COGITATIO

Media and Communication (ISSN: 2183–2439)
2022, Volume 10, Issue 2, Pages 66–80
https://doi.org/10.17645/mac.v10i2.5168

Article

Election Fraud and Misinformation on Twitter: Author, Cluster, and Message Antecedents

Ming Ming Chiu [1], Chong Hyun Park [2], Hyelim Lee [3], Yu Won Oh [4], and Jeong-Nam Kim [3,5,6,*]

[1] Analytics/Assessment Research Centre, The Education University of Hong Kong, Hong Kong
[2] School of Business, Sungkyunkwan University, Republic of Korea
[3] Gaylord College of Journalism and Mass Communication, University of Oklahoma, USA
[4] Department of Digital Media, Myongji University, Republic of Korea
[5] Debiasing and Lay Informatics, USA
[6] Data Institute for Societal Challenges, University of Oklahoma, USA

* Corresponding author (layinformatics@gmail.com)

Submitted: 6 December 2021 | Accepted: 14 March 2022 | Published: 29 April 2022

Abstract
This study determined the antecedents of diffusion scope (total audience), speed (number of adopters/time), and shape (broadcast vs. person-to-person transmission) for true vs. fake news about a falsely claimed stolen 2020 US Presidential election across clusters of users that responded to one another's tweets ("user clusters"). We examined 31,128 tweets with links to fake vs. true news by 20,179 users to identify 1,069 user clusters via clustering analysis. We tested whether attributes of authors (experience, followers, following, total tweets), time (date), or tweets (link to fake [vs. true] news, retweets) affected diffusion scope, speed, or shape, across user clusters via multilevel diffusion analysis. These tweets showed no overall diffusion pattern; instead, specific explanatory variables determined their scope, speed, and shape. Compared to true news tweets, fake news tweets started earlier and showed greater broadcast influence (greater diffusion speed), scope, and person-to-person influence. Authors with more experience and smaller user clusters both showed greater speed but less scope and less person-to-person influence. Likewise, later tweets showed slightly more broadcast influence, less scope, and more person-to-person influence. By contrast, users with more followers showed less broadcast influence but greater scope and slightly more person-to-person influence. These results highlight the earlier instances of fake news and the greater diffusion speed of fake news in smaller user clusters and by users with fewer followers, so they suggest that monitors can detect fake news earlier by focusing on earlier tweets, smaller user clusters, and users with fewer followers.

Keywords
diffusion; elections; fake news; situational theory of problem-solving; social networks

Issue
This article is part of the issue "Networks and Organizing Processes in Online Social Media" edited by Seungyoon Lee (Purdue University).

© 2022 by the author(s); licensee Cogitatio (Lisbon, Portugal). This article is licensed under a Creative Commons Attribution 4.0 International License (CC BY).

1. Introduction

Donald Trump and his followers falsely claimed that he won the 2020 US presidential election, sparking many of his supporters to repeat this fake news on social media (e.g., Twitter). Moreover, 88% of Trump supporters said that they would take action (e.g., protest; Pennycook & Rand, 2021), and thousands of them joined the Capitol Insurrection, resulting in five deaths and over 140 casualties (Guynn, 2021).

Malevolent authors intentionally write false information (*disinformation*) for ideology or profit (paid per viewer or ad-click; Braun & Eklund, 2019), but unwitting traffickers can further disseminate it (*misinformation*;

Hilary & Dumebi, 2021). Indeed, laypeople have gathered in social spaces to share thoughts, consume and react to them, seek cooperation, and mobilize others for over two millennia at the agora (Athens in Ancient Greece), dinner parties, coffee houses, salons, reading circles, and now social media (*publics*; Dewey & Rogers, 2012; Grunig & Kim, 2017). Social media (e.g., Twitter) accelerates this process, enabling people to share ideas much faster than before, with many more people in larger networks with weak ties (Fuchs, 2014). Especially concerning, fake news can spread faster than true news via social media (Vosoughi et al., 2018)—and many people rely on social media for accurate news (Walker & Matsa, 2021), with sometimes devastating consequences, such as the Capitol Insurrection.

Like opinion leaders (e.g., politicians, celebrities) in traditional realms (Rogers & Cartano, 1962), online influencers can quickly broadcast information or (relatively slowly) cascade information person-to-person, potentially influencing audience activities and opinions (Mittal & Bhatia, 2019; Rossman et al., 2008). However, some discussions without influencers (low activity, few followers) still virally spread ideas (Rosenthal, 2014). These diffusion differences suggest differences across groups of users who respond to one another's messages (*user cluster*).

No published study has determined the antecedents of diffusion scope (*maximum adopters* or N_{max}), speed (adopters over time or *adoption rate*), and shape (broadcast vs. person-to-person; or *external influence vs. internal influence*; Rossman et al., 2008) for true vs. fake news about a topic across different user clusters. Hence, we do so for 31,128 tweets with links to fake vs. true news about a stolen 2020 US presidential election shared by 20,179 users in 1,069 user clusters via multilevel diffusion analysis (MDA; Rossman et al., 2008). Specifically, we test whether attributes of authors (experience, followers, following, total tweets), time (date), or tweets (link to fake [vs. true] news, retweets) affect diffusion scope, speed, or shape.

2. Theoretical Framework of Diffusion Antecedents

First, we define diffusion scope, speed, and shapes (broadcast vs. person-to-person). Then, grounded in the situational theory of problem-solving (STOPS; Kim & Grunig, 2011), we examine motives for seeking, selecting, and sharing/forwarding a tweet, especially of fake vs. true news regarding a stolen 2020 US presidential election.

2.1. Diffusion

After a person invents an idea, product, or procedure, it may or may not spread to more users within a population (*diffusion*; Rossman et al., 2008). Diffusion can vary in scope, speed, and shape. The total number of users is diffusion *scope*. How quickly more people become users (the number of users divided by time) is diffusion *speed*.

Diffusion shapes differ in their extents of broadcast and person-to-person transmission. Many users might quickly engage with a tweet, with fewer additional people doing so over time, yielding a logarithmic-like cumulative distribution curve that rises quickly and then tapers off (*broadcast/external influence;* Rossman et al., 2008; see Figure A1 of the Supplementary File). Tweets by an influential person or institution typically show broadcast diffusion (e.g., Donald Trump, BBC news, etc.). By contrast, few initial adherents might engage with an attractive tweet by a low influence person, but as they proselytize it to others, its influence accelerates until the message saturates its target population, resulting in a cumulative distribution S-curve (*person-to-person/internal influence*; Rossman et al., 2008; see also Figure A2 of the Supplementary File).

2.2. Situational Theory of Problem-Solving

According to the STOPS (Kim & Grunig, 2011), humans ignore/discard most information, so they attend to and share only *relevant* novel information with their audience (Kim & Krishna, 2014). Their subjectivity in judging the relevance and integrity of true vs. false news hinders accurate detection. Even with training, most humans cannot identify fake news (Lutzke et al., 2019), especially as alternative media (e.g., *209 Times*) can publish 99% real news mixed with 1% fake news (Shaw & Natisse, 2021). People with less online media literacy are even less likely to accurately identify true vs. fake news (e.g., Brashier & Schacter, 2020).

2.2.1. Cognitive Progression vs. Cognitive Retrogression

When facing a problem, a person can follow a scientific method: start with a minor premise and gather information/evidence to construct/determine a suitable solution/conclusion (evidence → conclusion: *cognitive progression*; Kim & Grunig, 2011). Or a person can begin with a solution/conclusion (belief) and gather confirming information/evidence (conclusion → evidence: *cognitive retrogression*; see Kim & Grunig, 2021; for *confirmation bias* see Knobloch-Westerwick et al., 2020). As cognitive retrogression includes both true and false evidence that mutually reinforce each other, the true parts help shield the false parts, thereby strengthening its overall credibility.

When a problem solver *improvises* conclusions (e.g., wishful or willful end state) or *activates recyclable* conclusions, facts, or solutions, cognitive retrogression is more likely than cognitive progression. Cognitive retrogression is the default cognitive mode in problem-solving (Fiske & Taylor, 1991; Kim & Grunig, 2021; Oakhill & Johnson-Laird, 1985). Cognitive retrogression in problem-solving explains why people continue to accumulate evidence that supports their beliefs (e.g., stolen election) and resist evidence that violates them (*cognitive arrest*; Kim & Grunig, 2011). So, cognitive arrest drives fake news (e.g., cognitively arrested issue publics like QAnon or anti-vaxxers) and obstructs the cognitive progression of active publics.

2.2.2. Information Behaviors

Consider a Twitter user reading a tweet saying that Martians have landed in Tokyo and were chatting with his mom. Surprised and concerned about his mom, he imagines her deluged with tweets, forwards it to his siblings, and calls her—eventually finding that her friend wrote it to get her children to call her. According to the STOPS (Kim & Grunig, 2011; Kim et al., 2010), the user recognized a credible discrepancy between the tweet information and his experience/expectation (people had not previously tweeted that Martians chatted with his mom, *problem recognition*), his relation to this discrepancy (mom, *involvement recognition*), and few obstacles to addressing it (potential deluge of tweets, *constraint recognition*). All of these factors increased his epistemic motivation to increase problem-related communicative actions to seek and share information (call mom, forward to siblings; Kim et al., 2010).

2.2.2.1. Problem Salience: Fake News Vs. True News

STOPS (Kim & Grunig, 2011; Kim et al., 2010) suggests three motives for seeking, selecting, and sharing/forwarding a tweet: problem salience, relationship, and scale. When a person perceives a greater sense of discrepancy between the current information and past experiences/future expectations (*problem salience*, cf. indeterminate situation; Dewey, 1910), this information might have a greater impact (whether potential benefit or threat), so they are more likely to disseminate this information to their user cluster who might also share the benefit or help address a threat.

As fake news typically differ more than true news from humans' experiences, people are more likely to share/forward fake news than true news to more people and do so more quickly via both broadcast and person-to-person diffusion. For example, as food poisoning in popular food franchises can harm a person's health, people are more likely to share such news with others (Lee et al., 2021). Indeed, fake news spreads to exponentially more people within a user cluster compared to true news (Abilov et al., 2021; Bodaghi & Oliveira, 2022; Bovet & Makse, 2019). Hence, we propose hypothesis H1:

> H1: A tweet linked to a fake news story (rather than a true one) ignites *more* user cluster tweets on this topic (total users).

Compared to true news, such fake news (e.g., food poisoning) often elicits greater urgency, as indicated by more replies with surprise, fear, or disgust. Indeed, false information can spread 10 times faster than true information (Vosoughi et al., 2018). Also, a small number of influencers in a network often spread most of the fake news (Grinberg et al., 2019; Sharma et al., 2020). Together, these studies suggest that fake news diffuse faster via broadcast transmission, compared to true news.

> H2: A tweet linked to a fake news story (rather than a true one) *quickly* ignites tweets on this topic within its user cluster (broadcast transmission).

In addition to immediate broadcast action on fake news, we propose that users are more likely to share the often-alarming fake news with family members, friends, and acquaintances (person-to-person transmission).

> H3: A tweet linked to a fake news story (rather than a true one) elicits more person-to-person sharing.

2.2.2.2. Relationship

At the cluster level, the number of people in a user cluster (size) can also affect diffusion scope, speed, and shape. As larger user clusters have more people who respond to one another's messages, more people are likely to engage with a specific tweet.

> H4: A tweet in a larger user cluster ignites *more* tweets on this topic within its user cluster (total users).

In smaller user clusters, people have closer relationships (e.g., immediate family members), so they often engage with one another's concerns quickly (Kim & Grunig, 2011). In smaller user clusters, members can devote more time and attention to each member (vs. attention dilution in larger user clusters) and care more about each person. Thus, they are more likely to engage with one another's concerns and do so quickly.

> H5: A tweet in a smaller user cluster *quickly* ignites tweets on this topic within its user cluster (broadcast).

By contrast, people in larger user clusters are less likely to respond immediately. Instead, we propose that as more people in a large user cluster engage with a tweet, person-to-person engagement increases.

> H6: A tweet in a larger user cluster elicits more person-to-person sharing.

2.2.2.3. Scale

At the user-level, an author with more Twitter followers (*scale*) has greater motivation to send them tweets to maintain their followers (Kim et al., 2010). Given the larger number of followers compared to other authors, more of them are likely to engage.

H7: A tweet by an author with more followers ignites *more* tweets on this topic within its user cluster.

However, these many tweets might dilute the value of each tweet, so any specific tweet might be less likely to be relevant to each person, resulting in less immediate engagement.

H8: A tweet by an author with more followers slowly ignites tweets on this topic within its user cluster.

Instead, followers are more likely to wait for others to engage before they do. As more people engage with a tweet, their participation suggests that the tweet has greater value, which in turn elicits greater engagement from more user cluster members.

H9: A tweet by an author with more followers elicits more person-to-person sharing.

2.2.3. Other Explanatory Variables

As omitting significant explanatory variables from a statistical model can cause *omitted variable bias* (Cinelli & Hazlett, 2019), we also model these available variables: followers, following, tweets, author experience, total date, and retweets. As noted above, users with more followers often send out more tweets, so these variables are likely highly correlated. Users with more experience (days since user account creation date) might have more status, credibility, and authority, which suggests more total engagement, faster broadcast diffusion, and less person-to-person diffusion (Chiu, 2008).

H10: A tweet by an author with more experience ignites *more* tweets on this topic within its user cluster.

H11: A tweet by an author with more experience *quickly* ignites tweets on this topic within its user cluster.

H12: A tweet by an author with more experience elicits less person-to-person sharing.

As the value of news degrades over time, late tweets on later days might attract less engagement, with unclear effects on diffusion speed or shape (broadcast or person-to-person).

H13: A tweet at a later date ignites *fewer* tweets on this topic within its user cluster.

As retweets, replies, and new tweets on a topic are possible substitutes for one another, the effect of total retweets is unclear. See the summary of hypotheses in Table 1.

3. Method

To address our research questions, we identified tweets regarding the election, downloaded tweets linked to them, identified subsequent tweets that engaged with each original tweet within user clusters and analysed their diffusion patterns.

3.1. Data

To create the Twitter election fraud data set, we first identified true vs. fake news articles regarding election fraud in the 2020 US Presidential Election from October 24 to December 18, 2020. We first selected the news items identified as false or mostly false on *Snopes* (https://www.snopes.com), which included the archived links of fake news sources. Then, we identified true news articles from mainstream news websites. These results yielded 48 related news articles from news media such as *The New York Times*, *AP News*, *Reuter*, and *USA Today* (true news) and 43 from *Snopes* (identified fake news). We downloaded tweets during October 24 to December 18, 2020, with their URLs (linked to these news articles) and their replies, which capture interactions within user clusters. For example, each tweet contains the ID information of users who have retweeted. Through this process, we collected 3,340 tweets about true news articles on election fraud and 3,410 tweets about fake news articles on the same topic.

Table 1. Diffusion hypotheses (all supported except the ~~strikethrough~~ one).

		Expected Outcome		
Theory	Explanatory Variable	Scope	Speed/Broadcast Shape	Person-to-Person Shape
Problem salience (H1, H2, and H3)	Fake news	More	Faster	More
Relationship (H4, H5, and H6)	Larger user cluster	More	Faster	More
Scale (H7, H8, and H9)	Author has more followers	More	Slower	More
Author experience (H10, H11, and H12)	More experience	~~More~~	Faster	Less
Date (H13)	Later date	Fewer		

Notes: The results supported all hypotheses except for greater author experience yielding more scope; we have no hypotheses regarding Date's effects on diffusion speed or person-to-person shape.

3.1.1. User Cluster Detection

For this article, we broadly operationalize a user cluster as users who interact on a specific issue on a social media network (Leicht & Newman, 2008). So, we specify how we used *clustering* to identify each user cluster that interacts and reacts to fake (or true) news on the 2020 election fraud.

3.1.1.1. Transform Data to Determine User Clusters

First, we transform Twitter data into a suitable format to represent network structures (see Table 2). The "tweet_id" is a unique value identifying a tweet. Similarly, "user," "text," and "retweeted_user" indicate its author, its text message, and a user who retweeted it, respectively. Also, an author refers to a specific user in a message via the @ symbol in the "text" field. These data also include dates and time.

3.1.1.2. Construct the Weighted, Directed Network

We divided tweet interactions into three categories: mention, retweet, and self (see Table 3). A tweet can name a specific user in its text via "@" (*mention*). Also, a user can retweet a tweet. A user can respond to one's prior tweet (*self*). As this study examines diffusion across people, we excluded self-tweets. Table 4 shows the number of interactions between users (excluding self-tweets) as the sum of mentions and retweets. The above data transformation enables identification of weighted, directed social networks of user *nodes*, and interaction *edges* (Fortunato, 2010), as shown in Figure A4 of the Supplementary File. Each node represents a user, and arrows indicate source-to-target relations, with thicker arrows reflecting more interactions.

3.1.1.3. Clustering Analysis

We detected broadly defined user clusters by decomposing them into smaller subsets of interrelated users (Fortunato & Castellano, 2007) via their network structure information (see review by Azaouzi et al., 2019; some studies use community quality indicators, but we lack this information). Node i is in our weighted, directed user cluster c_i, and the strength of edges within a user cluster compared to other edges (*modularity*; Arenas et al., 2007) is:

$$Q = \frac{1}{2m} \sum_{\forall i,j} \left(A_{ij} - \frac{k_i^{\text{out}} k_j^{\text{in}}}{2m} \right) \delta \left(c_i, c_j \right) \tag{1}$$

Table 2. Sample Twitter data.

Tweet_id	User_id	Text	Retweeted_user
100	user1	to @user2 and @user3	user3, user5
101	user6	no mention	None
102	user1	to @user3	None
103	user9	no mention	user10

Table 3. Interactions between users.

Source	Target	Type
user1	user2	mention
user1	user3	mention
user1	user3	retweet
user1	user5	retweet
user6	user6	self
user1	user2	mention
user9	user10	retweet

Table 4. Merged edges for each user relationship.

Source	Target	Count
user1	user2	2
user1	user3	2
user1	user5	1
user9	user10	1

The weight of the edges between i and j is A_{ij}. The total weight from node i is $k_i^{out} = \sum_{\forall j} A_{ij}$. The total weight to node j is $k_j^{in} = \sum_{\forall i} A_{ij}$. For nodes i and j within a user cluster, the indicator function $\delta\left(c_i, c_j\right)$ has value 1; otherwise, 0. The total strength is $m = \frac{1}{2}\sum_{\forall i,j} A_{ij}$. When the actual edges in a user cluster exceed their expected number of randomly distributed edges (see Equation 1), modularity is positive.

Optimizing clustering by maximizing modularity detects user clusters (Srinivas & Rajendran, 2019). As exact optimization of larger networks requires exponentially more time, we use Blondel et al.'s (2008) heuristic via *Gephi* software (Cherven, 2015; see Figure 1). Users 1, 2, 3, and 5 are in one group, and users 9 and 10 are in another group.

$$A_{12} = 2, A_{13} = 2, A_{15} = 1, A_{910} = 1,$$
$$m = \frac{1}{2}\left(A_{12} + A_{13} + A_{15} + A_{910}\right) = 3,$$

$$k_1^{out} = 5, k_2^{out} = k_3^{out} = 2, k_5^{out} = k_9^{in} = k_{10}^{out} = 1,$$

So, optimal modularity Q^* is 0.278.

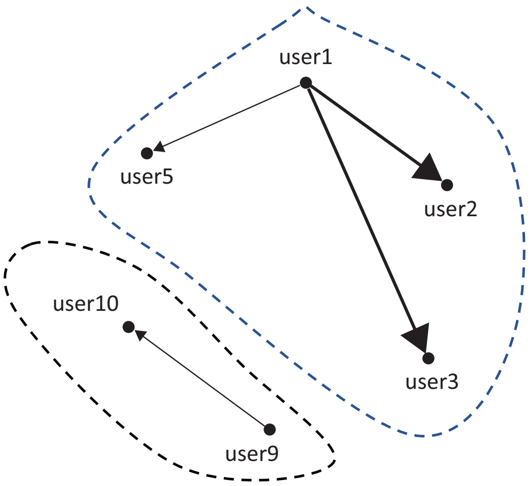

Figure 1. Support for and institutionalization of direct democracy. Source: Geissel (2016).

3.1.1.4. Online User Clusters

In tweets about true news articles, 12,241 users formed 655 user clusters. In the tweets about fake news articles, 7,938 users formed 414 user clusters. See visualization of the interactions among users in Figure 2 for a view of the overall network structure. Dots represent users, and those in the same cluster have the same color. These clustering results identify the online community of each user.

If a tweet was only visible on two days during this period, there are two days in which others can respond to it (two *tweet-days*). For each subsequent day (1–55) of each of the 6,750 initial tweets (resulting in 235,088 tweet-days), we counted the daily number of references to it.

3.1.2. Statistical Power

Statistical power differs across levels. For $\alpha = 0.05$ and a small effect size of 0.1, statistical power is 0.91 for 1,096 user clusters, and exceeds 0.99 for 20,179 users, 31,128 tweets, 6,750 initial tweets, and 235,088 tweet-days (Konstantopoulos, 2008).

3.2. Variables

Cumulative tweets is the number of tweets engaging with an initial tweet, inclusive, to date. We also computed its squared term *cumulative tweets*^{2}. Both are needed for a diffusion analysis. Author variables include author experience, total tweets, followers, and following. *Author experience* is computed as the number of days between the author creation date on Twitter and the date of the last tweet in the dataset (December 19, 2020). As total tweets, followers, and following have non-normal distributions, we computed log (total tweets + 1), log (followers + 1), and log (following + 1). The followers and following reflect the size of the user cluster. *Date* is the number of days from the first tweet in the data set (first date = 1). *Fake* indicates a tweet about fake (vs. true) news, in which the original tweet in this thread linked to a news article identified as fake on *Snopes*. *Retweets* is the number of retweets of the first tweet in a thread.

3.3. Multilevel Diffusion Analysis

To address our research questions with these data, we integrated *diffusion analysis* and *multilevel analysis* into MDA (Rossman et al., 2008). Diffusion analysis models the scope, speed, and shape (broadcast vs. person-to-person) of the dissemination of a tweet (Franz & Nunn, 2010). As tweets in the same user cluster likely resemble one another more than those in different user clusters (*nested data*), a traditional diffusion analysis underestimates the *standard errors*, so we use a multilevel analysis (Hox et al., 2017), specifically an MDA (Rossman et al., 2008).

3.3.1. Explanatory Model

MDA simultaneously models (a) diffusion of multiple tweets within multiple user clusters, (b) the expected total diffusion of a tweet (total adopters), (c) the extent of its broadcast transmission (*external influence*) vs. its person-to-person transmission (*internal influence*), and (d) explanatory variables at user cluster-, tweet-, and time-levels. We begin with a *variance components* model.

$$N_{k(t+1)i} - N_{kti} = A_k + e_{kti} + f_{ki} + g_k \qquad (2)$$

N_{kti} and $N_{k(t+1)i}$ are vectors of the numbers of members in user cluster k that have sent tweet i by day t and day $t + 1$, respectively, so the difference $N_{k(t+1)i} - N_{kti}$ is the number of new tweets sent on day $t + 1$. The grand mean

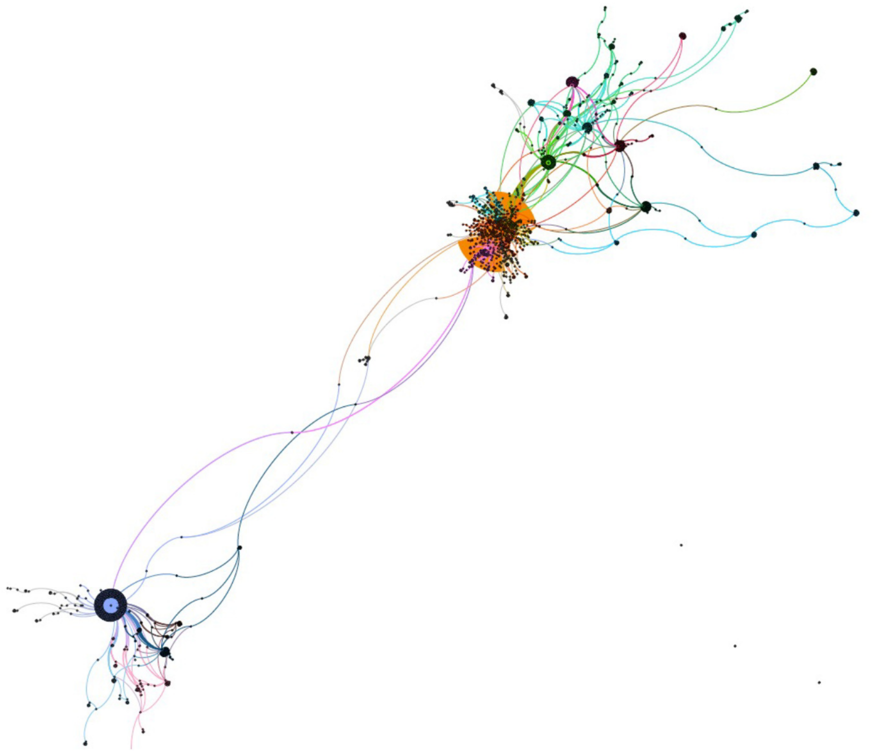

Figure 2. Identified user clusters with fake tweets (414 user clusters).

intercept is A_k with unexplained components (*residuals*) at the time-, tweet-, and user cluster-levels: e_{kti}, f_{ki}, and g_k. To model the diffusion shape (broadcast vs. person-to-person), we add the linear term N_{kti} and its quadratic term N_{kti}^2 in the following equation:

$$N_{k(t+1)i} - N_{kti} = \left(A_k + e_{kti} + f_{ki} + g_k\right) + \left(B_{k1i}\right) N_{kti} + \left(C_{k2i}\right) N_{kti}^2 \quad (3)$$

B_{k1i} and C_{k2i} are regression coefficients of N_{kti} and N_{kti}^2, respectively. The internal influence (b) in user cluster k of tweet i is as follows:

$$b_{ki} = -C_{k2i} \quad (4)$$

We compute the expected total diffusion $\left(N_{\max}\right)$ in user cluster k of a tweet i as follows:

$$N_{\max,ki} = -B_{k1i}/2C_{k2i} \pm \left(B_{k1i}^2 - 4 \times A_k \times C_{k2i}\right)^{0.5} /2C_{k2i} \quad (5)$$

We compute the external influence (a) in user cluster k of tweet i as follows:

$$a_{ki} = \left(A_k \times 2 \times C_{k2i}\right) / \left(-B_{k1i} \pm \left[B_{k1i}^2 - 4 \times A_k \times C_{k2i}\right]^{0.5}\right) \quad (6)$$

Next, we add explanatory variables:

$$\begin{aligned} N_{k(t+1)i} - N_{kti} = {} & \left(A_k + e_{kti} + f_{ki} + g_k + \pi_w \mathrm{AUTHOR}_k \right. \\ & \left. + \phi_{kzi}\mathrm{TIME}_{kti} + \alpha_{kx}\mathrm{TWEET}_{ki}\right) + \left(B_{k1i} + \theta_w \mathrm{AUTHOR}_k \right. \\ & \left. + \kappa_{kzi}\mathrm{TIME}_{kti} + \beta_{kx}\mathrm{TWEET}_{ki}\right) N_{kti} + \left(C_{k2i} + \rho_w \mathrm{AUTHOR}_k \right. \\ & \left. + \lambda_{kz}\mathrm{TIME}_{kti} + \gamma_{kx}\mathrm{TWEET}_{ki}\right) N_{kti}^2 \end{aligned} \quad (7)$$

AUTHOR_k, TIME_{kti}, and TWEET_{ki} are vectors of explanatory variables that might influence the diffusion in user cluster k of tweet i, with regression coefficients: π_w, α_{kx}, ϕ_{kzi}, θ_w, β_{kx}, κ_{kzi}, ρ_w, γ_{kx}, and λ_{kz}. AUTHOR captures the characteristics of the author of the initial tweet on this topic (in this case, stolen US presidential election in 2020): twitter experience (days), log (followers + 1), log (following + 1), and log (total tweets + 1). TIME is the date of the initial tweet of this topic. TWEET includes the following attributes: link to a fake news article (vs. true one), and log (retweets + 1). To test the robustness of our results, we repeated the above analyses on the following subsets: (a) user clusters with at least two tweets, (b) user clusters with at least 50 members, and (c) user clusters with at least 100 members.

4. Results

These 20,179 users in 1,069 user clusters sent 31,128 total tweets (see Table 5). There were 6,750 initial tweets (3,340 linked to fake news, 3,410 linked to true news) that ignited conversations. The mean length of these conversations lasted 35 days (6,750 tweets × ~35 days ≈ 235,088 tweet-days). For most days in these user clusters, there were no additional tweets on this stolen election topic (M = 0.029), and the number of cumulative tweets on this topic to date was small (M = 1.075). The author of the first tweet in a user cluster about this topic averaged 6.8 years (M = 2,489 days) of experience on Twitter, 32,595 total tweets, 5,713 followers, and 2,078 followings. A tweet was retweeted slightly more than

Table 5. Summary statistics (N = 235,088 days across tweets or *tweet-days*).

Variable	Mean	SD	Min	Median	Max
Additional tweets today	0.029	0.674	0	0	185
Cumulative tweets today	1.075	9.693	0	1	798
Author days of experience	2,489.223	1489.202	22	2,763	5,256
Total tweets	32,595.298	65,459.397	1	11,427	1,040,402
Followers	5,712.902	45,571.701	0	468	2,101,420
Following	2,077.980	6,157.396	0	757	195,749
Log (total tweets + 1)	9.196	1.760	0.693	9	13.855
Log (followers + 1)	6.144	2.214	0	6	14.558
Log (following + 1)	6.526	1.606	0	7	12.185
Date [a]	39.756	11.102	1	19	55
Fake	0.588	0.492	0	1	1
Retweets	1.157	14.101	0	0	610
Log (retweets + 1)	0.188	0.556	0	0	6.415
Isolated tweet	0.745	0.436	0	1	1

Notes: 31,128 total tweets with 6,750 initial tweets (3,340 fake, 3410 true) across ~35 days in 1,069 user clusters with 20,179 users (6,750 tweets × ~35 days ≈ 235,088 tweet-days); [a] the first possible date was October 24, 2020 (October 24 = 1; October 25 = 2; etc.).

once on average (M = 1.157). Nearly 60% of these tweets were linked to fake news articles. On any given day, over 25% of these tweets had at least one reply or retweet.

Users with more experience tweeted earlier than other users and had somewhat more tweets, followers and following (correlations [*r*] = 0.27, 0.31, 0.38, and 0.32 respectively; see correlation matrix in Table 6), showing more influence than users with less experience. Users with many followers often followed many others ($r = 0.67$) and wrote many tweets ($r = 0.77$). Initial tweets about fake news were sent earlier than those with true news ($r = 0.33$); otherwise, no other attributes were linked to fake news.

4.1. Explanatory Model

Most of the differences in diffusion of tweets varied across dates within a user cluster (89%), with significant differences across user clusters (11%; see Table 7). The multilevel diffusion regression showed that both *cumulative tweets* and its squared term *cumulative tweets*^{2} were significantly linked to additional tweets today (on the topic of the stolen US presidential election 2020; see Table 7). Also, nearly all their interactions with the explanatory variables—author days of experience, log (followers + 1), log (following + 1), and log (total tweets + 1), date, fake, log (retweets + 1)—were significant. All interactions of fake news with log (followers + 1) and log (following + 1) were not significant.

Thus, we enter these significant regression coefficients into our above diffusion equations to yield the results shown in Table 8. These results project an overall mean of 233 tweets for each original tweet, indicating that 233 subsequent tweets mentioned the original message author, retweeted, or replied to each original message, on average. Both broadcast and person-to-person diffusion were small overall, with much larger impacts of other explanatory variables on both types of diffusion. Together, they indicate that these tweets have no overall, common diffusion pattern. Instead, author, date, and tweet differences determine diffusion scope, speed, and shape (broadcast or person-to-person).

4.1.1. Scope

Author, date, and tweet attributes were linked to the expected total tweets on the topic of a stolen 2020 US presidential election. Authors with more experience ignited far fewer expected total tweets on this topic in their user cluster (−0.205 per day of Twitter experience, 75 fewer tweets per year of Twitter experience), rejecting hypothesis H10 (see Tables 1 and 8). By contrast, authors with more tweets, more followers, or following more users ignited slightly more expected total tweets on this topic in their user cluster (0.829, 0.068, or 0.726, respectively), supporting H4 and H7. Tweets igniting this topic in a user cluster on later dates yielded fewer expected total tweets (−0.222 per day, ~ −7 per month), supporting H13. Tweets with links to fake news rather than true news yielded over 32 more expected total tweets, supporting H1. Additional retweets of the original tweet on this topic in a conversation yielded slightly fewer expected total tweets (−0.011).

Table 6. Correlation-variance–covariance matrix of key variables in the lower left, diagonal, and upper right matrices.

	Variable	1	2	3	4	5	6	7	8	9	10
1	Number of tweets (t + 1)	**0.454**	1.405	664.110	1.945	0.013	0.012	0.016	−0.003	−0.035	0.005
2	Cumulative tweets	0.215	**93.951**	55,602.938	39.201	0.410	0.380	0.530	−0.038	−3.913	0.169
3	Cumulative tweets^2	0.158	0.922	**38,701.959**	41,245.653	178.170	157.846	190.561	−41.264	−121.612	−9.179
4	Days of experience	0.002	0.003	0.004	**2,217.713**	806.838	772.117	1,265.506	−196.298	986.039	87.345
5	Log (total tweets)	0.011	0.024	0.016	0.308	**3.097**	1.641	2.616	−0.094	0.144	0.180
6	Log (following)	0.011	0.024	0.016	0.323	0.580	**2.580**	2.737	−0.121	0.177	0.199
7	Log (followers)	0.011	0.025	0.014	0.384	0.671	0.770	**4.904**	−0.242	0.511	0.551
8	First date	−0.008	−0.008	−0.013	−0.268	−0.109	−0.153	−0.222	**0.242**	−1.827	−0.030
9	Fake	−0.005	−0.036	−0.002	0.060	0.007	0.010	0.021	−0.334	**123.261**	−0.083
10	Log (retweets)	0.012	0.031	−0.003	0.106	0.184	0.223	0.448	−0.109	−0.013	**0.309**

Table 7. MDA results (with 1,000 multiplier).

	Regressions predicting additional tweets today							
Explanatory variable	Model 1		Model 2		Model 3		Model 4	
Cumulative tweets	27.940	***	27.970	***	−80.910	***	−526.100	***
	(0.604)		(0.605)		(7.985)		(13.130)	
Cumulative tweets^2	−0.060	***	−0.060	***	−0.191	***	69.290	***
	(0.001)		(0.001)		(0.001)		(0.646)	
Author days of experience			−0.002		−0.019	***	0.024	***
			(0.004)		(0.004)		(0.004)	
Log (followers + 1)			0.283		4.202		−9.987	*
			(4.431)		(5.255)		(4.475)	
Log (following + 1)			2.282		−9.231		−28.000	***
			(4.914)		(5.830)		(4.978)	
Log (total tweets + 1)			2.765		−12.150	**	−19.690	***
			(3.847)		(4.556)		(3.879)	
Date			0.336		−0.704		−0.329	
			(0.469)		(0.556)		(0.471)	
Fake			−14.060		96.570	***	21.960	*
			(11.090)		(13.160)		(11.090)	
Log (retweets + 1)			−6.855		69.730	***	92.010	***
			(10.230)		(12.190)		(10.370)	
Author days of experience × Cumulative tweets					0.021	***	−0.034	***
					(0.001)		(0.001)	
Log (total tweets + 1) × Cumulative tweets					18.340	***	29.710	***
					(0.498)		(1.170)	
Log (followers + 1) × Cumulative tweets					−12.450	***	7.599	***
					(0.576)		(1.190)	
Log (following + 1) × Cumulative tweets					17.680	***	39.740	***
					(0.628)		(1.322)	
Date × Cumulative tweets					−2.612	***	2.766	***
					(0.160)		(0.192)	
Fake × Cumulative tweets					−149.800	***	4.531	
					(2.861)		(3.591)	
Log (retweets + 1) × Cumulative tweets					−18.210	***	−51.570	***
					(0.590)		(1.341)	
Author days of experience × Cumulative tweets^2							0.001	***
							(0.000)	
Log (total tweets + 1) × Cumulative tweets^2							0.192	***
							(0.014)	
Log (followers + 1) × Cumulative tweets^2							−0.307	***
							(0.011)	
Log (following + 1) × Cumulative tweets^2							(0.356)	***
							(0.011)	
Date × Cumulative tweets^2							(1.667)	***
							(0.015)	
Fake × Cumulative tweets^2							(29.260)	***
							(0.261)	
Log (retweets + 1) × Cumulative tweets^2							0.381	***
							(0.018)	
Variance at each level								
User cluster (11%)	0.000		0.000		0.000		0.000	
Date (89%)	0.037		0.037		0.117		0.180	
Total variance explained	0.033		0.033		0.104		0.160	

Notes: To aid the reading of small values, all regression coefficients and standard errors were multiplied by 1,000; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 8. Diffusion parameter results.

	Expected Total Tweets $(N_{\max})$	Broadcast × 100 (a, external) [a]	Person-to-Person × 100 (b, internal) [a]
Overall	232.807	0.001	0.006
Author experience (days)	−0.205	0.253	−0.018
Log (total tweets + 1)	0.829	−0.551	−0.018
Log (followers + 1)	0.068	−0.221	0.019
Log (following + 1)	0.726	−0.571	0.023
Date	−0.222	0.017	0.661
Fake	32.483	0.124	2.916
Log (retweets + 1)	−0.011	0.065	−0.001

Note: [a] As some broadcast and person-to-person influences were small, all results in this column were multiplied by 100 to aid reading.

4.1.2. Speed/Broadcast

Author, date, and tweet attributes were linked to broadcast diffusion of this topic in their user cluster. Authors with more experience yielded the fastest diffusion (broadcast; +0.00253 per day of Twitter experience, +0.923 per year of Twitter experience), supporting H11. By contrast, authors with more tweets, more followers, or following more users showed slightly less broadcast diffusion on this topic in their user cluster (−0.00551, −0.00221, or −0.00571, respectively), supporting H5 and H8. Tweets initiating this topic in a user cluster on later dates yielded slightly more broadcast diffusion (0.00017 per day). Tweets with links to fake news rather than true news yielded slightly more broadcast diffusion (0.00124), supporting H2. Lastly, additional retweets of the original tweet on this topic in a conversation yielded slightly more broadcast diffusion (0.00065).

4.1.3. Person-to-Person

Author, date, and tweet attributes were also linked to person-to-person diffusion of this topic in their user cluster. Authors with more experience showed less person-to-person diffusion (−0.00018 per day of Twitter experience, −0.0657 per year of Twitter experience), supporting H12. Likewise, authors with more tweets showed slightly less person-to-person diffusion (−0.00018). By contrast, authors with more followers or following more users showed slightly more person-to-person diffusion (0.00019 or 0.00023, respectively), supporting H6 and H9. Tweets starting this topic in a user cluster on later dates yielded the largest person-to-person diffusion (0.00661 per day, 0.19830 per month). Tweets with links to fake news rather than true news yielded much more person-to-person diffusion (0.02916) than broadcast diffusion (0.00124), supporting H3. Lastly, additional retweets of the original tweet on this topic in a conversation yielded slightly less person-to-person diffusion (−0.00001). Analyses of data subsets yielded similar results, suggesting their robustness.

5. Discussion

This is the first study to determine the antecedents of diffusion scope (total audience), speed (audience/time), and shape (broadcast vs. person-to-person) for true vs. fake news about a topic (stolen 2020 US presidential election) across different user clusters. Grounded in STOPS (Kim & Grunig, 2011), we hypothesized that fake (vs. true) news, user cluster size, followers, user experience, and date affect diffusion scope, speed, and shape. After examining 31,128 tweets, we identified 1,096 user clusters via *clustering analysis* (Srinivas & Rajendran, 2019), and tested our hypotheses with MDA (Rossman et al., 2008), thereby showcasing a new methodology for studying diffusion of messages (such as fake news) within user clusters. Our results showed an expected diffusion of each of these tweets to 233 people but no overall diffusion speed or shape for tweets. Instead, the above explanatory variables account for differences in scope, speed, and shape, mostly supporting our hypotheses (the results did not support significant interactions between fake news and user cluster size).

5.1. Fake News

Tweets linked to fake news started earlier, showed much greater diffusion scope, faster dissemination (broadcast), and more person-to-person transmission than tweets linked to true news. These results not only support those of earlier studies (e.g., Abilov et al., 2021; Vosoughi et al., 2018) but also extend them via more accurate measures of diffusion shape (some broadcast with mostly person-to-person transmission) and controlling for the impacts of other author, user cluster, date, or tweet attributes. Together, they show the many advantages of fake news tweets over true news tweets and highlight the need for pro-active measures to counter-act diffusion of fake news by focusing on earlier tweets. As no other user, user cluster, or tweet attributes were correlated with fake news (all $|r| < 0.02$), we need future studies with other explanatory variables that might affect fake news diffusion.

5.2. User Cluster Size

The results for numbers of followers and following aligned with our hypotheses that smaller user clusters show more intimacy and urgent concerns, resulting in faster broadcast diffusion but less scope and less person-to-person diffusion (Kim & Grunig, 2011). These results pinpoint a size trade-off between greater diffusion scope against slower diffusion speed. Furthermore, they suggest that the effects of social media user cluster size on interactions and diffusion resemble those of face-to-face user cluster size (Dunbar, 1996). User cluster size was not related to likelihood of fake news, so both fake news and true news tend to diffuse faster in smaller user clusters than in larger user clusters. Hence, monitors aiming for early detection of fast-spreading fake news should focus on smaller user clusters rather than larger user clusters.

5.3. Scale

The results supported the scale hypotheses that users with more followers send them more tweets to maintain their followers (Kim et al., 2010), and more of their followers engage with them but are less likely to immediately engage with any specific tweet (slower diffusion speed, less broadcast) and more likely to wait for other followers to engage before engaging themselves (more person-to-person engagement). Like user cluster size, more followers show a trade-off between greater diffusion scope against slower diffusion speed. These results apply for both fake and true news. Hence, monitors seeking early detection of quickly diffusing fake news should focus on users with fewer followers rather than those with many followers.

5.4. User Experience and Date

Authors with more experience showed greater diffusion speed (broadcast) and less person-to-person transmission (supporting both hypotheses) but had substantially smaller diffusion scope (rejecting our hypothesis). The greater broadcast diffusion and less person-to-person diffusion cohered with status effects (Chiu, 2008). The surprisingly smaller diffusion scope might stem from the illegitimacy of this topic of a falsely claimed stolen election. Future studies can test whether higher status, experienced people are less likely to engage substantially with an illegitimate topic and more likely to do so with a legitimate topic.

As expected, tweets on later dates showed less scope, supporting the claim that they lose audience to earlier tweets. Later tweets showed a slightly faster diffusion speed (broadcast) and the largest person-to-person diffusion of these explanatory variables. Future studies on other topics over longer time spans can test whether this result applies more generally across topics and discern its mechanism(s).

5.5. Limitations and Future Research

This study's limitations include its single topic, limited user clusters, single social media platform, limited time period, and limited explanatory variables. This study examined diffusion scope, speed, and shape for only one topic across a limited set of user clusters on one social media platform, Twitter, for 55 days; so, future studies can examine more topics, more user clusters, on more platforms for longer time periods. As this study tested few explanatory variables regarding each tweet, user, or user cluster, future studies can gather and test more information about each tweet, user, or user cluster. For example, this study did not consider whether subsequent tweets supported or rejected the original tweet, so future studies can examine whether supportive versus opposing tweets differ in their diffusion scope, speed, or shape. Also, this study tested few user attributes or behaviors, so future studies can do so in fine-grained detail. Likewise, future studies can collect more data on each user cluster and determine more structural attributes (e.g., degree of centrality). Adding these attributes to our model can improve our understanding of the antecedents of diffusion scope, speed, and shape.

6. Conclusion

Diffusion of tweets regarding a falsely claimed stolen 2020 US presidential election showed no overall diffusion pattern; instead, specific explanatory variables determined these tweets' diffusion scopes, speeds, and shapes. Tweets linked to fake news rather than true news started earlier, showed much greater diffusion scope, faster dissemination (broadcast), and more person-to-person transmission, highlighting the importance of pro-active countermeasures for fake news by focusing on earlier tweets, smaller user clusters, and users with fewer followers.

Smaller user clusters showed less scope and less person-to-person diffusion but faster broadcast diffusion. A user with many followers typically sends them many tweets, but with only slightly more scope, less speed, and slightly more person-to-person diffusion. Hence, both larger user cluster size and more followers trade off greater diffusion scope for slower diffusion speed. Authors with more experience showed greater diffusion speed (broadcast) and less person-to-person transmission but smaller diffusion scope. Tweets on later dates showed less diffusion scope, slightly faster diffusion speed (broadcast), and more person-to-person transmission.

Notably, these results highlight the greater diffusion speed of fake news in smaller user clusters and by users with fewer followers. Hence, they imply that monitors seeking to detect fake news early should focus on earlier tweets, smaller user clusters, and users with fewer followers.

Acknowledgments

We appreciate the statistics assistance of Yik Ting Choi. The Data Institute for Societal Challenges (DISC) at the University of Oklahoma has partially funded for this research.

Conflict of Interests

All authors declare no conflict of interests.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Abilov, A., Hua, Y., Matatov, H., Amir, O., & Naaman, M. (2021). *VoterFraud2020: A multi-modal dataset of election fraud claims on Twitter*. ArXiv. https://doi.org/10.48550/arXiv.2101.08210

Arenas, A., Duch, J., Fernández, A., & Gómez, S. (2007). Size reduction of complex networks preserving modularity. *New Journal of Physics*, *9*(6), Article 176. https://doi.org/10.1088/1367-2630/9/6/176

Azaouzi, M., Rhouma, D., & Romdhane, L. B. (2019). Community detection in large-scale social networks: State-of-the-art and future directions. *Social Network Analysis and Mining*, *9*, Article 23. https://doi.org/10.1007/s13278-019-0566-x

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*(10), Article P10008. https://doi.org/10.1088/1742-5468/2008/10/P10008

Bodaghi, A., & Oliveira, J. (2022). The theater of fake news spreading, who plays which role? A study on real graphs of spreading on Twitter. *Expert Systems with Applications*, *189*, Article 116110. https://doi-org.ezproxy.lib.ou.edu/10.1016/j.eswa.2021.116110

Bovet, A., & Makse, H. A. (2019). Influence of fake news in Twitter during the 2016 US presidential election. *Nature Communications*, *10*(1), Article 7. https://doi.org/10.1038/s41467-018-07761-2

Brashier, N. M., & Schacter, D. L. (2020). Aging in an era of fake news. *Current Directions in Psychological Science*, *29*(3), 316–323. https://doi.org/10.1177/0963721420915872

Braun, J. A., & Eklund, J. L. (2019). Fake news, real money: Ad tech platforms, profit-driven hoaxes, and the business of journalism. *Digital Journalism*, *7*(1), 1–21. https://doi.org/10.1080/21670811.2018.1556314

Cherven, K. (2015). *Mastering Gephi network visualization*. Packt.

Chiu, M. M. (2008). Flowing toward correct contributions during groups' mathematics problem solving: A statistical discourse analysis. *Journal of the Learning Sciences*, *17*(3), 415–463. https://doi.org/10.1080/10508400802224830

Cinelli, C., & Hazlett, C. (2019). Making sense of sensitivity: Extending omitted variable bias. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *82*(1), 39–67. https://doi.org/10.1111/rssb.12348

Dewey, J. (1910). *How we think*. Heath and Company.

Dewey, J., & Rogers, M. L. (2012). *The public and its problems: An essay in political inquiry*. Penn State University Press.

Dunbar, R. I. M. (1996). *Grooming, gossip, and the evolution of language*. Harvard University Press.

Fiske, S. T., & Taylor, S. E. (1991). *Social cognition*. McGraw-Hill.

Fortunato, S. (2010). Community detection in graph. *Physics Reports*, *486*(3/4/5), 75–174. https://doi.org/10.1016/j.physrep. 2009.11.002

Fortunato, S., & Castellano, C. (2007). *Community structure in graphs*. ArXiv. https://doi.org/10.48550/arXiv.0712.2716

Franz, M., & Nunn, C. L. (2010). Investigating the impact of observation errors on the statistical performance of network-based diffusion analysis. *Learning & Behavior*, *38*, 235–242. https://doi.org/10.3758/LB.38.3.235

Fuchs, C. (2014). Social media and the public sphere. *tripleC*, *12*(1), 57–101. https://doi.org/10.31269/triplec.v12i1.552

Grinberg, N., Joseph, K., Friedland, L., Swire-Thompson, B., & Lazer, D. (2019). Fake news on Twitter during the 2016 U.S. presidential election. *Science*, *363*(6425), 374–378. https://doi.org/10.1126/science.aau2706

Grunig, J. E., & Kim, J.-N. (2017). Publics approaches to segmentation in health and risk messaging. In R. Parrott (Ed.), *Encyclopedia of health and risk message design and processing*. Oxford University Press.

Guynn, J. (2021, February 12). "Burn down DC": Violence that erupted at Capitol was incited by pro-Trump mob on social media. *USA Today*. https://www.usatoday.com/story/tech/2021/01/06/trump-riot-twitter-parler-proud-boys-boogaloos-antifa-qanon/6570794002

Hilary, I. O., & Dumebi, O. O. (2021). Social media as a tool for misinformation and disinformation management. *Linguistics and Culture Review*, *5*(S1), 496–505. https://doi.org/10.21744/lingcure.v5nS1.1435

Hox, J. J., Moerbeek, M., & van de Schoot, R. (2017). *Multilevel analysis* (3rd ed.). Routledge.

Kim, J.-N., & Grunig, J. E. (2011). Problem solving and communicative action: A situational theory of problem solving. *Journal of Communication*, *61*(1), 120–149. https://doi.org/10.1111/j.1460-2466.2010.01529.x

Kim, J.-N., & Grunig, J. E. (2021). Lost in informational paradise: Cognitive arrest to epistemic inertia in problem solving. *American Behavioral Scientist*, *65*(2), 213–242. https://doi.org/10.1177/0002764219878237

Kim, J.-N., Grunig, J. E., & Ni, L. (2010). Reconceptualizing the communicative action of publics: Acquisition, selection, and transmission of information in problematic situations. *International Journal of Strategic Communication*, *4*(2), 126–154. https://doi.org/10.1080/15531181003701913

Kim, J.-N., & Krishna, A. (2014). Publics and lay informatics: A review of the situational theory of problem solving. *Annals of the International Communication Association*, *38*(1), 71–105. https://doi.org/10.1080/23808985.2014.11679159

Knobloch-Westerwick, S., Mothes, C., & Polavin, N. (2020). Confirmation bias, ingroup bias, and negativity bias in selective exposure to political information. *Communication Research*, *47*(1), 104–124. https://doi.org/10.1177/0093650217719596

Konstantopoulos, S. (2008). The power of the test for treatment effects in three-level cluster randomized designs. *Journal of Research on Educational Effectiveness*, *1*(1), 66–88. https://doi.org/10.1080/19345740701692522

Lee, H., Kim, J., & Kim, J.-N. (2021). Mechanics of rumor mills and epistemic motivational processes of food-related rumour spread: Interplay between attitude and issue motivation. *Health Communication*, *36*(6), 722–730. https://doi.org/10.1080/10410236.2020.1712518

Leicht, E. A., & Newman, M. E. (2008). Community structure in directed networks. *Physical Review Letters*, *100*(11), Article 118703. https://doi.org/10.1103/PhysRevLett.100.118703

Lutzke, L., Drummond, C., Slovic, P., & Árvai, J. (2019). Priming critical thinking: Simple interventions limit the influence of fake news about climate change on Facebook. *Global Environmental Change*, *58*, Article 101964. https://doi.org/10.1016/j.gloenvcha.2019.101964

Mittal, R., & Bhatia, M. P. S. (2019). Classifying the influential individuals in multi-layer social networks. *International Journal of Electronics, Communications, and Measurement Engineering*, *8*(1), 21–32. https://doi.org/10.4018/IJECME.2019010102

Oakhill, J. V., & Johnson-Laird, P. N. (1985). Rationality, memory and the search for counterexamples. *Cognition*, *20*(1), 79–94. https://doi.org/10.1016/0010-0277(85)90006-X

Pennycook, G., & Rand, D. G. (2021). Research note: Examining false beliefs about voter fraud in the wake of the 2020 Presidential election. *Harvard Kennedy School Misinformation Review*, *2*(1), 1–15.

Rogers, E. M., & Cartano, D. G. (1962). Methods of measuring opinion leadership. *The Public Opinion Quarterly*, *26*(3), 435–441.

Rosenthal, S. (2014). Detecting influencers in social media discussions. *XRDS: Crossroads*, *21*(1), 40–45. https://doi.org/10.1145/2659889

Rossman, G., Chiu, M. M., & Mol, J. M. (2008). Modeling diffusion of multiple innovations via multilevel diffusion curves: Payola in pop music radio. *Sociological Methodology*, *38*(1), 201–230. https://doi.org/10.1111/j.1467-9531.2008.00201.x

Sharma, K., He, X., Seo, S., & Liu, Y. (2020). *Network inference from a mixture of diffusion models for fake news mitigation*. ArXiv. https://doi.org/10.48550/arXiv.2008.03450

Shaw, Y., & Natisse, K. M. (Hosts). (2021, April 29). The chaos machine: An endless hole [Audio podcast episode]. NPR. https://www.npr.org/programs/invisibilia/992214107/the-chaos-machine-an-endless-hole

Srinivas, S., & Rajendran, C. (2019). Community detection and influential node identification in complex networks using mathematical programming. *Expert Systems With Applications*, *135*, 296–312. https://doi.org/10.1016/j.eswa.2019.05.059

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, *359*(6380), 1146–1151. https://doi.org/10.1126/science.aap9559

Walker, M., & Matsa, K. E. (2021). *News consumption across social media in 2021: More than half of Twitter user get news on the site regularly.* Pew Research Center. https://www.pewresearch.org/journalism/2021/09/20/news-consumption-across-social-media-in-2021

About the Authors

Ming Ming Chiu is chair (distinguished) professor of analytics and diversity. He invented (a) the artificial intelligence program *Statistician*, (b) statistical discourse analysis to model chats/conversations, (c) multilevel diffusion analysis to detect corruption, and (d) online detection of sexual predators. His 67 grants (US$14 million) yielded 255 publications (166 journal articles, more than 11,000 citations, 13 keynote speeches, five television broadcasts, 17 radio broadcasts, and 168 news articles in 21 countries. He creates automatic statistical analyses for big data.

Chong Hyun Park is an assistant professor at the School of Business at Sungkyunkwan University. His research interests include the mathematical programming and the machine learning modeling. He conducts interdisciplinary research to solve various social problems. He recently developed a machine learning algorithm that can detect manipulated opinion spams in comments sections. He has published research papers in *Production and Operations Management*, *European Journal of Operations Research*, and *American Behavioral Scientist*.

Hyelim Lee is a doctoral student at the University of Oklahoma's Gaylord College of Journalism and Mass Communication. She studied political communications and big data analysis at Seoul National University. Her doctoral research explores how computational social science methods can inform theories of public relations. She also studies conspiratorial public issues in public relations. Lee recently joined the Debiasing and Lay Informatics (DaLI) lab in the Center for Applied Social Research at the University of Oklahoma where she researches fake news detection and social group interaction in social media through machine learning and computational text analysis. In 2021, with co-author Lisa Tam, Lee received the International Communication Association Public Relations Division Top Faculty Paper Award.

Yu Won Oh (PhD, University of Michigan, 2015) is an assistant professor in the Department of Digital Media at Myongji University, Republic of Korea, and the associate director of the Debiasing and Lay Informatics (DaLI) lab in Norman, Oklahoma. Her research interests include the intersection of new media and political communication with an emphasis on opinion expression, misinformation, issue development, and big data analytics. Oh's research has been published in top-ranked journals and she has received best paper awards from major communication conferences including the National Communication Association and the World Association for Public Opinion Research.

Jeong-Nam Kim (PhD, University of Maryland, 2006) is Gaylord Family Endowed Chair of Strategic Communication at the University of Oklahoma and the founding director of the Debiasing and Lay Informatics (DaLI) lab. Kim studies communicative action and informatics among lay problem solvers (cf. expert/scientific problem solvers). He constructed the situational theory of problem solving (STOPS) and a model of cognitive arrest and epistemic inertia among lay problem solvers with James E. Grunig. His lab, DaLI, seeks solutions for information problems such as pseudo-information, public biases, and failing information markets.