European Journal of Operational Research xxx (xxxx) xxx

Contents lists available at ScienceDirect

European Journal of Operational Research

journal homepage: www.elsevier.com/locate/ejor

Innovative Applications of O.R.

Interplay of rumor propagation and clarification on social media during crisis events - A game-theoretic approach

Puneet Agarwal, Ridwan Al Aziz, Jun Zhuang[1],*

Department of Industrial and Systems Engineering, University at Buffalo, Buffalo, NY 14260-2050, USA

ARTICLE INFO

Article history:
Received 24 September 2020
Accepted 29 June 2021
Available online xxx

Keywords:
Game theory
Social media
Rumor propagation
Rumor control
Emergency communication

ABSTRACT

For a rapid dissemination of information during crisis events, official agencies and disaster relief organizations have been utilizing social media platforms, which are susceptible to rumor propagation. To minimize the impact of rumors with limited time and resources, the agencies and social media companies not only need to wisely choose the cases to clarify amongst the numerous cases, but they should also make an informed decision on the timing of clarification. Reacting fast can be misjudged as an obvious best policy as partial/imprecise information may fail to contain the impact of the rumors. On the other hand, investment in terms of time, effort, and money to clarify with more complete information also allows the rumors to spread with their full force during the learning phase, thereby making the process of decision-making very challenging. The objective of this paper is to determine the optimal strategies for the official agencies and social media companies by developing two novel sequential game-theoretic models, namely "Rumor Selection for Clarification" and "Learning for Rumor Clarification", that can help decide which rumor to clarify and when to clarify, respectively. Results from this study indicate that posting verified information on social media reduces the uncertainties involved in rumor transmission, thereby enabling social media users to make informed decisions on whether to support or oppose the rumor being circulated. This verification needs to be obtained within reasonable limits of time and cost to keep the learning process worthwhile.

1. Introduction

Rumor is defined as an "informally improvised news" (Shibutani, 1966) that can affect individuals and their communities in the time of crisis situations (Prasad, 1935). According to Zhao, Resnick, & Mei (2015), there are two salient characteristics of a rumor: (i) it generally occurs in situations where its truth value appears to be uncertain and ambiguous to the public, and (ii) it may or may not be always false, that is, it may contain truthful information which is yet to be verified by the authentic sources.

Since World War II, psychology of rumors and how to contain them have attracted significant attention from the research community (Festinger et al., 1948; Knapp, 1944). Books, newspapers, magazines, and interviews were the sources of rumor data collection for the early studies. In the current era of online social media, any piece of information can be diffused by online users without censorship (Kwon, Cha, Jung, Chen, & Wang, 2013). The harmful impacts of false rumors on any organization or individuals have received attention in both research and society; and it is often argued that rumors are generally generated and propagated in situations that are important, uncertain, threatening, uncontrollable, and produce anxiety (DiFonzo, 2008; DiFonzo & Bordia, 2007). For example, rumors may often be generated in wars or crises because these are life or death situations, and are certainly threatening, uncontrollable, and anxiety-producing. Accurate and complete information may be completely unavailable or available beyond the limits of a reasonable time frame for clarification, which in turn, is the perfect condition for rumor generation and transmission (Rubin, 2017).

Social media has been infamously dubbed as a "rumor mill" for diffusing false rumors and misinformation during crisis situations, which has the potential to promote large scale panic and financial loss (Oh, Kwon, & Rao, 2010). Spread of misinformation on online platforms was ranked first by the World Economic Fo-

* Corresponding author.
E-mail addresses: puneetag@buffalo.edu (P. Agarwal), ridwanal@buffalo.edu (R. Al Aziz), jzhuang@buffalo.edu (J. Zhuang).

[1] All authors contributed equally and the author names are listed alphabetically by last name. Jun Zhuang is the corresponding author. This research was supported by the National Science Foundation under Award No. 1762807. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. We thank the editor and two anonymous reviewers for helpful comments.

https://doi.org/10.1016/j.ejor.2021.06.060
0377-2217/© 2021 Elsevier B.V. All rights reserved.

rum among the top future global risks (Howel, 2013). For example, false rumors such as: “Mandatory evacuations are underway in the City of Houston” (Bennet, 2017), “Immigration status has to be checked before you are allowed to enter a shelter” (Bennet, 2017), and “Residents could not return to the coastal city until all critical services were restored” (Alfonso, 2017) during Hurricane Harvey, caused great confusion, panic, and anger among people in the affected areas. In lieu of these events that continue to occur frequently, rumor control and management on social media requires more attention from both researchers and practitioners.

During large-scale crises, the mainstream media often cover incidents that are initially reported by local spectators (Oh, Agrawal, & Rao, 2011). For example, according to Twitter, within the first 60 seconds after the 2011 Virginia Earthquake, there were 40,000 tweets related to that incident (Indvik, 2011). As a result, rumors or misinformation could have been propagated widely by the time social debunking and verification information was available (Ozturk, Li, & Sakamoto, 2015). Therefore, in such situations, social media users assess the veracity of information by themselves before taking an action to spread, ignore or debunk the piece of information being circulated (Ozturk et al., 2015). On the other side, people like to spread rumors because of importance, social responsibility (Luttrell, 2015), awareness of adverse consequences or personal norms (Zhao, Yin, & Song, 2016). Research shows that due to truth-biased characters, people are prone to believe the false rumors and propagate them as true information (Rubin, 2017; Wang & Zhuang, 2017; 2018; Zubiaga, Liakata, Procter, Hoi, & Tolmie, 2016). Social media platforms are also being increasingly abused by bots that mislead, exploit, and manipulate users by spreading rumors, misinformation, disinformation, spam, and malware (Ferrara, Varol, Davis, Menczer, & Flammini, 2016). The bots are able to inflate the popularity of a post, irrespective of its accuracy, thus exerting significant influence on users during critical events such as elections, disasters, and pandemic. In the recent years, it has been found that the behavior of bots is becoming increasingly human-like that makes their detection more difficult (Hwang, Pearce, & Nanis, 2012).

Official agencies and disaster relief organizations often use online social media as informational support tools to disseminate critical information to social media users about activities such as evacuation routes, aid distribution, and sheltering, during crisis events (Yan & Pedraza-Martinez, 2019). In case of rumor dispersion on social media platforms, major government, news, non-governmental (NGO), social media companies, and emergency management agencies make statements and post to social media platforms in order to clarify the rumor and provide the public with accurate content. For example, following the false rumors during Hurricanes Harvey and Irma, the U.S. Federal Emergency Management Agency (FEMA) created “Rumor Control” pages (Federal Emergency Management Agency, 2017; 2017) on their website in order to dispel the inaccuracies and provide updated and thorough communications. FEMA used their Twitter account to disseminate this web page by posting twelve different tweets over the course of one week, with all of these tweets having a direct link to the Rumor Control page. In a recent event of coronavirus outbreak in Wuhan, Hubei province, China, it was reported that in early January 2020, the Wuhan police had arrested eight people accused of spreading false information about a mysterious pneumonia that caused serious complications. When the spread of coronavirus made national headlines, the journalists reporting on the outbreak were detained or threatened to be arrested. The steps taken by the Chinese government to contain the spread of misinformation without sufficient information about the outbreak have received strong criticism from the international community that deemed it as a major “cover up” (Ashley Collman, 2020). This particular incident corroborates the necessity of using verified information by the official agencies for an effective rumor clarification. Hence, in order to clarify rumors, agencies must expend human resources and time in order to locate rumors on social media, track the rumors in order to understand their reach and impact, and formulate effective clarification and debunking messages.

Once a rumor case is identified within social media networks, the subsequent online communications associated with the rumor have to be monitored in order to take timely actions and contain its spread. Deciding the balance between how quickly to respond and how much time to invest in gathering verified information before clarifying becomes more crucial in the context of social learning. During emergency situations, a lot of novice users also rely on the information found online and how others are reacting to those information. Their perception towards the truth of rumor is acquired through a cognitive process of observing and sometimes imitating others in the social context. Such behavioral responses have received significant attention in the recent operations management literature. Among such works, Papanastasiou & Savva (2017) and Crapis, Ifrach, Maglaras, & Scarsini (2017) focused on how the optimal pricing policies are influenced by the customer reviews; while Feldman, Papanastasiou, & Segev (2019) showed that the social learning may contribute in decreasing the quality of new experience goods. Hu, Milner, & Wu (2016) considered the effects of social influence on optimal inventory decisions and Gao & Su (2017) considered whether offering the option between buying online and picking up in store is beneficial to the retailers. Papanastasiou (2020) deployed a sequential model to study the problem of dynamically choosing whether to conduct a æfact-checkg of an article whose veracity is not known beforehand.

Over the last few years, the problem of determining the veracity of the information that an individual user posts with respect to the detected case of rumor has attracted many studies (Chen, Zheng, & Ceran, 2016; Hamidian & Diab, 2015; Lee, Qiu, & Whinston, 2018; Qazvinian, Rosengren, Radev, & Mei, 2011; Zeng, Starbird, & Spiro, 2016; Zhang, Gupta, Kauten, Deokar, & Qin, 2019; Zubiaga, Kochkina, Liakata, Procter, & Lukasik, 2016). Numerous studies also have characterized the emergence and propagation of rumors in social media platforms. Liao & Shi (2013) explored the dynamics of rumor transmission in China’s largest microblogging system, Sina Weibo, and identified four major categories that describe how users intervene in rumor discussions: providing information, expressing emotions, sharing opinions, and analyzing and interpreting situations. Zubiaga et al. (2016) analyzed a dataset of 330 rumor threads associated with 9 newsworthy events to understand the role of different types of users in rumor propagation and clarification process throughout the life cycle of a rumor. Cheng, Liu, Shen, & Yuan (2013) found that the diffusion of rumors in online social networks is a function of the strength of ties between users, where the possibility of a rumor spreading is more likely across strong ties in a network. Studies conducted by Oh, Agrawal, & Rao (2013) on rumor mongering show that the effect of source ambiguity (the lack of an official source) on rumoring is much more significant than that of content ambiguity (lack of persuasive statements in Twitter posts), and anxiety. Vosoughi, Roy, & Aral (2018) analyzed the diffusion dynamics of true and false rumors and found that false rumors propagated significantly faster and deeper as compared to true rumors in all categories of information; namely political news, terrorism, natural disasters, science, urban legends, entertainment, and financial information. Roozenbeek & Van Der Linden (2019) developed a fake news game to evaluate its effectiveness on educating the public in fighting and managing the risks posed by fake news. In this experiment, the participants were trained to recognize fake news tactics by assuming different characters in order to provide a broad level resistance to the transmission of fake news.

As evident from the review of current works in the domain of rumor propagation and clarification, a wide range of studies are focused on analyzing the propagation dynamics of rumors on social media coupled with the behaviors of users with respect to these rumors. As a result, a significant surfeit of research exists that takes into account the strategic interactions between official agencies and social media users in the process of rumor propagation and clarification. In the past, numerous studies have used game-theoretic approaches to model the interactions between an official agency and a private entity in different application domains. Cheung & Zhuang (2012) analyzed the strategic interactions between the government and the oil spill companies by formulating game-theoretic models with different attributes such as one-company/two-company, with/without competition. Agarwal, Hunt, Srinivasan, & Zhuang (2020) developed centralized and decentralized game-theoretic models to study the strategic behaviors of fire inspection agencies and building owners in the process of fire safety code inspection and compliance. Bier & Haphuriwat (2011); Bier, Haphuriwat, Menoyo, Zimmerman, & Culpen (2008); Shan & Zhuang (2014) developed game-theoretic models to analyze the retaliation efforts of defenders (official agencies) and attack strategies of smugglers (individual/groups of people) in the context of nuclear smuggling. Strategic interactions between the hackers and defenders in the context of cyber security problems have also successfully attracted the attention of game theory enthusiasts (Rao et al., 2016; Ten, Manimaran, & Liu, 2010). Game theory and the concept of Nash equilibrium have also been utilized for identifying the equilibrium strategies for attacker (for example, terrorist organizations) and defender (for example, government defense agencies) in the context of disaster management. In this application domain, Zhuang & Bier (2007) developed simultaneous and sequential game models that provided critical insights to government agencies for allocating defensive investment between terrorism and natural disasters.

To the best of our knowledge, no previous research has studied the strategic interactions between official agencies and social media users during rumor propagation and clarification process. This paper fills this gap by modeling the strategic behaviors of the players during rumor transmission using two novel game-theoretic models. The objective of this study is to analyze the impact of rumor clarification and verification strategies of the agencies and social media companies on decisions of the social media users during rumor propagation. The first model "Rumor Selection for Clarification" serves as a decision support tool for the emergency agencies to take a crucial decision on selection of rumor cases for clarification, and subsequently convey correct information to the population by effective utilization of available resources. The second model "Learning for Rumor Clarification" helps to determine the optimal strategy for the agencies and social media companies by addressing the trade-offs between reacting fast with partial/imprecise information and reacting later with verified information.

The remaining sections of the paper are organized as follows: Section 2 provides an overview on rumor clarification and verification processes in real life situations. Section 3 presents Model 1 by introducing its notations, assumptions and problem definitions, provides the analytic and numerical results of the model, and finally delivers prescriptive insights that are derived from the model results. Section 4 introduces Model 2 by its notations, assumptions and objective functions, provides the analytic and numerical results of the model, and presents prescriptive insights that are derived from the model results. Section 5 presents the validation of the propositions based on Twitter data from seven real life rumor cases and Section 6 provides the summary of the paper and proposes future research directions. The appendix provides proofs for the propositions.

2. Background on rumor clarification and verification

2.1. Rumor clarification

Acknowledging the destructive effect of rumors during crisis events, official agencies and social media companies have used rumor clarification as one of the major strategies for restraining rumors in social media (Wen et al., 2014). For example, during the 2016 Louisiana Floods, the American Red Cross published a blog named "Top Questions About Louisiana Flood Relief" to disseminate critical information about resources and raise situational awareness (American Red Cross, 2016). This blog was widely used by the digital volunteers and online supporters of the Red Cross to spread correct information among the public. The Red Cross Social Engagement team also created a secret Facebook group to channel important updates, flag urgent issues, and collaborate with other teams. Situational updates through informational videos have been an integral tool to combat misinformation for the Red Cross team (U.S. Department of Homeland Security, 2018). During the 2017 Hurricane Harvey, FEMA requested support froma Virtual Operation Support Team (VOST) in three mission areas: (i) tracking and delivery of large donations, (ii) tracking the recruitment of volunteers and their locations, and (iii) tracking donation scams that redirected funds allocated for the victims to funds unrelated to Hurricane Harvey (U.S. Department of Homeland Security, 2018).

For clarification of rumors on social media, a host of different practices can be used by the official agencies as prescribed by the official report released by the U.S. Department of Homeland Security (U.S. Department of Homeland Security, 2018). In addition to online social media, platforms such as local television, radio, and news media can be leveraged to propagate necessary information and debunk false information. As seen from the above case studies, official organizations such as FEMA and the Red Cross can seek support fromVOSTs or other digital volunteers on different areas such as tracking and monitoring social media platforms, identifying false rumors and investigating their potential sources, and collaborating with the official agencies for controlling the spread of rumors. Seeking such assistance also requires training of volunteers and responders on how to pinpoint false rumors and misinformation, and what practices should be used to respond to harmful information. Agencies can also identify and leverage trusted crowd sources or influencers to disseminate rumor correction information. Using these practices usually demand significant effort and time, thereby limiting the number of rumors that can be effectively contained and clarified in time (Wang et al., 2019).

2.2. Rumor verification

Clarifying rumor with insufficient or unverified information contributes to the rapid spread of rumors. While a lack of verification resources is a very common reason attributed to insufficient information, a lack of authority to release information or ambiguity in the responsibilities are also pervasive factors behind insufficiency. Not clarifying the rumor by covering all aspects can leave room for speculation and lead to serious detrimental effects. The 2017 Orville Dam Evacuation rumor is a case where insufficient information failed to contain the large scale panic among people. The National Weather Service (NWS) Sacramento station distributed a tweet that showed an image of the rainfall flooded area encompassing Sacramento County along with the Oroville Dam. Around 6 p.m. (local time), @LNU_fire_photo doctored that image by incorporating evacuation information, as shown in Fig. 1 and distributed in Twitter. People seeing the image believed they are under an evacuation order. At 6:30 p.m., @JDLazo tried to clarify the rumor, but the clarification attempt with insufficient information could not convince the mass people and calls began flooding Sacramento

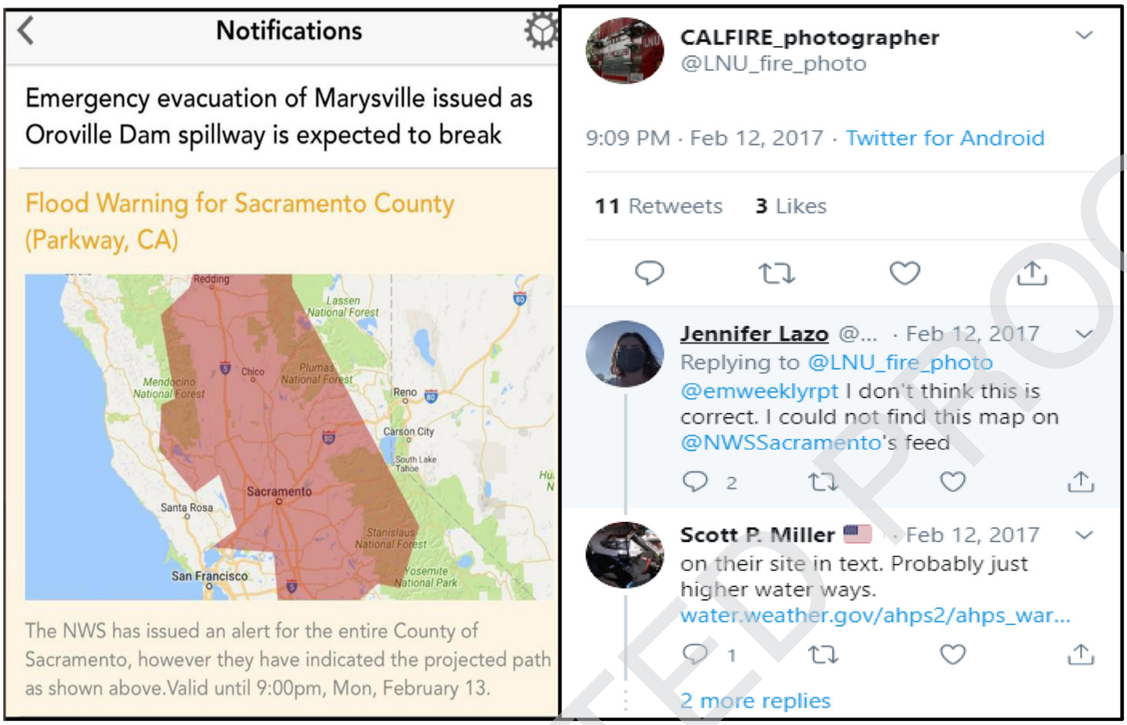

Fig. 1. Unsuccessful rumor verification attempt on Twitter due to insufficient information during Oroville Dam crisis.

County 9-1-1 dispatch. At 8:50 p.m., staff from Sacramento County Emergency Operations Center (EOC) shot a Facebook Live Video (sac, 2017) to correct misinformation with verified information and uploaded it. Almost immediately, news media including television and iHeartRadio took the initiative to quote the live feeds in their broadcasts and the videos were circulated. The radio stations also joined by broadcasting the audio portion of the live video and television stations utilized their anchors to quote information from the Sacramento County EOC. After the broadcasts, the call volume in County's 9-1-1 dispatch center returned to normal.

For verifying the rumored information, a host of different practices can be used as prescribed by the official report released by the U.S. Department of Homeland Security (U.S. Department of Homeland Security, 2018). Verifications can be obtained by checking on the primary and supporting sources. The author of the social media content can be contacted to get more information. Eye witnesses and first responders can be contacted to get verified information on the disseminated content on social media. The legitimacy of information disseminated on social media platforms can be cross-checked using websites such as Factcheck.org (Factcheck, 2020) or Snopes.com (Snopes, 2020). These websites initially communicate with the source of the claim for explanation and supporting information. They also try to communicate with the individuals and organizations who possess relevant expertise in the subject of interest. News articles, scientific and medical journal articles, books, interview transcripts and statistical sources on the topic are often looked into for completeness of the information. Verification can also be performed by conducting reverse image searches using Google. Google's Search by Image (Google Images, 2020) is a feature that uses reverse image search and allows users to search for related images just by uploading an image or image URL. Geofence and/or Twitter searches for geolocations can also be used to mine and filter the real information from false information.

3. Model 1: Rumor selection for clarification

During crisis events, multiple rumor cases propagate that may vary with respect to different factors such as the impact on the public, extent of diffusion in social networks, and efforts required to control their propagation. In such time-sensitive situations, the official agencies and social media companies must make critical decisions in order to minimize the spread of rumors, thereby preventing widespread panic and confusion among the misinformed users. Due to the limitation of resources, the official agencies and social media companies must strategically choose the specific rumor case(s) for clarification, while considering the potential trade-offs between the cost of clarification of rumor and impact of rumor on the affected community.

Model 1 identifies the strategic interactions between two clusters of users, namely User A (she) and User B (he), in the context of rumor selection for clarification. A decision maker is defined as User A if she has authority, responsibility, and resources to clarify and/or verify a rumor. Agencies such as FEMA, Red Cross, and Department of Defense, social media companies, news organizations, and fact-checking websites such as FactCheck.org, Snopes, and Politifact fall under the category of user type A. On the other hand, User B is defined as a social media account that is not necessarily responsible for rumor clarification and/or verification. User B may engage in the process of rumor clarification and/or verification by supporting, opposing or showing neutral participation. The objective of this model is to study the impact of the User A's rumor clarification strategies on the User B's decision to support, oppose, or show neutral participation for a specific rumor. This is achieved by modeling the scenario of rumor selection using a sequential game model. This model is regarded as a first approximation to the problem of rumor selection for clarification by considering a case of two rumors.

3.1. Notations, assumptions and description of model

Notations for Model 1 are introduced and defined in Table 1, that include two decision variables, eleven parameters, and two functions.

In the sequential game that is illustrated in Fig. 2, User A is assumed to minimize her expected loss, L_{1A}, while User B is assumed to maximize his expected utility, U_{1B}. User A is assumed to be the

Table 1
Notations used in Model 1.

Decision Options of User A	
D_i	Clarify rumor i, where $i = 1, 2$
Decision Options of User B	
K_j	Oppose rumor j, where $j = 1, 2$
S_j	Support rumor j, where $j = 1, 2$
N_j	Neutral participation with respect to rumor j, where $j = 1, 2$
Decision Variables	
x_i	Whether User A decides to clarify rumor i, where $i = 1, 2$; $x_i \in \{0, 1\}$ and $\sum_i x_i = 1$
y_{ik}	Whether User B decides to choose option k given that User A clarifies rumor i, where $k \in \{K_j, S_j, N_j\}$; $i = 1, 2$; $i \neq j$; $y_{ik} \in \{0, 1\}$ and $\sum_k y_{ik} = 1$, $\forall i = 1, 2$
Parameters	
r_i^D	Cost of clarification of rumor i, where $i = 1, 2$
r_i^H	Impact of rumor i, where $i = 1, 2$
d_i	Clarification index for rumor i, where $d_i \in [0, 1], i = 1, 2$
u_i	Index for spreading true information by User B, where $u_i \in [0, 1]$, $i = 1, 2$
v_i	Index for spreading false information by User B, where $v_i > 1$, $i = 1, 2$
p_i	Probability of rumor i being true, where $i = 1, 2$
f	Number of followers of User B
a_i	Benefit of spreading true information to each follower regarding rumor i, where $i = 1, 2$
c_i	Cost of spreading false information to each follower regarding rumor i, where $i = 1, 2$
r_i^S	Engagement rate obtained by User B due to importance of the event, where $i = 1, 2$
r_i^P	Engagement rate obtained by User B due to popularity of the rumor spreader, where $i = 1, 2$
Functions	
$L_{1A}(x_i, y_{ik})$	Expected loss of User A
$U_{1B}(x_i, y_{ik})$	Expected utility of User B

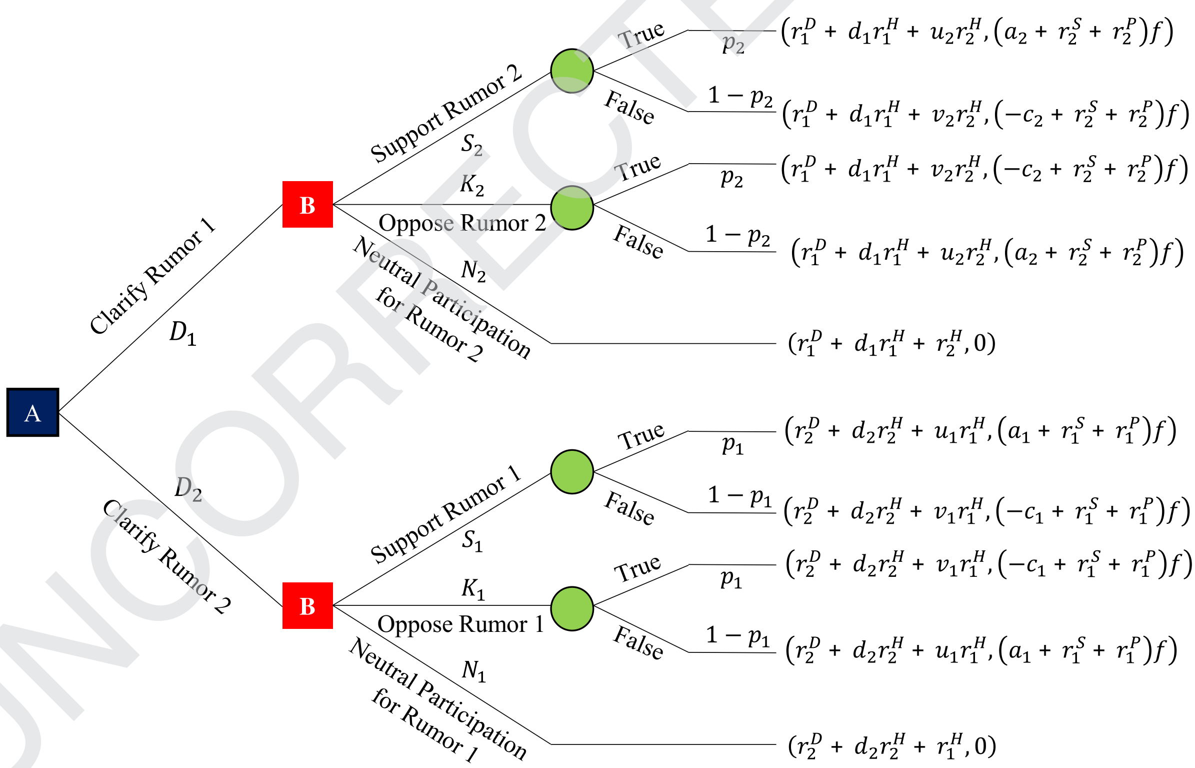

Fig. 2. Sequence of moves of players in a rumor selection for clarification game (a case of 2 rumors).

first mover who chooses her strategy first regarding the selection of rumor to clarify. In this model, each player has a different set of objectives: User A seeks to minimize the cost of rumor clarification and the impact of rumor transmission, while User B desires to maximize his influence and credibility ratings in the social networks. The veracity of a rumor i is modeled using a chance event with probability p_i that rumor i is true. The value of p_i is assumed to be independent of the strategies taken by the players and their corresponding subjective assessments.

Given a case of two rumors, User A can choose to clarify rumor 1 (D_1) or rumor 2 (D_2). On the other hand, User B can choose to support (S_j), oppose (K_j) or engage in neutral participation (N_j) by posting comments and questions regarding rumor j. When User A chooses to clarify a particular rumor, it is assumed that there is no incentive for User B to make a move regarding that rumor, while he can still choose to support, oppose or engage in neutral participation with respect to the other rumor. If User B supports a true rumor or opposes a false rumor, he earns a benefit $a_i f$ for spreading true information to his followers. While if he chooses to support a false rumor or oppose a true rumor, he bears a cost $c_i f$ for spreading false information to his followers. No benefit or cost is associated with the User B's decision of neutral participation. The engagement rate obtained by User B while supporting or opposing a rumor also depends on the importance of the event (r_i^S) and the popularity of the rumor spreader (r_i^P). For User A, there is a cost of clarification r_i^D for rumor i. The impact of the rumor r_i^H that User A seeks to minimize depends on the strategy of User B. If User B supports a false rumor or opposes a true rumor, the impact of rumor increases by a factor ν_i, while if he supports a true rumor, or opposes a false rumor, the impact of rumor decreases by a factor u_i. The impact of rumor that User A clarifies is dependent on the corresponding quality of rumor clarification, d_i. For example, if User A chooses to clarify rumor 1, the expected impact of rumor will be $d_1 r_1^H$, where d_1 is the quality of clarification for rumor 1. In Model 1, the objective of User A is to minimize her expected loss L_{1A} by choosing x_i to clarify rumor i. The objective of User B is to maximize his expected utility U_{1B} by choosing y_{ik} to support, oppose or engage in neutral participation regarding rumor j, given that User A chooses to clarify rumor i. Therefore, the optimization functions of both players in Model 1 can be written as shown below:

$$
\begin{aligned}
\min_{x_i} L_{1A}(x_i, y_{ik}) = x_1 \Big(& \big(p_2 u_2 + (1-p_2)\nu_2 - 1\big) r_2^H y_{1S_2} \\
& + \big(p_2 \nu_2 + (1-p_2) u_2 - 1\big) r_2^H y_{1K_2} + r_1^D + d_1 r_1^H + r_2^H \Big) \\
& + (1-x_1)\Big(\big(p_1 u_1 + (1-p_1)\nu_1 - 1\big) r_1^H y_{2S_1} \\
& + \big(p_1 \nu_1 + (1-p_1) u_1 - 1\big) r_1^H y_{2K_1} + r_2^D + d_2 r_2^H + r_1^H \Big) \\
\max_{y_{ik}} U_{1B}(x_i, y_{ik}) = x_1 \Big(& y_{1S_2} f\big(a_2 p_2 - c_2(1-p_2) + r_2^S + r_2^P\big) \\
& + y_{1K_2} f\big(-c_2 p_2 + a_2(1-p_2) + r_2^S + r_2^P\big)\Big) \\
& + (1-x_1)\Big(y_{2S_1} f\big(a_1 p_1 - c_1(1-p_1) + r_1^S + r_1^P\big) \\
& + y_{2K_1} f\big(-c_1 p_1 + a_1(1-p_1) + r_1^S + r_1^P\big)\Big)
\end{aligned} \tag{1}
$$

3.2. Best response of user B

Since User B is assumed to be the second mover in Model 1, we first derive the best response of User B, $\hat{y}_n$, which is defined as follows:

$$\hat{y}_n \equiv \underset{y_{ik}\in\{0,1\}}{\operatorname{argmax}}\ U_{1B}(x_i, y_{ik}), \text{ where } n = 1, 2 \tag{2}$$

The best response function of User B enables us to obtain his optimal strategy to maximize his expected utility, U_{1B}, with respect to the option chosen by the User A (either to clarify rumor 1 or rumor 2).

Proposition 1. *The best response function of User B, $\hat{y}_n$, is given by:*

$$
\hat{y}_n \equiv \begin{cases}
S_j & \text{if } p_j \geq \max\left(\frac{c_j - (r_j^S + r_j^P)}{c_j + a_j}, \frac{1}{2}\right), \ \forall j = 1, 2 \\
K_j & \text{if } p_j \leq \min\left(\frac{a_j + (r_j^S + r_j^P)}{c_j + a_j}, \frac{1}{2}\right), \ \forall j = 1, 2 \\
N_j & \text{if } p_j \in \left(\frac{a_j + (r_j^S + r_j^P)}{c_j + a_j}, \frac{c_j - (r_j^S + r_j^P)}{c_j + a_j}\right), \ \forall j = 1, 2
\end{cases} \tag{3}
$$

Remark. Proposition 1 identifies the boundary conditions for different decision options of User B. The best response of User B as a function of $p_1, p_2, a_1, a_2, c_1, c_2, r_1^S, r_2^S, r_1^P$, and r_2^P are shown graphically in Fig. 3. For numerical illustrations, the baseline values of the parameters used in this model are assumed as follows: $p_1 = 0.6$, $p_2 = 0.4$, $a_1 = 0.9$, $a_2 = 0.7$, $c_1 = 2.3$, $c_2 = 2.0$, $r_1^S = 0.3$, $r_2^S = 0.2$, $r_1^P = .2$, $r_2^P = .3$, $f = 250$, $r_1^D = 7.0$, $r_2^D = 6.0$, $r_1^H = 5.0$, $r_2^H = 4.0$, $d_1 = 0.75$, $d_2 = 0.5$, $u_1 = 0.8$, $u_2 = 0.3$, $\nu_1 = 1.5$, and $\nu_2 = 1.1$. The average number of favorites/likes, retweets/shares, and positive comments/replies received by User B by sharing true information on social media can be used to estimate the values of a_i and u_i. Similarly, the values of r_i^S and r_i^P can be quantified based on the average number of favorites/likes, retweets/shares, and positive comments/replies that User B gets due to the importance of event and popularity of rumor spreader, respectively. On the other hand, the values of c_i and ν_i can be determined from the average number of negative comments/replies that User B receives by sharing false information on social media. A higher value of the number of negative comments/replies for each user sharing false information will result in a higher c_i and lower ν_i. The value of parameter f can be obtained from the user profiles in social media platforms. The values of probabilities p_i for rumor cases being true or false can be derived from the historical database of rumors that have similar profiles in terms of diffusion, impact, and type of content spread. Costs of clarification r_i^D depend on the type of resources that are utilized by User A, which in turn, is

directly dependent on the profile of the specific rumor case being considered for clarification. The impact of rumors r_i^H can be determined by following their spread in both online and offline social environments. In addition to this, surveys related to the potential direct and/or indirect damage caused by a rumor can prove to be an effective tool in quantifying the impact created by that specific rumor. The values of parameters d_i can be determined using the number of users sharing false/true information before and after the clarification is made by User A.

As evident in Fig. 3(a) and 3(b), when the probability of a rumor being true is low, User B chooses to oppose the rumor. At sufficiently high values of this probability, he shifts his strategy to supporting the rumor. In Fig. 3(c) and 3(d), it is observed that when the benefit of spreading true information for a rumor is low, User B engages in neutral participation, while a higher benefit motivates him to either support or oppose the rumor based on the probability of that rumor being true. Fig. 3(e) and 3(f) show that when the cost of spreading false information for a rumor ranges from low to moderate, User B is likely to support or oppose that rumor based on its probability being true. A higher value of cost shifts his strategy to neutral participation in order to maximize his expected utility. In Fig. 3(g) to 3(j), it is observed that a higher engagement rate due to the importance of event or popularity of the spreader motivates User B to change his strategy from neutral participation to either supporting or opposing the rumor based on the probability of rumor being true.

3.3. Equilibrium solutions

Definition 1. A pair of User A's and User B's strategies (x^*, y^*) is called a subgame-perfect Nash equilibrium (SPNE) if and only if:

$$x^* = \underset{x\in\{0,1\}}{\operatorname{argmin}}\ L_{1A}(x, \hat{y}_n), \text{ where } n \in \{1, 2\} \tag{4}$$

$$y^* = \hat{y}_n(x^*) = \underset{y_n\in\{0,1\}}{\operatorname{argmax}}\ U_{1B}(x^*, y_n), \text{ where } n \in \{1, 2\} \tag{5}$$

The SPNE solutions are obtained using backward induction technique (Agarwal et al., 2020; Ho & Su, 2013).

Proposition 2. *The SPNE solutions of the selection of rumor model along with the optimal expected loss/utility of each player are provided in Table 2, where P_m, $m = 1, 2, \ldots, 6$ are the optimal conditions*

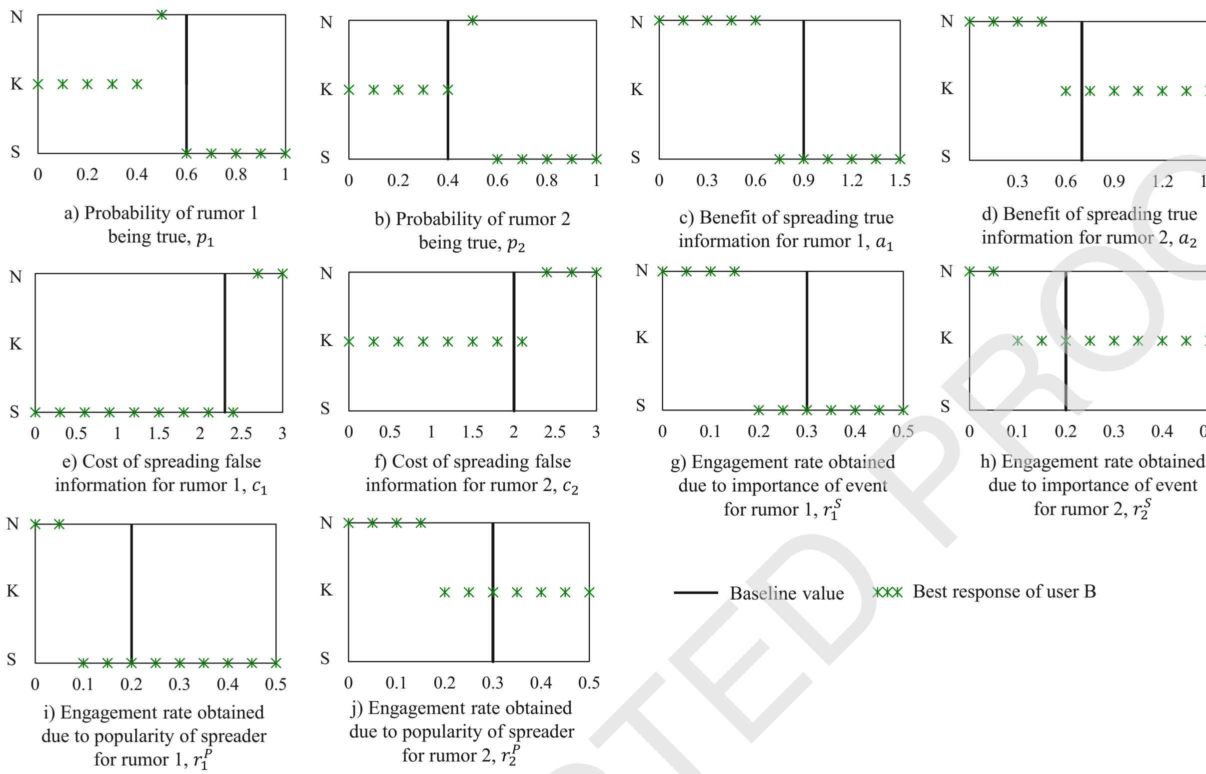

Fig. 3. Best response of User B as a one-way function of the parameters used in the rumor selection for clarification model.

Table 2
Equilibrium solutions of the selection of rumor model.

Case No.	Conditions	$(x^*; y^*)$	L_{1A}^*	U_{1B}^*
1	P_1	$(D_1; S_2)$	$r_1^D + d_1 r_1^H + (p_2 u_2 + (1-p_2) v_2) r_2^H$	$(p_2 a_2 - (1-p_2) c_2 + r_2^S + r_2^P) f$
2	P_2	$(D_1; K_2)$	$r_1^D + d_1 r_1^H + (p_2 v_2 + (1-p_2) u_2) r_2^H$	$(-p_2 c_2 + (1-p_2) a_2 + r_2^S + r_2^P) f$
3	P_3	$(D_1; N_2)$	$r_1^D + d_1 r_1^H + r_2^H$	0
4	P_4	$(D_2; S_1)$	$r_2^D + d_2 r_2^H + (p_1 u_1 + (1-p_1) v_1) r_1^H$	$(p_1 a_1 - (1-p_1) c_1 + r_1^S + r_1^P) f$
5	P_5	$(D_2; K_1)$	$r_2^D + d_2 r_2^H + (p_1 v_1 + (1-p_1) u_1) r_1^H$	$(-p_1 c_1 + (1-p_1) a_1 + r_1^S + r_1^P) f$
6	P_6	$(D_2; N_1)$	$r_2^D + d_2 r_2^H + r_1^H$	0

defined in Appendix A.2. L_{1A}^* and U_{1B}^* are the optimal expected loss and utility for User A and User B, respectively.

Remark. Proposition 2 shows six possible SPNE strategies for User A and User B. User A chooses to clarify rumor 1 ($x^* = D_1$) at equilibrium in cases 1, 2, and 3, while she chooses to clarify rumor 2 ($x^* = D_2$) in cases 4, 5, and 6. User B supports the rumor ($y^* = S_j$) at equilibrium in cases 1 and 4, opposes in cases 2 and 5 ($y^* = K_j$), and engages in neutral participation ($y^* = N_j$) in cases 3 and 6.

3.4. Sensitivity analyses of equilibrium solutions

In this section, we study how the equilibrium solutions are sensitive to the changes in the parameters used in the rumor selection for clarification model. In order to present a consistent comparison between the objective functions of the players in sensitivity analysis, we convert the expected loss function of User A into an expected utility function, U_{1A}.

Fig. 4 (a) and 4(c) show the sensitivity in the equilibrium strategies of the players relative to parameters a_1 and c_1. It is observed that User B exhibits contrasting behaviors with respect to a_1 and c_1. At low values of a_1 and high values of c_1, he engages in neutral participation, while a high value of a_1 and a low value of c_1 increases his expected utility, which in turn, motivates him to oppose the rumor. In Fig. 4(b) and 4(d), it is observed that a high a_2 and a low c_2 encourages User B to oppose rumor 2, given that the baseline value of p_1 is higher than that of p_2. This particular strategy of User B provides an opportunity to User A to shift her focus entirely on rumor 1 for clarification.

Fig. 4 (e) and 4(g) show that User B demonstrates similar equilibrium behaviors with respect to parameters r_1^S and r_1^P. At low values of r_1^S and r_1^P, he engages in neutral participation, while higher values of r_1^S and r_1^P increases his expected utility, thus motivating him to shift his strategy to opposing the rumor. In Fig. 4(f) and 4(h), it is observed that at low values of r_2^S and r_2^P, User B focuses his attention on rumor 1 by supporting it. With increase in the values of r_2^S and r_2^P, he shifts his focus to rumor 2 and chooses to oppose it. This transition in the strategy of User B allows User A to focus on rumor 1 for clarification.

Fig. 4 (i) and 4(j) illustrates how sensitive the equilibrium strategies of the players are with respect to parameters p_1 and p_2. It is observed that at initial low values of p_i for rumor i, User B chooses to oppose the rumor fearing the risk of high damage caused by supporting a false rumor. At moderate values of p_i, User B engages in neutral participation while a sufficiently high value of p_i motivates him to support the rumor in order to increase his social influence and credibility rating. In addition to this, it is also observed that User B changes his equilibrium decision to support or oppose the other rumor in strategic regions that provide him

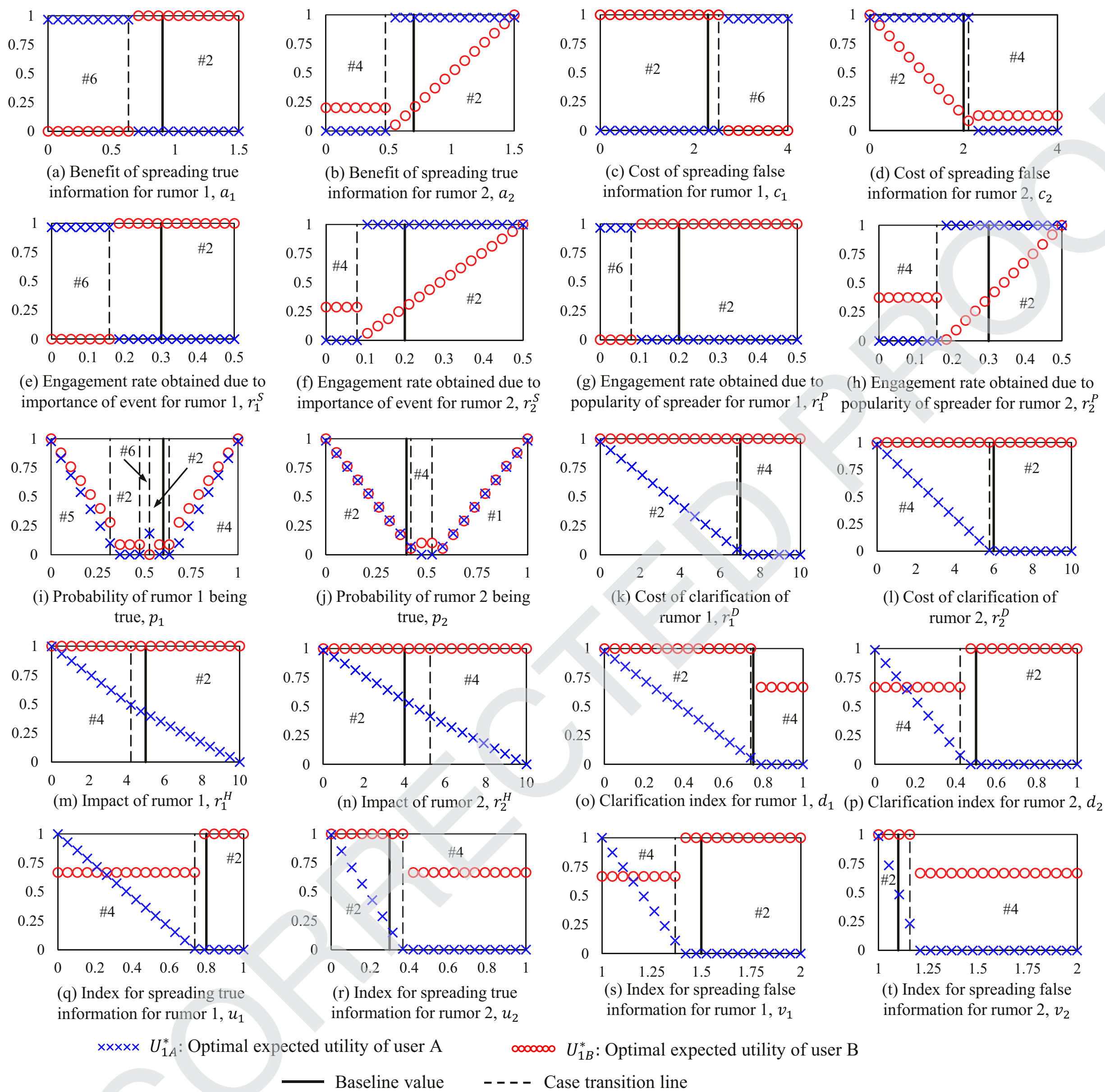

Fig. 4. Sensitivity analysis of the optimal strategies of the players and their expected utilities as one-way functions of the parameters used in the selection of rumor model.

better probabilistic opportunities. For example, in Fig. 4(i), when p_1 crosses 0.4, the baseline value of p_2, he shifts his strategy to opposing rumor 2.

Fig. 4 (k) and 4(m) show the equilibrium behaviors of players with respect to parameters r_1^D and r_1^H. It is observed that a low r_1^D and a high r_1^H motivates User A to clarify rumor 1 while a high r_1^D and a low r_1^H motivates her to clarify rumor 2. A similar pattern in the behaviors of User A is observed with respect to parameters r_2^D and r_2^H.

In Fig. 4(o), 4(r), and 4(t), similar strategic profiles are observed for User A. At low values of d_1, u_2, and v_2, User A chooses to clarify rumor 1. A low value of d_1 means that the quality of clarification for rumor 1 is comparatively better than that of rumor 2. Additionally, at low values of u_2 and v_2, User B chooses to focus his attention on rumor 2 by opposing it which decreases the expected impact of rumor 2. This provides an opportunity to User A to focus on rumor 1 for clarification. At high values of d_1, u_2, and v_2, User A shifts his attention to rumor 2 because of the decrease in the quality of clarification for rumor 1 and the change in the strategy of User B to support rumor 1 that decreases its corresponding impact, given that the baseline value of p_1 is higher than that of p_2. An opposite pattern in the equilibrium behaviors of the players is observed with respect to parameters d_2, u_1, and v_1.

3.5. Analytical implications

When the impact of a rumor is high, the cost of spreading false information for the rumor is also high, which in turn poses a high risk for User B for being criticized if they choose to support the rumor. This possibility guides User B to engage in neutral participation due to which User A's costs associated with the control of rumor propagation decreases. User B is also motivated to engage in the process of supporting or opposing a rumor based on the engagement rates that he obtains due to the importance of event and popularity of rumor spreader. User A's strategy of selecting a rumor for clarification depends on whether User B is spreading false or true information to his followers and the quality of clarifications provided by User A to the social network. To deter User B from spreading false information, User A should post verified information on social media platforms that can reduce the extent of uncertainties involved in the process of rumor diffusion, thereby enabling User B to take informed decisions. With a decrease in the extent of uncertainties involved in rumor spreading and clarification, the expected utilities of both players increase (as shown in Fig. 4(i) and 4(j)). This phenomenon bolsters the need of posting verified information on social media platforms for minimizing the transmission of false rumors. With this motivation, we develop Model 2 that incorporates the response of User A for providing verified information leading to a subsequent reduction of the impact of false rumor transmission on social media platforms.

4. Model 2: Learning for rumor clarification

During rumor propagation, User A can react fast to minimize the spread of rumor with the available information on hand. But if the available information is unverified/unproven and does not clarify all aspects of the rumor, it can leave room for speculation and lead to serious detrimental effects such as widespread panic and confusion among people. In some cases, investment in terms of effort, time and money to completely learn and verify the details of rumor for effective clarification may also allow the rumor to spread with its full force during the learning phase. In this model, we determine the equilibrium clarification strategy for User A so that she can minimize the spread of rumors during crisis events by addressing the trade-offs between reacting fast with partial/unverified information and reacting at a later stage with verified information. Model 2 identifies the strategic interactions between User A and User B during the learning phase for rumor clarification. The objective of this model is to study the impact of the User A's rumor verification strategies on User B's decision to support, oppose, or show neutral participation for a specific rumor. This is achieved by modeling the scenario of rumor verification using a sequential game model.

4.1. Notations, assumptions and description of model

Notations for Model 2 are introduced and defined in Table 3, that include three decision variables, fourteen parameters, and three functions.

In this model, User A is assumed to minimize her expected loss, L_{2A}, while the User B is assumed to maximize his expected utility, U_{2B}. The sequence of moves of players is illustrated in Fig. 5. The objectives of players in Model 2 are same as that of Model 1.

In this model, User A can either choose to clarify (D) or disregard (ND) rumor immediately, or she can choose to enter into a learning phase (L) in order to get verified information for rumor clarification. Given that User A chooses to clarify rumor immediately, User B can then decide to disseminate (Q) this information to his social network or may engage in neutral participation (N). The strategy of information dissemination by User B provides him with a benefit af, while no benefit or cost is associated with his decision of neutral participation. The engagement rate obtained by User B while supporting or opposing a rumor depends on the importance of the event (r^S) and the popularity of the rumor spreader (r^P). The effectiveness of clarification provided by User A is modeled using clarification index, d. A higher value of d signifies that the clarifications provided by User A are not sufficient to prevent User B from spreading false information to his followers, thus increasing the impact of rumor. The impact of rumor that User A seeks to minimize is dependent on the strategy of User B. When User B chooses to disseminate the clarified information posted by User A, the impact of rumor decreases by a factor du, while in case of neutral participation shown by User B, its impact decreases by a factor d.

When User A decides to disregard the rumor, User B can decide amongst three options: oppose (K), support (S), or engage in neutral participation (N). If User B supports a true rumor or opposes a false rumor, he earns benefit af for spreading true information to his followers. While if he chooses to support a false rumor or oppose a true rumor, he bears a cost cf for spreading false information to his followers. No benefit or cost is associated with the User B's decision of neutral participation. When User B supports a true rumor or opposes a false rumor, the impact of rumor decreases by a factor u. Whereas, if he chooses to support a false rumor or oppose a true rumor, the impact of rumor increases by a factor v.

In the learning phase, the probability p_v models the uncertainties that exist while obtaining verified information for rumor clarification. In addition to the costs associated with rumor clarification and the impact of rumor, there exists a time-dependent cost $r^L g(t)$ for User A to get verified information, where $g(t)$ is function of her learning period t. If User A manages to obtain verified information, she will choose to clarify the rumor using this information. In this case, User B can decide to disseminate the information to his social network or may engage in neutral participation. The effectiveness of verified information in reducing the impact of rumor is modeled using verification index, l. A higher value of l denotes that the verification provided by User A is not sufficient to convince User B, due to which User B will continue to spread false information to his followers, thus increasing the impact of rumor. Finally, if User A does not get verified information, she can choose to clarify or disregard the rumor.

In Model 2, the objective of User A is to minimize her expected loss L_{2A} by choosing x_i and $z_{j|q}$ to clarify rumor, with/without entering into the learning phase. The objective of User B is to maximize his expected utility U_{2B} by choosing y_k to disseminate, support, oppose or engage in neutral participation. Therefore, the optimization functions of both players in Model 2 can be written as shown below:

$$
\begin{aligned}
&\min_{x_i, z_{j|g}} L_{2A}(x_i, z_{j|q}, y_k) \\
&= x_D\Big((1 - y_Q + uy_Q)dr^H + r^D\Big) + x_L\Big(p_v z_{D|V}\big((1 - y_Q + uy_Q)dlr^H \\
&\quad + r^D + r^L g(t)\big) + (1 - p_v)\Big(z_{D|NV}\big((1 - y_Q + uy_Q)dr^H \\
&\quad + r^D + r^L g(t)\big) + z_{ND|NV}\big(y_K\big(pv + (1 - p)u - 1\big)r^H \\
&\quad + y_S\big(pu + (1 - p)v - 1\big)r^H + r^L g(t)\big)\Big)\Big) \\
&\quad + x_{ND}\Big(y_K\big(pv + (1 - p)u - 1\big)r^H + y_S\big(pu + (1 - p)v - 1\big)r^H\Big)
\end{aligned}
\tag{6}
$$

$$
\begin{aligned}
&\max_{y_k} U_{2B}(x_i, z_{j|q}, y_k) \\
&= x_D y_Q f(a + r^S + r^P) + x_L f\Big(p_v\Big(z_{D|V} y_Q (a + r^S + r^P)\Big) \\
&\quad +(1 - p_v)\Big(z_{D|NV} y_Q (a + r^S + r^P) + z_{ND|NV}\Big(y_K\big((1 - p)a - pc \\
&\quad +r^S + r^P\big) + y_S\big(pa - (1 - p)c + r^S + r^P\big)\Big)\Big)\Big) \\
&\quad + x_{ND} f\Big(y_K\big((1 - p)a - pc + r^S + r^P\big) \\
&\quad +y_S\big(pa - (1 - p)c + r^S + r^P\big)\Big)
\end{aligned}
\tag{7}
$$

4.2. Best response of user B

Since User B is assumed to be the second mover in Model 2, we first derive the best response of User B, $\hat{y}_n$, which is defined as

Table 3
Notations used in Model 2.

Decision Options of User A	
D	Clarify
ND	Disregard
$L_{D,D}$	Learn and clarify irrespective of getting verification
$L_{D,ND}$	Learn and clarify when verified (V) and disregard when unverified (NV)
Decision Options of User B	
Q	Disseminate
K	Oppose
S	Support
N	Do Nothing
Decision Variables	
x_i	Whether User A decides to choose option i, where $i \in \{D, ND, L\}$; $x_i \in \{0, 1\}$ and $\sum_i x_i = 1$
$z_{j\|q}$	Whether User A decides to choose option j given q while learning, where $j \in \{D, ND\}$; $q \in \{V, NV\}$ and $z_{j\|q} \in \{0, 1\}$
y_k	Whether User B decides to choose option k, where $k \in \{Q, K, S, N\}$ and $y_k \in \{0, 1\}$
Parameters	
r^D	Cost of rumor clarification
r^L	Learning cost per unit time period of User A
r^H	Impact of rumor
d	Clarification index where $d \in [0, 1]$
l	Verification index where $l \in [0, 1]$
p_v	Probability that User A will get verified information about the rumor
p	Probability of rumor being true
u	Index for spreading true information by User B where $u \in [0, 1]$
v	Index for spreading false information by User B where $v \geq 1$
f	Number of followers of User B
a	Benefit of spreading true information to each follower
c	Cost of spreading false information to each follower
r^S	Engagement rate obtained by User B due to importance of the event
r^P	Engagement rate obtained by User B due to popularity of the rumor spreader
Functions	
$L_{2A}(x_i, z_{j\|q}, y_k)$	Expected loss of User A
$U_{2B}(x_i, z_{j\|q}, y_k)$	Expected utility of User B
$g(t)$	Function of learning period t for User A

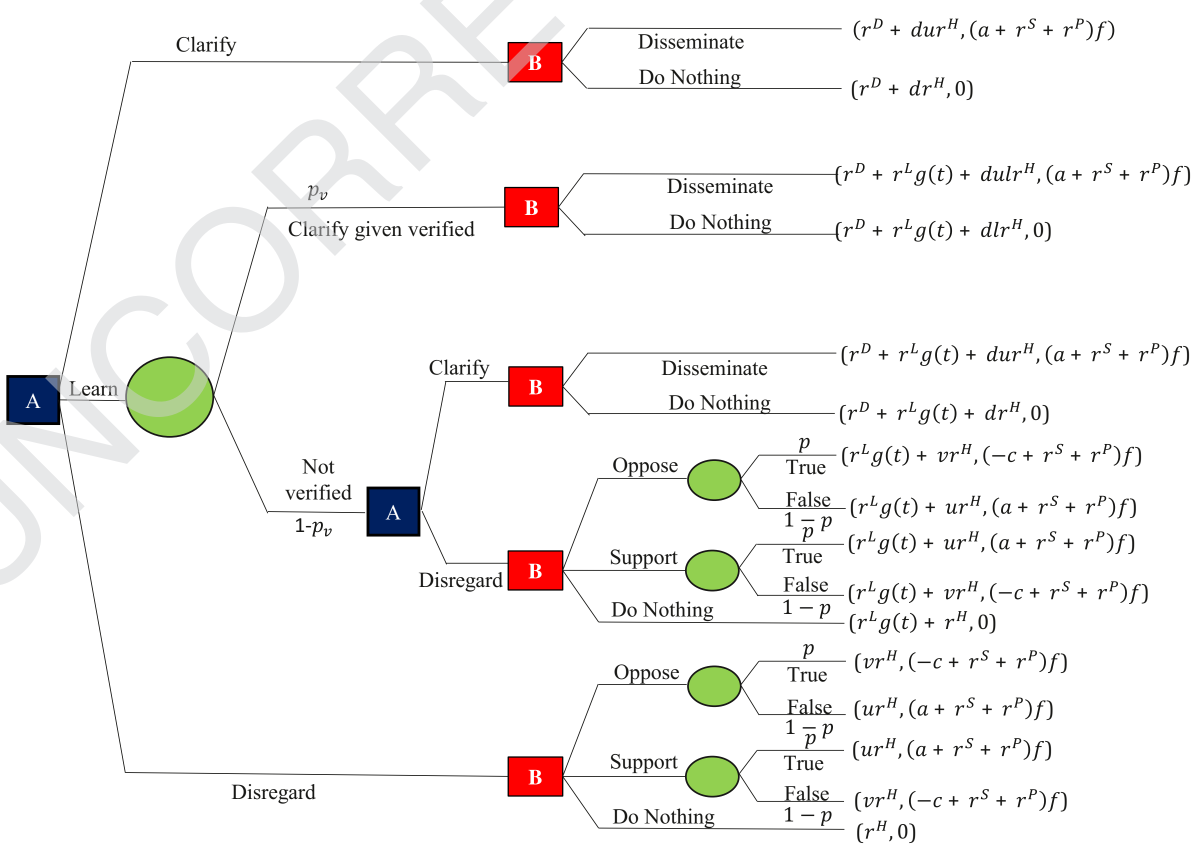

Fig. 5. Sequence of moves of players in learning for rumor clarification game.

Table 4
Equilibrium solutions of learning for rumor clarification model.

Case No.	Conditions	$(x^*; y^*)$	L^*_{2A}	U^*_{2B}
1	R_1	$(D; Q)$	$r^D + dur^H$	$(a + r^S + r^P)f$
2	R_2	$(L_{D,D}; Q, Q)$	$(p_v(l-1)+1)dur^H + r^D + r^L g(t)$	$(a + r^S + r^P)f$
3	R_3	$(L_{D,ND}; Q, K)$	$p_v(dulr^H + r^D) + (1-p_v)$ $(pv + (1-p)u - 1)r^H + r^L g(t)$	$p_v fa + (1-p_v)f$ $((1-p)a - pc) + r^S + r^P$
4	R_4	$(L_{D,ND}; Q, S)$	$p_v(dulr^H + r^D) + (1-p_v)$ $(pu + (1-p)v - 1)r^H + r^L g(t)$	$p_v fa + (1-p_v)f$ $(pa - (1-p)c) + r^S + r^P$
5	R_5	$(L_{D,ND}; Q, N)$	$p_v(dulr^H + r^D) + (1-p_v)r^H + r^L g(t)$	$p_v f(a + r^S + r^P)$
6	R_6	$(ND; K)$	$(pv + (1-p)u)r^H$	$(((1-p)a - pc) + r^S + r^P)f$
7	R_7	$(ND; S)$	$(pu + (1-p)v)r^H$	$((pa - (1-p)c) + r^S + r^P)f$
8	R_8	$(ND; N)$	r^H	0

follows:

$$\hat{y}_n \equiv \underset{y_k \in \{0,1\}}{\text{argmax}}\ U_{2B}(x_i, z_{j|q}, y_k), \ \text{where } n = 1, 2, \ldots, 5 \tag{8}$$

Proposition 3. *The best response function of User B, $\hat{y}_n$, is given by:*
For $n = 1, 2,$ and 3, the strategy Q is strictly dominant over N, therefore,

$$\hat{y}_n \equiv Q \tag{9}$$

For $n = 4$ and 5,

$$\hat{y}_n \equiv \begin{cases} S & \text{if} \quad p \geq \max\left(\frac{c-(r^S+r^P)}{c+a}, \frac{1}{2}\right) \\ K & \text{if} \quad p \leq \min\left(\frac{a+(r^S+r^P)}{c+a}, \frac{1}{2}\right) \\ N & \text{if} \quad p \in \left(\frac{a+(r^S+r^P)}{c+a}, \frac{c-(r^S+r^P)}{c+a}\right) \end{cases} \tag{10}$$

Remark. Proposition 3 identifies the boundary conditions for different response strategies of User B. When User A chooses to clarify the rumor ($n = 1, 2,$ and 3), User B's strategy is to disseminate this information to his social network, irrespective of the variations in model parameters. On the other hand, when User A decides to disregard the rumor ($n = 4$ and 5), User B can then choose amongst three options, that is, support, oppose or engage in neutral participation. The boundary conditions for these response strategies of User B is similar to the ones explained in Proposition 1.

4.3. Equilibrium solutions

Proposition 4. *The SPNE solutions of the learning for rumor clarification model along with the optimal expected loss/utility of each player are provided in Table 4, where R_m, $m = 1, 2, \ldots, 8$ are the optimal conditions defined in Appendix A.4. L^*_{2A} and U^*_{2B} are the optimal expected loss and utility for Users A and B, respectively.*

Remark. In Proposition 4, eight possible SPNE strategies for Users A and B are shown. User A chooses to clarify rumor instantaneously at equilibrium ($x^* = D$) in case 1, she chooses to disregard rumor instantaneously ($x^* = ND$) in cases 6, 7, and 8, and she chooses to enter into the learning phase to obtain verified information ($x^* = L$) in cases 2, 3, 4, and 5. If User A's equilibrium strategy is to clarify immediately, User B chooses to disseminate the information posted by User A ($y^* = Q$). When User A disregards the rumor immediately at equilibrium, User B can choose to oppose ($y^* = K$), support ($y^* = S$), or engage in neutral participation ($y^* = N$). In the learning phase, User A can choose to clarify at equilibrium irrespective of getting verified information ($x^* = L_{D,D}$), in which case User B chooses to disseminate the information provided by User A ($y^* = Q, Q$). The other equilibrium strategy of User A in learning phase is to clarify when verification is obtained and disregard when the rumored information remains unproven ($x^* = L_{D,ND}$). In these cases, there are three possible combinations of equilibrium strategies for User B, that is, ($y^* = Q, K$), ($y^* = Q, S$), and ($y^* = Q, N$).

4.4. Sensitivity analyses of equilibrium solutions

In this section, we study how the equilibrium solutions are sensitive to the changes in the parameters of learning for rumor clarification model. For a consistent comparison between the objective functions of the players in sensitivity analysis, we convert the expected loss function of User A into an expected utility function, U_{2A}. For numerical illustrations, the baseline values of the parameters used in this model are assumed as follows: $p = 0.6$, $p_v = 0.8$, $a = 0.8$, $c = 2.3$, $r_S = 0.3$, $r_P = 0.2$, $f = 250$, $r^D = 4.0$, $r^H = 5.0$, $r^L = 0.2$, $d = 0.5$, $l = 0.3$, $t = 2$, $u = 0.8$, and $v = 1.5$.

Fig. 6 (a) and 6(b) show the sensitivity in the equilibrium strategies of the players relative to parameters a and c. In these figures, it is observed that the behavior of User B with respect to a is in complete contrast to that of c. At higher a and lower c, User B chooses to support the rumor, given that User A disregards the rumor when no verification is obtained. This equilibrium strategy of the User B arises due to the baseline value of p being 0.6.

Fig. 6 (c) and 6(d) show that User B demonstrates similar equilibrium behaviors with respect to parameters r^S and r^P. At low values of r^S and r^P, he engages in neutral participation, while higher values of r^S and r^P increases his expected utility, thus motivating him to shift his strategy to supporting the rumor.

Fig. 6 (e) illustrates the sensitivity in the equilibrium strategies of the players relative to parameter p. At low or high p, the extent of uncertainties involved regarding the nature of rumor being true or false is low, thereby encouraging User A to react faster without entering into learning process. Moderate values of p result into greater uncertainties about the nature of rumor, which in turn motivates User A to obtain verified information via learning process.

In Fig. 6(k), the strategies of the players are analysed with respect to l, an index that quantifies the effectiveness of learning process on the impact of rumor. At low values of l, User A is highly likely to enter into the learning phase to obtain verified information, while a high l switches her strategy to completely disregard the rumor. The strategy of User A to enter into the learning phase is also dependent on parameters p_v and r^L, as shown in Fig. 6(f) and 6(i), respectively. It is observed that the equilibrium behaviors of the players with respect to r^L are similar to that of l, while with respect to p_v, the results are in complete contrast with that of l.

Fig. 6 (g) and 6(h) show the equilibrium behaviors of players with respect to parameters r^D and r^H. It is observed that a low r^D and a high r^H motivate User A to clarify the rumor by entering into learning phase to obtain verified information, while a high r^D and a low r^H change her strategy to react faster without going into the process of learning.

In Fig. 6(j), the equilibrium behaviors of the players are analyzed with respect to variations in d. This index quantifies the ef-

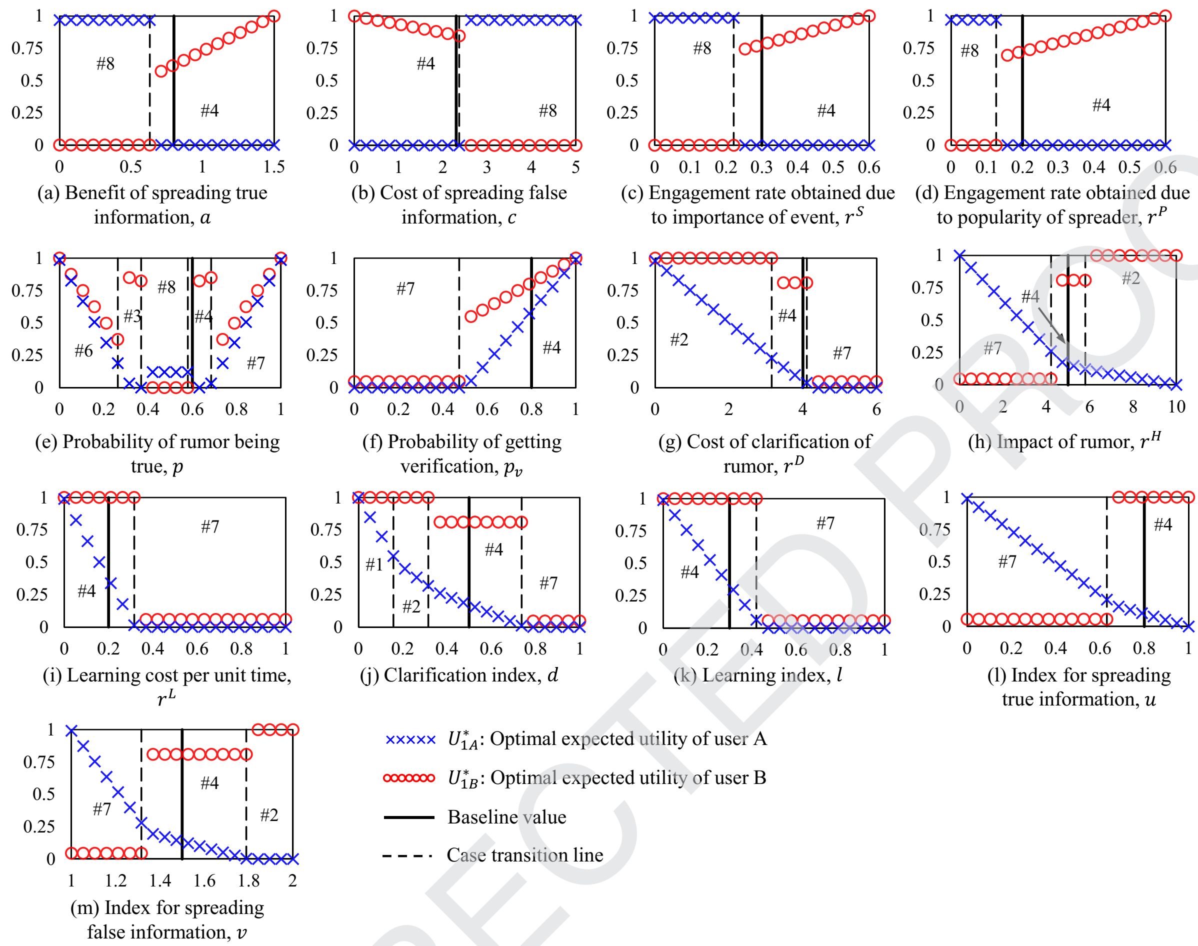

Fig. 6. Sensitivity analysis of the optimal strategies of the players and their expected utilities as one-way functions of the parameters used in the learning for rumor clarification model.

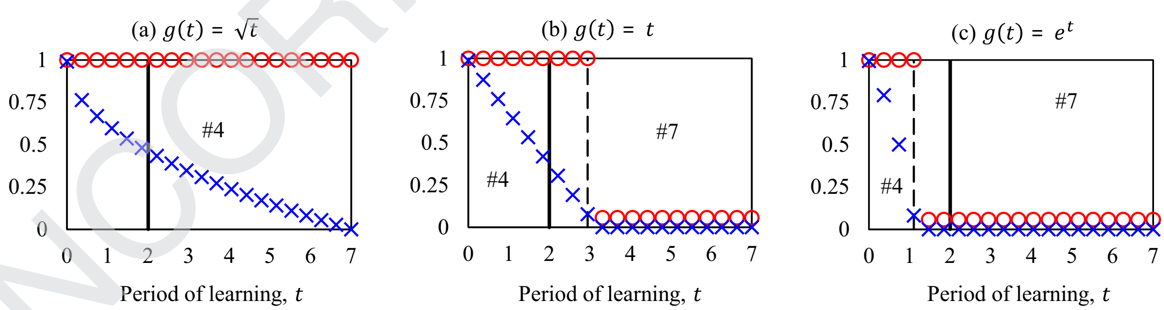

Fig. 7. Sensitivity analysis of the optimal strategies of the players and their expected utilities with respect to different functional forms of learning period t.

fectiveness of clarification on reducing the impact of rumor. It is observed that at low values of d, User A chooses to clarify the rumor instantaneously; at moderate values of d, she enters into a learning phase to get verified information for clarification; and at high values of d, she completely disregards the rumor.

Fig. 6 (l) and 6(m) shows the changes in the equilibrium strategies of players with respect to variations in u and v. A high value of u and v increases the impact of rumor, thereby acting as factors of motivation for User A to enter into learning process for obtaining verified information.

In Fig. 7, the equilibrium behaviors of the players are analyzed with respect to variations in the functional forms of $g(t)$, where the rate of change of t is decreasing, constant, and increasing in Fig. 7(a), 7(b), and 7(c), respectively. In Fig. 7(b) where $g(t) = t$, it is observed that at low values of t, User A enters into the learning phase to obtain verified information, while a high value of t switches her strategy to completely disregard the rumor. On moving from $g(t) = \sqrt{t}$ to $g(t) = e^t$, the value of t at which the strategy of User A shows transition also decreases. This shows that if the time to get verification grows exponentially, User A is highly unlikely to spend time in the learning process.

Fig. 8 shows that at low values of u, User A chooses to clarify or disregard rumor instantaneously since a low value of u corresponds to a lower expected impact of the rumor. At higher values of u, the impact of rumor as a result of User B's behavior increases, thus motivating User A to obtain verified information by entering into the learning phase. At low and high values of p, the uncertainty regarding the rumor being true or false is low due to which

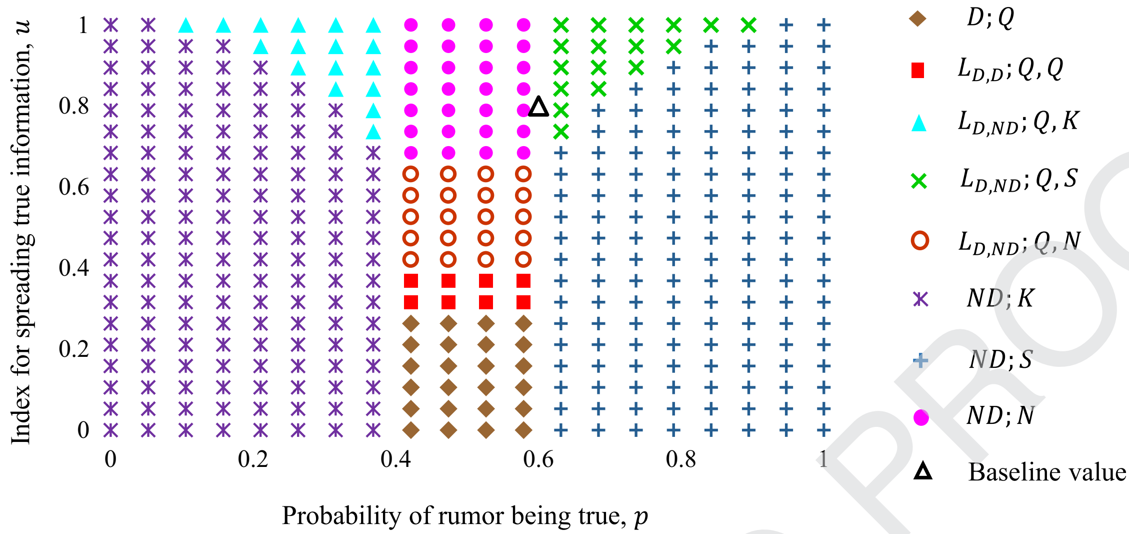

Fig. 8. Sensitivity analysis of the optimal strategies of the players as two-way functions of p and u.

User A chooses to react fast without engaging in learning. On the other hand, for moderate values of p, User A is motivated to spend time in the learning process for obtaining verified information.

4.5. Analytical implications

When the impact of rumor is high, the results show that User A should engage herself in the learning process to get detailed verified information about the rumor that is being spread on social media platforms. This strategy minimizes the risk of significant rumor transmission as a result of a faster reaction with unverified/unproven information. When User A clarifies the rumor with verification obtained from different sources, User B becomes much more confident in disseminating the information posted by User A. However, this engagement of User A in the learning process is constrained on two factors, that is, the cost of learning and period of learning. If the cost and period of learning are beyond the reasonable limits, the reduction obtained on the impact of rumor using verified information may not be significant enough to motivate User A to engage in the learning process. The rate of increase/decrease of the time taken to get verified information also influences the decision of User A to participate in the learning process. Our results also show that entering into the learning phase also helps User A deal with the uncertainties regarding the nature of rumor being true or false.

5. Validation of the strategies of social media user

In this section, we analyze the strategies of social media users using data collected from Twitter across seven different rumor cases, as shown in Table 5. The criteria for choosing and collecting these datasets were based upon their large-scale news coverage and the availability of the data on Twitter. The false rumor from the Boston Marathon bombing was broadcast across the online environment, and was identified through news outlets and social media platforms (Sager, 2013). For the Hurricane Harvey and Hurricane Irma rumors, the cases were identified on FEMA's Rumor Control pages (Federal Emergency Management Agency, 2017). News from the 2018 Hawaiian incoming missile and Tsunami warning false alerts were broadcasted online, on the radio, and on television. A brief description about these cases are provided in Appendix A.5.

Twitter's Search API (Twitter Search API, 2020) was used for collecting all of the tweets in this research. Twitter's Standard Search API returns tweets from the previous seven days based on user-specified search criteria. Data used in this study only contains a sample of tweets that were returned by the API based on our search queries since the API does not provide an exhaustive list of tweets. To resolve this issue and collect more comprehensive datasets, the tweets were collected over a 28-day window for every case, with collection performed every three days by using the same search criteria every time. In total between the seven cases, we collected 18,314 tweets. The queries used for all cases were a combination of case insensitive keywords and hashtags (e.g., immigration and #harvey; hawaii and #missile). The queries were selected following an extensive Twitter Advanced Search (Twitter Advanced Search, 2020) to find the pertinent keywords and hashtags that identified tweets related to the rumors being studied in this paper.

Utilizing latent content analysis (Hunt, Wang, & Zhuang, 2020; Wang & Zhuang, 2017) and following the rules suggested by Krippendorff (2013) and Landis & Koch (1977), the text of each tweet was coded to identify the stance of the user with respect to the specific rumor case. Three researchers ("coder 1," "coder 2," and "coder3") participated in the coding process for all of the tweets. The coders were required to become familiar with all seven cases of rumor in this study before coding began. Coders 1 and 2 independently coded all of the tweets into the following three mutually exclusive classes: support, oppose, and neutral participation. After coders 1 and 2 completed the datasets, coder 3 then cross-validated all of the tweets in which 1 and 2 disagreed on the class. In this study, we analyze the tweets related to classes, support and oppose, since these rumor cases were clarified as being true or false by the official agencies. This clarification act as a complete information regarding the values of the probabilities for rumor cases being true.

Table 6 shows the description of rumor clarifications provided by different official accounts for each rumor case. The criteria used to select the first clarifying post were - (a) it must be posted by an official account of a verified user, and (b) it must be the most shared post (in terms of the number of retweets and likes) among the posts that were made on the first day of clarification. As shown in this table, a majority of these clarifications are provided by the news agencies while in cases such as Hurricane Harvey Hiring and

Table 5
Summary statistics, collection dates, and total tweets collected for the seven rumor cases.

Rumor Case	Rumor Type	# Likes per Unique Tweet	# Retweets per Unique Tweet	Collection Began	Collection Ended	Collected Tweets
Hurricane Irma Fuel Demand	True	33	20	September 01, 2017	September 28, 2017	432
Hurricane Harvey Hiring	True	5	6	August 25, 2017	September 21, 2017	435
Hurricane Irma Immigration	False	26	16	September 09, 2017	October 06, 2017	594
Boston Marathon Bombing Donation	False	1	11	April 18, 2013	May 15, 2013	650
Tsunami Warning False Alert	False	3	2	February 04, 2018	March 03, 2018	7,478
Hawaii Missile False Alert	False	17	7	January 14, 2018	February 10, 2018	6,691
Hurricane Harvey Immigration	False	210	102	August 28, 2017	September 24, 2017	2,034
Total Collected Tweets						**18,314**

Table 6
Description of clarification provided by the official accounts for each rumor case.

Rumor Case	Official Account	Account Description	Clarification Posted On	Retweets for Clarification Post	Likes for Clarification Post	First Tweet Posted On
Hurricane Irma Fuel Demand	Reuters	News Agency	September 06, 2017	109	82	September 05, 2017
Hurricane Harvey Hiring	WCraigFugate	Former FEMA Administrator	August 30, 2017	115	71	August 28, 2017
Hurricane Harvey Immigration	washingtonpost	News Agency	August 25, 2017	561	1496	August 22, 2017
Hurricane Irma Immigration	NPR	News Agency	September 06, 2017	303	618	September 05, 2017
Hawaii Missile False Alert	ABCWorldNews	News Agency	January 13, 2018	50	64	January 13, 2018
Tsunami False Alert	NBCNews	News Agency	February 06, 2018	637	677	February 06, 2018
Boston Marathon Bombing Donation	darrenrovell	News Reporter	April 15, 2013	1957	158	April 15, 2013

Boston Marathon Bombing Donation, the clarification message is posted by individuals associated with the official agencies. It is observed that the official agencies usually take one to three days to verify and provide clarifications. In highly sensitive cases such as Hawaii Missile and Tsunami False Alerts, the clarification is provided within the same day of the first rumored post to prevent widespread panic and confusion.

Propositions 1 and 3 claim that when the probability of a rumor being true is low, User B chooses to oppose the rumor, while at high values of this probability, the best response strategy of User B is to support the rumor. These propositions are further validated using real life rumor cases during crisis events, as shown in segment *Overall Response* of Fig. 9. In this segment, it is observed that for true rumors, most of the users chose to support these rumors. Whereas, in case of false rumors, majority of the users responded by opposing the rumor. The importance of verifying rumors and providing clarified information as prescribed by the model results are further validated by segments *Before Clarification* and *After Clarification* of Fig. 9. For false rumors, the percentage of users showing opposition increases up to 44.4% after the first clarifying post is made by an official account. In case of true rumors, the increase in the percentage of users showing support is found to be up to 6.7%. This shows that the influence of a clarification post is relatively more impactful for false rumors as compared to true rumors.

6. Conclusions and future research directions

In times of crises, millions of people turn to social media for breaking news updates, evacuation planning, situational awareness, safety protocols, among many other emergency needs. Although there are many significant benefits associated with social media platforms, there are also certain characteristics which can lead to a dangerous social environment. Unfortunately, due to the unmoderated nature of social media platforms, rumors often spread, reaching and influencing people around the world. In order to clarify rumors, official agencies and social media companies must expend human resources and time in order to locate rumors on social media, track the rumors in order to understand their reach and impact, and formulate rumor clarification messages.

In a rumor propagation and clarification process, the players make different strategic decisions by taking into account the trade-offs between the cost involved while spreading/clarifying, and the impact of rumor in terms of social reach and losses. Given the dearth of existing game-theoretic works on rumor propagation and clarification, we develop two novel game-theoretic models to study the strategic interactions that take place between Users A and B in a rumor transmission process. In these models, we determine the SPNE strategies of the players and identify the equilibrium conditions that motivate/demotivate the players to engage in a rumor transmission and clarification process. We also perform the numerical sensitivity analyses of the equilibrium strategies of the players and their expected utilities as functions of the parameters used in the models. Results of the sensitivity analyses help us to identify the relative threshold at which the strategies taken by the players undergo transition. The results from the models indicate that posting verified information on social media reduces the uncertainties involved in rumor transmission, thereby enabling User B to take informed decisions on whether to support or oppose the rumor being circulated. This verification should be obtained within reasonable limits of time and cost in order to motivate User A to engage in the learning process. The prescriptive insights obtained from this paper will be useful to inform decision makers about the behaviors of Users A and B in a rumor transmission and clarification process under different strategic conditions, which in turn will improve the rumor information dissemination and control practices during crisis events.

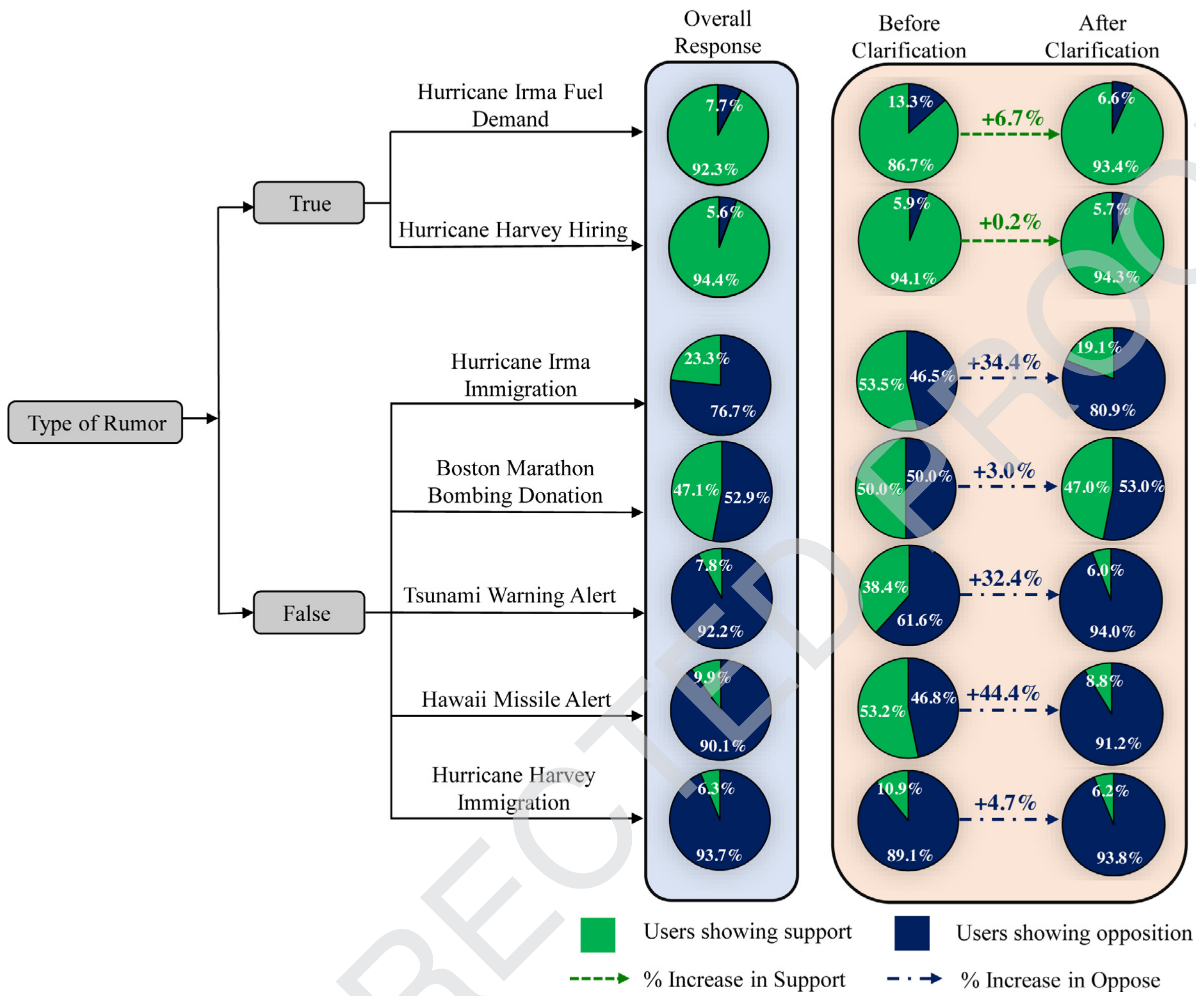

Fig. 9. Validation of the strategies of social media users based on real life rumor cases.

Social media companies are implementing policies in their fight against rumors, misinformation, and disinformation by taking down inauthentic behavior, labeling misleading information, working closely with civil society groups, and engaging with researchers and governments. Before 2020 U.S. elections, LinkedIn, Pinterest, Reddit, Verizon Media, and theWikimedia Foundation joined Google, Facebook, Twitter, and Microsoft to coordinate with the U.S. intelligence community to identify disinformation campaigns (Ben Nimmo, 2020; Isaac & Conger, 2020). Twitter uses a framework to label and remove manipulated or synthetic media and misleading information intended to undermine public trust on democratic policies and events such as elections (Dawn.com, 2020). In September 2020, Twitter built a U.S. Election hub that provided credible and verified news and voting resources to the social media users for making informed decisions during elections (Gadde & Beykpour, 2020). Model 1 will help the social media companies to select the cases/posts to label/remove based on the users' behaviors, the impact of the cases on users, the importance of event, and the popularity of rumor spreader. Model 2 will help the agencies and social media companies to make decisions regarding engaging in the learning process to provide verified and credible information during rumor clarification.

In this paper, it is assumed that the players have complete knowledge of their opponent's objectives, payoffs, beliefs, and possible actions. However, in real life, different types of social media users may have different set of objectives and beliefs while participating in rumor clarification and verification processes. In future, this can be addressed by using Adversarial Risk Analysis framework (Rios Insua, Ríos, & Banks, 2009) where probabilistic measures are used to define and assess the players' possible actions based on the uncertainties about the players' decision problem.

In the models that are developed in this paper, the players are assumed to be risk neutral. To build a more robust model, the players could be allowed to have risk preferences. For example, an official agency or a social media user with a large network of followers is likely to have a risk averse profile since any controversial post/statement made by him/her on social media can draw a lot of criticism and a subsequent fall in social influence. Whereas, a user account with a small network of users will be more willing to take risks while supporting or opposing a rumor. One of the ways of incorporating these risk preferences into the expected utility functions of the players is by considering a power utility function of the form: $U(x) = x^{\beta}$, where $0 < \beta < 1$, $\beta = 1$, and $\beta > 1$ covers risk-averse, risk-neutral, and risk-seeking behaviors of the players (Payyappalli, Zhuang, & Jose, 2017).

In this paper, we model the decisions of User A to clarify/verify a rumor using binary variables. But in the real world, User A may select levels of effort to clarify two or more rumors simultaneously and to verify a particular rumor. In that case, the level of effort used by User A might influence the decisions of User B to oppose/support a given rumor. A high level of effort might also increase User B's trust on User A's ability to clarify and verify a rumor being circulated on social media. In real-life situations, User A may receive multiple rumor cases pertaining to a disaster event with a possibility of rumor spreading by multiple user accounts. These models can be more realistic by incorporating N rumor cases and the interaction of User A with N user accounts. In this paper, we assume that the probability of a rumor being true is independent of the strategies of the players and their corresponding subjective assessments. However, in reality, each player is likely to have a different set of beliefs regarding the value of the probability that can be considered as a future research direction.

In this study, User B is assumed to maximize his influence and credibility ratings in the social networks. While in practical scenarios, an unscrupulous User B may have no concerns about veracity and may engage due to different objectives such as monetizing by running advertisements, trying to seed polarization or misinformation on purpose, etc. The variability in User B's objectives can be modeled by identifying different user types (for example - trolls, reputed personalities) that are active in social media platforms. Future research can also extend the current study to a multi-stage game model to consider continued review by the players as content becomes more viral and information is gradually revealed. Similarly, the models could also consider repeated game interactions between the players where the present course of actions of User A affects its reputation and ability to influence User B in the future.

The efforts taken by User A to clarify a particular rumor is also dependent on the strength of connections between different users in a social network. The stronger the strength of connections between the users, it is likely that the spread of rumor will be faster due to which User A needs to invest more time and resources in clarifying and verifying the rumor effectively. In future, the strength of ties between the users in a given social network can be considered to study its impact on the propagation patters of rumors, behaviors of users, and clarification and verification strategies of official agencies and social media companies. In addition, future research could also consider unsupervised techniques such as sentiment analysis to separate the data into different clusters. Research and developments in this domain could remove the need for labeling the data if unsupervised machine learning approaches could automatically identify and create the different classes.

Appendix A

A1. Proof of Proposition 1

Expected utility of the social media user while supporting the jth rumor,

$$U_{1B}(y_n = S_j) = \left(p_j a_j - (1-p_j)c_j + r_j^S + r_j^P\right) f \tag{12}$$

Expected utility of the social media user while opposing the jth rumor,

$$U_{1B}(y_n = K_j) = \left(-p_j c_j + (1-p_j)a_j + r_j^S + r_j^P\right) f \tag{13}$$

Expected utility of the social media user while doing nothing regarding the jth rumor,

$$U_{1B}(y_n = N_j) = 0 \tag{14}$$

For $\hat{y}_n = S_j$, the following condition must hold:

$$U_{1B}(y_n = S_j) \geq U_{1B}(y_n = K_j), \text{ and} \tag{15}$$

$$U_{1B}(y_n = S_j) \geq U_{1B}(y_n = N_j) \tag{16}$$

On solving inequalities (15) and (16), we get the following condition:

$$p_j \geq \max\left(\frac{c_j - (r_j^S + r_j^P)}{c_j + a_j}, \frac{1}{2}\right) \tag{17}$$

For $\hat{y}_n = K_j$, the following condition must hold:

$$U_{1B}(y_n = K_j) \geq U_{1B}(y_n = S_j), \text{ and} \tag{18}$$

$$U_{1B}(y_n = K_j) \geq U_{1B}(y_n = N_j) \tag{19}$$

On solving inequalities (18) and (19), we get the following condition:

$$p_j \leq \min\left(\frac{a_j + (r_j^S + r_j^P)}{c_j + a_j}, \frac{1}{2}\right) \tag{20}$$

For $\hat{y}_n = N_j$, the following condition must hold:

$$U_{1B}(y_n = N_j) \geq U_{1B}(y_n = S_j), \text{ and} \tag{21}$$

$$U_{1B}(y_n = N_j) \geq U_{1B}(y_n = K_j) \tag{22}$$

On solving inequalities (21) and (22), we get the following condition:

$$p_j \in \left(\frac{a_j + (r_j^S + r_j^P)}{c_j + a_j}, \frac{c_j - (r_j^S + r_j^P)}{c_j + a_j}\right) \tag{23}$$

A2. Proof of Proposition 2

We substitute the best response function of the social media user defined in Eq. (3) into the expected loss function of the agency defined in Eq. to obtain the following expressions for L_{1A} in terms of $\hat{y}_n$:

$$L_{1A}(x = D_1, \hat{y}_n = S_2) = r_1^D + d_1 r_1^H + \left(p_2 u_2 + (1-p_2)v_2\right) r_2^H, \tag{24}$$

$$\text{subject to } C_1 \equiv p_2 \geq \max\left(\frac{c_2 - (r_2^S + r_2^P)}{c_2 + a_2}, \frac{1}{2}\right)$$

$$L_{1A}(x = D_1, \hat{y}_n = K_2) = r_1^D + d_1 r_1^H + \left(p_2 v_2 + (1-p_2)u_2\right) r_2^H, \tag{25}$$

$$\text{subject to } C_2 \equiv p_2 \leq \min\left(\frac{a_2 + (r_2^S + r_2^P)}{c_2 + a_2}, \frac{1}{2}\right)$$

$$L_{1A}(x = D_1, \hat{y}_n = N_2) = r_1^D + d_1 r_1^H + r_2^H, \tag{26}$$

$$\text{subject to } C_3 \equiv p_2 \in \left(\frac{a_2 + (r_2^S + r_2^P)}{c_2 + a_2}, \frac{c_2 - (r_2^S + r_2^P)}{c_2 + a_2}\right)$$

$$L_{1A}(x = D_2, \hat{y}_n = S_1) = r_2^D + d_2 r_2^H + \left(p_1 u_1 + (1-p_1)v_1\right) r_1^H, \tag{27}$$

$$\text{subject to } C_4 \equiv p_1 \geq \max\left(\frac{c_1 - (r_1^S + r_1^P)}{c_1 + a_1}, \frac{1}{2}\right)$$

$$L_{1A}(x = D_2, \hat{y}_n = K_1) = r_2^D + d_2 r_2^H + \left(p_1 v_1 + (1-p_1)u_1\right) r_1^H, \tag{28}$$

$$\text{subject to } C_5 \equiv p_1 \leq \min\left(\frac{a_1 + (r_1^S + r_1^P)}{c_1 + a_1}, \frac{1}{2}\right)$$

$$L_{1A}(x = D_2, \hat{y}_n = N_1) = r_2^D + d_2 r_2^H + r_1^H, \tag{29}$$

$$\text{subject to } C_6 \equiv p_1 \in \left(\frac{a_1 + (r_1^S + r_1^P)}{c_1 + a_1}, \frac{c_1 - (r_1^S + r_1^P)}{c_1 + a_1}\right)$$

For $(x^*, y^*) = (D_1, S_2)$, the following condition must hold:

$$P_1 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = S_2) \le L_{1A}(x = D_2, \hat{y}_n = S_1) \big) \cap \big(C_1 \cap C_4 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = S_2) \le L_{1A}(x = D_2, \hat{y}_n = K_1) \big) \cap \big(C_1 \cap C_5 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = S_2) \le L_{1A}(x = D_2, \hat{y}_n = N_1) \big) \cap \big(C_1 \cap C_6 \big) \Big\} \Big\} \quad (30)$$

For $(x^*, y^*) = (D_1, K_2)$, the following condition must hold:

$$P_2 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = K_2) \le L_{1A}(x = D_2, \hat{y}_n = S_1) \big) \cap \big(C_2 \cap C_4 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = K_2) \le L_{1A}(x = D_2, \hat{y}_n = K_1) \big) \cap \big(C_2 \cap C_5 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = K_2) \le L_{1A}(x = D_2, \hat{y}_n = N_1) \big) \cap \big(C_2 \cap C_6 \big) \Big\} \Big\} \quad (31)$$

For $(x^*, y^*) = (D_1, N_2)$, the following condition must hold:

$$P_3 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = N_2) \le L_{1A}(x = D_2, \hat{y}_n = S_1) \big) \cap \big(C_3 \cap C_4 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = N_2) \le L_{1A}(x = D_2, \hat{y}_n = K_1) \big) \cap \big(C_3 \cap C_5 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_1, \hat{y}_n = N_2) \le L_{1A}(x = D_2, \hat{y}_n = N_1) \big) \cap \big(C_3 \cap C_6 \big) \Big\} \Big\} \quad (32)$$

For $(x^*, y^*) = (D_2, S_1)$, the following condition must hold:

$$P_4 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = S_1) \le L_{1A}(x = D_1, \hat{y}_n = S_2) \big) \cap \big(C_4 \cap C_1 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = S_1) \le L_{1A}(x = D_1, \hat{y}_n = K_2) \big) \cap \big(C_4 \cap C_2 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = S_1) \le L_{1A}(x = D_1, \hat{y}_n = N_2) \big) \cap \big(C_4 \cap C_3 \big) \Big\} \Big\} \quad (33)$$

For $(x^*, y^*) = (D_2, K_1)$, the following condition must hold:

$$P_5 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = K_1) \le L_{1A}(x = D_1, \hat{y}_n = S_2) \big) \cap \big(C_5 \cap C_1 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = K_1) \le L_{1A}(x = D_1, \hat{y}_n = K_2) \big) \cap \big(C_5 \cap C_2 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = K_1) \le L_{1A}(x = D_1, \hat{y}_n = N_2) \big) \cap \big(C_5 \cap C_3 \big) \Big\} \Big\} \quad (34)$$

For $(x^*, y^*) = (D_2, N_1)$, the following condition must hold:

$$P_6 \equiv \Big\{ \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = N_1) \le L_{1A}(x = D_1, \hat{y}_n = S_2) \big) \cap \big(C_6 \cap C_1 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = N_1) \le L_{1A}(x = D_1, \hat{y}_n = K_2) \big) \cap \big(C_6 \cap C_2 \big) \Big\} \cup \Big\{ \big(L_{1A}(x = D_2, \hat{y}_n = N_1) \le L_{1A}(x = D_1, \hat{y}_n = N_2) \big) \cap \big(C_6 \cap C_3 \big) \Big\} \Big\} \quad (35)$$

From Eqs. (24)–(29), it is observed that $C_1 \cap C_2$, $C_1 \cap C_3$, $C_2 \cap C_3$, $C_4 \cap C_5$, $C_4 \cap C_6$, and $C_5 \cap C_6$ yield empty set, due to which they are not taken into consideration while determining the optimal conditions for the SPNE solutions in the selection of rumor model.

A3. Proof of Proposition 3

Expected utility of the social media user while disseminating the clarified information,

$$U_{2B}(y_n = Q) = (a + r^S + r^P) f \quad (36)$$

Expected utility of the social media user while supporting the rumor,

$$U_{2B}(y_n = S) = \big(pa - (1-p)c + r^S + r^P \big) f \quad (37)$$

Expected utility of the social media user while opposing the rumor,

$$U_{2B}(y_n = K) = \big(-pc + (1-p)a + r^S + r^P \big) f \quad (38)$$

Expected utility of the social media user while doing nothing regarding the rumor,

$$U_{2B}(y_n = N) = 0 \quad (39)$$

For $\hat{y}_n = S$, the following condition must hold:

$$U_{2B}(y_n = S) \ge U_{2B}(y_n = K), \ \text{and} \quad (40)$$

$$U_{2B}(y_n = S) \ge U_{2B}(y_n = N) \quad (41)$$

On solving inequalities (40) and (41), we get the following condition:

$$p \ge \max\left(\frac{c - (r^S + r^P)}{c + a}, \frac{1}{2} \right) \quad (42)$$

For $\hat{y}_n = K$, the following condition must hold:

$$U_{2B}(y_n = K) \ge U_{2B}(y_n = S), \ \text{and} \quad (43)$$

$$U_{2B}(y_n = K) \ge U_{2B}(y_n = N) \quad (44)$$

On solving inequalities (43) and (44), we get the following condition:

$$p \le \min\left(\frac{a + (r^S + r^P)}{c + a}, \frac{1}{2} \right) \quad (45)$$

For $\hat{y}_n = N$, the following condition must hold:

$$U_{2B}(y_n = N) \ge U_{2B}(y_n = S), \ \text{and} \quad (46)$$

$$U_{2B}(y_n = N) \ge U_{2B}(y_n = K) \quad (47)$$

On solving inequalities (46) and (47), we get the following condition:

$$p \in \left(\frac{a + (r^S + r^P)}{c + a}, \frac{c - (r^S + r^P)}{c + a} \right) \quad (48)$$

A4. Proof of Proposition 4

We substitute the best response function of the social media user defined in Eqs. (9) and (10) into the expected loss function of the agency defined in Eq. (6) to obtain the following expressions for L_{2A} in terms of $\hat{y}_n$:

$$L_{2A}(x = D, \hat{y}_n = Q) = r^D + dur^H \quad (49)$$

$$L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) = p_v(dulr^H - dur^H) + dur^H + r^D + r^L t \quad (50)$$

$$L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) = p_v(dulr^H + r^D + r^L t) + (1 - p_v)(pvr^H - pur^H + ur^H - r^H + r^L t),$$

$$\text{subject to } F_1 \equiv p \leq \min\left(\frac{a}{c+a}, \frac{1}{2}\right) \tag{51}$$

1082

$$L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) = p_v(dulr^H + r^D + r^L t) + (1 - p_v)(pur^H - pvr^H + vr^H - r^H + r^L t),$$
$$\text{subject to } F_2 \equiv p \geq \max\left(\frac{c}{c+a}, \frac{1}{2}\right) \tag{52}$$

1083

$$L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) = p_v(dulr^H + r^D) + r^L t,$$
$$\text{subject to } F_3 \equiv p \in \left(\frac{a}{c+a}, \frac{c}{c+a}\right) \tag{53}$$

1084

$$L_{2A}(x = ND, \hat{y}_n = K) = pvr^H - pur^H + ur^H,$$
$$\text{subject to } F_4 \equiv p \leq \min\left(\frac{a}{c+a}, \frac{1}{2}\right) \tag{54}$$

1085

$$L_{2A}(x = ND, \hat{y}_n = S) = pur^H - pvr^H + vr^H,$$
$$\text{subject to } F_5 \equiv p \geq \max\left(\frac{c}{c+a}, \frac{1}{2}\right) \tag{55}$$

1086

$$L_{2A}(x = ND, \hat{y}_n = N) = r^H, \quad \text{subject to } F_6 \equiv p \in \left(\frac{a}{c+a}, \frac{c}{c+a}\right) \tag{56}$$

1087 For $(x^*; y^*) = (D; Q)$, the following condition must hold:

$$\begin{aligned} R_1 \equiv \Bigg\{ & \Big\{ L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \Big\} \\ \cap & \Big\{ \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \big) \cap F_1 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \big) \cap F_2 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \big) \cap F_3 \Big\} \Big\} \\ \cap & \Big\{ \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = ND, \hat{y}_n = K) \big) \cap F_1 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = ND, \hat{y}_n = S) \big) \cap F_2 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = D, \hat{y}_n = Q) \leq L_{2A}(x = ND, \hat{y}_n = N) \big) \cap F_3 \Big\} \Big\} \Bigg\} \end{aligned} \tag{57}$$

1088 For $(x^*; y^*) = (L_{D,D}; Q, Q)$, the following condition must hold:

$$\begin{aligned} R_2 \equiv \Bigg\{ & \Big\{ L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = D, \hat{y}_n = Q) \Big\} \\ \cap & \Big\{ \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \big) \cap F_1 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \big) \cap F_2 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \big) \cap F_3 \Big\} \Big\} \\ \cap & \Big\{ \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = ND, \hat{y}_n = K) \big) \cap F_1 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = ND, \hat{y}_n = S) \big) \cap F_2 \Big\} \\ \cup & \Big\{ \big(L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \leq L_{2A}(x = ND, \hat{y}_n = N) \big) \cap F_3 \Big\} \Big\} \Bigg\} \end{aligned} \tag{58}$$

For $(x^*; y^*) = (L_{D,ND}; Q, K)$, the following condition must hold: 1089

$$\begin{aligned} R_3 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_1 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_1 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \leq L_{2A}(x = ND, \hat{y}_n = K) \big) \cap F_1 \Big\} \Bigg\} \end{aligned} \tag{59}$$

For $(x^*; y^*) = (L_{D,ND}; Q, S)$, the following condition must hold: 1090

$$\begin{aligned} R_4 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_2 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_2 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \leq L_{2A}(x = ND, \hat{y}_n = S) \big) \cap F_2 \Big\} \Bigg\} \end{aligned} \tag{60}$$

For $(x^*; y^*) = (L_{D,ND}; Q, N)$, the following condition must hold: 1091

$$\begin{aligned} R_5 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_3 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_3 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \leq L_{2A}(x = ND, \hat{y}_n = N) \big) \cap F_3 \Big\} \Bigg\} \end{aligned} \tag{61}$$

For $(x^*; y^*) = (ND; K)$, the following condition must hold: 1092

$$\begin{aligned} R_6 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = K) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_1 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = K) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_1 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = K) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, K) \big) \cap F_1 \Big\} \Bigg\} \end{aligned} \tag{62}$$

For $(x^*; y^*) = (ND; S)$, the following condition must hold: 1093

$$\begin{aligned} R_7 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = S) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_2 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = S) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_2 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = S) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, S) \big) \cap F_2 \Big\} \Bigg\} \end{aligned} \tag{63}$$

For $(x^*; y^*) = (ND; N)$, the following condition must hold: 1094

$$\begin{aligned} R_8 \equiv \Bigg\{ & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = N) \leq L_{2A}(x = D, \hat{y}_n = Q) \big) \cap F_3 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = N) \leq L_{2A}(x = L_{D,D}, \hat{y}_n = Q, Q) \big) \cap F_3 \Big\} \\ \cap & \Big\{ \big(L_{2A}(x = ND, \hat{y}_n = N) \leq L_{2A}(x = L_{D,ND}, \hat{y}_n = Q, N) \big) \cap F_3 \Big\} \Bigg\} \end{aligned} \tag{64}$$

A5. Brief descriptions of rumor cases considered 1095

1096

Table A.5
A brief description of rumor cases considered in this study.

Case	Brief Description
Hurricane Irma Fuel Demand	As evacuations were taking place in the State of Florida after Hurricane Irma, a rumor was spread on September 8th, 2017, stating that there was a high demand for fuel in Florida. This rumor was found to be true and the Florida Emergency Operations Center confirmed that demand in some areas has increased five times above normal levels.
Hurricane Harvey Hiring	After Hurricane Harvey, there was a post on Twitter and Facebook on August 12th, 2017, that claimed the Federal Emergency Management Agency (FEMA) is hiring field inspectors and paying $4,000 to $5,000 weekly. This rumor was confirmed to be true.
Hurricane Harvey Immigration	During Hurricane Harvey, a false rumor spread on social media claiming that the City of Houston would conduct routine checks of immigration status at evacuation sites and relief centers such as shelters and food banks. This rumor was debunked on social media with various tweets from the City of Houston and U.S. Immigration and Customs Enforcement (ICE). The rumor was also addressed on FEMA's rumor control page on Hurricane Harvey.
Hurricane Irma Immigration	On September 6th, 2017, a Sheriff from Polk County posted a tweet which read "If you go to a shelter for #Irma, be advised: sworn LEOs will be at every shelter, checking IDs. Sex offenders/predators will not be allowed." This tweet caused anger and panic among citizens and undocumented immigrants as they inferred that he was checking IDs to primarily scare undocumented immigrants from seeking safety in Polk County shelters.
Hawaii Missile False Alert	On January 13th, 2018, Hawaii's Emergency Management Agency sent out an emergency alert to cell phones, televisions, and radio stations stating that a ballistic missile was headed towards the islands. A second alert that was sent 38 minutes later notified the public that this was a false alert, and there was no incoming missile.
Tsunami False Alert	Emergency alarms began to wail on Hawaii's Oahu and Maui islands on February 6th, 2018, with warning of a potential tsunami. The alert turned out to be a false alarm, as confirmed by various authorities including the Honolulu Weather Center and the state Emergency Management System.
Boston Marathon Bombing Donation	Following the 2013 Boston Marathon bombing, a fake Twitter account named @BostonMarathon was responsible for spreading a false rumor. The account posted a tweet which read "For every retweet we receive we will donate $1.00 to the #BostonMarathon victims."

References

Agarwal, P., Hunt, K., Srinivasan, S., & Zhuang, J. (2020). Fire code inspection and compliance: a game-theoretic model between fire inspection agencies and building owners. *Decision Analysis, 17*(3), 208–226.

Alfonso, F. (2017). Corpus Christi debunks a second Hurricane Harvey rumor. Retrieved August, 2020 from https://www.houstonchronicle.com/news/houston-weather/hurricaneharvey/article/hurricane-harvey-rumor-fake-false-facebook-corpus-11982732.php.

American Red Cross (2016). 7 Top Questions about Louisiana Flood Relief. Retrieved August, 2020 from https://www.redcross.org/about-us/news-and-events/news/7-Top-Questions-about-Louisiana-Flood-Relief.html.

Ashley Collman (2020). Wuhan Coronavirus Outbreak. Retrieved August, 2020 from https://www.businessinsider.com/china-information-crackdown-on-wuhan-coronavirus-2020-1.

Ben Nimmo, C. S. E. L. R., Camille FranAois(2020). Facebook Takes Down Small, recently created network linked to internet research agency. Retrieved December 2020 from https://graphika.com/reports/ira-again-unlucky-thirteen/.

Bennet, N. (2017). 7 false rumors about Harvey. Retrieved August, 2020 from https://www.khou.com/article/news/local/7-false-rumors-about-harvey/285-468792969.

Bier, V. M., & Haphuriwat, N. (2011). Analytical method to identify the number of containers to inspect at US ports to deter terrorist attacks. *Annals of Operations Research, 187*(1), 137–158.

Bier, V. M., Haphuriwat, N., Menoyo, J., Zimmerman, R., & Culpen, A. M. (2008). Optimal resource allocation for defense of targets based on differing measures of attractiveness. *Risk Analysis, 28*(3), 763–770.

Chen, H., Zheng, Z., & Ceran, Y. (2016). De-biasing the reporting bias in social media analytics. *Production and Operations Management, 25*(5), 849–865.

Cheng, J.-J., Liu, Y., Shen, B., & Yuan, W.-G. (2013). An epidemic model of rumor diffusion in online social networks. *The European Physical Journal B, 86*(1), 29.

Cheung, M., & Zhuang, J. (2012). Regulation games between government and competing companies: Oil spills and other disasters. *Decision Analysis, 9*(2), 156–164.

Crapis, D., Ifrach, B., Maglaras, C., & Scarsini, M. (2017). Monopoly pricing in the presence of social learning. *Management Science, 63*(11), 3586–3608.

Dawn.com (2020). In a first, Twitter flags BJP leader's post as manipulated media. Retrieved December 2020 from https://www.dawn.com/news/1593615/in-a-first-twitter-flags-bjp-leaders-post-as-manipulated-media.

DiFonzo, N. (2008). *The watercooler effect: A psychologist explores the extraordinary power of rumors*. Penguin.

DiFonzo, N., & Bordia, P. (2007). *Rumor psychology: social and organizational approaches*: vol. 750. American Psychological Association Washington, DC.

Factcheck. Retrieved August, 2020 from https://www.factcheck.org/.

Federal Emergency Management Agency (2017). Harvey Rumor Control. Retrieved August, 2019 from https://www.fema.gov/disaster/4332/updates/rumor-control.

Federal Emergency Management Agency (2017). Hurricane Irma Rumor Control. Retrieved August, 2019 from https://www.fema.gov/hurricane-irma-rumor-control.

Feldman, P., Papanastasiou, Y., & Segev, E. (2019). Social learning and the design of new experience goods. *Management Science, 65*(4), 1502–1519.

Ferrara, E., Varol, O., Davis, C., Menczer, F., & Flammini, A. (2016). The rise of social bots. *Communications of the ACM, 59*(7), 96–104.

Festinger, L., Cartwright, D., Barber, K., Fleischl, J., Gottsdanker, J., Keysen, A., & Leavitt, G. (1948). A study of a rumor: Its origin and spread. *Human Relations, 1*(4), 464–486.

Gadde, V., & Beykpour, K. (2020). Additional steps we're taking ahead of the 2020 US Election. Retrieved December 2020 from https://blog.twitter.com/en_us/topics/company/2020/2020-election-changes.html.

Gao, F., & Su, X. (2017). Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science, 63*(8), 2478–2492.

Google Images. Retrieved August, 2020 from https://images.google.com/.

Hamidian, S., & Diab, M. T. (2015). Rumor detection and classification for Twitter data. In *Proceedings of the Fifth International Conference on Social Media Technologies, Communication, and Informatics (SOTICS)* (pp. 71–77).

Ho, T.-H., & Su, X. (2013). A dynamic level-k model in sequential games. *Management Science, 59*(2), 452–469.

Howel, L. (2013). Digital wildfires in a hyperconnected world. Global Risks Report. *World Economic Forum, 3*, 15–94.

Hu, M., Milner, J., & Wu, J. (2016). Liking and following and the newsvendor: Operations and marketing policies under social influence. *Management Science, 62*(3), 867–879.

Hunt, K., Wang, B., & Zhuang, J. (2020). Misinformation debunking and cross-platform information sharing through Twitter during Hurricanes Harvey and Irma: a case study on shelters and ID checks. *Natural Hazards, 103*(1), 861–883.

Hwang, T., Pearce, I., & Nanis, M. (2012). Socialbots: Voices from the fronts. *Interactions, 19*(2), 38–45.

Indvik, L. (2011). East coasters turn to Twitter during Virginia earthquake. Retrieved August, 2020 from https://mashable.com/2011/08/23/virginia-earthquake/.

Isaac, M., & Conger, K. (2020). Google, Facebook and Others Broaden Group to Secure U.S. Election. Retrieved December 2020 from https://www.nytimes.com/2020/08/12/technology/google-facebook-coalition-us-election.html.

Knapp, R. H. (1944). A psychology of rumor. *Public Opinion Quarterly, 8*(1), 22–37.

Krippendorff, K. (2013). *Content Analysis: An Introduction to its Methodology*. Sage.

Kwon, S., Cha, M., Jung, K., Chen, W., & Wang, Y. (2013). Prominent features of rumor propagation in online social media. In *Proceeding of the IEEE 13th International Conference on Data Mining* (pp. 1103–1108). IEEE.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lee, S.-Y., Qiu, L., & Whinston, A. (2018). Sentiment manipulation in online platforms: An analysis of movie tweets. *Production and Operations Management, 27*(3), 393–416.

Liao, Q., & Shi, L. (2013). She gets a sports car from our donation: Rumor transmission in a chinese microblogging community. In *Proceedings of the Conference on Computer Supported Cooperative Work* (pp. 587–598). ACM.

Luttrell, A. (2015). The psychology of rumors: 6 reasons why rumors spread. Retrieved August, 2020 from http://socialpsychonline.com/2015/09/psychology-why-rumors-spread/.

Oh, O., Agrawal, M., & Rao, H. R. (2011). Information control and terrorism: Tracking the Mumbai terrorist attack through Twitter. *Information Systems Frontiers, 13*(1), 33–43.

Oh, O., Agrawal, M., & Rao, H. R. (2013). Community intelligence and social media services: A rumor theoretic analysis of tweets during social crises.. *MIS Quarterly, 37*(2).

Oh, O., Kwon, K. H., & Rao, H. R. (2010). An exploration of social media in extreme events: Rumor theory and Twitter during the Haiti Earthquake 2010.. In *Proceedings of the ICIS: vol. 231* (pp. 7332–7336).

Ozturk, P., Li, H., & Sakamoto, Y. (2015). Combating rumor spread on social media: The effectiveness of refutation and warning. In *Proceedings of the 48th Hawaii International Conference on System Sciences* (pp. 2406–2414). IEEE.

Papanastasiou, Y. (2020). Fake news propagation and detection: a sequential model. *Management Science, 66*(5), 1826–1846.

Papanastasiou, Y., & Savva, N. (2017). Dynamic pricing in the presence of social learning and strategic consumers. *Management Science, 63*(4), 919–939.

Payyappalli, V. M., Zhuang, J., & Jose, V. R. R. (2017). Deterrence and risk preferences in sequential attacker–defender games with continuous efforts. *Risk Analysis, 37*(11), 2229–2245.

Prasad, J. (1935). The psychology of rumour: A study relating to the great Indian earthquake of 1934. *British Journal of Psychology. General Section, 26*(1), 1–15.

Qazvinian, V., Rosengren, E., Radev, D. R., & Mei, Q. (2011). Rumor has it: Identifying misinformation in microblogs. In *Proceedings of the Conference on Empirical Methods in Natural Language Processing* (pp. 1589–1599). Association for Computational Linguistics.

Rao, N. S., Poole, S. W., Ma, C. Y., He, F., Zhuang, J., & Yau, D. K. (2016). Defense of cyber infrastructures against cyber-physical attacks using game-theoretic models. *Risk Analysis, 36*(4), 694–710.

Rios Insua, D., Ríos, J., & Banks, D. (2009). Adversarial risk analysis. *Journal of the American Statistical Association, 104*(486), 841–854.

Roozenbeek, J., & Van Der Linden, S. (2019). The fake news game: actively inoculating against the risk of misinformation. *Journal of Risk Research, 22*(5), 570–580.

Rubin, V. L. (2017). Deception detection and rumor debunking for social media. In *The SAGE Handbook of Social Media Research Methods* (p. 342). SAGE.

(2017). Sacramento County update on Oroville. Retrieved August, 2020 from https://www.facebook.com/SacramentoOES/videos/1352512824812204/.

Sager, J. (2013). 10 Boston Marathon Bombing rumors that need to be stopped immediately. Retrieved August, 2019 from https://thestir.cafemom.com/crime/154242/10_boston_marathon_bombing_rumors.

Shan, X., & Zhuang, J. (2014). Modeling credible retaliation threats in deterring the smuggling of nuclear weapons using partial inspection - a three-stage game. *Decision Analysis, 11*(1), 43–62.

Shibutani, T. (1966). *Improvised news: A sociological study of rumor*. Ardent Media.

Snopes, Retrieved August, 2020 from https://www.snopes.com/.

Ten, C.-W., Manimaran, G., & Liu, C.-C. (2010). Cybersecurity for critical infrastructures: Attack and defense modeling. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans, 40*(4), 853–865.

Twitter Advanced Search. Retrieved August, 2020 from https://twitter.com/search-advanced?lang=en.

Twitter Search API. Retrieved August, 2020 from https://developer.twitter.com/en/docs/tweets/search/api-reference/get-search-tweets.

U.S. Department of Homeland Security (2018). Countering False Information on Social Media in Disasters and Emergencies. Retrieved August, 2020 from https://www.dhs.gov/sites/default/files/publications/SMWG_Countering-False-Info-Social-Media-Disasters-Emergencies_Mar2018-508.pdf.

Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science, 359*(6380), 1146–1151.

Wang, B., & Zhuang, J. (2017). Crisis information distribution on tWitter: A content analysis of tweets during Hurricane Sandy. *Natural Hazards, 89*(1), 161–181.

Wang, B., & Zhuang, J. (2018). Rumor response, debunking response, and decision makings of misinformed Twitter users during disasters. *Natural Hazards, 93*(3), 1145–1162.

Wang, J., Xie, Z., Li, Q., Tan, J., Xing, R., Chen, Y., & Wu, F. (2019). Effect of digitalized rumor clarification on stock markets. *Emerging Markets Finance and Trade, 55*(2), 450–474.

Wen, S., Jiang, J., Xiang, Y., Yu, S., Zhou, W., & Jia, W. (2014). To shut them up or to clarify: Restraining the spread of rumors in online social networks. *IEEE Transactions on Parallel and Distributed Systems, 25*(12), 3306–3316.

Yan, L., & Pedraza-Martinez, A. J. (2019). Social media for disaster management: Operational value of the social conversation. *Production and Operations Management, 28*(10), 2514–2532.

Zeng, L., Starbird, K., & Spiro, E. S. (2016). # unconfirmed: Classifying rumor stance in crisis-related social media messages. *Proceedings of the Tenth International AAAI Conference on Web and Social Media*.

Zhang, C., Gupta, A., Kauten, C., Deokar, A. V., & Qin, X. (2019). Detecting fake news for reducing misinformation risks using analytics approaches. *European Journal of Operational Research, 279*(3), 1036–1052.

Zhao, L., Yin, J., & Song, Y. (2016). An exploration of rumor combating behavior on social media in the context of social crises. *Computers in Human Behavior, 58*, 25–36.

Zhao, Z., Resnick, P., & Mei, Q. (2015). Enquiring minds: Early detection of rumors in social media from enquiry posts. In *Proceedings of the 24th International Conference on World Wide Web* (pp. 1395–1405).

Zhuang, J., & Bier, V. M. (2007). Balancing terrorism and natural disasters - Defensive strategy with endogenous attacker effort. *Operations Research, 55*(5), 976–991.

Zubiaga, A., Kochkina, E., Liakata, M., Procter, R., & Lukasik, M. (2016). Stance classification in rumours as a sequential task exploiting the tree structure of social media conversations. In *Proceedings of the COLING 26th International Conference on Computational Linguistics: Technical Papers* (pp. 2438–2448).

Zubiaga, A., Liakata, M., Procter, R., Hoi, G. W. S., & Tolmie, P. (2016). Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS One, 11*(3).