Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
An entropy-based method to control COVID-19 rumors in online social networks using opinion leaders

Lokesh Jain

Department of Computer Science & Engineering, India
Delhi Technological University, New Delhi, India

ARTICLE INFO

Keywords:
COVID-19 rumors
Opinion leader
Online social network
Entropy
Reputation
Health misinformation

ABSTRACT

In the ongoing COVID-19 pandemic, people spread various COVID-19-related rumors and hoaxes that negatively influence human civilization through online social networks (OSN). The proposed research addresses the unique and innovative approach to controlling COVID-19 rumors through the power of opinion leaders (OLs) in OSN. The entire process is partitioned into two phases; the first phase describes the novel Reputation-based Opinion Leader Identification (ROLI) algorithm, including a unique voting method to identify the top-T OLs in the OSN. The second phase describes the technique to measure the aggregated polarity score of each posted tweet/post and compute each user’s reputation. The empirical reputation is utilized to calculate the user’s trust, the post’s entropy, and its veracity. If the experimental entropy of the post is lower than the empirical threshold value, the post is likely to be categorized as a rumor. The proposed approach operated on Twitter, Instagram, and Reddit social networks for validation. The ROLI algorithm provides 91% accuracy, 93% precision, 95% recall, and 94% F1-score over other Social Network Analysis (SNA) measures to find OLs in OSN. Moreover, the proposed approach’s rumor controlling effectiveness and efficiency is also estimated based on three standard metrics; affected degree, representer degree, and diffuser degree, and obtained 26%, 22%, and 23% improvement, respectively. The concluding outcomes illustrate that the influence of OLs is exceptionally significant in controlling COVID-19 rumors.

1. Introduction

Since the last few times, the COVID-19 pandemic has spread swiftly at an alarming rate. The disease is considered a global crisis from the health and economic view. On February 15, 2020, World Health Organization director-General Tedros Adhanom Ghebreyesus stated, “We’re not just fighting an epidemic; We’re fighting an infodemic,” about COVID-19 at the virtual Munich Security Conference [1-3]. The term “infodemic” means facing complications in finding the answer to a problem due to an excess quantity of propaganda and rumors spread online and offline over the networks. So, the rumor leads to misinformation and produces mental fear, distress, and social disorder [4]. It deliberately makes agonize about the density of the COVID-19. COVID-19 rumors spread rapidly on social media without any authentication and legal verification. Even a single bit of illegal rumor dreadfully affects the entire society and instigates health consciousness [5,6]. The negative rumors spread fear and anxiety in society and impact human emotions and beliefs [7]. Eventually, WHO came forward has decided to make a myth-buster that clarifies and diffuses most of the myths about the COVID-19. It also includes various public advice, videos, and preventive measures to control the disease [8].

As we know that social media plays a critical role in widespread public events and activities. Sometimes they provide the white side of the picture and unknowingly or deliberately promote an image’s dark side [5]. So it is social media’s monopoly to decide whether to filter out the complete information or make it viral. They also provide the facility to access vast information to achieve a unified goal [9]. Whenever a specific emergency happens in the real world, different tweets, posts, and messages are circulated on social media without knowing the veracity [10]. Even though the national government agencies are ready to face the crisis and support providing all necessary services, all facilities are brought down and suspected because of various false rumors [11,12]. The main reason behind the rumor spreading is blind trust without knowing the facts. Rumors spreading in those communities are very rapid. They align with the previously presented values, i.e., people already believe in the community’s other populace to a certain degree. So, whenever people struggle for basic needs and the existing condition is crucial, such a situation stimulates rumors. Besides, one of the leading
factors affecting human psychological and mental decisions is trust. A trust provides the opportunity for the user with whom a user can share and accept information [13].

Any unreliable messages can quickly target a trusted system. In OSNs, it is challenging to identify the reliable and trusted users on which other users can rely without any uncertainty [14]. In Ref. [15], a three-valued subjective logic (3VSL) model is addressed that models the uncertainties in the graph using Dirichlet-Categorical (DC) distribution and network topologies. In Ref. [16], a trust management system is designed for a social reviewing system based on the indication theory and fuzzy logic. This research also considered time-dependent and content-dependent crown agreement for decision making. Similarly, other researchers also proposed multiple approaches based on trust value, trust score, trust degree, trust path, and various other factors [17–20].

In the 1950s, OL’s character in building and recognizing the individuals’ sentiments can be followed back to specific works. Researchers suggested that a person is not just selecting to acquire the prominent administration position dependent on their characteristics; besides, they ought to have a few attributes that coordinate with a specific condition of the community and the bunch of common interests shared by its individuals. According to this theory, a person or set of people has the significant capacity to influence the community assessments on some recognizable and applicable issues in contrast with the others. Katz proposed the ‘two-step flow of communication’ that significantly addressed OL’s power to transition sentiments and views from social media to the public community [22,23]. They also considered the influence integration based on three main components; quality of values, capability, and domain of interest. Leader’s and follower’s character and decisions are likely to change unusually over time and domain applicability [25]. In the era of digital advertising, at whatever point an OL is associated with a specific item or material, these items’ buying power expands sooner when contrasted with different things. An OL impacts another user’s conduct by turning over more information about the items through their insight and knowledge [26]. Thus, an OL is an individual or coalition of people who significantly impact the social network’s human acceptance process. They can differentiate the valid and invalid things through their skills and technical knowledge. They have effective communication and a vibrant attitude to convince other human beings surrounding misinformation and rumors.

Companies and organizations utilize OL’s credibility to design and shape the conceptual business model. The abstract business model provides the plan and growth for small and medium businesses and recommends more values and benefits to potential customers [27]. Suppose the businessperson and marketer know their potential customer who follows a particular opinion leader on social media. In that case, they can adopt or contact the opinion leader to influence and advise them to choose the product manufactured by their company [28]. The significance of OLs is also very vibrant in knowledge graphs, where databases are made from various homogeneous and heterogeneous entities interconnected with semantic relations [29]. OLs can support identifying the most suitable techniques and procedures through digital marketing to spread the product information into large communities. An OL is supposed to be a social media model that motivates and propagates marketing to spread the product information into large communities. An OL is a primary user who controls and transfers the appropriate information intelligently to the other communities in the knowledge graph. So, it is rather significant to transmit only relevant information that is effective to a particular group of users. OLs support and recommend trying to choose those customers who recommend authentic and correct products so that others do not face any problems due to the quality of the product. They also help resolve the cold start problem in the recommended system [33,34].

In the same way, the power of OLs can modify the perception and awareness of the severity of COVID-19 [35,36]. They rigorously examine the veracity by measuring the source information and inspecting the rumor’s authenticity. Also, it is essential to scrutinize the influence and information dissemination of OLs in the social network to triumph in the battle between COVID-19 and potential epidemic outbreaks [37]. Thus, the following research objectives are identified based on the above discussion.

1.1. Research objectives

The leading research objectives (ROs) of the article are as follows:

- RO1: Analyzing and evaluating the power and significance of OLs to control COVID-19 rumors based on various rumor controlling metrics.
- RO2: Discover a unique and innovative approach to find the top-T OLs in OSN based on the user’s reputation.
- RO3: Identifying the veracity of the COVID-19 posts by measuring entropy.
- RO4: Inspecting and measuring the implication of reputation and trust in OL detection (see Fig. 1).

To achieve the mentioned ROs, we have presented an exceptional and innovative technique in which OL behaves like a repressor who can control spreading rumors as much as possible. OLs substantially impact and prevent their supporters and believers [38,39]. So, if they follow any pattern or provide an opinion on a specific matter for society’s benefit, their followers probably also pursue the same guidelines as shown in Fig. 2. When the users interact either with the rumor diffuser or OLs, it may be possible that they either transform their viewpoint or are consistent with their beliefs.

Initially, a large number of COVID-19 tweets have extracted from social networks. Next, we have preprocessed all the tweets by handling negations, punctuations, degree modifiers, word shapes, emoticons, emojis, slang words, and sentiments-laden. Further, the reputation of each node is measured to figure out trust. Trust’s central significance is verifying the tweet’s trueness so that if any user spreads any information on the social network, the OL can confirm the tweet’s authenticity [40]. The ROLI algorithm targets the list of top-T OLs responsible for controlling rumors [41]. Finally, we measure each tweet’s entropy based on the probability that the other user responds to the tweets and estimated trust. If the tweet’s entropy is less than the pre-defined threshold value λ, it seems to be a rumor. A report is generated against it; otherwise, it transmits in the network, usually per the other user’s opinion.

1.2. Practical applications of research in society

In this research, we have identified the top-T OLs who can control the transmission or rumor as much as possible. This research’s practical implication is very much in the real world because the diffusion of any negative news may develop the mystification related to COVID-19 treatment and diagnosis that affects human health and the inner sentiment support system [42]. The influence of OL is not only limited to preventing misinformation or rumors. Still, it is significant in various
A novel ROLI algorithm addresses to find the OLs in OSN.

Tweet’s aggregated polarity is used to compute the user’s reputation.

Trust-based entropy is measured to substantiate the authenticity of the tweet.

The entire experiment is performed on Twitter, Instagram, and Reddit OSNs.

ROLI algorithm produces 91% accuracy, 93% precision, 95% recall, and 94% F1-score over other SNA measures

2. Analysis of related work

The role of OLs is very prominent and data-centric in social media. They can insulate users against risks such as rumor or misinformation exposure, anxiety, mental illness, and many more; and create a layer of trust that provides healthy and accurate information. They can also condense the influence of rumors by their knowledge and experience. The COVID-19 disease has completely spread worldwide, and many people have already lost their lives due to this disease. Social media is emerging as a strapping platform for communicating trustworthy health-related advisory and guidelines. Many people came closer than ever via social media to triumph in the battle against this infection [44, 45]. Thus the responsibility of OLs is now more essential and worthy to control the dissemination of fake news [46, 47]. In this research, we concentrated on two different kinds of study. In the initial investigation,
| Caption | Details | Supports | Constraints |
|---------|---------|----------|-------------|
| Reviews of Apple’s iPhone | The relationship in the user’s cluster is calculated based on various characteristics, and also OLs are detected using a text mining approach. | Straight-forward; Support in the recommended system. | Limited dataset; Used in online communities and forums only. |
| Epinions dataset | The reliable users are extracted by eliminating the unreliable user-based capacity-first maximum flow procedure. | Suitable for the recommended system; Control authenticity. | Restricted connections used; Only applied trust metrics. |
| Synthesized dataset | The propagative decision, spreading speed, and the count of the number of adopters determine the efficient OLs. | Support for marketing operations; Significant in the indirecst proposal system. | Need initial adopters for spreading information; Bounded centralities. |
| Educational blogs dataset, Parent-child forum dataset, Amazon.com dataset | User’s status, links, behaviors, and response time are measured to find the OLs. | Reinforced online blogs; Reduced execution time. | Restricted scope and domain; Applied Topic-specific constraints. |
| Facebook dataset, citation network dataset (Google scholar & DBLP) | A mathematical formula is derived from creating a series of OLs’ actions and influence. A universal series is also designed to normalize the result. | Assist with promotions; Support authority. | Needed initial user’s concern as input; Used in bounded and regulated space. |
| Web-based Chinese stock Forum dataset | Clustering algorithms and the posted text are used to identify users’ actions. Sentiment analysis-based case study discussed to find price trends. | | Applicable to the limited domain. |
| Synthesized dataset | Popularity and competency scores are measured to find the chances of the OLs based on the specific domain. | | Limited to online forums; Only PageRank centrality is used for comparison. |
| Epinions dataset | A novel approach used the trust value to find the OLs. Hybrid centrality measures are used to calculate trust values. | Simple calculation; Better results for hybrid centralities. | Vulnerable to the topic of interest; Not focused on retweets. |
| Mobile01 Forum | A new algorithm suggested finding the OL by reducing candidates using overlapping communities and user relationships. | More scalable; Reduced overlapping influence issues. | Few trust metrics like trust spread, trust maturity, and trust penetration are used for comparison. |
| turnbackhoax.id | The proposed approach focused on edge and centrality power. Edge power is used to find the relationship, while centrality power is used to decide each one’s proportion to find the OLs. | Found affected path; Inflated accuracy. | The entire procedure is complex and requires lots of computation. |
| Sina micro-blog dataset | The topic of interest and user’s transmission features are identified to define the user’s control. The IC-based propagation model simulation identifies the OLs based on the number of infected nodes. | Significant for information spreading; Presented an agreement algorithm. | Applicable to the limited domain. |
| Synthesized dataset | A trust model produced used the concept of fuzzy logic to measure reputation. The critical user with the highest status is selected as an OL. | Implement human reasoning; Produce results with more fidelity. | Fewer rules are used; A limited number of fuzzy rules are used. |
| Chinese Sina BBS | An augmented algorithm suggest based on domain sensitivity used for clusters identification while temporal attributes computed the OL’s effect during a time. | Easily implementable; Better outcomes over PageRank. | Primary domain and topic information required. Compared the outcomes with the PageRank algorithm. |
| Facebook, Twitter, Google+ | An innovative, dynamic model is proposed based on discrete-time. A key user is selected depending on the equivalent opinion vector of the tightly connected elements. | Unique vigorous, dynamic model; Highly efficient Clustering outcome. | Complicated to understand; Not handling overlapping cliques. |
| Twitter | A new rank of centrality measure was invented based on user interest and specialty on some topic. OLs are selected based on maximum centrality. | New Rank centrality is defined; No need for a prior relationship. | Limited topic considered; Tweet’s monitoring needed to check outcomes. |
| Local Motor community | Statistical components such as multi-featured are used to find the OLs. Also, the value and interest parameters are used to upgrade accuracy. | Reliable; Simple; Easy to implement. | Shortage of promotional degree, Restricted dataset. |
| slashdot dataset | Local and global OLs addressed based on the upgraded firefly algorithm and attractiveness score. | Acknowledge easily; Higher performance. | Static dataset; Few features are used for measurement. |
| travel community dataset | User’s actions and their effects are used to acknowledge consumer’s decisions. OLs are selected based on action-specific attributes. | Integrate SNA with Virtual travel groups; Significantly measured influencer affect. | Bounded dataset; Limited parameters are used for computation. |
| Wiki-vote and synthesized datasets | An upgraded whale optimization algorithm addressed to find OLs based on various objective functions and prestige. Also, an adjoining nodes-based procedure is used to identify the clique. | Highly optimized results; Unique application of the nature-inspired algorithm. | Lesser number of datasets are used; Differentiate outcomes with few algorithms. |
| Higgs Boson data from Twitter | Louvain method implemented for community detection. Next Betweenness centrality based approach is used to find OLs. | Straight forward; Easily implementable. | Limited centralities used; Implemented on few datasets. |
| Wiki-vote and Bitcoin OTC trust weighted network | Conditional probability-dependent groups of OLs are identified using the Game theory approach. For individual users, a shapely value is calculated to find the payoff in each group. Groups with maximum synergy are considered for selection. | Produced highly accurate outcomes; Powerful application of game theory. | The user’s initial trust score is required; Specific elements are used for measuring distance. |
we have considered the various OL detecting techniques, while in the latter study, we focused on various rumor control strategies. So, initially, we discussed the OLs recognition mechanism in brief.

According to Ref. [22], recognizing OLs is a two-way procedure involving information gain and followers’ adoption. Initially, OLs gathered the information from the various resources and shared it with the associates after analyzing it. Over time, multiple methodologies and theories have come into the picture related to graph theory, mathematics, data mining, machine learning, game theory, metaheuristic algorithms, behavioral science, ontologies, etc. [48–53]. So, the concise description of OL identification approaches is exposed in Table 1.

Now, we addressed the few studies related to rumor spreading that supported the foundation of this research. In Ref. [75], an innovative model is proposed based on the mean-field concept to find the communities’ substantial threshold value. The user worked as a mobile agent with some probability responsible for the inter-group long-range movement and disease-free clusters. The proposed approach also determined the transmission effect of the disease in the network. In Ref. [76], a memory-based model is suggested that explains the impact of memory over time. Also, they have used a few attributes and memory rates that have more influence on rumor transmission. According to Ref. [77], a unique 8-phase ICSAR model is proposed covering different features that find rumor dissemination. They have also considered the eight reasonable elements to evaluate the rumor, including information interest, rumor goal, user inclination, trust scale on social media, transmission rate, augmented component, block score, and specialist influence. In Ref. [14], researchers addressed a modified trust-based SIR model that defined a few mathematical equations representing the SIM model dynamics in the network. They also stated that the trust component could reduce the network population by limiting an evaluative threshold to control rumor transmission momentum. In Ref. [78], the analyst investigated that followers are the leading player responsible for the rumor/antitumor diffusion or maintenance. So they studied the network of users who already have faith in rumors and can influence others through their strong beliefs. As per [79], researchers established a new Susceptible–Hesitating–Affected–Resistant (SHAR) model that evaluates users’ behaviors against rumors and local and universal rumor concern equilibrium. They also computed the model’s reproduction and sensitivity of rumors towards parameter change. They have also analyzed the three different human viewpoints for experimental purposes. In Ref. [80], the authors investigated the importance and briefness of weak ties in rumor spreading. They argued that weak ties’ role is not influential for rumor propagation, but the diffusion rate depends on identifying weak links. They have also defined a probability based on a weak-tie-dependent function.

In [81], analysts proposed two transmission channels; the first is for point-to-point transmission, and the second is for global rumor transmission. A modified SIR model is also addressed using the mean-field formula. A geometric function and simulation have been performed to authenticate the proposed approach and verify the impact of rumor spreading. In Ref. [82], a novel rumor transmission model has been proposed based on the time interval and non-linear procedure. Initially, the primary reproduction number is calculated based on the next-generation matrix. Next, the researchers analyzed the rumor stability and factors needed for survival using experimental simulation. According to Ref. [83], the researchers developed an innovative SIR-based SKIR model that integrated rumor and anti-rumor-related factors. By using the proposed approach, they have also identified the active factors that stimulate the people to accept the rumors and the anti-rumors. They have also found the internal and external user characteristics that support the power of rumors using regression analysis. Also, the user’s attitude and dynamic behavior are considered for rumor propagation. In Ref. [84], a rumor reputation effectiveness index (REI) is developed to identify the factors that support the denial of rumors. They have also proposed four regression models to discover the dependence between REI and components to reject the rumors. The user’s interest and topic sensitivity are also considered when implementing the proposed system. In Ref. [85], the NLP model is presented using the deep learning-based DistilBERT and SHAP models. Initially, they integrated many datasets, performed back-translation, and found more accurate outcomes using DistilBERT. The SHAP makes the model more observable under three different experimental provisions to enhance trust.

Thus, various approaches are put forward for rumor controlling from the human and social perspectives. Most techniques are based on epidemic models, user response time, distribution speed, rumor transmission rate, content, behavior, attitude, and opinions [86]. The major pitfall with these approaches is the lack of user trust, reputation, and content accountability. Therefore, collective challenges are identified in the previously developed approaches, and how the proposed approach support overcoming those challenges is described as follows:

- Barely optimized result: The previously developed approaches can barely give the more optimized and enhanced results for all kinds of datasets. However, few approaches produced better results in a specific domain but lacked in other kinds of OSNs. In this article, a novel ROLI algorithm provides highly accurate results w.r.t. precision, recall, F1-score, and accuracy metrics. In addition, three performance metrics, diffuser degree, repressor degree, and affected degree, authenticate the proposed approach’s effectiveness in controlling rumors.
- Support few datasets: Most prior studies evaluated the online forums, the review communities, microblogs, online review forums, Twitter, Facebook, and other online platforms. Only a few approaches used multiple datasets to validate their work and detect OLs. The projected study supports mainly all types of OSNs as a user and network-specific features desirable to compute the user’s centralities, trust, and reputation.
- Extremely convoluted system: The preceding approaches are complex, and composite contains many exponential functions, multi-dimensional vectors, composite logical structure, inconsistent flows, and logarithmic equations and functions. Moreover, many hypotheses and procedures needed too much execution time and a highly advanced machine for implementation. On the other side, the coherent organization of the proposed approach is simple and understandable that does not comprise any multifaceted and varying materialization. We have used centrality measures, voting-based procedures, and mathematical equations to compute OLs and the veracity of rumors.
- Leverage of data mining-based techniques: It has been observed that most of the former approaches leverage the data mining-based techniques and approaches like supervised, unsupervised learning, cluster analysis, text mining, fuzzy logic, probabilistic methods, network structure, and user’s centrality and many more techniques to identify OLs ad rumor control. Though data mining-based methods are attained better results for various real-world applications, due to the masked and dynamic nature and high volume of information in OSNs, these techniques are not much more efficient and appropriate for social networking. This research does not deploy any data mining techniques. Only new innovative algorithms are designed for analysis purposes.
- Insufficiency of user trust and reputation: It has been observed that most of the approaches are lacking in considering user trust and reputation criteria for OL identification. Although few researchers ruminated on trust and reputation, reasonable outcomes are not produced because they only focused on the methods to compute trust. In our approach, trust and reputation are considered and focus on the method of describing the propagation of rumors. Thus, our approach concentrated on three tasks: identifying the OLs in OSN, validating the tweet’s veracity, and controlling the spreading of tweets.
- Deficiency of tweet’s entropy: Most previous approaches did not consider the tweet’s veracity by computing the entropy. Few
techniques used the modified SIR models, network topologies, user features, tweets, and other elements to validate the legitimacy of the tweet. Although they have made lots of efforts and experiments to authenticate their work, entropy should also be considered for checking the validity of the tweet. In our approach, an entropy-based method is addressed to verify the validity of the tweet using the user’s trust and probability to determine the chances of a retweet.

- **Absence of user agreements:** Most proposed approaches used user’s centralities, network structure, topologies, tweet response time, tweet polarity, user behavior, user communities, and many other characteristics to discover the appropriate OLs. No study has focused on user agreement and voting practice for identifying the OLs. We suggest an innovative voting-based algorithm whose computational complexity is also very low to identify the OLs in OSNs. Besides, as the number of users in the network is enhanced, it is also conceivable to get more optimized results efficiently.

Although a few techniques also worked on the discussed issues but either fulfilled only a few limitations or could not demonstrate the power of OLs in rumor control. So in this research, we may overcome most of the constraints and present an approach that successfully gave the novel solution by validating the procedure through experimental outcomes.

3. Background

Social networks play a significant role in information dissemination and misinformation control between people in the current electronic era. It is considered a rumor if certain information is flooded without official legal verification. In other words, a rumor contains faithless information that changes rapidly once transmitted from one person to another. The topology and dynamics of social networks have great significance in generating rumors. Various methods have been proposed to control the rumor spreading on the OSNs. The rumor diffusion model’s behavior is often similar to spreading the epidemic [80,87]. The disease transmission in epidemic models is dependent on the variety of agents and types of disease [88]. In this research, the core behavior of the rumor propagation process is based on the classical SIR model, making it more practical and rational. The SIR model is the elementary epidemiological model that illustrates the community’s theoretical model of disease transmission.

3.1. Social Network Analysis

A social network illustrates as a graph \( G = \{V, E\} \) that includes the collection of nodes \( V \) and links \( E \). A node shows the individual, group, community, or organization, while a link illustrates a relationship among the individuals. Graph theory concepts are used to characterize the dynamics of the social network. The clustering coefficient, density, diameter, network size, structural hole, strong-weak ties, and many other related concepts are derived from the graph theory. Social network analysis involves measuring the node’s centralities, community identification, homophily, bridges, triangles, link prediction, node classification, clique identification, and other related components [89–92].

- **In-degree:** Calculate the number of nodes related to the node.
- **Out-degree:** Calculate the number of nodes linked from the node.
- **Clustering coefficient:** Explore the scale to which a node is inclining towards making a cluster with the other nodes in the network.
- **Network density:** Explore the fraction between the existing links and the network’s full potential links.
- **Homophily:** Explore the feature that the nodes may likely connect or union with other nodes in the future.
- **Population size:** Count the total number of nodes in the network.

- **Closeness Centrality** (CC): States the mean distance of the node from other nodes, i.e., the node can transmit the information rapidly as it is near the other nodes in the network.
- **Betweenness Centrality** (BC): This indicates the fraction between the entire shortest link through the node and the shortest path possible among the network nodes.
- **Degree Centrality** (DC): Find the sum of in-degree and out-degree, i.e., the total count of links that a node encounters in the network.
- **Eigenvector Centrality** (EC): States the node’s power on the connected neighbors and allocate a value to the node based on the neighbors’ relative high or low value. The higher eigenvector centrality node is connected with neighbors with higher eigenvector values. The particular case of EC is the Page-Rank (PR) algorithm used by Google to rank web content.

3.2. SIR model

The primary theory of the SIR epidemic model is based on the intensity of infection, dissemination pattern, and network topology [88]. The SIR model has three phases: Susceptible, Infected, and Recovered. A Susceptible (S) phase defines the stipulation of the still not contaminated nodes but expected to be infected. An Infected (I) phase represents the total number of contaminated nodes that can spread the infection to other network nodes. A Recovered (R) phase states the total number of recovered nodes tainted from the disease. Also, the node is recuperated from the contagion and may or may not distribute the infection depending upon the neighbor’s interaction. Different researchers use a more comprehensive range of epidemic models such as SI, SIR, SIS, SEIR, SEIS, and SWIR for immunization, vaccination, infection measurement, information maximization, controlling policies, and others in different domains [93–97]. A node may exist in any one of the states at a particular time and can toggle its state over time. The total amount of users in the social network may vary, i.e., each time, some new nodes generate, and some old nodes disappear from the social network. The model uses \( o \) and \( f \) parameters to determine the origin and fatality degree. \( \gamma \) depicts the revival degree while \( \beta \) represents the transmission degree through which disease is distributed in the population of size \( N \) at time \( t \). Hence, the following differential equations, Eq. (1), Eq. (2), and Eq. (3) used to explain the model.

\[
\frac{dS}{dt} = oN - \beta IS - fS... (1)
\]

\[
\frac{dI}{dt} = \beta IS - \gamma I - fI... (2)
\]

\[
\frac{dR}{dt} = \gamma I - fR... (3)
\]

Often, the transmission degree is relatively low compared to the revival degree, so the disease’s effect depreciates after some time. If the disease’s intensity is very soaring, the probability of infection is very high, and a node may progress the phase from being susceptible to infection. In contrast, if the disease’s intensity is very low, a node may shift from infected to the recovered stage.

4. Proposed methodology materials-methods

The entire network is categorized into three types of nodes; oblivious node, diffuser node, and repressor node. A diffuser node is a user who propagates a rumor from one person to another. They directly connect with the other users and are involved intensely during rumor diffusion. When an individual establishes a relationship with a diffuser node, the probability of affection becomes high. An Oblivious node is a user who has neutral behavior during the rumor spreading, i.e., they do not take any action on the rumor and neither stop nor forward them. The repressor node is a user who never transmitted it to others whenever
they received the rumor and tried to stop its transmission.

For example, consider the network, as shown in Fig. 2. There are three colors node; The grey color nodes indicate the oblivious node. A green node represents the diffuser nodes, while red nodes depict the repressor nodes in the network. When the repressor node influences the diffuser node by its intelligence, it becomes a repressor node. In the same way, the complete knowledge about the rumor veracity is transmitted in the network gradually. Sometimes, few nodes are not much influenced by the repressor node and are firm on their decisions. Thus, it is very complicated to divert others’ decisions through social media; Still, OLs try to persuade others’ attitudes through their expertise [39,98].

In this research, we have illustrated the complete efforts performed by OLs and explained the procedure to find the OLs. So, all the computation and operations are done statistically, but in real work, they would be performed by OLs. So, we initially fetched many tweets from the three social media to effectively identify the population’s polarity and sentiments. The ROLI algorithm is explored to find the effective top-T OLs in OSN. Also, a VADER python library is used to find the tweet’s polarity covering various features and sentiments [99]. Each user’s reputation is computed based on the polarity scores received from the neighbors and other nodes. Next, trust is measured to validate the posts’ trueness, ensuring whether the posted tweet is a rumor. Algorithm 1 shows the pseudocode, and Fig. 3 shows the flow chart of the proposed approach.

Algorithm 1: Opinion Leader-based Rumor Detection (OLRD) Algorithm

| Input: | 1. Rumor threshold \( \lambda \)  
| 2. Total m number of tweets posted by n number of users |
| Output: | Decision about rumor spreading |
| Steps: | 1. Apply the ROLI algorithm to find the top-T OLs.  
| 2. For \( \forall t \in m \) do  
| Preprocess \( t \) by removing iterating characters, hashtags, hashtags, and URLs.  
| 3. For \( \forall j \in n \) do  
| Compute the degree of trust \( T_j \).  
| 4. For \( \forall j \in n \) do  
| Calculate the polarity-based reputation \( s \) of each tweet.  
| 5. For \( \forall j \in n \) do  
| Measure the entropy \( E_t(x_t) \) of each tweet.  
| 6. For \( \forall t \in m \) do  
| if \( (E_t(x_t) < \lambda) \)  
| Report the tweet as a rumor and discontinue the rumor spreading.  
| Else  
| Transmit the tweet typically. \( r_t \).  
| 7. End if;  
| 8. End for;  

4.1. Opinion leader identification

This research has proposed a new ROLI algorithm that identifies the most critical OL in the OSN based on the highest reputation. The concept of the ROLI algorithm came from product awareness in product marketing. The product manager’s main motive is to promote the product by recognizing the users with many social media followers as a business strategy. The product managers try to attract these information spreaders by providing a few rewards or benefits. So it is a very classical problem to choose such a spreader that can maximize the total selling of the product regardless of network topology and other competitive business strategies. Hence such types of users are considered OLs who can control other nodes. Generally, most OLs identification approaches are based on various distance-based centralities measures. Over the last few years, The PageRank-based procedures have provided better outcomes in most cases and more deliberation towards solving the significant user identification problems.

A user with a higher reputation in the network has more chances to get more votes from its neighbors. So, identifying the user’s reputation is critical and depends on the total number of tweets and retweets posted by the user’s neighbors. The general concept of the ROLI algorithm originated from the SIR epidemic model in which a user exists in any of the three stages: Susceptible, Infected, and Recovered. Many researchers have observed that if the disease’s infection rate is high compared to the recovery rate, it is complicated to eliminate the condition that formed a pandemic. Every time an infected node attempted to spread the disease with the rate \( \beta \) to any of its neighbors.

Similarly, a node can be recuperated with the rate \( \gamma \) over time. Hence, we have used the ROLI algorithm in which a node voted to its neighbors based on their reputation score. Each node’s reputation is initially calculated based on its total degree, i.e., the complete number of connected nodes. It is also observed that most social networks follow the richer gets richer phenomenon, i.e., the most potent nodes tend to magnetize more other nodes in the network. So, in the real world, a node with a higher level of trust or weighted relationship with its neighbor or other nodes would prefer to vote for that node. In this algorithm, if we want to choose the top-T OLs, every player has a chance to vote \( t \) times. That node does not participate in further voting rounds if a particular user is selected as an OL in one round. The main reason behind this strategy is to avoid biasing among the nodes because the spreading power of the elected node may influence or affect the decision of other nodes. So as the elected node would not be involved in voting, the subsequent node’s neighbors and their neighbor’s power also shrink. Hence the selected nodes are separated from the entire process and can not use their control unnecessarily with the neighbors. Finally, after \( t \) number of rounds, top-T OLs have been selected successfully.

In the ROLI Algorithm, a value pair \((v_s, v_r)\) is associated with every node, where \( v_s \) depicts the voting score received from the node’s neighbors and \( v_r \) indicates the voting ability of the node, i.e., the total number of votes that the node can grant to its neighbors. Initially, all the nodes have the same capacity in the first round. Every node can vote for its neighbors, and all neighbors can also vote for the subsequent node. The voting score of the node is the collective aggregation of the votes that its neighbors have given, i.e., if a node receives a total of five votes from its neighbors based on its reputation, the voting score of the node would be five. After each round, the node with the highest voting score would be declared the OL. It also has noticed that the elected OL would not participate and set its voting score to zero in the subsequent round.

All the nodes connected with the OL have to update their voting capacity in the upcoming phases. The node updates its voting capacity by \((v_f - f)\) until the value of \( v_r \) reaches to zero. In this mechanism, \( f \) indicates the diminishing variable whose value lies between 0 and 1. For simplicity, \( f \) is measured as \( \frac{1}{d} \), where \(< d >\) is in the mid-degree of the network. The same are iterated the \( t \) number of times or until the required number of OLs are identified in the network. Thus, the entire structure of the ROLI algorithm is as shown in algorithm 2.

Algorithm 2: Reputation-based Opinion Leader Identification (ROLI) Algorithm

| Input: | The total A number of users in the network |
| Output: | Top-T OLs |
| Steps: | 1. Identify the initial reputation \( r_x \) of each node \( x \).  
| 2. Assign the value to \((v_s, v_r) = (r_x, 1)\) to each node \( x \).  
| 3. O \( \cup \) Set of OLs  
| 4. While \((i < T)\)  
| 5. For \( x \in A \) do  
| 6. if \( x \notin O \)  
| 7. set \( v_s < 0 \) and \( v_r < 0 \)  
| else if \( x \in N(j) \) and \( j \notin O \)  
| 8. set \( v_s < \sum_{x', j} r_x \) and \( v_r < (v_s - f) \)  
| Else  
| 9. set \( v_s < r_x \) and \( v_r = 1 \)  
| end if;  
| 10. Find the node \( j \) with max \((v_s)\) and set \( O = O \cup \{j\} \) and \( A = A \setminus \{j\} \)  
| 11. \( i = i + 1 \)  
| end for;  
| end while;  
| 12. Find the list of top-T OLs return by O  
| 13. End; |
To understand the ROLI algorithm more clearly, consider the following network having nine nodes. Initially, the voting capacity of each node is set to be one, and the reputation of each node is calculated based on the aggregated degree, i.e., the total number of links connected with nodes. Fig. 4(a) depicts the first round outcomes, and node #4 is elected as an OL. In the next round, all the neighbor nodes, i.e., #1, #2, #3, #5, and #6, connected with node #4, update their voting ability and voting score according to the mentioned rule. The average degree of the network is 2.22, so the value of f becomes 0.45. Therefore, each connected node reduced its voting ability by the factor of f, and the new voting capability of the node becomes 0.55. In the next round, the voting capability of node #4 becomes 0. In Fig. 4(b), the reputation of the nodes is measured, and node #7 is selected as OL. The same procedure is carried out for the remaining successive rounds until the identified number of OLs.

4.2. Computational complexity

In this article, two algorithms, (i) OLRD and (ii) ROLI, respectively, are proposed to identify and validate the efficiency of OLs. The first algorithm OLRD is a decisive algorithm that produces the outcomes of the rumor. This algorithm mainly involves three tasks. The first task is to find out the OLs using the ROLI algorithm, the second is to compute the polarity-based reputation, and the third is to measure each user’s trust value in the network. The procedure to measure the time complexity of the ROLI algorithm is divided into three steps; The first step involved the time \( t_1 \) required to measure the initial reputation and voting ability. The second step included the time \( t_2 \) needed to select the node with the highest reputation score. The third step had the time \( t_3 \) required to modify the reputation score. There are total \( n \) nodes and \( e \) edges in the network. The total of \( O(n) \) time is needed to measure the initial reputation of each node. Similarly, the initial voting ability of each node is 1. So, only \( O(e) \) time is needed to assign initial voting ability. Thus, the whole time complexity \( t_1 \) is \( O(n) + O(e) \approx O(e) \). If we choose the optimal procedure to find the node with the maximum reputation score, the time complexity \( t_2 \) of the second step would be \( O(n) \). In the third step, the reputation score of only those nodes would be updated that are one or two units far away from the repressor nodes that have been selected OLs in the previous round. The mid-degree of the network is \( \langle d \rangle \). Thus the time complexity of step 3 is \( O(\langle d \rangle^2) = O(\frac{e^2}{n}) \). If we want to select \( t \) number of OLs, the overall complexity would be \( O(e + t + t\frac{e}{n}) \). If the network is sparse, i.e., \( n \gg e \) and \( n \gg t \), the time complexity of the algorithm can be \( O(n) \) for the network.

A simple mathematical equation that includes a simple science and iterative function is successfully deployed to compute each user’s polarity-based reputation and trust. First-order predicate logic is used to measure reputation. Therefore the complexity of measuring the trust and reputation of each node is \( O(n) \).

4.3. Stages of rumor transition

4.3.1. Diffusion stage

In this stage, the diffuser user attempted to spread the rumor with the rate \( \beta \) in the network. The user finds the appropriate piece of information that is to be transmitted. Sometimes, the diffuser may modify some of the information according to the current circumstances and spread the indistinct information to its neighbors. If the nodes do not want to participate in the transmission or have null facts, they ignore the transmission process.

4.3.2. Recognition stage

In this stage, the node decides whether to accept the rumor with the probability \( q_d(t) \). If the degree of trust is very high among the other
users, the node may accept the rumor and propagate it to its neighbors, but the node rejects the rumor if the degree of trust is low.

4.4. Probability of retweeting/forwarding the tweet

In the social network, millions of users posted various posts and retweets related to different events every second of a day at the rate of \( \theta \). The user finds various new posts whenever the users refresh and reload the social media interface. The user decides to read, forward, retweet, or ignore the posts. Let \( u_1 \) be the total number of first-order followers, and \( u_2 \) be the second-order followers of a user \( x \). The second-order followers are considered only those who retweet or read the posts or retweeted by first-order followers. Let, \( u_k \) indicates the total number of followers in the individual network, and \( q_d(t) \) depicts the probability of finding the total number of followers who pursue, read, and retweet the tweet at level \( d \) after its posting within time \( t \). We have assumed that most users who read and retweet the post are equal to \( p \). Let \( R_d(t) \) represents the total number of followers who read or retweet the posts at level \( d \) after \( t \). Therefore, the value of \( R_d(t) \) can be computed using Eq. (4).

\[
R_d(t) = u_k(t)q_d(t)p \ldots \quad (4)
\]

In this process, the total number of tweets and the appearance of the total number of \( I \)-order users are considered exponential. So, Poisson distribution is used to measure the total number of users who randomly read and retweet the posts at time \( t \), represented using Eq. (5).

\[
\int_0^t f(\tau)d\tau = \int_0^t \theta e^{-\theta t}d\tau = 1 - e^{-\theta t} \ldots \quad (5)
\]

So, the probability \( q_1(t) \) or first-order users to retweet can be represented using Eq. (6).

\[
q_1(t) = 1 - e^{-\theta t} \ldots \quad (6)
\]

Now, we have formulated the Poisson distribution for second-order users. Here, one interesting fact is that the second-order users only retweet the post that is retweeted by the first-order users on or after time \( t \). Let \( r \) consider the time on or before the first-order user read the tweets such that \( r \leq t \). So, the probability \( q_1(t-r) \) for first-order users to retweet, the post between \( t \) and \( r \) can be depicted using Eq. (7).

\[
q_1(t-r) = 1 - e^{-\theta (t-r)} \ldots \quad (7)
\]

So, the probability \( q_2(t) \), for second-order users to retweet can be represented using Eq. (8).

\[
q_2(t) = \int_0^t \theta e^{-\theta t} p q_1(t-r)dr
\]

\[
= \int_0^t \theta e^{-\theta t} p (1 - e^{-\theta (t-r)}) q_1(t-r)dr
\]

\[
= p (1 - e^{-\theta t} - \theta t e^{-\theta t}) \ldots \quad (8)
\]

Correspondingly, the probability \( q_3(t-r) \) can be shown using Eq. (9).

\[
q_3(t-r) = p (1 - e^{-\theta (t-r)} - \theta (t-r) e^{-\theta (t-r)}) \ldots \quad (9)
\]

Similarly, the probability \( q_3(t) \) of retweet for third-order users is represented using Eq. (10).

\[
q_3(t) = \int_0^t \theta e^{-\theta t} p q_2(t-r)dr
\]

\[
= \int_0^t \theta e^{-\theta t} p (1 - e^{-\theta (t-r)} - \theta (t-r) e^{-\theta (t-r)}) d\tau
\]

\[
= p e^{-\theta t} - \theta t e^{-\theta t} - \frac{1}{2} \theta^2 t^2 e^{-\theta t} \ldots \quad (10)
\]

Thus, the general formula to calculate the probability by the user at level-\( d \) is shown using Eq. (11).

\[
q_d(t) = p^{d-1} \left( 1 - \sum_{a=0}^{d-1} \left( \frac{(\theta t)^a e^{-\theta t}}{a!} \right) \right) \ldots \quad (11)
\]

If an individual receives a bit of information from the other users in the network, it depends on the degree of trust that the user perceives from the other user. So, \( q_d(t) \) measures the probability of reading and retweeting the other users’ posts published in the network.

4.5. Tweet preprocessing and reputation calculation

There are two types of users in the social network, active and passive, based on their actions. Most active users respond to the tweets according
to their domain of interest and topic. On the other hand, the passive user rarely responds to the tweet regardless of their topic interest. During the analysis, we only considered the active users’ actions who have commented and retweeted on COVID-19-related posts. Generally, a tweet is significantly affected by the active user’s response, environment, content, time interval, and domain [100]. The other user’s tweet also influences the tweet’s content and schism over time. So, we have redesigned our dataset in which each tweet is represented by the tuple <#user_id, content, time>. To measure each tweet’s polarity, we have used the VADER (Valence Aware Dictionary and Sentiment Reasoner) python tool that categorized each tweet into the positive, negative, and neutral categories. This tool also defines positive, negative, or neutral polarity potency based on character case sensitivity, emojis, punctuation, emoticons, and slang. The primary elements of the tools include degree modifiers, conjunctions, and n-grams. The modified reputation measuring course of action calculates a user’s reputation in the social network. Thus the reputation, emoticons, and slang. The primary elements of the tools include degree modifiers, conjunctions, and n-grams. The modified reputation measuring course of action calculates a user’s reputation in the social network.

4.7. Tweet’s entropy calculation

A tweet’s entropy signifies the importance and amount of information a tweet perceives during the transmission. Promotion and advertisement-like tweets have a low entropy, while news and informative-like tweets likely have high entropy. Generally, the user retweets or forwards only those tweets with some new and unique information or originated from some authorized source.

There is a probability \( q(t) \) to determine the chances of retweeting on user \( x \) by other users. So, the user, \( x \), i-th tweet overall entropy \( E_t(x_i) \) at time \( t \) is calculated using Eq. (16).

\[
E_t(x_i) = - \sum_{n=1}^{m} \eta(r_i(t)) \log(q(t)) \log(q_0(t))) \ldots
\]

The measured entropy is interestingly helpful to encode the hidden information in OSN as the network is hugely limber and changing. Thus, exhaustively finding the coalitions, synergy, and other network dynamics through entropy is convenient.

4.8. Parameters setting

The ideal environment is chosen to determine the optimal values of all the heuristic parameters to implement the proposed approach. The value of a few variables like network size, density, clustering coefficient, the total number of relationships, and centralities can be calculated using statistical and graph theory-based formulas. For the certifying purpose, 70% of the entire data set is selected for the testing task, and the remaining 30% of the dataset is used to validate the parameter’s value. Initially, specific random parameter values are selected based on the specified constraints and conditions.

Let \( m \) be the total number of users in the network, and \( j \) is the absolute number of iterations needed to find optimal values. Thus the amalgamation of \( m^j \) created a problem space to find out the optimum value. Moreover, If the size of \( m^j \) is enormous, the chances of obtaining the best values are very high for the heuristic parameters. So, considering this fact, we have chosen the Twitter dataset for analysis as the dataset size is enormous. In this research, we have used the four heuristic variables \( x_1, x_2, x_3 \), and \( x_4 \). A linear model Analysis of Variance (ANOVA) approach is used to identify the optimal values for all the parameters. This approach assists in discovering the relationship among the variables that support approving or rejecting the null hypothesis. In this experiment, four groups of size \( m^j \) are formed. There are 50 K users whose information is processed over 100 iterations in the first group. Similarly, the total size of the second, third, and fourth groups is 100 K*100, 200 K*100, and 500 K*100, respectively. Table 2 shows the outcomes of each group along with its Mean Square (MS), Sum of Squares (SS), F-statistics, and P-test values.

5. Result analysis and performance evaluation

This part of the research discusses the descriptions of extracted posts/comments from the Twitter, Instagram, and Reddit social networks. Further, various attributes are gathered during the fetching of tweets like a timestamp, user id, text, source, geo, coordinate, place, contributor, retweet count, sensitivity, favorite count, and many more. Next, we experimented with the proposed ROLLi algorithm and demonstrated the list of top-T (≤10) OLS. We also conferred the significance and importance of the proposed approach over other SNA measures.
through appraising the various rumor-controlling metrics.

5.1. Dataset

5.1.1. Twitter dataset

Since the COVID-19 pandemic, social media have practiced a large volume of data and comments. Twitter also plays an essential role in extracting information about the pandemic from the last few times. This research has pulled many tweets from Twitter to find people’s opinions and sentiments. The main reason for choosing Twitter is its popularity, and usability drastically increased during the pandemic. Some unauthorized sources have posted different kinds of rumors and misinformation. Later, Twitter added the fact-checking tie with the tweets to check the tweet’s authenticity.

So, in this research, we have also extracted Twitter’s tweets dataset for analysis purposes. A Twitter streaming API is used to pull out the COVID-19 tweets. Twitter’s API supports access to tweets, users, messages, trends, links, etc. Due to memory and CPU constraints, we have collected 65.3 M tweets from January 1, 2020, to March 30, 2020. We have used the five search keywords for obtaining the tweets. We have gathered the tweet in Tweet, userId, time, country form to demonstrate the single tuple.

5.1.2. Instagram dataset

Instagram is another widely used social media for images, videos, and views sharing purposes. This research has also extracted the Instagram posts related to COVID-19. An Instagram API collects the posts from January 1, 2020, to March 30, 2020. Initially, a token is generated, and POST and GET requests are built through HTTP. However, the total number of posts is relatively more minor than the Twitter data set but contains fascinating facts, rumors, and misinformation about COVID-19. We have collected around 9.8 K posts, and 37 K comments originated from the different validated user accounts.

5.1.3. Reddit dataset

Reddit is an online discussion medium that integrates news, posts, comments, conversations, views, images, and queries. Although the recognition of Reddit is not so wider, it covers more rich content and information. Reddit contains many communities that provide information in a very creative and innovative way. We have used the python-based Reddit API to collect the content based on five searching words. We have assembled around 15.7 K comments derived from the various user communities and specific posts. Thus, the statistical description of the datasets is shown in Table 3.

A Java-based network analyzer tool, Gephi, is used to analyze the network. The device supports finding the relation among the users and measuring the network’s clustering coefficient. Next, we have used the VADER python library to get the sentiments of each tweet. We have classified each tweet into three classes: positive, negative, and neutral. Further, each user’s reputation is measured based on the aggregate polarity. Trust is also calculated to predict whether the users retweet, forward, or reject the message.

Next, we applied the ROLL algorithm to find the top-T OLs in the network. We found the list of users who have posted maximum tweets or posts, and other people also commented on those tweets in the network. These tweets contain valuable information and provide a powerful direction toward controlling COVID-19. Numerous users have liked, disliked, retweeted, and shared these tweets. Further, we have calculated the entropy of each tweet based on measured trust.

Moreover, a filtering operation discards those tweets whose users have low reputation scores. So this operation reduced the whole entropy processing time. Finally, we obtained the records of users whose tweets are mostly retweeted using SNA measures. We suggested reputation scores along with their SNA measures, as shown in Table 4, Table 5, and Table 6 for all three datasets.

In the above results, each table consists of twelve columns, and all the columns are in pairs, i.e., one column complements another adjacent column. The first column indicates the node id, and the adjacent second column depicts the corresponding SNA measures value of that node. For example, in Table 4, the degree centrality of the top node #3782784 is 0.0353824, and the closeness centrality of the top node #4882929 is 0.0099734. Similarly, Tables 5 and 6 identify the top-10 OLs and their SNA measure. The last column of each table identifies the reputation of top nodes in chronological order for all three datasets: Twitter, Instagram, and Reddit, respectively, as per the proposed approach. These tables represent the value of standard SNA centrality measures and show that different nodes have different centrality measures. The same node is not mandatory, securing the top position for all the centrality measures as the complex structure of the social network. As per literature, various techniques and methods have been proposed to discover the OLs, and most researchers used only these SNA measures for analysis purposes.

Once the top-T (10) OLs are identified in the network, the next step is choosing the optimum threshold value to calculate the entropy. Selecting a particular threshold value i for declaring a tweet as a rumor is a critical task. Therefore, after various analyses and experiments, we have chosen i = 0.95, i.e., if the entropy of the tweet is less than 0.95, the tweet would be reported as a rumor; otherwise, OLs may only forward or add their comments to the post and forward to the followers and other users in the network.

5.2. Performance metrics for rumour controlling

To evaluate the performance and effectiveness of the proposed approach for COVID-19 rumor controlling, we have used the performance metrics based on the behavior of the oblivious node n_o, diffuser node n_d, and repressor node n_r. As mentioned earlier, rumor spreading is similar to the disease-spreading behavior in the traditional SIR epidemic model. The authenticity of the proposed approach depends on how quickly it identifies and controls the total number of rumors in the network. Thus, we have used three metrics: diffuser degree, repressor degree, and affected degree, respectively, for measuring the approach’s performance.

5.2.1. Diffuser degree

The diffuser degree explains the behavior of the diffuser node in the network. If the rumors spreading rate β is high, most nodes might be infected or influenced by the rumors. It is mandatory to transmit the correct information in the network as early as possible to avoid any crisis. So, The diffuser degree f_d(t) is defined as the ratio between the total number of diffuser nodes and the total count of representer nodes, oblivious nodes, and diffuser nodes in the network at time t, as shown in Eq. (17).

\[ f_d(t) = \frac{n_d}{n_o + n_r + n_d} \]  

In Fig. 5, we can observe that as soon as the count of OLs increased, the diffuser degree reduced gradually; But due to some spreader nodes with a strong belief in the rumor, it is impossible to reach the zero level. Also, the proposed approach reduced the number of diffusers 26% faster than other SNA measures as the number of OLs increased gradually.

5.2.2. Repressor degree

Repressor degree states the control of the represer node in the network because eventually, the representer nodes are considered the OLs in the network. So, if the representer nodes identify the tweet’s trueness as soon as possible, they can help stop the rumor from spreading. Thus, a repressor degree f_r(t) is defined as the ratio between the total number of representer nodes and the total count of representer nodes, oblivious nodes, and diffuser nodes in the network at time t, as shown in Eq. (18).

\[ f_r(t) = \frac{n_r}{n_o + n_r + n_d} \]
Datasets statistical description.

**Table 3**

| Dataset (January 1-March 30, 2020) | Statistics     |
|-----------------------------------|----------------|
|                                   | Twitter | Instagram | Reddit |
| Total number of tweets            | 65.3 M  | 46.8 K     | 25.7 K  |
| % of the tweets in English         | 71.2%   | 89.3%      | 96.4%   |
| % of tweets in other and regional languages | 28.8%   | 10.7%      | 5.6%    |
| % of verified accounts             | 8.4%    | 23.5%      | 64.8%   |
| Total number of participating countries | 173     | 148        | 94      |
| Total number of searched keyword   | 5 ('Covid19', 'coronavirus', '#2019-ncov', '#covid_19', '#pandemic') | 5 ('Covid19', 'coronavirus', '#2019-ncov', '#covid_19', '#pandemic') | 5 ('Covid19', 'coronavirus', '#2019-ncov', '#covid_19', '#pandemic') |
| Density                           | 0.00945 | 0.00591    | 0.0139  |
| Clustering coefficient            | 0.000641| 0.000472   | 0.000863|

In Fig. 6, we can monitor the relationship between the represser degree and diffuser degree over time $t$. As soon as the represser nodes spread the actuality of the rumor with the rate $\gamma$, the represser degree increased, and the depressor degree decreased with time. Again, the proposed approach performed better, spreading the tweet’s veracity around 22% faster than other SNA measures.

### 5.2.3. Affected degree

An affected degree is a scale that explains the effect of OLs in the network. It measures the total number of users whom OLs influence. The represser degree $f_r(t)$ define as the fraction between the total number of influenced nodes and the total count of represser nodes, oblivious nodes, diffuser nodes, and affected nodes in the network at time $t$, as shown in Eq. (19).

$$f_r(t) = \frac{n_a}{n_r + n_i + n_o + n_d} \ldots \quad (19)$$

where $n_a$ depicts the number of users influenced by the OLs. Further in Fig. 7, we can infer that as represser nodes spread the rumor’s actuality with the rate $\gamma$, the affected degree increases with time, i.e., more number of users influenced by the OLs. The proposed approach produced better outcomes and impacted the users approximately 23% faster than other SNA measures.

Each OL has different followers, and rumor spreaders are also presented in the network. So, controlling the rumors among the general users varies depending on the total number of followers and network structure. Again, we can observe how the OLs influenced and controlled the rumors for all the three networks over other SNA measures. As the number of OLs increases progressively, the control of OLs on the other users also spreads rapidly. In Fig. 8(a), the red fragment shows the influence of OLs on the Twitter dataset, while the different color fragments (purple, blue, sky blue, green, and yellow) show the effect of OLs from lower to the higher transition state.

Similarly, Fig. 8(b) and (c) demonstrate the impact of OLs on the other datasets. We can observe that the effect of OLs is significantly higher as the number of OLs increase gradually in each network. Thus they can identify the integrity of the tweets as much as possible within a limited time interval. The red fragment in each part depicts the influence of OLs in the network.

### 5.3. Performance comparison of ROLI algorithm

Further, we have compared the results with the other SNA measures to authorize the ROLI algorithm’s performance. We compared the impact of the proposed ROLI algorithm with the multiple centralities measures, which are also used globally to find the prominent users in the network. The biggest challenge with the OSN is the lack of ground truth about the key users. The various approaches use different contexts, network-dependent parameters, and theories. So, these approaches do not follow universal principles; standard centrality measures are used to match up and assess the results. We utilized the four performance metrics, accuracy, precision, recall, and F1-score, to ensure the reaching of the approach. For analysis, four primary outcomes: True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN) are required. All these metrics can be calculated using the following formulas:

- **Accuracy** $= \frac{TP + TN}{TP + TN + FP + FN}$
- **Precision** $= \frac{TP}{TP + FP}$
- **Recall** $= \frac{TP}{TP + FN}$
List of top-10 OLs along with their reputation score and other SNS measures for the Twitter dataset.

| Node id | DC   | Node id | CC   | Node id | BC   | Node id | PR   | Node id | EC   | Node id | Reputation |
|---------|------|---------|------|---------|------|---------|------|---------|------|---------|------------|
| 1,782,784 | 0.0353824 | 4,882,929 | 0.0099734 | 5,781,775 | 0.0504885 | 4,938,103 | 0.0038952 | 937,482 | 0.0144372 | 758,380 | 0.1276728 |
| 1,636,273 | 0.0353823 | 837,321 | 0.0099732 | 184,773 | 0.0504883 | 5,062,765 | 0.0038951 | 2,827,292 | 0.0144372 | 4,791,216 | 0.1276728 |
| 466,738 | 0.0353821 | 3,877,392 | 0.0099730 | 2,390,202 | 0.0504882 | 3,773,291 | 0.0038949 | 837,174 | 0.0144371 | 829,252 | 0.1276727 |
| 1,046,730 | 0.0353819 | 174,992 | 0.0099731 | 734,218 | 0.0504882 | 638,383 | 0.0038948 | 84,983 | 0.0144371 | 940,251 | 0.1276726 |
| 473,721 | 0.0353815 | 1,062,525 | 0.0099730 | 78,022 | 0.0504881 | 194,482 | 0.0038947 | 2,839,290 | 0.014437 | 3,936,037 | 0.1276726 |
| 5,254,466 | 0.0353811 | 494,775 | 0.0099730 | 519,287 | 0.0504880 | 2,784,929 | 0.0038946 | 3,521,043 | 0.0144369 | 7385 | 0.1276726 |
| 904,537 | 0.0353809 | 105,829 | 0.0099729 | 3,992,801 | 0.0504879 | 574,922 | 0.0038944 | 449,293 | 0.0144367 | 1,289,504 | 0.1276724 |
| 3,029,229 | 0.0353808 | 2,593,920 | 0.0099728 | 1,820,378 | 0.0504877 | 1,383,092 | 0.0038942 | 1,588,391 | 0.0144367 | 5,827,403 | 0.1276725 |
| 2,372,722 | 0.0353804 | 59,201 | 0.0099726 | 3,629,048 | 0.0504876 | 292,739 | 0.0038941 | 449,293 | 0.0144367 | 1,289,504 | 0.1276724 |
| 375,981 | 0.0353804 | 429,322 | 0.0099727 | 947,324 | 0.0504876 | 5,417,321 | 0.0038941 | 814,871 | 0.0144366 | 683,692 | 0.1276724 |

Another significant aspect that affects the algorithm’s importance is the execution time. The previous section discussed the algorithm’s complexity, almost equivalent to O(n). We have considered only the time needed to discover the designated OLs from the network. We have also performed extra work to compute our approach’s reputation, trust, and entropy. Each method has its parameters to estimate the OLs and is very materialistic. In Fig. 10, we demonstrated the execution time required by each SNA measure and the proposed ROLI algorithm. From the analysis, we examined that the proposed algorithm needed a shorter time, comparatively lower than the other SNA measures for all three datasets. The proposed method consumes only 73 min for Twitter, 34 min for Instagram, and 17 min for the Reddit dataset for processing. Thus, we can strongly recommend that the proposed algorithm perfectly controls COVID-19 rumors and supports strong beliefs and reputation on the social network.

6. Theoretical implications of the proposed approach

Nowadays, most human decision-making is influenced by social media activities. The importance of OL is very vital and impactful for managing various kinds of rumors and misinformation transmitted over OSNs. In this research, we have explored the power of OL to control COVID-19 rumors. Thus, this research’s contribution is significant in various domains to prevent rumors and hoaxes comprehensively.

First, we have defined a reputation-based OL detection algorithm by calculating the polarity of the tweet. The impression of reputation is very worthy in OSN to determine the node’s influence on others and gain loyalty. Only limited approaches used reputation and trust to find OLs in limited domains in previous studies. This study’s main essence is effectively utilizing reputation and trust. Also, the polarity of each tweet is measured to identify the population’s sentiments about the COVID-19 epidemic.

Second, we did not apply any data mining approach in this study. Although data mining approaches outperformed well in real-world

- **F1-score = 2 * Precision * Recall / Precision + Recall**

In Fig. 9, we have compared the outcomes from the proposed ROLI algorithm w.r.t. mentioned performance metrics. For experiment purpose, we have chosen the value of β = 0.0055 and γ = 0.0085. So, all the results and experiments have been performed based on these parameters. We observe that the proposed algorithm provides an average of 91% accuracy, 93% precision, 95% recall, and 94% F1-score. If we deeply observed the results for all the three datasets, we found that the proposed ROLI method produces 90% accuracy, 97% precision, 92% recall, and 94% F1-score for the Twitter dataset. Similarly, for Instagram, 91% accuracy, 94% precision, 95% recall, and 95% F1-score are obtained; for the Reddit dataset, 92% accuracy, 97% precision, 94% recall, and 95% F1-score are achieved.

Another significant aspect that affects the algorithm’s importance is the execution time. The previous section discussed the algorithm’s complexity, almost equivalent to O(n). We have considered only the time needed to discover the designated OLs from the network. We have also performed extra work to compute our approach’s reputation, trust, and entropy. Each method has its parameters to estimate the OLs and is very materialistic. In Fig. 10, we demonstrated the execution time required by each SNA measure and the proposed ROLI algorithm. From the analysis, we examined that the proposed algorithm needed a shorter time, comparatively lower than the other SNA measures for all three
applications and produced improved outcomes, they are not appropriate for OSN due to the dynamic nature of the network. In this approach, we have merely used the software tools, statistical formulas, and methods to discover OLs and validate the integrity of the post. One of the principal merits of this approach is that it is very beneficial for large datasets. As the number of users steadily increased, more users would exchange their opinions and views. So, at that moment, each user can measure other users’ reputations and degree of trust more accurately and precisely.

Third, the proposed system’s complexity is straightforward and does not include any multifaceted structure and formulation. For OL detection, we have used the voting score-based approach in which the user assigned its vote to the user having a higher reputation. Trust and polarity-based formulas are used to measure each post’s entropy to validate the post. Thus, this approach uses a very elementary strategy.
with lesser complexity. Most previously developed methods involved complex, composite, and lengthy computations that are difficult to implement and understand.

Fourth, this study gives exceptional progress toward innovation in information science by exploring and demonstrating the power of OLs to control COVID-19 rumors in the current pandemic situation. Also, the operational behavior of the proposed ROLI approach is analogous to the SIR epidemic model that shows the spreading of disease in the real world. Such research also supports preventing a high amount of misinformation transition in OSNs. It is essential to identify the set of users who have an immense impact on their followers to improve the trustworthiness of OSNs. So this research contributes enormously to fulfilling this objective with higher accuracy and effectiveness.

7. Conclusion, limitation, and future scope

In the present pandemic condition, the COVID-19 disease has drastically affected the world. Social media’s role and power are vital regarding COVID-19-related rumors and misinformation. Although the World Health Organization (WHO) and other official government organizations have already circulated various guidelines and control measures to avoid the disease, numerous sources on social media [105, 106]. Thus, controlling such rumors and misinformation is essential to save public health. Since Twitter, Instagram, and other social media have drastically increased after this pandemic, consumers have posted different information without checking the source’s authenticity.

In this research, we have proposed an approach that can control the COVID-19-related rumors on social media up to certain limits as much as possible. We have extracted many tweets from Twitter, Instagram, and Reddit social networks for analysis purposes. We have also fetched the tweet’s user id, country, and time information to discover the list of maximum active users along with the region. After prepossessing the tweets, we have used the VADER python library to find each tweet’s polarity score. The aggregated polarity score measures each user’s reputation in the network. Next, we calculated the degree of trust used to calculate each tweet’s entropy. So, an entropy-based approach is addressed to verify the reality of the tweet’s source and content. We have also mentioned the ROLI algorithm to find the list of OLs. They matched each tweet’s entropy with the pre-decided threshold value and announced whether it was a rumor. The addressed ROLI method averages 91% accuracy, 93% precision, 95% recall, and 94% F1-score to identify OLs. It also reduced the number of diffusers by 26% faster, spreading the tweet’s veracity by around 22% and influencing users around 23% faster than other SNA measures. Thus, the experimental analysis demonstrated that the proposed approach enhanced effects over standard SNA measures w.r.t. various performance metrics.

Real-time monitoring of tweets, rumors, patterns, trends, and misinformation is required to avoid any trouble. Such control may identify only the verified information and ensure that only verified and trusted information would be transmitted to prevent dangerous consequences. One of the biggest challenges with the social network is the large dataset volume due to millions of frequently posted data by millions of users [107]. In the future, we might make an effort to design a model that can dynamically supervise the tweets and their origin under the constraints to be more vulnerable to the rumors. Also, we would attempt to discover some more advanced techniques and measuring tools to visualize and interpretative OLs recognition techniques that eventually avoid rumors and misinformation by analyzing a real-time dynamic dataset. The work might be extended to find OLs using users’ posts like images, audio, video, etc. Other social networks, like Facebook, YouTube, Flicker, Instagram, and others, can also be utilized for extracting the dataset to find the OLs. A new model might be designed that can dynamically oversee the tweets and their origin under the supervision of OLs so that other people would be more aware of the
Fig. 8. Visual representation of OLs impact for (a) Twitter, (b) Instagram, and (c) Reddit dataset.
Fig. 9. Analysis of Performance metrics between Proposed ROLI and standard SNA measures for (a) Twitter, (b) Instagram, and (c) Reddit dataset.

Fig. 10. Execution time analysis of ROLI algorithm with standard SNA measures.
COVID-19 rumors. Some more advanced deep-learning-based models can also be associated with other users’ multi-relational characteristics, such as user’s response time, geographical location, the domain of interest, etc., as upcoming directives for computing [108].

Author contribution

The author contributed to the study conception and design, material preparation, data collection and analysis.

Funding

The author declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability

Not applicable.

Declaration of competing interest

The author declare that there are no known conflicts of interest associated with this publication. There has been no significant financial support for this work that could have influenced its outcome.

Acknowledgements

Not applicable.

References

[1] World Health Organization, WHO Characterizes COVID-19 as a Pandemic, World Heal. Organ., 2020. https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19——11-march-2020.
[2] R. Radu, Fighting the ‘Infodemic’: LegalResponses to COVID-19 Disinformation, vol. 6, Soc. Media + Soc., 2020, 205630512049819, https://doi.org/10.1177/2056305120948190.
[3] M. Palacios Cruz, E. Santos, M.A. Veláquez Cervantes, M. León Juárez, COVID-19, una emergencia de salud pública mundial, Rev. Clínica Española 221 (2021) 55–61, https://doi.org/10.1016/j.rece.2020.03.001.
[4] M.S. Islam, T. Sarkar, S.H. Khan, A.H.M. Kamal, S. Kabir, D. Yesamin, M.A. Islam, K.I.A. Chowdhury, K.S. Anwar, A.A. Chughtai, H. Saadeh, M. Al-Hassan, B. Bremie, A. Albahlal, Time-aware domain-based social influence prediction, J. Big Data 7 (2020), https://doi.org/10.1186/s40537-020-0283-3.
[5] L. Jain, R. Katarya, S. Sachdeva, Opinion Leader Discovery Based on Text Analysis in Online Social Network, in: 4th Int. Conf. Inf. Syst. Comput. Networks, ISCON, 2019, https://doi.org/10.1007/978-3-030-19499-1_23.
[6] B. Abu-Salih, K.Y. Chan, D. Zhu, CredaSA: credibility rating of users in big social data incorporating semantic analysis and temporal factor, J. Int. Inf. Sci. 45 (2019), https://doi.org/10.1108/JIM-05-2018-0432.
[7] B. Abu-Salih, Domain-specific knowledge graphs: a survey, J. Netw. Comput. Appl. (2018), https://doi.org/10.1016/j.jnca.2020.102076.
[8] S. Meneghelli, N. Thompson, K. Lee, K.W. Wong, B. Abu-Salih, Unlocking social media and user generated content as a data source for knowledge management, Int. J. Knowl. Manag. 16 (2020), https://doi.org/10.1080/14737653.2020.1107240.
[9] B. Abu-Salih, K.Y. Chan, O. Al-Kadi, M. At-Talib, W. Dogramaci, C. Jeurah, Key opinion leaders in recommendation systems: opinion elicitation and diffusion, in: WSDM 2020 - Proc. 13th Int. Conf., Web Search Data Min., 2020, https://doi.org/10.1145/3336191.3371826.
[10] L. Jain, R. Katarya, S. Sachdeva, Role of Opinion Leader in the Diffusion of Products Using Epidemic Model in Online Social Network, in: 11th Int. Conf. Contemp. Comput. ICom, 2019, https://doi.org/10.1007/978-3-030-66698-2_27.
[11] K.G. Quinn, Applying the popular opinion leader intervention for HIV to COVID-19, AIDS Behav. 24 (2020) 3291–3294, https://doi.org/10.1007/s10461-020-02874-7.
[12] W. Fang, B. Gao, N. Li, Analysis of the influence of opinion leaders on public emergencies through microblogging, Open J. Soc. Sci. 8 (2020) 154–158, https://doi.org/10.4236/jss.2020.82016.
[13] A. Mohammed, A. Ferraz, Factors influencing user participation in social media: evidence from twitter usage during COVID-19 pandemic in Saudi Arabia, Technol. Soc. 66 (2021), https://doi.org/10.1016/j.techsoc.2021.102087.
[14] N. Keys, D.C. Thomsen, T.P. Smith, Opinion leaders and complex sustainability issues, Manag. Environ. Qual. Int. J. 21 (2010) 187–197, https://doi.org/10.1108/09593940970153883.
[15] M.J.H. Banaee, L. Foroozandehzadeh, N. Poormohammad, F. Tabrizi, Opinion Leader Identification in Social Network, in: 12th Int. Conf. Inf. Syst. Comput. Networks, ISCON, 2012, pp. 193–196.
[16] Y. Xing, X. Wang, C. Qi, L. Li, W. He, Research on opinion polarization by big data analytics capabilities in online social networks, Technol. Soc. 68 (2022), https://doi.org/10.1016/j.techsoc.2022.102190.
[17] N. Siyam, O. Alapayesi, S. Abdallah, Mining government tweets to identify and predict citizens engagement, Technol. Soc. 60 (2020), https://doi.org/10.1016/j.techsoc.2019.101211.
[18] J. Tang, H. Gao, X. Hu, H. Liu, Exploiting homophily effect for trust prediction, in: WSDM 2013 - Proc. 6th ACM Int., Conf. Web Search Data Min., 2013, pp. 55–62, https://doi.org/10.1145/2433936.2434050.
[19] Y.Q. Wang, X.Y. Yang, Y.L. Han, X.A. Wang, Rumor spreading model with trust mechanism in complex social networks, Commun. Theor. Phys. 59 (2013) 510–516, https://doi.org/10.1088/0253-6102/59/4/21.
S. Al-Oufi, H.N. Kim, A. El Saddik, A group trust metric for identifying people of
P. Domingos, M. Richardson, Mining the network value of customers, in: Proc.
N. Lin, Building a network theory of social capital, Connections 22 (1999) 28–51
Y. Li, S. Ma, Y. Zhang, R. Huang, Kinshuk, an improved mix framework for
J. Duan, J. Zeng, B. Luo, Identification of opinion leader in online social network using
Y. Hu, Q. Pan, W. Hou, M. He, Rumor spreading model with the different attitudes
towards rumors, Phys. A Stat. Mech. Its Appl. 502 (2018) 331–344, https://doi.org/10.1016/j.physa.2018.02.096.
J.J. Cheng, Y. Liu, B. Shen, W.G. Yuan, An epidemic model of rumor diffusion in online social networks, Eur. Phys. J. B 86 (2013) 29, https://doi.org/10.1140/epjb/e2013-40843-5.
P. Jia, C. Wang, G. Zhang, J. Ma, A rumor spreading model based on two propagation channels in social networks, Phys. A Stat. Mech. Its Appl. 524 (2019) 342–353, https://doi.org/10.1016/j.physa.2019.04.163.
P. Redaghi, S. Goliat, A NOVEL MODEL for RUMOR SPREADING on SOCIAL NETWORKS with CONSIDERING the INFLUENCE of DISSENTING OPINIONS, Adv. Complex Syst. 21 (2018), https://doi.org/10.1142/S021948781850011X.
Y. Xiao, D. Chen, S. Wei, Q. Li, H. Wang, M. Xu, Rumor propagation dynamic model based on evaluation and anti-rumor, Nonlinear Dyn. 95 (2019) 523–539, https://doi.org/10.1007/s11071-018-4579-1.
Z. Li, Q. Zhang, X. Yu, M. Ya, S. Wang, Social media rumor rumour effectiveness: evaluation, modelling and enhancement, Inf. Process. Manage. 58 (2021), 102569, https://doi.org/10.1016/j.ipm.2020.102569.
L. Xu, B. Zafarani, A Survey of Fake News, ACM Comput. Surv. 53 (2020) 1–40, https://doi.org/10.1145/3395046.
X. Chen, N. Wang, Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality, Sci. Rep. 10 (2020) 5887, https://doi.org/10.1038/s41598-020-62585-9.
T.G. Kurtz, 9. Epidemic models, in: Approx. Popul. Process., Society for Industrial and Applied Mathematics, 1981, pp. 63–67, https://doi.org/10.1137/1.9781611970333.ch9.
M.S. Granovetter, The strength of weak ties, Am. J. Sociol. 78 (1973) 1360–1380.
L.J.S. Allen, An introduction to stochastic epidemic models, Lect. Notes Math. 1908, https://doi.org/10.1007/978-3-030-22338-0_23.
T. Britton, Epidemic models on social networks, 2014, pp. 127–131, https://doi.org/10.1016/j.eswa.2012.05.084.
L. Jain, R. Katarya, Identification of opinion leader in online social network using firefly optimization algorithm in online social network, Expert Syst. Appl. 142 (2020), 113016, https://doi.org/10.1016/j.eswa.2019.113016.
L. Jain, R. Katarya, S. Sachdeva, Recognition of opinion leaders coalitions in online social networks using games theory, Soc. Netw. Anal. Min. 203 (2020), 100158, https://doi.org/10.1016/j.jknoys.2020.106158.
C. Xia, S. Sun, F. Rao, J. Sun, J. Wang, Z. Chen, SI model of epidemic spreading on dynamical networks with community, Front. Comput. Sci. China 3 (2009) 361–365, https://doi.org/10.1007/s11704-009-0057-8.
Y. Zhang, J. Xu, A rumor spreading model considering the cumulative effects of memory, Discret. Nat. Soc. (2015) 1–11, https://doi.org/10.1155/2015/531245.
N. Zhang, H. Huang, B. Su, J. Zhao, B. Zhang, Dynamic 8-state ICAR rumor propagation model considering official rumor refutation, Phys. A Stat. Mech. Its Appl. 415 (2014) 333–346, https://doi.org/10.1016/j.physa.2014.07.023.
A. Redaghi, S. Goliat, A NOVEL MODEL for RUMOR SPREADING on SOCIAL NETWORKS with CONSIDERING the INFLUENCE of DISSENTING OPINIONS, Adv. Complex Syst. 21 (2018), https://doi.org/10.1142/S021948781850011X.
Y. Hao, X. Chen, S. Wei, Q. Li, H. Wang, M. Xu, Rumor propagation dynamic model based on evaluation and anti-rumor, Nonlinear Dyn. 95 (2019) 523–539, https://doi.org/10.1007/s11071-018-4579-1.
Z. Li, Q. Zhang, X. Yu, M. Ya, S. Wang, Social media rumor rumour effectiveness: evaluation, modelling and enhancement, Inf. Process. Manage. 58 (2021), 102569, https://doi.org/10.1016/j.ipm.2020.102569.
L. Xu, B. Zafarani, A Survey of Fake News, ACM Comput. Surv. 53 (2020) 1–40, https://doi.org/10.1145/3395046.
X. Chen, N. Wang, Rumor spreading model considering rumor credibility, correlation and crowd classification based on personality, Sci. Rep. 10 (2020) 5887, https://doi.org/10.1038/s41598-020-62585-9.
T.G. Kurtz, 9. Epidemic models, in: Approx. Popul. Process., Society for Industrial and Applied Mathematics, 1981, pp. 63–67, https://doi.org/10.1137/1.9781611970333.ch9.
M.S. Granovetter, The strength of weak ties, Am. J. Sociol. 78 (1973) 1360–1380.
L.J.S. Allen, An introduction to stochastic epidemic models, Lect. Notes Math. 1908, https://doi.org/10.1007/978-3-030-22338-0_23.
COVID-19 around the world: susceptibility to COVID misinformation, R. Soc.
Open Sci. 7 (2020), https://doi.org/10.1098/rsos.201199, 201199.

[106] R. Kouzy, J. Abi Jaoude, A. Kraitem, M.B. El Alam, B. Karam, E. Adib, J. Zarka,
C. Traboulsi, E. Akl, K. Baddour, Coronavirus Goes Viral: Quantifying the COVID-
19 Misinformation Epidemic on Twitter, Cureus, 2020, https://doi.org/10.7759/
cureus.7255.

[107] S. Stieglitz, M. Mirbabaie, B. Ross, C. Neuberger, Social media analytics –
challenges in topic discovery, data collection, and data preparation, Int. J. Inf.
Manag. 39 (2018) 156–168, https://doi.org/10.1016/j.ijinfomgt.2017.12.002.

[108] Q. Tan, N. Liu, X. Hu, Deep Representation Learning for Social Network Analysis,
2019, https://doi.org/10.3389/fdata.2019.00002. ArXiv.