Online Social Networks and Media 9 (2019) 30–42

Contents lists available at ScienceDirect

Online Social Networks and Media

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

Source detection of rumor in social network – A review

Sushila Shelke[a,*], Vahida Attar[b]

[a] Ph.D. Research Scholar, Department of Computer Engineering & IT, College of Engineering, Pune (COEP), 411005, India
[b] Associate Professor, Department of Computer Engineering & IT, College of Engineering, Pune (COEP), 411005, India

ARTICLE INFO

Article history:
Received 14 September 2018
Revised 22 December 2018
Accepted 23 December 2018

Keywords:
Misinformation
Rumor
Social network
Source detection

ABSTRACT

The ubiquity of handheld devices provides straightforward access to the Internet and Social networking. The quick and easy updates from social networks help users in many situations like natural disasters, man-made disasters, etc. In such situations, individuals share information with the people in their network without checking the veracity of posts, which leads to the issue of rumor diffusion in a social network. Detection of rumor and source identification plays a vital role to control the diffusion of misinformation in a social network and also a good research domain in social network analysis. Source detection of such misinformation is often interesting and challenging task due to the fast diffusion of information and dynamic evolution of the social network. Accurate and quick detection of the rumor source is a very important and useful task in many application domains like source of disease in an epidemic model, start of virus spread, source of information or rumor in a social network. Most of the existing reviews which focused on source detection relate to various application domains and network perspective. But as per the need of current social networking usage and its influence on the society, it is a crucial and important topic to review the source detection approaches in the social network. The objective of this paper is to study and analyze the source detection approaches of rumor or misinformation in a social network. As an outcome of the literature study, we present the pictorial taxonomy of factors to be considered for the source detection approach and the classification of current source detection approaches in the social network. The focus has been given to various state-of-the-art source detection approaches of rumor or misinformation and comparison between approaches in social networks. This paper also focused on research challenges in current source detection approaches, public datasets and future research directions.
© 2018 Elsevier B.V. All rights reserved.

1. Introduction

Nowadays people are living in the society where everyone is connected to various networks like Social network, Internet, Biological network, Technological network, etc. [1] from which they acquire, process and share the information in a network which rabidly leads to an increasing amount of information propagation and diffusion [2]. Emergence and growth trend of social networking sites like Twitter, Facebook, and Reddit are proven as very helpful in disaster situations such as natural disasters (Flood, Storm, Earthquake), man-made disaster (Shootouts, Terrorist attacks) and emergencies [3,4]. The news and information diffusion across social sites got more research attention. This is because social media is a common means for disseminating trending discussions and breaking-news which may contain unproven information regarding events or incidents happened in the world. As per the statistical survey of social network, by the year 2021 there will be 3.02 billion monthly active users worldwide, approximately one third of the Earth's total population [5]. To deal with the analysis of such a huge data is very vital and demanding task.

The rise in interconnections of the network discloses a large range of hazards like viruses, misinformation, rumors with a frequently severe end result [6–8]. The latest example of rumor related to "600 Murders Take Place in Chicago during the second weekend of August 2018" presents a fear and an anxiety about such a large number of violence in the city [9]. The statistical statement was given in the Television show to just target the politicians and their promises during the election campaign. The actual truth after verification said that during the second week of August 2018 only a single murder was found and 600 was totally a misinformation because the city had not seen 600 murders in the entire 2018 up to the date. These types of rumor can spread widely in a social network and introduce many questions about the security of the people living in the city. The extensive spread of misinformation can lead to unacceptable, destructive [10] and negative impacts on individuals and society.

* Corresponding author.
E-mail addresses: sss17.comp@coep.ac.in (S. Shelke), vahida.comp@coep.ac.in (V. Attar).

https://doi.org/10.1016/j.osnem.2018.12.001
2468-6964/© 2018 Elsevier B.V. All rights reserved.

The information diffusion in social sites opened many research trends like detection of misinformation or rumor, checking recognition of social bots, monitoring the spread of fake news, prediction of future diffusion and source detection of rumors, etc. An unverified information which diffused very rapidly in the social network is referred to as rumor. Rumors may disseminate misinformation (wrong information) or disinformation (intentionally wrong information) [11] before, throughout and after the disasters or incidents [12]. After investigation of true and fake news diffusion in a social network [13], it has been observed that the diffusion of false news is faster and deeper than true news and news related to politics, disaster and critical situations are widely spread compared to conventional news. Many researchers have focused on information spread during or before an election for identifying the diffusion of astroturf [14] or detection of social bots [15], identification of rumors [16], misinformation [17], etc. Similarly, analysis of posts after disaster situations like Earthquake [18], Flood [19], Blasts [20], Shootouts [21] is one of the most popular research topics in social networks. In many cases the original text of a message is modified by different users by mistake or with intention, which leads to the diffusion of rumor in the network [22]. To ensure the wide spread of information during campaign [23], the social bots are induced into the social network for manoeuvring the data, such as information in stock market [24], where social bot is a program which behaves like a user and automatically creates contents and interacts with others in a social network. In rumor detection, many features like diffusion patterns and network structures, contents and sentiments features, temporal features are used for early identification of raising advertisement campaigns [25]. Also, the features of cognitive psychology (consistency, common acceptability and coherency of contents and credibility of source) are used for misinformation detection in a social network [26]. The experience of the contestants of DARPA-Twitter bot challenge [27] concludes that the various factors like behavioural and inconsistency modelling, network and text analysis along with semi-supervised machine learning techniques are useful for enhancing the accuracy of bot detection. The dissemination of rumor in network generates many risks like fear of epidemic virus among people, erroneous decisions in disaster situations and harming the reputation of individuals or an organizations. The prevention and control of large diffusion of rumors in social networks is very important. The rumor diffusion in a network can be controlled by early detection of rumors [28], checking the truthfulness of rumors [29] or misinformation, and identification of rumor source [30].

Non-credible information spreads rapidly in online social networks. Detecting fast and accurate source of rumors in a social network is a very challenging task due to complex diffusion process, real-time data, and dynamic changes in the network. In recent times, many web-based systems have been developed to detect and evaluate the rumors which include (i)TwitterTrails.com [31], a system which permits users to determine the features of propagated rumors and its falsification, (ii) TweedCred [32], a instantaneous system to judge trustworthiness of posts on Twitter, (iii) Hoaxy [33], a platform for tracking the misinformation in a social network, (iv) Emergent.info [34], a real-time rumor follower that focused on rising tales on the internet and observe their faithfulness, and (v) Snopes.com [35], factcheck.org [36], an admired websites archiving memes and urban myths. The reality checking abilities of these rumor detection systems validates the authentication of rumors on web and vary from entirely automatic to semi-automatic. But, these systems do not track or observe the diffusion progress and do not detect all possible source(s).

Source detection is very significant in various application domains such as Medical (to find the source of epidemic), Security (to detect the source of virus), Large interconnected network (to detect the flaws in power grid network, gas or water pipeline network), Social network (to identify the culprits who spread wrong information), Financial network (for checking the reasons of cascade failures), etc. Due to its wide scope in different applications, past two decades observed large improvements in source detection techniques. Major research has been done for source identification in different application areas like finding the first patient to control an epidemic of disease [37], source of virus [38], gas leakage source in wireless sensor network [39], source in email network [40], propagation sources in complex networks [41] and source of rumor or misinformation in a social network [42,43] which are directly or indirectly related to rumor source detection. From the existing work, it can be observed that very less review work has been done on source detection approaches for rumor or misinformation in the social network. Therefore, the aim of this survey paper is to understand and analyze the growth made by source detection approaches for rumors in the social networks.

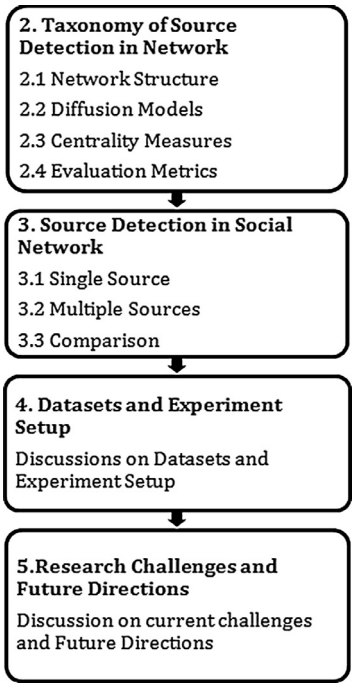

Fig. 1. Roadmap of the paper.

This paper provides a systematic study as follows:

- The pictorial taxonomy of factors to be considered for source detection has been proposed.
- Comprehensive survey of state-of-the-art source detection approaches of misinformation and rumor in a social network is presented.
- Publicly available datasets and experimental setup used for source detection approaches in a social network, the needs and challenges of source detection are explored.
- The current issues and potential future directions for source detection techniques of rumor in social networks are thoroughly presented.

Fig. 1 outlines the roadmap of this paper. Section 2 presents the taxonomy of factors required for source detection in the network. Section 3 deals with the techniques related to source detection of rumor in the social network and their comparison. Datasets and experimental setups are shortly overviewed in Section 4. Research issues and future directions in source detection of rumor are stated in Section 5. Section 6 concludes the paper.

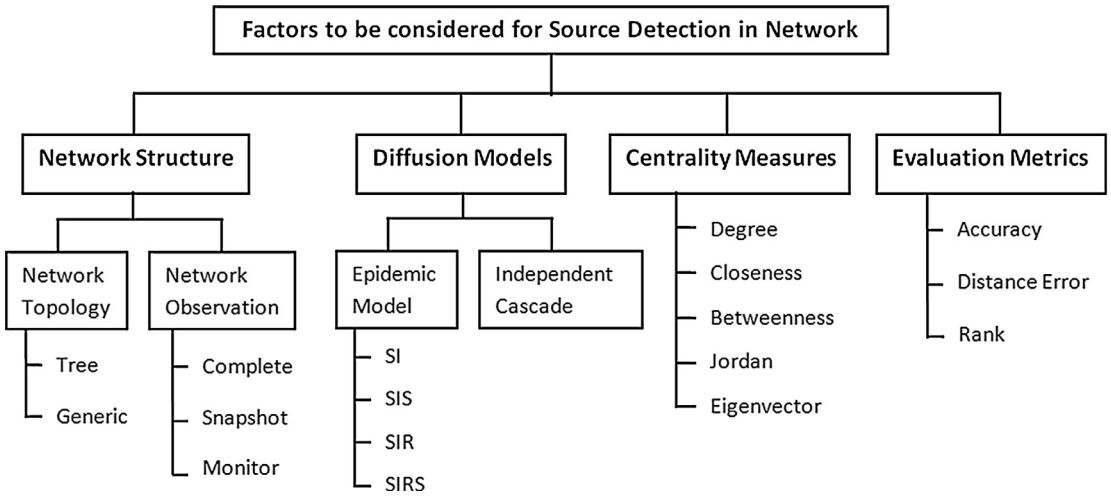

Fig. 2. Taxonomy of factors to be considered for source detection in network.

2. Taxonomy of source detection in network

Source detection is to find the person or location from where the entities like a viral disease, a virus in the network or a misinformation in the social network started initially. Different domains are represented by a network like the Network of computers, Network of peoples, Social Network, etc. where the source identification has been performed to find the root or origin. The taxonomy of source detection in the network is depicted in Fig. 2. The taxonomy is represented based on various aspects to be considered for identification of source in the network, which includes Network Structure, Diffusion Models, Centrality Measures and Evaluation Metrics.

2.1. Network structure

The Network structure is classified into network topology and observation of the network. Network topology is referred as the structure of a network in terms of tree or generic graph. The source detection is more complicated in the generic graph topology than the tree, as in the tree topology the source can be identified by traversing the parent nodes of nodes of the tree up to the root. According to the survey [41] networks are observed mainly in three categories as complete, snapshot and monitor based approach which gives knowledge about states of nodes during rumor propagation.

2.1.1. Network topology

In each domain like medical, security, etc. networks are normally represented by graphs which includes Network of computers, Social Network, Interconnected network of water, gas pipeline or power grid and network of people. These graphs are used to reconstruct two topologies as Tree and Generic network. In the beginning, the problem of source detection has been solved by considering the network as a random d-regular tree where every node has the similar degree as d in [44] and random tree in [45,46] for source detection. Generic network is the random graph in which Breadth First Search technique is used to construct the generic network into the tree referred as BFS tree. In generic network, the rumor source has been identified with the assumption that the source node S has been selected randomly and consider as the starting node for the BFS tree which spreads the rumor across BFS tree [30]. Tree topology does not consider cycles; therefore the generic network is mostly used in the research. The experiment of generic network is performed on a common synthetic dataset as small world network [47] and a scale free network [48]. Fig. 3 shows an illustration of network topologies. Tree topology is majorly used for synthetic dataset. However, in the real world, the graph is most suitable structure to represent the network.

2.1.2. Network observation

The results of source detection methods vary according to the selection of network observations. Network observation provides knowledge about the different states of the nodes in the network, such as Infected node, i.e., the node which receives and propagates inaccurate information or Susceptible node, i.e., the one whose neighboring nodes have received the information and therefore has higher chance to be infected or Recovered, i.e., the node that either ignores the information or prevents its spread.

A. *Complete observation.* Complete observation of network provides broad knowledge of the momentary states of a node in the network [41]. It provides the sufficient knowledge of network required for source detection at a particular time. However, owing to the large scale of the network, the complete observation is hardly possible.
B. *Snapshot observation.* Snapshot based observation provides a limited information about the network at a stated time, which includes details about infected nodes observed at the time of the snapshot, the infection probability of the nodes that have been contaminated with the information [49]. The snapshot observation provides the details of only infected nodes, but cannot discriminate between susceptible or recovered node. To overcome the problem of partial knowledge through single snapshot, researchers take multiple snapshots [50,51] at varying time slots to get sufficient knowledge of the network.
C. *Monitor observation.* In a monitor observation, initially monitor or sensor nodes are inserted into the network, which will work as an observer for evolutionary propagation in network [52,53]. These monitor nodes are controlled by an administrator where the information about states of the node, information received by monitor node and their infection time are gathered. Some misinformation might get ignored if it did not pass through the monitor nodes and also accuracy depends on number of monitor nodes placed in the network. Accuracy can be improved by adding more monitor nodes in the network, however, it may decrease the performance of the system.

Overall from the network observation studied, it is understandable that the snapshot and monitor based observation gives limited data about the network. Multiple numbers of snapshot and

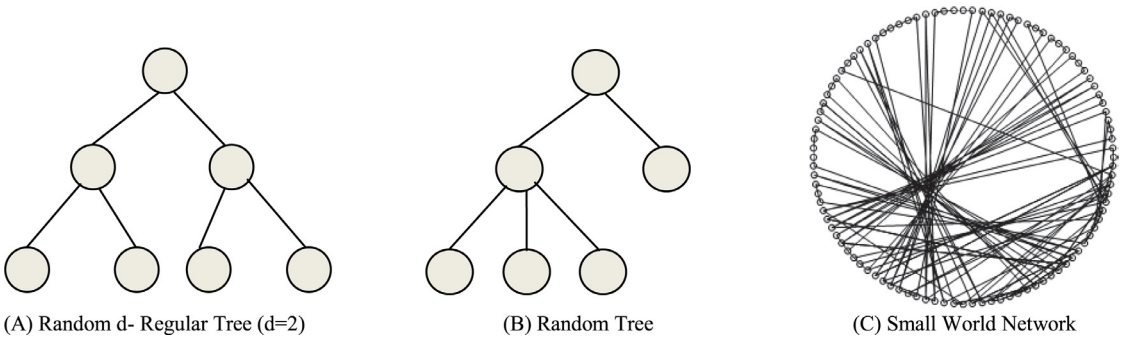

Fig. 3. Illustration of network topologies. (a) random d-regular tree; (b) random tree; (c) small world network as generic network.

monitor nodes can overcome the problem of less information up to some extent.

2.2. Diffusion models

Diffusion models are used to mine the data about where and when a part of information generated in the network and how rapidly the dissemination occurs [53]. Diffusion models are utilized to explain and reproduce the spread of information in the network. Epidemic models, the Independent cascade models are major examples of diffusion models used in source detection techniques. Epidemic models are the way infectious diseases are spread among the population. Epidemic models are mainly used in disease epidemic and are divided into four basic types as Susceptible-Infected (SI), Susceptible-Infected-Susceptible (SIS), Susceptible-Infected-Recovered (SIR) and Susceptible-Infected-Recovered-Susceptible (SIRS). The Independent cascade model is one of the information diffusion models, where the dissemination of an information moves from one person to another through their social relations and "infected" folks can, in turn, pass on the information to others.

2.2.1. Basic epidemic models

The epidemic models are basically used for finding the origin of viral disease and utilized for finding the sources of rumors because epidemic disease in the population is similar to rumor diffusion in a social network. Initially nodes are susceptible (S) and can be infected (I) with propagated rumors in the SI model. Susceptible nodes are uninfected nodes having infected neighbors and presenting therefore a higher probability of becoming infected due to contagion from neighbors and a contaminated node remains infected forever. In terms of rumor diffusion in social networks, infected node is the one who has received the rumor and susceptible node is the node which has not received any rumor, but due to its neighboring infected nodes, can become infected after receiving a rumor. SI model is not viable as it does not take into account that contaminated users can be cured after having been contaminated. In the SIS model, the probable states are yet again susceptible (S) and infected (I), but in this model, when susceptible nodes become contaminated they can cure back to susceptible after some period. In SIR model, the potential states are susceptible (S), infected (I), and recovered (R). The only variation between SIS and SIR model is that in the SIS model an infected node can become susceptible whereas in SIR model an infected node may recover by ignoring the message or not passing to neighbors. Recovered nodes remain in the recovered state further. In a social network, the recovered node is the one who is aware of a rumor, therefore, either it will delete the post or will not forward that post to a neighboring node. In SIRS model, a recovered node can again become a susceptible node with some probability. All these basic epidemic models are well explained in [54]. Fig. 4 shows a comparison of basic epidemic

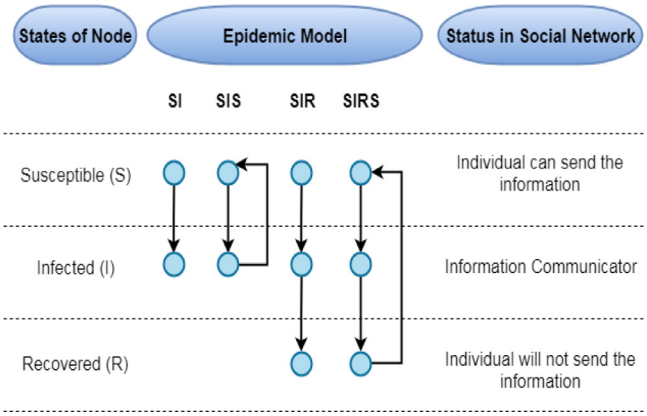

Fig. 4. Comparison of basic epidemic models.

models for information flow in social network and change in status of the node.

Epidemic models are utilized to determine the infection and recovery of rumor propagation in the networks and therefore are widely used in source detection. These models are also used for different motives such as SI for detection of infection source [55], SIS for identification of rumor source [56] and influential node [57], SIR for information source [58] and dissemination of topic in web forums [59] and SIRS model for analyzing botnet interactions in the network [60].

2.2.2. Independent cascade model (IC)

The information diffusion models are used to study the flow of information in the network, where the independent Cascade (IC) and the Linear Threshold (LT) models are two commonly used diffusion models. In both the models, the possible states are active or inactive. In IC model, the information passes in the network through cascades, where the active nodes are the one who received the rumor and become infected, other nodes are inactive [61].

In IC model, at one given instance the node has a only one chance to activate one of its inactive neighboring nodes. The illustration of IC model is shown in Fig. 5 where green nodes are active nodes, and yellow nodes are newly activated nodes by neighbors and white nodes are inactive nodes. As shown in the figure, at step 0 or at one instance, node 1 and 2 are active nodes, which spread the rumor cascade to their respective inactive neighbors as node 4 and 5 in step 1, thus becomes active at step 2 and spread the cascade to inactive nodes 3 and 4. In the final step 3, node 1 to 6 are active node, meaning those nodes are infected by the rumor, node 7 is newly infected node and node 8 remains inactive. The IC model describes the flow of information from source to other nodes in terms of a directed graph, which is helpful for finding highly influential users and identifying a source

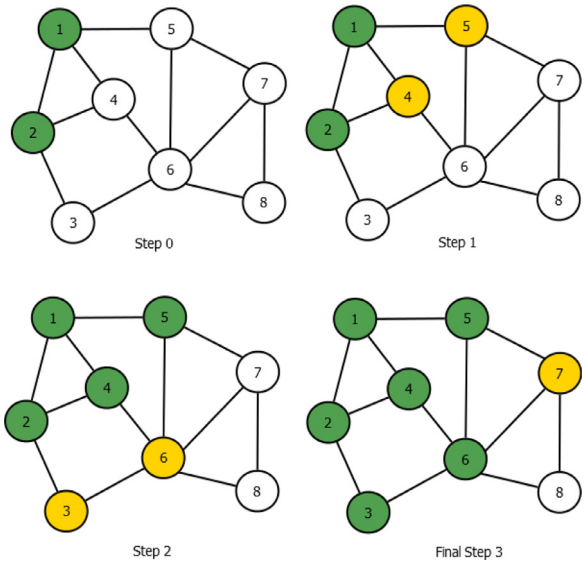

Fig. 5. Independent cascade model.

of misinformation by analyzing the diffusion network in reverse direction [42]. IC model is also used for finding the maximum influence [62], predicting the development of cascades [63] and understanding the structure of diffusion in social network [64].

In LT model, each node in the network selects a random threshold value, then at each time step inactive node gets influenced by all its active neighbors. The inactive node turns into active once the total weight of all its incoming neighbors reaches a threshold value. Similar to IC model, LT model is majorly utilized to maximize the influence of spread in network [65,66].

In diffusion models, the probabilities of infection and recovery of nodes are considered to understand the evolution of rumor propagation. The infection probability of a node is the rate at which node who has received the rumor may forward it to (infect) other nodes. The node who has been infected with the rumor and also finds that the post is rumor and then discards it (recovers) with a defined probability, referred to as recovered probability [67]. These infections and recovery rates are considered as uniformly constant or computed and are classified as homogeneous if all the neighboring nodes get infected with the same infection probability and heterogeneous if the infection probability assumed to be different among the neighboring nodes. In source detection of rumor, the infection-recovery probabilities with homogeneous [42,44] and heterogeneous probabilities [49,51] used with different diffusion models.

2.3. Centrality measures

Centrality measures are used to assign count to each node in the graph which put across the influence of nodes in the diffusion process [68,69]. The research in the domain of source detection has utilized the following six majorly used centrality measures. Fig. 6 illustrates the different centrality measures [39].

A. *Degree centrality*. It is referred as the number of edges linked to the node in the graph. In the real world, famous people are having a high degree of connectivity with other people in the network [70]. Fig. 6(A) shows the example of degree centrality, where the central node has a degree of 6.
B. *Closeness centrality*. It is the mean shortest distance between a node and all the other accessible nodes in the graph [71]. Fig. 6(B) shows that node C has a higher closeness centrality as it is close to all the remaining nodes in the graph.
C. *Betweenness centrality*. It describes the number of times the vertex appears in the shortest paths between any other nodes in the network and acts as a bridge. Researchers observed that the nodes with higher betweenness centrality which might not have maximum degree also play an essential role in the information propagation [72–74]. Fig. 6(C) clarifies that node D is having maximum betweenness centrality since it is present in maximum number of shortest paths.
D. *Jordan centrality*. It is characterized as a node having the smallest maximum distance to other contaminated and recovered nodes [75]. Similarly, the number of Jordan centers is equivalent to the radius of the graph [76]. Fig. 6(D) gives the example of Jordan centrality, where nodes A, D and C are having Jordan centrality 3.
E. *Eigenvector centrality*. It measures the centrality of a node as a sum of degree centrality of all its connected nodes or neighbors. It is the eigenvector of the adjacency matrix coupled with largest eigenvalue [77,78]. A node with higher eigenvector centrality has leading influence in the surrounding nodes in the network. In Fig. 6(E) nodes V_1 and V_3 have the maximum eigenvector centrality.

2.4. Evaluation metrics

Different evaluation metrics are used for source detection in the network such as Execution time (Time taken to identify the estimated source/s), F-measure and Precision (Metrics to check the accuracy of estimated or identified sources), Distance Error (The count of nodes between actual source and estimated source by algorithm) and Rank. Some evaluation metrics are defined in the following sections.

A. *Accuracy*. The accuracy of source detection is measured in terms of two metrics, F-Measure and Precision. The F-Measure metric calculates the overall accuracy as the ratio of correctly identified sources to the sum of all testing sources [79]. It can be defined as below:

$$F\ Measure = \frac{2 \times precision \times recall}{precision + recall} \tag{1}$$

Where, precision is the ratio of the number of correctly identified sources over the number of all retrieved sources which is defined in Eq. (2) and recall is the ratio of the number of correctly identified sources over the ground truth sources, described in Eq. (3).

$$precision = \frac{|\{retrived\ sources\} \cap \{true\ sources\}|}{|\{retrived\ sources\}|} \tag{2}$$

$$recall = \frac{|\{retrived\ sources\} \cap \{true\ sources\}|}{|\{true\ sources\}|} \tag{3}$$

B. *Rank*. The rank is the location of the actual source in the list of nodes sorted in decreasing order by the score. If the real source has accurately the same score as any other node (or nodes), the true source is always below that node (these nodes) in the score list sorted in descending order [80]. The ranking measure is well suited for identifying a small group of nodes among which the source node is present.
C. *Distance error*. The distance error is referred as the shortest distance of hops between the accurate source and estimated source found by an algorithm [51].

3. Source detection in social network

Identifying the origin or source where the propagation started is a generic problem in any network. However, identifying the accurate rumor sources at the earliest [81] in a social media network

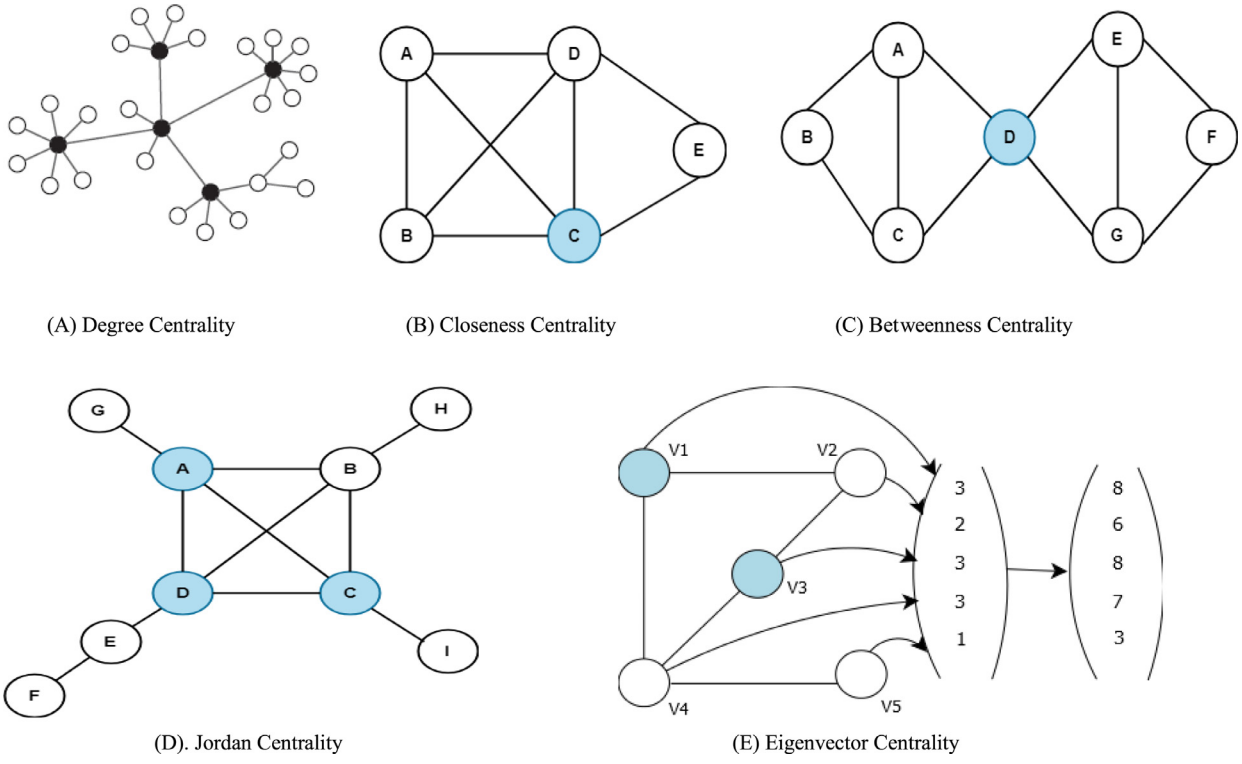

Fig. 6. Illustration of centrality measures (a) degree centrality; (b) closeness centrality; (c) betweenness centrality; (d) jordan centrality; (e) eigenvector centrality.

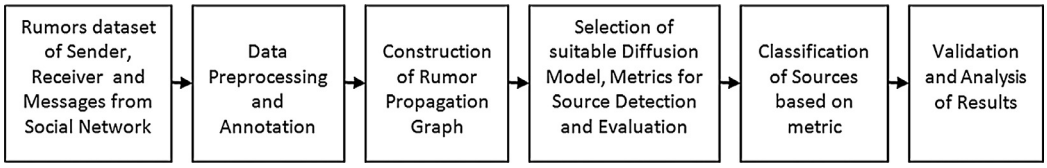

Fig. 7. Process for source detection of rumor in social network.

is very important to prevent and control the diffusion and destructive effects caused by rumors. The normal process for source detection of misinformation or rumor when considering the Twitter social network is shown in Fig. 7 and can be described in terms of the following steps.

1. Identify the rumors and collect the data for rumors in the social network with respect to sender, receiver and the post or message send by the user.
2. Preprocess the collected data, e.g. removal of urls, hashtags, stopwords etc. in twitter and carry out data annotation.
3. Construct the propagation graph for rumor or misinformation.
4. Select the suitable diffusion model like Epidemic models, IC model, etc. and metrics for source detection like centrality measures and metrics for evaluation like accuracy in terms of precision, F-measure, distance error etc.
5. Classify the sources based on source detection metrics.
6. Evaluate the result with actual and estimated sources.

The classification of source detection approaches in social networks is depicted in Fig. 8. This paper classifies the source detection into two major categories as a single source and multiple source detection and major focus of the paper is for source detection in a social network.

3.1. Single source detection

The majority of research work has contributed to finding a single source in social networks, which can be further classified based on network observation, query and anti-rumor based approach.

3.1.1. Network observation based approach

Based on three types of network observations studied in Section 2.1.2, source detection approaches in social networks are classified as follows.

Complete observation. Shah and Zaman [44] first proposed the work of rumor source identification in a tree like network, where they assume that every node in the network receives information from one of its neighbor nodes. Rumor centrality metric is used for source estimation and considers SIR model for information diffusion. The rumor centrality of a node is described as a number of definite propagation paths starting from the origin node. The node having higher rumor centrality is the source of information propagation. The efficacy of the rumor centrality measure is further analyzed in [82]. They perform experiments on the synthetic dataset for random d-regular tree, random tree and general network. For general network, the original network graph is converted into a tree referred as a BFS tree using Breadth First Search (BFS) technique, in which any node in the graph assumed as source node is considered as starting node for BFS. BFS tree is used with infection probability p to find the origin. The same methods and results are

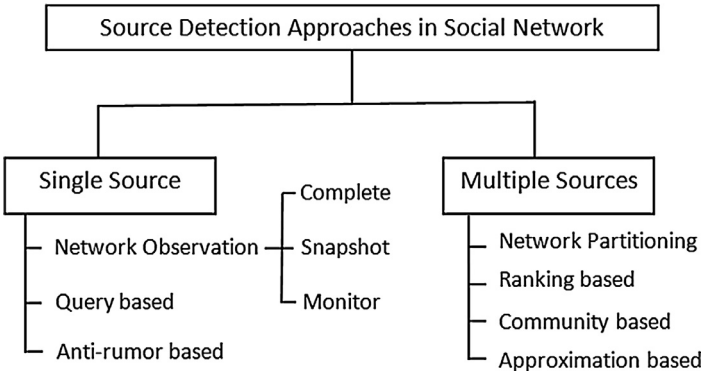

Fig. 8. Classification of source detection approaches in social networks.

utilized for source identification using local rumor center metric by Dong et al. [83]. By considering the susceptible nodes as finite and end vertices (the node which receives the rumor, but does not forward further) as a boundary node for source detection using rumor centrality metric is proposed by Yu et al. [84]. They consider a finite graph for source detection and use the message passing approach to reduce the search of vertices for maximum likelihood estimation.

The node with the maximum rumor centrality compared to the other suspicious infected node is referred as local rumor center and is further considered as origin node. The local rumor center is beneficial as it reduces the searching scale to a set of suspected nodes only. They required some information about the main network in advance, such as a rumor contaminated subgraph and a set of nodes as the "suspect" nodes. In the situation, where the time taken to detect the accurate source goes outside a limit, it would miss out the significant harm caused by the misinformation in the network. By understanding the need of fast detection of rumor source, Jain et al. [85] proposed a method using hitting time information about substitute random walk process with complete network and discrete time SI model. They randomly select the infection probability from [0, 1] and also require the infected graph and observation time for computation.

Snapshot observation. The rumor centrality metric has been applied for single source detection by Wang et al. [50] in a tree like network. Although, the snapshot provides a partial information about the network, the author proves that multiple independent snapshot of network improves the performance. The same measure is used when the node gives the infection probability μ [86] independently of each other and also proves that the rumor centrality method assures a good accuracy for large value of μ. Zhu and Ying [58] derived a sample path based method with homogeneous infection probability on the network to identify the propagation source. The source along with the best possible sample path, i.e. the one which most probably leads to the observed network snapshot was proven as the Jordan center of the contagion graph and is measured as a source. They also examine the performance of a reverse infection process for regular tree and conclude that, irrespective of the number of contaminated nodes and time of network observation, the distance between true source and identified source remains constant. With the same sample path estimator and the Jordan centrality measure, the work was extended to the heterogeneous SIR model in [49] for single source and tree network. The heterogeneous SIR model considers that the infection and recovery probabilities among some of the two neighboring nodes vary from each other, in which they demonstrate that, the source node coupled with the optimum sample path is the Jordan center of infinite tree. They required a small set of infected nodes in prior and identify the source in sparse observations of infected networks. The same approach is used by Luo et al. [55] in SI and SIS [87] diffusion model. Cai et al. [51] proposed the source detection with heterogeneous infection probability using the continuing in time SI model for single source with a dependent snapshot of the network. The cost of taking many sequential snapshots may affect the computation time required for source detection.

The problem of source identification along with rumors detection was proposed by Sahana et al. [88] using twitter network where they focused only on time of the posted tweet and rumor text. They majorly aim for detection of rumors and proposed a method which provides account information of users who propagated identified rumors in twitter. Krol and Wisniewska [89] perform the source detection in the Twitter network for varying sizes of tweets and conclude that in the small network it is difficult to be discriminating the early rumor users from most influential spreaders, contrary to the case for the large network.

Monitor observation. Xu and Chen [43] proposed a source detection approach by introducing monitor or sensor nodes in the network and without using textual information of a rumor. In this approach, the monitor nodes are induced into the network to find the ranking of actual source using rumor quantifier metric. They analyzed the different cascades for rumor and non-rumor posts and use a dynamic IC model to identify the source of a rumor. They reveal that the IC model is represented by directed graph and cascades in IC model are automatically depicted as directed Tree. In this, the accuracy of rumor source detection depends on number of sensor nodes. Jiang et al. [52] proposed a rumor source detection in time varying or temporal network using SIR model. They designed a novel Maximum Likelihood (ML) based estimator to check the dynamic evolution in the network. The Time varying network is converted into a set of static networks with discrete time windows. The SIR diffusion model with heterogeneous infection probability is considered for source detection. They perform experiments on the snapshot and monitor based network observation where it is inferred that sensor based observation gives good accuracy for source detection. This paper gives the future direction for source detection in continuous time varying social network. Pinto et al. [90] proposed the PTVA (Pinto, Thiran and Vetterli Algorithm derived for the initials of the authors), a method with number of monitor nodes and receiving time of posts at observer nodes and also, they assume that the rumor spreads along the BFS tree. In this, they observe the data from all the monitor nodes, which is time consuming. By ignoring the monitor nodes with poor quality information and selecting only important sources with maximum likelihood Paluch et al. [80] improvise a method of source detection for scale free network in terms of accuracy and time as compared to [90] using the SI model.

3.1.2. Query based approach

The query based approach is proposed by Choi et al. [91] for source detection. They generate a network based on queries and answers. Firstly, they used a simple batch query where prior information related to the amount of questions, the candidate set, estimators for a given propagation snapshot and the models of answers is required. Secondly, they used the interactive query where directions are based on the answers of query "Who spreads the rumor to you?" and the nodes are directed at the network. It considers the rumor centrality measure with the SI diffusion model. They guarantee the detection performance when there is a regular tree-like structure and perform experiments on the real world Facebook network.

3.1.3. Anti-rumor based approach

The monitor based network observation is used to spread anti-rumor information in the network for source identification by Choi et al. [92]. They examine the detection performance by injecting hidden agents or monitors called as protectors to send the "anti-rumor" messages. The job of this guard node is to broadcast the true information in opposition to the rumor, like doctors protect people against contagious disease using vaccine. Maximum-A-Posterior-Estimator (MAPE) is used to understand the distribution parameters in source and protector node to detect the source. They use regular tree structure and a variation of the SI model, namely SIP where P stands for protected status.

3.2. Multiple source detection

Most of the researchers assume that there is a single source of misinformation or rumor in the network, but the information may propagate from single or multiple sources to increase the diffusion rapidly. Few researchers have focused on identifying multiple sources and generic approaches for detecting single or multiple sources. Wang et al. [30] has extended their work of single source using rumor centrality on the synthetic dataset for multiple sources. In this, they follow breadth-first-search (BFS) approach to get the final tree from several observations. The guaranty of source detection probability depends on the number of network observations. Multiple source identification has been classified into following four approaches: (i) network partitioning, (ii) ranking based, (iii) community based and (iv) approximation based.

3.2.1. Network partitioning

The first time rumor centrality metric is evaluated for a set of nodes to find multiple sources by Luo et al. [81] in two phases. In the first phase, they consider a number of infected nodes as sources. Voronoi partition method [93] is used to divide all the infected nodes into various partitions on these sources. The source in each partition is identified using the rumor center method. In the second phase, identified sources are adjusted by the two-source estimator among any two adjacent partitions. These two phases are repeated awaiting the determined sources become stable. New metric of effective distance [94] for the complex diffusion process is used by Jiang et al. [95] to find multiple rumor sources. They used Capacity Constrained Network-Voronoi Diagram (CCNVD) [96] approach to partition the network. The SI model that does not consider the recovery of nodes is used and also the knowledge of infected node is required in prior.

3.2.2. Ranking based

The way of finding multiple sources of misinformation is proposed by finding the top k suspects in the network using rank based and optimization based approach by Nguyen et al. [42]. They use a greedy approximation approach and the reverse diffusion process in the IC model to identify the suspects from the set of already infected nodes. They evaluated the approach for attacks or misinformation from multiple sources and for incomplete data. Kumar and Geethakumari [97] proposed a method to identify the multiple sources using cognitive psychology principle and the Gini coefficient metric (a measure to check the distribution of messages among peoples in the network). Also, they identify the sources those are colluding with each other for rapid diffusion of misinformation in the network. The process of cognitive psychology validates the factors as consistency, coherency and general acceptability of the message and credibility of source to identify the acceptance level of the messages. The PageRank algorithm is used to determine the normal acceptance level of the user, and messages send by users and classify the sources as credible or non-credible based on acceptance level and propagation pattern. The collusion between sources for misinformation diffusion in the network by retweeting each other's posts could be determined using core-periphery and community detection algorithm.

3.2.3. Community based

Zang et al. [98] proposed a community partitioning method to recognize multiple sources in each community. They follow SIR propagation model and experiment on the synthetic network generated by NetworkX, python based software package further explained in Section 4. The reverse diffusion approach is employed to identify the unseen and recovered infected nodes, then community detection method is applied to group the infected nodes into various partitions. The problem of multiple source locating problem is solved by applying the single source locating problem into different communities. Zang et al. [67] extended their work of multiple source detection in a real world dataset using divide and conquer strategy for reducing the computation complexity. They follow the SIR model with eigenvector based metric.

3.2.4. Approximation based

Nguyen et al. [99] proposed an approximation algorithm for source detection using the IC model and heterogeneous infection probability which is suitable only for progressive models, that means once a node turn into infected node, it stays infected forever. The prior knowledge of infected nodes is not required for this model. They identify the seed set which minimizes the difference between seed set and set of infected nodes. They consider the SI diffusion model and work on general graphs. It will not directly apply to non-progressive models. The novel approach for multiple source detection using set resolving set (SRS) is proposed by Zhang et al. [100]. The SRS is a subset of nodes with least cardinality. On the network graph G, V is a set of all nodes, then node set $K \subseteq V$ is called as SRS, if any detectable node sets $A, B \in V$ are distinguishable by K. The set A and B are differentiated by K if x, y $\in K$ such that

$$r_A(x) - r_A(y) \neq r_B(x) - r_B(y)$$

Wherein, $r_A(x)$ is the receiving time of rumor at node x from set A.

They devise a polynomial time greedy algorithm for discovering least SRS, so that the sources can be uniquely identified by the infected times of nodes in SRS set.

3.3. Comparison

Table 1 shows the comparative study of source detection of rumor in a social network. The legends used in Table 1 as SI, SIR, IC, etc. are information diffusion models explained in Section 2.2. From Table 1, it can be observed that the source detection approaches present variations based on the detection of the number of sources. This table suggests that a research for multiple source identification with heterogeneous infection probability and the SIR

Table 1
Comparative study of source detection of rumor in social network.

Ref.	No. of sources		Network topology		Infection probability		Diffusion models	
	Single	Multiple	Tree	Generic	HM	HT	Epidemic	IC
Nguyen et al. [42]		√		√		√		√
Wang et al. [50]	√		√		√		SI	
Zhu and Ying [49]	√		√			√	SIR	
Zang et al. [98]		√		√	√		SIR	
Xu and Chen [43]	√			√		√		√
Wang et al. [30]		√	√		√		SI	
Jiang et al. [95]		√		√		√	SI	
Zhu and Ying [58]	√		√		√		SIR	
Jain et al. [85]	√			√	√		SI	
Nguyen et al. [99]		√		√		√		√
Jiang et al. [52]	√			√		√	SIR	
Cai et al. [51]	√		√			√	SI	

Table 2
Comparative study of evaluation metrics (distance error and time complexity) (results taken from respective papers).

Ref. no.	Network observation	Source detection measures	Evaluation metric	Assumption	Real world dataset	Distance error				Time complexity
						0–1	1–2	2–3	3–4	
[43]	Monitor	Rumor Quantifier	ADE	ADE decreases as the monitor nodes increases	Twitter	–	–	√	–	–
[95]	Complete	Effective Distance	ADE	For Number of Real Sources = 2	Facebook [110]	–	–	–	√	–
[67]	Snapshot	Eigenvector	ADE	For Number of Real Sources = 2	Ego-Facebook [106]	–	–	√	–	–
[52]	Monitor	ML	DE	Data Sampling ≥ 20%	Facebook [110]	–	√	√	–	–
[85]	Complete	Random Walk	FZE, MEE, execution time	Complexity is based on Infection probability and the structure of the graph	Ego-Facebook [106]	–	–	–	–	$O(\|V_i\|\ (\|V\|+\|E\|))$
[99]	Snapshot	Symmetric Difference	F Measure, execution time	Complexity is based on Number of links from infected set to the external world	NetHEPT [111]	–	–	–	–	$O(m\Delta\Lambda/\|E_s\| + \Lambda^2)$
[80]	Monitor	ML	Accuracy, rank, DE	Focused on the small number of Monitors	Gnutella [106]	–	–	–	–	$O(N^2\ \log(N))$

diffusion model has not yet been explored. In a dynamic social network this combination of (Multiple Source – Heterogeneous Infection Probability – SIR diffusion model) may improve the accuracy of detection, as these terms look more realistic in the social network.

The comparative study of evaluation metrics with respect to time complexity and distance error, i.e. number of hops between actual and estimated source is shown in Table 2. The values stated in Table 2 are taken from results mentioned in the related research papers for the respective datasets and methodology. The legends used in distance error are ADE for Average Distance Error and DE for Distance Error which are compared based on the dataset, detection metrics like rumor quantifier, Maximum Likelihood (ML), effective distance, etc. are already described in respective citations and are compared to show the overall distance error in source detection approaches in social networks. The legends used in time complexity, V and E, stand for the number of vertices and edges in a graph or network. The other legends used in comparison of time complexity are MEE for Mean Error in Estimation i.e. average number of hops between the identified and actual source and FZE for Frequency of Zero Error i.e. the frequency of estimated source is identical to actual source. This table considers only respected metrics and real world dataset for the purpose of comparison.

Datasets are briefly described in Section 4. "√" indicates the presence of the corresponding feature in the corresponding paper. From the comparison shown in Table 2, it can be observed that the overall computation time is large and accuracy calculated in terms of distance error should be as minimum as possible, where it shows between 0 and 3 hops distance between actual and estimated source.

4. Datasets and experimental setup

Datasets used for source detection in social networks are classified as a synthetic dataset which includes tree and graph network and the real-world dataset from a social network.

4.1. Synthetic datasets

The Synthetic datasets are mainly structured in terms of tree and graph. The Tree networks are represented by Random d-regular tree. *Small-world (SW)* networks and scale free networks are basically used for graph network. The SW [47] networks are a kind of graphs in which majority of nodes can be traversed from one another in a few hops. The Scale free network (SF) [48] follows a power law distribution and is also referred as Barabasi Albert (BA) model which is used to create a random scale-free networks. The network starts with an underlying associated network of nodes. New nodes are added to the network one node at a time. *Erdos–Renyi (ER)* model [101] produces random graphs with subjective degree distributions. In this model, a network is constructed by associating nodes randomly. Each edge is incorporated in the graph with probability p which is independent of any other edge in the network. A python language software package for creation, manipulation, investigation of the formation, dynamics and functions of complex networks are called as NetworkX [102] and the same is used for synthetic data generation. R-Mat [103] is a recursive matrix based synthetic graph generator which is useful for identifying community structure in a graph.

Recently, researchers have designed a few synthetic graph generators such as Darwini [104] and Datasynth [105]. Darwini is an extensible synthetic graph generator which can utilize the

Table 3
Real world datasets used by source detection approaches in social networks.

Dataset used by papers	Dataset	Description	$\|V\|$	$\|E\|$	Diameter	Average clustering coefficient
[112]	Sina weibo [108]	Chinese micro blogging website	570	6574	–	–
[67,85,91]	Ego-Facebook [106]	This includes a circles or companions records from Facebook also referred as an ego-Facebook network. The information was gathered from members review utilizing Facebook application.	4039	88,234	8	0.6055
[80]	Gnutella [106]	Peer to Peer file sharing network	6301	20,777	9	0.0109
[60,67]	Wiki-vote [106]	This includes the data of who-votes-to-whom from the Wikipedia network.	7115	103,689	7	0.1409
[52]	Enron Email [113]	This includes the data of email discussions between 143 users in 2001.	36,692	183,831	11	0.497
[52,79,95]	Facebook [110]	This involves the communications between 45,813 users collected from 29th Dec 2008 to 3rd Jan 2009 from Facebook.	45,813	370,532	–	–
[42,67]	Epinion [105]	A who-trust-whom online interpersonal, organization of a common customer audit website Epinions.com. Individuals from the website can choose whether to "believe" each other and the edges speak about the interconnection between individuals.	75,879	508,837	14	0.1378
[51]	Higgs-Twitter [106]	A popular social network where individual can share information, pictures, audio, and video.	456,626	14,855,875	9	0.1887

Table 4
Experimental setups used by source detection approaches in social networks.

Ref.	Number of sources	Datasets	Platform	Implementation
[80]	Single	Gnutella	AMD FX-8350 4 GHz processor	Java 7
[92]	Single	Facebook Ego Network	–	MATLAB
[89]	Single	Twitter	Microsoft Windows with a 3.2 GHz quad-core Intel Core i7 and 16GB memory	–
[79]	Single	Facebook	Microsoft Windows Server 2008 with 8 CPUs, 32GB memory	C++ 2010 and MATLAB 2012
[52]	Single	Enron Email, Facebook	Microsoft Windows 7 with 2 CPUs, 4GB memory	C++, MATLAB 2012
[112]	Multiple	Sina Weibo	Ubuntu Server with 4 × 2.4 GHz CPU, 32GB memory	C++
[67]	Multiple	Facebook, Epinion, Wiki-Vote	Ubuntu 11.10 Server with 1400 M Hz six-core CPU, 32 GB memory	Python

metrics of real world social networks like degree, clustering coefficient. DataSynth is a framework for a synthetic graph generation with customized features like user features, graph features and correlations between structure and properties of graphs.

4.2. Real datasets

Most of the real datasets like Facebook, Twitter, Wiki-vote are freely accessible on Stanford Large Network Dataset Collection [106]. Power grid dataset (PG) [107] representing a power grid network of the Western States of the USA is also available. Similar to Twitter, there is a popular Chinese micro blogging network as Sina Weibo [108] used for rumor detection and source identification.

Table 3 shows the study of real world datasets which are publicly available and used by source detection approaches in the social network. The datasets in Table 3 are arranged in ascending order of number of nodes or users in the respective datasets. Table 3 summarizes information about the dataset, its description and details of network, which include number of vertices $|V|$, number of edges $|E|$, diameter of the graph (largest distance between any two pairs of node) and average clustering coefficient (a measure of the degree to which nodes in a graph can be clustered together). The Twitter network provides a search API to access the data. The TAGS [109] tool is used for collection of searched results on Twitter by Kumar and Geethakumari [97].

Table 4 gives the details of the experimental setup and programming languages used by source detection techniques in the social network and a number of sources (Single or Multiple) detected.

5. Research challenges and future scope

Identification of misinformation sources in a social network is very important to prevent and control the misinformation diffusion in the network. Research challenges that need to be addressed for efficient detection of sources in the social networks are identified as follows.

1. *Complex network structure*. Current research, majorly considered a tree like network for source detection, which does not consider cycles, whereas in the real world the information propagated by sources can be received by the same nodes, i.e., sources again from neighboring nodes. Though the source detection is quite easy in a tree like network using BFS strategy, a graph is best suitable topology to represent the real world network. In the category of network observation, snapshot and monitor based approaches provide partial information of the network. Therefore, utilizing the minimum information of network to give better accuracy can be investigated in the future.
2. *Real-time data collection*. Quick and real time detection of sources is useful to control the spread of rumor and reduce the harsh impact on society. The data sets considered for source detection must be collected in a real time. There is a scope for automatic detection of rumors and finding its sources for further investigation.
3. *Dynamic network evolution*. The social networks are dynamic in nature. Individual behavior can strongly affect the temporal dynamics of rumor propagation. Evolution of network must be taken into consideration while examining the network and selecting the diffusion model. Nowadays, researchers are taking network evolution into account for event detection [114] and community detection [115] in a dynamic social network. However, there is a scope for source detection by considering the evolutions in the network.
4. *Heterogeneous diffusion of information*. Many researchers assume that neighboring nodes get infected with the same probability, however, in reality the information sharing probability is different among the people and depends on their relationship. There-

fore, the heterogeneous infection probability should be considered to improve the accuracy of the source in the future.

5. *Number of propagation sources*. Many researchers assume that the misinformation or rumor originates from a single source. However, the purpose of source nodes is to spread the information as fast as possible, so they might often propagate the information from multiple sources. Currently the methodology of source detection depends on a number of sources to be identified and there is no uniform model for single and multiple source detection in the network. There is scope for wide research to design a single model to detect source/s irrespective of their number.
6. *Detection of communities involved in information diffusion*. While identifying multiple sources, few researches follow community detection methods, which is also a good topic for further study. The misinformation might spread by multiple sources that form their own community. Along with the source detection, the community can be identified.
7. *Source detection across interconnected social network*. Users are having their accounts on various social networking sites like Facebook, Twitter, etc. Sometimes they spread the rumor across their different social networks. A single interconnected network can be constructed by using the relationship of users in the different social networks. Therefore, to generate the linkable or interconnected social network and identify the true sources in a linkable social network can be a vast and complex area for further research.

6. Conclusion

The proliferation of data generated by a social network generates a number of real-world problems to be solved and rumor source identification is one of them. The source detection techniques are used in many domain specific networks like wireless sensor network, network of virus spread, epidemic network, etc. This paper aims to analyze and summarize the approaches for source detection of rumor and misinformation in social network and provides an intense research contribution for further exploration of source detection of rumor in a social network.

Based on the different factors impacting on source detection like centrality measure, network observation and diffusion models, it can be understood that the current methods of source detection in social network presents a large variation in accuracy. This research domain shows potential, hence, it will act as a foundation in the social aspect and will help to reduce the risk or damage caused by misinformation in a social network. As rumors play with the sentiments of the person and their spread introduces negative impact in the society, it is very vital to control such rumor diffusion using source detection.

Considering the source detection in a social network, the majority of the research focuses on a single source, whereas in reality information spreads from many sources. So there is much scope to improve source detection of rumor in social networks irrespective of number of sources. There is also an interesting direction towards considering the dynamic nature of a social network during source detection.

References

[1] M.E.J. Newman, The structure and function of complex networks, SIAM Rev. 45 (2003) 167–256, doi:10.1137/S003614450342480.

[2] B. Albert-Laszlo, Linked: how everything is connected to everything else and what it means for business, science, and everyday life, New York Plume Ed. (2003).

[3] S. Luna, M. Pennock, Social media in emergency management advances, challenges and future directions, in: Proceedings of the 2015 Ninth Annual IEEE International Systems Conference (SysCon), 2015, pp. 792–797, doi:10.1109/SYSCON.2015.7116847.

[4] I. Kotsiopoulos, Social media in crisis management: Role, potential, and risk, in: Proceedings of the 2014 IEEE/ACM Seventh International Conference on Utility and Cloud Computing, 2014, pp. 681–686, doi:10.1109/UCC.2014.110.

[5] Social Network Statistics, https://www.statista.com/topics/1164/social-networks/.

[6] I. Moya, M. Chica, J.L. Saez-Lozano, O. Cordon, An agent-based model for understanding the influence of the 11-M terrorist attacks on the 2004 Spanish elections, Knowled. Based Syst. 123 (2017) 200–216, doi:10.1016/j.knosys.2017.02.015.

[7] M. Sun, H. Zhang, H. Kang, G. Zhu, X. Fu, Epidemic spreading on adaptively weighted scale-free networks, J. Math. Biol. 74 (2017) 1263–1298, doi:10.1007/s00285-016-1057-6.

[8] F. Fu, N.A. Christakis, J.H. Fowler, Dueling biological and social contagions, Sci. Rep. 7 (2017) 43634, doi:10.1038/srep43634.

[9] Example of Latest Rumor -https://www.snopes.com/fact-check/600-murders-in-chicago/.

[10] H. Zhang, M.A. Alim, X. Li, M.T. Thai, H.T. Nguyen, Misinformation in online social networks: detect them all with a limited budget, ACM Trans. Inf. Syst. 34 (2016) 18.

[11] V. Qazvinian, E. Rosengren, D.R. Radev, Q. Mei, Rumor has it : identifying misinformation in microblogs, in: Proceedings of the Conference on Empirical Methods in Natural Language Processing, 2011, pp. 1589–1599.

[12] M. De Domenico, A. Lima, P. Mougel, M. Musolesi, The anatomy of a scientific rumor, Sci. Rep. 3 (2013) 2980.

[13] S. Vosoughi, D. Roy, S. Aral, The spread of true and false news online, Science 359 (2018) 1146–1151.

[14] J. Ratkiewicz, M. Conover, M. Meiss, B. Gonçalves, S. Patil, A. Flammini, F. Menczer, Truthy: Mapping the spread of astroturf in microblog streams, in: Proceedings of the Twentieth International Conference on Companion World Wide Web (WWW '11), 2011, pp. 249–252, doi:10.1145/1963192.1963301.

[15] C. Shao, G.L. Ciampaglia, O. Varol, A. Flammini, F. Menczer, The spread of fake news by social bots, ArXiv Prepr. Arxiv: 1707.07592. (2017).

[16] Z. Jin, J. Cao, H. Guo, Y. Zhang, Y. Wang, J. Luo, Detection and analysis of 2016 US presidential election related rumors on Twitter, in: Proceedings of the International Conference on Social Computing, Behavioral-Cultural Modeling & Prediction and Behavior Representation in Modeling and Simulation, 2017, pp. 14–24.

[17] C. Shao, P.-M. Hui, L. Wang, X. Jiang, A. Flammini, F. Menczer, G.L. Ciampaglia, Anatomy of an online misinformation network, PLoS One 13 (2018) e0196087.

[18] M. Miyabe, A. Nadamoto, E. Aramaki, How do rumors spread during a crisis? Analysis of rumor expansion and disaffirmation on Twitter after 3.11 in Japan, Int. J. Web Inf. Syst. 10 (2014) 394–412.

[19] T. Mondal, P. Pramanik, I. Bhattacharya, N. Boral, S. Ghosh, Analysis and early detection of rumors in a post disaster scenario, Inf. Syst. Front. (2018) 1–19, doi:10.1007/s10796-018-9837-8.

[20] K. Starbird, J. Maddock, M. Orand, P. Achterman, R.M. Mason, Rumors, false flags, and digital vigilantes: misinformation on twitter after the 2013 Boston Marathon Bombing, in: Proceedings of the iConference 2014, 2014, doi:10.9776/14308.

[21] K. Starbird, Examining the alternative media ecosystem through the production of alternative narratives of mass shooting events on Twitter., in: Proceedings of the International AAAI Conference on Web and Social Media, ICWSM, 2017, pp. 230–239.

[22] A. Friggeri, L. Adamic, D. Eckles, J. Cheng, Rumor Cascades, in: Proceedings of the International AAAI Conference on Web and Social Media, ICWSM, 2014, pp. 101–110. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM14/paper/download/8122/8110.

[23] O. Varol, I. Uluturk, Deception strategies and threats for online discussions, First Monday (2018) 22.

[24] E. Ferrara, O. Varol, C. Davis, F. Menczer, A. Flammini, The rise of social bots, Commun. ACM 59 (2016) 96–104.

[25] O. Varol, E. Ferrara, F. Menczer, A. Flammini, Early detection of promoting campaigns on social media, EPJ Data Sci. (2017) 6, doi:10.1140/epjds/s13688-017-0111-y.

[26] K.P.K. Kumar, G. Geethakumari, Detecting misinformation in online social networks using cognitive psychology, Human-Centric Comput. Inf. Sci. 4 (2014) 14.

[27] V.S. Subrahmanian, A. Azaria, S. Durst, V. Kagan, A. Galstyan, K. Lerman, L. Zhu, E. Ferrara, A. Flammini, F. Menczer, The DARPA Twitter bot challenge, Comput. 49 (2016) 38–46, doi:10.1109/MC.2016.183.

[28] Z. Zhao, P. Resnick, Q. Mei, Enquiring minds: Early detection of rumors in social media from enquiry posts, in: Proceedings of the Twenty-forth International Conference World Wide Web, 2015, pp. 1395–1405.

[29] S. Vosoughi, N. Mohsenvand, D.K. Roy, Rumor Gauge : Predicting the Veracity of Rumors on Twitter, 11, ACM Trans. Knowl. Discov. from Data (TKDD). Press., 2017, pp. 1–38, doi:10.1145/3070644.

[30] Z. Wang, W. Dong, W. Zhang, C.W. Tan, Rooting our rumor sources in online social networks: the value of diversity from multiple observations, IEEE J. Sel. Top. Signal Process. 9 (2015) 663–677.

[31] P.T. Metaxas, S. Finn, E. Mustafaraj, Using twittertrails. com to investigate rumor propagation, in: Proceedings of the Eigteenth ACM Conference Companion Computer-Supported Cooperative Work and Social Computing, 2015, pp. 69–72.

[32] A. Gupta, P. Kumaraguru, C. Castillo, P. Meier, Tweetcred: real-time credibility assessment of content on twitter, in: Proceedings of the International Conference on Social Informatics, 2014, pp. 228–243.

[33] C. Shao, G.L. Ciampaglia, A. Flammini, F. Menczer, Hoaxy: a platform for tracking online misinformation, in: Proceedings of the Twenty-fifth International Conference Companion World Wide Web, 2016, pp. 745–750.
[34] Real Time Rumor Tracker http://www.emergent.info/.
[35] Snopes.com https://www.snopes.com/.
[36] factcheck.org https://www.factcheck.org/.
[37] N. Antulov-Fantulin, A. Lancic, T. Smuc, H. Stefancic, M. Sikic, Identification of patient zero in static and temporal networks: robustness and limitations, Phys. Rev. Lett. 114 (2015) 248701.
[38] D. Shah, T. Zaman, Detecting sources of computer viruses in networks: theory and experiment, in: ACM SIGMETRICS Perform, Eval. Rev. (2010) 203–214.
[39] L. Shu, M. Mukherjee, X. Xu, K. Wang, X. Wu, A survey on gas leakage source detection and boundary tracking with wireless sensor networks, IEEE Access 4 (2016) 1700–1715.
[40] C.H. Comin, L. da Fontoura Costa, Identifying the starting point of a spreading process in complex networks, Phys. Rev. E. 84 (2011) 56105.
[41] J. Jiang, S. Wen, S. Yu, Y. Xiang, W. Zhou, Identifying propagation sources in networks: State-of-the-art and comparative studies, IEEE Commun. Surv. Tutorials. 19 (2017) 465–481.
[42] D.T. Nguyen, N.P. Nguyen, M.T. Thai, Sources of misinformation in Online Social Networks: Who to suspect? in: Proceedings of the IEEE Military Communications Conference MILCOM, 2012, doi:10.1109/MILCOM.2012.6415780.
[43] W. Xu, H. Chen, Scalable rumor source detection under independent cascade model in online social networks, in: Proceedings of the 2015 Eleventh International Conference on Mobile Ad-hoc and Sensor Networks (MSN), 2015, pp. 236–242.
[44] D. Shah, T. Zaman, Rumors in a network: Who's the culprit? IEEE Trans. Inf. Theory 57 (2011) 5163–5181, doi:10.1109/TIT.2011.2158885.
[45] D. Shah, T. Zaman, Finding rumor sources on random trees, Oper. Res. 64 (2016) 736–755. ttps://doi.org/10.1287/opre.2015.1455.
[46] M. Fuchs, P.D. Yu, Rumor source detection for rumor spreading on random increasing trees, Electron. Commun. Probab. 20 (2015) 1–12, doi:10.1214/ECP.v20-3743.
[47] D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature 393 (1998) 440.
[48] A.-L. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[49] K. Zhu, L. Ying, A robust information source estimator with sparse observations, Comput. Soc. Netw. 1 (2014) 3.
[50] Z. Wang, W. Dong, W. Zhang, C.W. Tan, Rumor source detection with multiple observations: fundamental limits and algorithms, ACM SIGMETRICS Perform. Eval. Rev. (2014) 1–13.
[51] K. Cai, H. Xie, J. Lui, Information spreading forensics via sequential dependent snapshots, IEEE/ACM Trans. Netw. 26 (2018) 478–491.
[52] J. Jiang, W.E.N. Sheng, S. Yu, Y. Xiang, W. Zhou, Rumor source identification in social networks with time-varying topology, IEEE Trans. Dependable Secur. Comput. (2016).
[53] A. Guille, H. Hacid, C. Favre, D.A. Zighed, Information diffusion in online social networks: a survey, ACM Sigmod Rec. 42 (2013) 17–28, doi:10.1145/2503792.2503797.
[54] M. Li, X. Wang, K. Gao, S. Zhang, A survey on information diffusion in online social networks: models and methods, Information 8 (2017) 118.
[55] W. Luo, W.P. Tay, M. Leng, How to identify an infection source with limited observations, IEEE J. Sel. Top. Signal Process. 8 (2014) 586–597.
[56] Z. Wang, W. Zhang, C.W. Tan, On inferring rumor source for SIS model under multiple observations, in: Proceedings of the 2015 IEEE International Conference on Digital Signal Processing (DSP), 2015, pp. 755–759.
[57] K. Saito, M. Kimura, H. Motoda, Discovering influential nodes for SIS models in social networks, in: Proceedings of the International Conference on Discovery Science, 2009, pp. 302–316.
[58] K. Zhu, L. Ying, Information source detection in the SIR model: A sample–path-based approach, IEEE/ACM Trans. Netw. 24 (2016) 408–421.
[59] J. Woo, H. Chen, Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog, Springerplus 5 (2016) 66.
[60] L.P. Song, Z. Jin, G.Q. Sun, Modeling and analyzing of botnet interactions, Phys. A Stat. Mech. Appl. 390 (2011) 347–358.
[61] G. D'Angelo, L. Severini, Y. Velaj, Influence maximization in the independent cascade model, in: Proceedings of the International Conference on Teaching and Computational Science (ICTCS), 2016, pp. 269–274.
[62] A. Kumari, S.N. Singh, Online influence maximization using rapid continuous time independent cascade model, in: Proceedings of the 2017 Seventh International Conference Cloud Cloud Computing, Data Science & Engineering, 2017, pp. 356–361.
[63] J. Cheng, L. Adamic, P.A. Dow, J.M. Kleinberg, J. Leskovec, Can cascades be predicted? in: Proceedings of the Twenty-third International Conference on World Wide Web, 2014, pp. 925–936.
[64] X. Yu, T. Chu, Learning the structure of influence diffusion in the independent cascade model, in: Proceedings of the 2017 Thirty-sixth Chinese Control Conference (CCC), 2017, pp. 5647–5651.
[65] D. Kempe, J. Kleinberg, É. Tardos, Maximizing the spread of influence through a social network, in: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2003, pp. 137–146.
[66] F. Gursoy, D. Gunnec, Influence maximization in social networks under deterministic linear threshold model, Knowled. Based Syst. 161 (2018) 111–123.
[67] W. Zang, P. Zhang, C. Zhou, L. Guo, Locating multiple sources in social networks under the sir model: a divide-and-conquer approach, J. Comput. Sci. 10 (2015) 278–287.
[68] D.J. Klein, Centrality measure in graphs, J. Math. Chem. 47 (2010) 1209–1223, doi:10.1007/s10910-009-9635-0.
[69] F. Boudin, A comparison of centrality measures for graph-based keyphrase extraction, in: Proceedings of the International Joint Confrence on Natural Language Processing, 2013, pp. 834–838.
[70] R. Albert, H. Jeong, A.-L. Barabasi, Error and attack tolerance of complex networks, Nature 406 (2000) 378.
[71] L.C. Freeman, Centrality in social networks conceptual clarification, Soc. Netw. 1 (1978) 215–239.
[72] P. Holme, B.J. Kim, C.N. Yoon, S.K. Han, Attack vulnerability of complex networks, Phys. Rev. E 65 (2002) 56109.
[73] Y.Y. Liu, J.J. Slotine, A.L. Barabasi, Controllability of complex networks, Nature 473 (2011) 167–173.
[74] A. Louni, K.P. Subbalakshmi, A two-stage algorithm to estimate the source of information diffusion in social media networks., in: Proceedings of the INFOCOM Work, 2014, pp. 329–333.
[75] A.H. Dekker, et al., Centrality in social networks: theoretical and simulation approaches, Proc. SimTecT 2008 (2008) 12–15.
[76] K. Miura, D. Takahashi, S. Nakano, T. Nishizeki, A linear-time algorithm to find four independent spanning trees in four-connected planar graphs, in: Proceedings of the Twenty-second International Workshop on Graph-Theoretic Concepts in Computer Science, 1998, pp. 310–323.
[77] P. Bonacich, Power and centrality: a family of measures, Am. J. Sociol. 92 (1987) 1170–1182.
[78] M.E.J. Newman, The mathematics of networks, New Palgrave Encycl. Econ. 2 (2008) 1–12.
[79] D. Wang, S. Wen, Y. Xiang, W. Zhou, J. Zhang, S. Nepal, Catch Me If You Can: detecting compromised users through partial observation on networks, in: Proceedings of the IEEE 2017 Thirty-seventh International Conference on Distributed Computing Systems (ICDCS), 2017, pp. 2417–2422.
[80] R. Paluch, X. Lu, K. Suchecki, B.K. Szymanski, J.A. Hołyst, Fast and accurate detection of spread source in large complex networks, Sci. Rep. 8 (2018) 2508, doi:10.1038/s41598-018-20546-3.
[81] W. Luo, W.P. Tay, M. Leng, Identifying infection sources and regions in large networks, IEEE Trans. Signal Process 61 (2013) 2850–2865.
[82] D. Shah, T. Zaman, Rumor centrality: a universal source detector, in: Proceedings of the ACM SIGMETRICS Performance Evaluation Review, 2012, pp. 199–210.
[83] W. Dong, W. Zhang, C.W. Tan, Rooting out the rumor culprit from suspects, in: Proceedings of the IEEE International Symposium on Information Theory, 2013, pp. 2671–2675, doi:10.1109/ISIT.2013.6620711.
[84] P. Yu, C.W. Tan, H. Fu, Rumor source detection in finite graphs with boundary effects by message-passing algorithms, (2017) 86–90.
[85] A. Jain, V. Borkar, D. Garg, Fast rumor source identification via random walks, Soc. Netw. Anal. Min. 6 (2016) 62.
[86] N. Karamchandani, M. Franceschetti, Rumor source detection under probabilistic sampling, in: Proceedings of the 2013 IEEE International Symposium on Information Theory (ISIT), 2013, pp. 2184–2188.
[87] W. Luo, W.P. Tay, Finding an infection source under the SIS model, in: Proceedings of the 2013 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), 2013, pp. 2930–2934.
[88] V.P. Sahana, A.R. Pias, R. Shastri, S. Mandloi, Automatic detection of rumoured tweets and finding its origin, in: Proceedings of the 2015 International Conference on Computing and Network Communications (CoCoNet), 2015, pp. 607–612.
[89] D. Krol, K. Wisniewska, On rumor source detection and its experimental verification on Twitter, in: Proc. Asian Conf. Intell. Inf. Database Syst. (ACIIDS), 2017, pp. 110–119, doi:10.1007/978-3-319-54472-4_11.
[90] P.C. Pinto, P. Thiran, M. Vetterli, Locating the source of diffusion in large-scale networks, Phys. Rev. Lett. 109 (2012) 68702.
[91] J. Choi, S. Moon, J. Woo, K. Son, J. Shin, Y. Yi, Rumor source detection under querying with untruthful answers, in: Proceedings of the INFOCOM 2017 IEEE Conference on Computer Communications, IEEE, 2017, pp. 1–9.
[92] J. Choi, S. Moon, J. Shin, Y. Yi, Estimating the rumor source with anti-rumor in social networks, in: Proceedings of the 2016 IEEE Twenty-fourth International Conference on Network Protocols (ICNP), 2016, pp. 1–6.
[93] S.L. Hakimi, M.L. Labbe, E. Schmeichel, The Voronoi partition of a network and its implications in location theory, ORSA J. Comput. 4 (1992) 412–417.
[94] D. Brockmann, D. Helbing, The hidden geometry of complex, network-driven contagion phenomena, Science 342 (80) (2013) 1337–1342.
[95] J. Jiang, S. Wen, S. Yu, Y. Xiang, W. Zhou, K-center: An approach on the multi–source identification of information diffusion, IEEE Trans. Inf. Forensics Secur. 10 (2015) 2616–2626.
[96] K. Yang, A.H. Shekhar, D. Oliver, S. Shekhar, Capacity-constrained network-voronoi diagram: a summary of results, in: Proceedings of the International Symposium on Spatial and Temporal Databases, 2013, pp. 56–73.
[97] K.P.K. Kumar, G. Geethakumari, Identifying sources of misinformation in online social networks, Advances in Signal Processing and Intelligent Recognition Systems, Springer, 2014, pp. 417–428.
[98] W. Zang, P. Zhang, C. Zhou, L. Guo, Discovering multiple diffusion source nodes in social networks, Procedia Comput. Sci. 29 (2014) 443–452.
[99] H.T. Nguyen, P. Ghosh, M.L. Mayo, T.N. Dinh, Multiple infection sources identification with provable guarantees, in: Proceedings of the Twenty-fifth ACM

International Conference on Information & Knowledge Management, 2016, pp. 1663–1672.
[100] Z. Zhang, W. Xu, W. Wu, D.-Z. Du, A novel approach for detecting multiple rumor sources in networks with partial observations, J. Comb. Optim. 33 (2017) 132–146.
[101] P. Erdös, A. Rényi, On random graphs I, Publ. Math. 6 (1959) 290–297.
[102] NetworkX software for synthetic network- https://networkx.github.io/.
[103] D. Chakrabarti, Y. Zhan, C. Faloutsos, R-MAT: A recursive model for graph mining, in: Proceedings of the 2004 SIAM International Conference on Data Mining, 2004, pp. 442–446.
[104] S. Edunov, D. Logothetis, C. Wang, A. Ching, M. Kabiljo, Darwini: Generating realistic large-scale social graphs, CoRR, [Online]. Available: http://arxiv.org/abs/1610.00664. (2016).
[105] A. Prat-Pérez, J. Guisado-Gámez, X.F. Salas, P. Koupy, S. Depner, D.B. Bartolini, Towards a property graph generator for benchmarking, in: Proceedings of the Fifth International Workshop on Graph Data-management Experiences & Systems, 2017, p. 6.
[106] Stanford Large Network dataset- http://snap.stanford.edu/data/.
[107] Power grid datasets- http://www-personal.umich.edu/~mejn/netdata/.
[108] Sina Weibo - http://weibo.com/.
[109] TAGS (Twitter Archiving Google Sheet)- https://tags.hawksey.info/.
[110] B. Viswanath, A. Mislove, M. Cha, K.P. Gummadi, On the evolution of user interaction in Facebook, in: Proceedings of the Second ACM Workshop on Online Social Networks – WOSN'09, 37, 2009, doi:10.1145/1592665.1592675.
[111] B.A. Prakash, J. Vreeken, C. Faloutsos, Spotting culprits in epidemics: how many and which ones? in: Proceedings of the 2012 IEEE Twelfth International Conference on Data Mining (ICDM), 2012, pp. 11–20.
[112] P. Zhang, J. He, G. Long, G. Huang, C. Zhang, Towards anomalous diffusion sources detection in a large network, ACM Trans. Internet Technol 16 (2016) 2.
[113] J. Shetty, J. Adibi, The Enron email dataset database schema and brief statistical report, Information Sciences Institute Technical Report, University of Southern California, 4 (2004) 120–128.
[114] H. Wang, W. Hu, Z. Qiu, B. Wu, An event detection method for social networks based on evolution fluctuations of nodes, IEEE Access 6 (2018) 12351–12359.
[115] T. Puranik, L. Narayanan, Community detection in evolving networks, in: Proceedings of the 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2017, ACM, New York, NY, USA, 2017, pp. 385–390, doi:10.1145/3110025.3110067.

Sushila Shelke is currently a Ph.D. research scholar in the Department of Computer Engineering and IT at College of Engineering, Pune, India. She received her B.E. and M.E. in Computer Engineering from Pune University, India. She is the Assistant Professor at Cummins College of Engineering, Pune, India. Her research interests are Social Network Analysis, Machine Learning and Graph Mining. She has 8 years of teaching experience, 2 years of industry and 1.5 years of research experience.

Dr. Vahida Z. Attar is the Associate Professor at the Department of Computer Engineering and IT at College of Engineering Pune, India. She received her Ph.D. degree in Computer Engineering with specialization in Data Mining from Pune University, India in 2012. She has more than 20 years of teaching experience and 7 years of industry and research experience. More than 50 national and international research paper publications are to her credit. Her research interests are Data Mining, Machine Learning and Social Network Analysis. She is a Member of IEEE and Life Member of IE, ISTE, and CSI. She also worked as Principal Investigator for projects funded by All India Council for Technical Education (AICTE), Board of Research in Nuclear Sciences (BRNS) India.