalism Bimonthly*, 4(119):1–14, 2021.
- Yong Hu, Heyan Huang, Anfan Chen, and Xian-Ling Mao. Weibo-cov: A large-scale covid-19 social media dataset from weibo. *arXiv preprint arXiv:2005.09174*, 2020.
- Institute of Sociology, Chinese Academy of Social Sciences. Center for Social Psychology Research. The more effective the epidemic prevention and control, the more positive the social mentality: A survey on social mentality changes during the epidemic. *Beijing Daily*, 2020.
- Bruno Latour. How to follow scientists and engineers through society. *Cambridge, Mass*, 1987.
- Peihuan Li and Chunxia Shao. Counter-agenda setting in emergent events and its governance: From the perspective of actor-network theory. *Contemporary Communication*, 2:77–81, 2023.

- Peng Liu. "the whole world is talking": Research on user generated journalism cases via digital media during the covid-19 epidemic. *Chinese Journal of Journalism Communication*, 42(9):62–84, 2020.
- Maxwell E McCombs and Donald L Shaw. The agenda-setting function of mass media. *Public opinion quarterly*, 36(2):176–187, 1972.
- Wen Shi and Changfeng Chen. Issue salience and association building: The construction of covid-19 discussions by twitter social bots. *Modern Communication (Journal of Communication University of China)*, 42(10):50–57, 2020.
- Shaojing Sun, Fan Wang, Zhiyuan Liu, et al. Weibo discourse expression of diverse media in the context of the covid-19 pandemic. *Journalism University*, (3):16–30+117, 2020.
- Peter LM Vasterman. Media-hype: Self-reinforcing news waves, journalistic standards and the construction of social problems. *European Journal of Communication*, 20(4):508–530, 2005.
- H. Wang and L. Wu. Chinese intellectual influencers and popular nationalism during covid-19: A study based on jinri toutiao video blogs. *Communication and Society*, 65:63–92, 2023.
- Rong Wang, Alvin Zhou, and Tabitha H Kinneer. Moral framing and issue-based framing of# stopasianhate campaigns on twitter. *Chinese Journal of Communication*, pages 1–19, 2023.
- Rui Wang and Zhiruo Chang. Information overload and cognitive constraints: Examining factors that affect people’s judgementt and diffusion of true and false information during the covid-19 pandemic. *Journalism Communication*, 30(1):17–33+126, 2023.
- Wikipedia. Timeline of the covid-19 pandemic in 2019. Online, 2023. URL
...