Governance in Smart City: An Approach Based on Social Network

Paramita Dey and Sarbani Roy

Abstract The governance for smart cities will be more citizen-centric and government policies will be based on the demand of the citizen. Social network has the potential of elevating the governance process to new levels. It enables government for instantaneous transmission of information to the targeted citizen, processing large scale data available through social media and can enable to take decisions based on that data in judicial way to increase transparency and accountability. For smart governance, it is required that within a limited budget government can propagate information to the maximum people. Social media analysis provides knowledge to ensure maximize the influence. Similarly for misinformation/rumour, influence can be minimized and SNA helps to identify communities and locations affected by this information. Moreover it can be used to identify those influential users who are able to spread information. This knowledge will empower Government agencies to take necessary precautionary measures such as targeted campaigning against the rumour, identified the source of information, block the source node or delayed rumour propagation. In this chapter we will study an conceptual framework for generating a governance system which will be able to identify the citizens those have maximum influence in the social network, campaigning of government policies through them for the maximum propagation to the citizen with minimum budget, detect rumour, found the source of misinformation, block the source node of rumour, detection of community, generating anti-spreading model for already propagated rumour.

Keywords Smart city · Online social network · Centrality measures · Influence maximization · Influence minimization · Community detection

P. Dey (✉) Government College of Engineering & Ceramic Technology, Kolkata, India e-mail: paramitadey@gcect.ac.in

S. Roy Jadavpur University, Kolkata, India e-mail: sarbani.roy@jadavpuruniversity.in

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021 M. Khan et al. (eds.), Smart Cities: A Data Analytics Perspective, Lecture Notes in Intelligent Transportation and Infrastructure, https://doi.org/10.1007/978-3-030-60922-1_4
1 Introduction

In today’s scenario, due to easy access to the user and faster reachability, social media is widely used for generating opinion in society. In recent past, these data play a pivotal role in Brexit and U.S. presidential elections for same reasons. Even in India, for prime minister Election, social media plays an important role. Different organizations use this huge user base for different purposes like advertising, campaigning, viral marketing etc. Smart city administration can use social network information for knowledge about public opinion on laws and policies and based on the feedback counter measures can be taken. In time of natural calamity like earthquakes, tsunamis can collect information in real time and planned appropriate strategy to handle the situation. Mainstream media, government and private agencies and other organizations can use the online social network data for information feed. They spend their money for advertising their policy/product in social media. Now an advertising through a famous person like footballer Leonel Messi in social media with huge fan-base can able to create a large promotion instantaneously but it will not be cost effective in terms of money. As social media is stored in the form of a graph, it is advantageous for some users due to their hierarchical positions to propagate the data in a more efficient manner. Moreover community detection over some criterion can identify groups of people who are more interested in that particular topic. Therefore a strategy to influence maximum number of people with minimum cost is important. A topic aware circulation of data can be more effective to reach intended users of the social network. Rumour spreading in social network is an important issue in social media. As there is no existing censorship for generating information in social network, it is important to resist misinformation in smart city framework which may cause harm to the society. For instance, in recent scenario of covid-19 pandemic, different rumours are circulating in social media like in hot and humid weather the virus cannot persist, which was later proven false. It is important to track the flow of rumour and can take measurements to resist the propagation and find the source of rumour to make a control over it. The main constraints of social network data are the large volume and transient nature of information. Therefore it is required to generate the infrastructure which can use the paradigm like distributed computing, big data analytics and cloud computing.

Rest sections are organized in following manner. In Sect. 2 literature review and some critical issues about this topic are presented. An overview of system hierarchy for governance in smart city is discussed with the help of a flow diagram in Sect. 3. Data collection procedure is elaborated in Sect. 4. In Sect. 5, sentiment analysis is briefly discussed. In Sect. 6, information propagation techniques to reach maximum number of citizens are explained. Rumour detection and control strategies are discussed in Sect. 7. Conclusion and future research scopes are derived in Sect. 8.
2 Literature Review and Issues

2.1 Literature Review

The smart city is an idea or concept where citizen would be facilitated with all infrastructural amenities. The governance for smart cities must improve democratic processes in terms of transparency and citizen feedback centric governance. Development of the cities and political strategies will be based on the view of citizen. For this purpose, the government needs to collect information and feedback from the citizen.

Lots of research works have been performed to generate a framework for smart governance [1, 2]. Though a large number of infrastructure facilities are available in smart cities (already generated as pilot projects in different cities), there is no effective system in smart city infrastructure where government can interact with citizen in an interactive manner.

Online Social media platform empowers government to use common platform to share views with citizens of smart city and can reach to maximum number of peoples with minimum expenditure. Due to voluminous data of social network, it is trivial to collect and process online social network data. This chapter proposes a framework for smart governance through social media platform and uses some popular methodologies in social network analysis like influence maximization [3, 4], centrality measurements [5, 6], community detection [7, 8], Rumour detection [9] and control strategies [10]. For smart city administration framework, all existing approaches are generally one directional, i.e. people can interact with government, but in proposed framework interaction is bidirectional and government would proactively participate in social media. Moreover it can reach to the maximum number of peoples with their policies in minimal budget of advertising.

2.2 Issues

For generating an infrastructure for smart governance, it is important to concentrate on some issues which would be taken care of.

- Social network data are mostly unstructured data. Data collected using different crawlers are generally in the form of CSV (comma separated values). Therefore data has to store in NSQL (Not only SQL) database like Mongo DB.

- Generating a usage pattern from past history of the users is an important aspect. All users in the network are not participating in the communication in same way. Users, who is represented by the node of social network graph show different nature of participation like active (generating or sharing lots of post) or stifler (does not participate in communication), ignorant (the information cannot reach to the user due to poor connectivity of the node in network). Moreover this nature can be topic
specific. For example, who is very interested in politics cannot be interested in entertainment related news. Therefore an active node in one situation can be a stiffler node in other situation.

As mentioned, social network can be used as a tool for controlling the influence of the information like maximization or minimization. Now it is not possible to choose a person with lots of follower to use for that purpose. Again it is not always effective as it generally passes through one or two hop from the source node. Finding influential nodes or links using graph properties can be more effective for this purpose. This nodes and links are generally denoted as centrality measures and it is important to identify them.

Another issue of the analysis is large volume of social network data. Most of the computation complexity in social network is in the power of nodes count in the social graph. Therefore it is required to generate tools and algorithms for large scale data analytics. In the incubation period of social network analysis graph sampling was widely used to solve the problem but this methodology suffers with valuable information loss. Now with the paradigm like big data analytics, cloud computing it is become easier to process large scale data with large computation, time and storage complexity. Figure 1 describes different methodologies for large scale data analysis of online social network.

3 System Hierarchies for Smart Governance

An overview of system hierarchy is presented through the flow diagram as Fig. 2. Online social network data will be extracted from social network through network crawling programmes. For generating feedback, identification of keywords plays an important role. To reach maximum number of people, influence maximization methodology can be used. Similarly, for misinformation techniques for influence minimization is important. Government policies can be modified depending on positive or negative feedback of the citizen of smart cities.

4 Collection of Data for Reviewing Government Policies

Government policies implemented for the people can have positive, negative or neutral effect. We can collect temporal and spatial online social network data from Twitter, Facebook about the governance of smart city and can find the impact of government policies. Online social networks generate a platform for human to share information at an unprecedented large scale. It provides access to vast amount of information which is of significant value to government agencies. Online social network data is characterized with 4 V: volume, variety, veracity and velocity. Volume denotes the large scale data, variety define different kind of data like text, audio, video etc., veracity indicates noise in data and velocity indicates transient nature of data.
Mining and analyzing the characteristics of information in online social networks provides valuable insights for decision making. Some positive feedback of government policy can be considered as incentive to work in that direction whereas negative feedback may cause reconsideration of government policies. Information spreading through the posts of users in the social media may cause harmful impact in the society also. It is important for news agencies and government organizations to verify the impact of in-formation and create necessary measures.
Fig. 2  SNA based system hierarchy for governance in smart city

4.1 Identification of Keywords

Social network data contains features such as textual, temporal, spatial and network information. Keyword detection, which are spread across the network and become popular, is of considerable interest to government agencies, for e.g. trending topics in Facebook, WhatsApp or Twitter data. These topics can represent a piece of information, meme or emoticon in a transient period of time. Generally temporal behaviour of trending topics in online social media show a burst nature of characteristics in short period of time. For example, when there was occurrence of some important sports event like world cup, it becomes more trending keyword.
Online social network (OSN) is a source of voluminous data with a lot of inbuilt meta-information such as time, location etc. in addition to the textual content. For instance, Twitter provides information on the user, time and location of the Tweet. Typically, topic models consider only textual information and to some extent the temporal information. But for a smart city application, topic models will consider spatial information related to that city in addition to the temporal and textual information. It is important to distinguish location and time at which an event happened along with the textual content. For e.g. the posts of the people celebrating Christmas across the world have similar textual content but considering temporal and spatial information will help one to distinguish events happening in that particular region.

Traditional topic detection techniques used in network data may not work for social network data streams where temporal data show the nature of topic shifting and bursty data traffic. In [11, 12], the handling of bursty topic and tweet along with background topic modelling are considered respectively. For online infrastructure this topic detection modelling is rectified [13].

Detection of back topic or subtopic of a trending topic is far more difficult as it contain similar types of keywords. In [14, 15], A sub-topic detection methodology is presented where a graph of sequence of words is considered. All these algorithms are applicable if both topic and subtopics are equally trending. But for low tweeted sub-topic detection this algorithms cannot work efficiently. Hierarchical clustering algorithms can be used for social network topic detection related to their subtopics. As mentioned, online social network data are temporal in nature and for time dependent topic detection there exist topic detection model [16], where this temporal nature is taken care of. One disadvantage of this model is considering discrete time data whereas in online social network, continuous time data is generated. Therefore topic extraction along with subtopic detection requires more modelling.

The steps for trending topic detection from online social net-work can be described in following steps-

4.1.1 Data Extraction

OSN data are extracted through crawling of the social graph and derived data are formalized with their relative frequencies over a span of time. For example, Twitter data can be crawled over a period of time can be extracted in the form of metadata generally represented in CSV format.

4.1.2 Formation of Knowledge Network

Network can be generated through the active users based on their social relationships. Active users would be considered as the node of the network, similar kind of contents the relationship among the users of the network and relative frequency will define the edge of the network and thus knowledge network will be formed.
4.1.3 Temporal Features of Data

For each term, life cycle of the topic will be studied according to the relevance of that topic for specified time interval. After a certain time interval the features will be old and lost the significance. A set of emerging keywords would be generated based on their life span.

4.1.4 Spatial Features of Data

Generally online social network data provides us with spatial data. For example, in Twitter, it provides us with the latitude and longitude coordinate. For a smart city application of spatial features extraction will be twofold:

- From the citizen collecting information about the impacts of the government policies.
- The reaction from outside word.

Though reactions from the citizen have direct impact in the implementation of the policy but the outside reaction also should be counted.

4.1.5 Subtopic Detection

A topic graph that links the extracted emerging topics with their subtopics and background topic is required in order to generate a perspective about trending topic. Generally hierarchical clustering is a very good method for subtopic detection.

4.2 Feedback Collection

Data can be extracted from social network by crawling into network. Various softwares are available for that purpose. Data are generally extracted through two procedures:

- Using Keyword: Using identified keywords as hashtag or searching in the content, network information can be extracted from the network using the crawler like Tweepy. For example, from Twitter network data can be extracted through I-graph using # as shown in Fig. 3.
- Using user account: Networks are extracted based on some user accounts. For example, Facebook data can be extracted through Netvizz in the form of csv file [17] as shown in Fig. 4.
Fig. 3 Data extracted through hashtag(#) Scatter network B. Ego network

Fig. 4 Data extracted through user profile [18]
4.3 Classification of Information

The information extracted from online social network can be broadly categorized into five taxonomies.

4.3.1 Message-Based/content Based Features

This feature considers characteristics of the information content of the online post. Content can be used to find the following pattern from the online post:

- **Lexical patterns and part-of-speech patterns**: Generally it provides some information about the origin of the post. This feature varies with place, profession, religion, country etc.
- **Multimedia data**: used to check whether the post contains pictures, videos, or audios.
- **URL**: This is used to check whether the post contains URL to an external source to support the content of the post.
- **Time interval**: The time difference between the time of posting and user registration.
- **Sentiment**: It analyzes the emotions of the content. Sentiment can be positive, negative or neutral.

4.3.2 User-Based Features

These features analyze the user characteristics whoever is generating the post or sharing it. These features can include

- **Registration age**: As some post is more relevant to particular age group, it can be important features for classification.
- **Number of followers**: The user who have more followers in the online network have greater influence in the network.
- **Number of friends**: Number of friends connected to the user. Number of user posts: Number of post signifies whether user is active or not in the social media.
- **Number of Retweet**: Whether the post is shared by the followers of the user several times (measuring influence of the post).
- **Number of sharing**: Measuring the count of shared information in the network (may not be directly connected with the user).
- **Number of comments**: Measuring the count of comments on the post, may be positive, negative or neutral.
- **User profile**: It includes gender, personal information like organization, real name etc.
• **Spatial information**: Spatial information signifies the location where the event mentioned by posts occurred.

• **Account credibility**: Whether the identity of the user is verified by the social network site.

4.3.3 Topic Based Features

Topic based features includes the features of the most discussed topics in the social network, and consider all content based and user based features for the analytics.

4.3.4 Propagation Based Features

Propagation based features consider network analysis if the propagated information. As the information is shares by several users, it generates several clusters within the shared network.

4.3.5 Spatial Features

Spatial features of the user can generate knowledge based community structures in the social network. It is obvious that impact of one incident can cause different impact in different places. So based on spatial features information can be classified.

5 Sentiment Analysis

Sentiment analysis is an important tool used for the different applications of extracted information. This analysis is generally generated using natural language processing and measures similarity of sentiment about a particular keyword. Sentiment would be positive when citizen are agreed with the government policies, neutral sentiment would be reflected by the zero value and negative value reflects disagreement. In Fig. 5, a sentiment analysis about GST (a tax system imposed in India in recent time) is shown as a scatter diagram. If the sentiment is negative about a policy, it is required to find whether there is some misinformation about this topic and take necessary actions about that. But if it is not a rumour, good governance can rethink or reconsider that policy.
For influence maximization influential nodes can be used as the source node or propagator of the propagation. In online social network influential nodes can be identified through centrality measures.

### 6.1 Nodes Classification

Nodes, which signify the user in a network, can be classified according to their nature of participation in the network.

#### 6.1.1 Daley-Kendal Model

According to Daley-Kendal (DK) model online social network users can be grouped into three categories:

- **Ignorant**: the user who does not know about that particular information.
- **Spreaders**: the user who know about the information.
- **Stifler**: the users who know about the information but do not want to spread the information in the network.
6.1.2 Behavioural Model

According to the behaviour of the user present in online social network, nodes can be classified into two categories:

- **Active**: The user who interacts with other users more often.
- **Dormant**: The user with minimal interaction.

6.1.3 Flow Model

According to the flow of the information of the user present in online social network, nodes can be classified into three categories:

- **Source**: the source node denotes the node who originates the data for particular flow.
- **Sink**: The end node of the flow is indicated by sink node. This criterion can occur in two situations:
  - The node is stifler in nature.
  - The information lost its significance along with time.
- **Participatory**: All other nodes besides source and sink node, participated in that information flow are considered as participatory node.

6.1.4 Case Study

A case study on classification of Twitter data are described in this subsection [19]. Information is extracted as the set of all Tweets containing a particular #hashtag.

Using #hashtag, a data filtering mechanism is used with algorithmic efficiency and statistical robustness at the processing of large metadata available in Twitter.

We define a directed Twitter User Graph. Its vertex set consists of every user on Twitter. Its edge set consists of all Following links: directed from every user, the followees, to each of their followers.

We distinguish users as Participants based on their role in rumour propagation through the network. We designate any user tweeting with a given #hashtag to be an active Participant in the rumour. Passive Participants include those not tweeting that particular #hashtag. To study the propagation of rumours on Twitter data, we define a directed Tweet Rumour Graph. Its vertex set corresponds to that of the Twitter User Graph. Stiflers and ignorant participants are designated as inactive nodes, while sources and propagators are designated active nodes, as shown in Fig. 6. We further label all active nodes with the timestamp of their 1st Tweet containing the specified #hashtag. The edge set denotes when both nodes are active. We get source node as user that has a lower timestamp than the other nodes in the in a connected component of the Tweet network. Thus, every connected component has its own source node in the Twitter graph.
6.2 Centrality Measures

For derivation of influential nodes in online social network, different centrality measures are widely used [20]. In a graph, centrality indicates the nodes that are mostly connected to other nodes in the network. Thus centrality measures in social graph indicate the most influential persons in the network.

6.2.1 Node Degree Centrality

Degree of a node \( v \in V \) can be defined as degree(v) = \(|\{u: (v; u) \in E\}|\). This centrality measure [5] informs us of the number of connectivity among the vertices of the social graph. Nodes with a greater degree are connected to a large population within the graph, and are hence act as good seed candidates for information propagation.

6.2.2 Betweenness Centrality

Betweenness centrality of a node \( v \in V \) is defined by between(v) = \( \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \) where \( \sigma_{st} \) represents all possible shortest paths from node s to node t and \( \sigma_{st}(v) \) denotes the number of shortest paths pass through node v. This centrality measure [18] indicates the capability of a node to quickly transfer information through the network.
6.2.3 Closeness Centrality

The closeness of a node \( v \in V \) can be defined as \( \text{close}(v) = \frac{1}{\sum_{n \in V} \delta(n,v)} \), where \( \delta(n,v) \) is the shortest distance from \( n \) to \( v \). The closeness centrality [18] indicates how quickly information can spread from the node to the rest of the network.

We note here that if any node is \( v \) is unreachable from the node \( n \), then \( \text{close}(v) = 0 \). This is a well-known problem with closeness centrality. As experts suggest, we thus redefine our implementations to use the mean harmonic distance: 
\[
\frac{1}{|V|-1} \sum_{v \neq n \in V} \frac{1}{\delta(n,v)}.
\]

6.2.4 Clustering Coefficient

Clustering coefficient for a node \( v \in V \) is defined as \( \text{cluster}(v) = \frac{\lambda(v)}{\tau(v)} \), where \( \lambda(v) \) denotes the number of triangles and \( \tau(v) \) denotes the number of triplets containing \( v \). It [5] reflects the property of the graph to cluster around some particular nodes.

6.2.5 K-core

K-core centrality of a network \( G \) denotes the maximal connected sub-graph of \( G \) where all nodes are assigned with minimum \( k \) number of other nodes. In other nodes after k-core decomposition, all nodes with node degree < \( k \) would be deleted from the graph [10].

6.2.6 PageRank

The PageRank of a node \( v \in V \) is defined as \( prank(v) = \frac{1-d}{|V|} + d \sum_{n:(v,n) \in E} \frac{prank(u)}{|\{n:(v,u) \in E\}|} \), where \( d \) is a dampening factor. It is a weighted centrality which reflects a node is connected to the nodes those who have more importance in the graph [21].

6.2.7 Case Study

A Twitter graph is extracted and three centrality values are derived. Distribution graph of three centrality measures are shown in Fig. 7.
Fig. 7 Centrality measure from user distribution graph a Betweenness Centrality b Closeness Centrality c Node degree Centrality
6.3 Spreading Techniques

The maximum information propagation will be beneficial for smart governance as it ensures to reach maximum number of citizens. Some information propagation simulator is discussed in this subsection. A case study based on random walk is also presented.

6.3.1 Breadth First Search

In 1958 E. F. Moore propose Breadth-first search (BFS) for finding the shortest path among a grid network, similarly at almost same time in 1961 same algorithm was independently proposed by C. Y. Lee for wire routing algorithm [18]. BFS is used for propagation or searching in data structures like tree or graph. BFS starts from the seed node selected randomly from the network and propagate to the immediate neighbour with a probability. The neighbours then hopped to their next level neighbours and the process is continued. One of the disadvantage of BFS is that, nodes with a higher degree would be traversed more frequently which in term causes more biasing to the nodes with higher degree and results in local maximization with higher degree node. BFS is most studied and applied in online social network for finding user behaviour pattern, measurement and topological characterization of social network.

6.3.2 Forest Fire

This model is based on the concept of cellular automata. For a grid with dimension d and length l, the propagation follows in ld in the network using the following criterion:

- The burning node turns will be converted into empty node
- A node will be effected if at least one neighbor of the node is burning
- A node will ignite with the probability p even if no neighbor node is burning
- An empty space will fills with a node with probability s

6.3.3 Random Walk

This propagation is based on a path that consists of a continuation of randomly chosen advancement following some mathematical space such as the integers. Random walk is generally associated with Markov chains or Markov processes or Markov model or other variants of the model. Random walk started with the seed node and propagates to other nodes randomly with a probability p and traversed through the network.
6.3.4 Susceptible-Infectious-Recovered (SIR)

It is a probabilistic method with three states: Susceptible: no knowledge about the information, Infectious: have knowledge and want to spread and Recovered: Aware but not interested in spreading. Three states are more analogous to practical scenario and widely used to simulate the propagation.

6.3.5 Case Study

Using three highest centrality measures of previous result, information propagation is measured using random walk. Figure 8 represent the affected value in the network. This simulation is a small case study to reflect the utilization of centrality measure as source node of information propagation.

7 Rumour Detection & Influence Minimization

As discussed in earlier section the spread of malicious or rumour in online social network, can have negative effects in society. In that situation smart governance should provide a mechanism which can identify rumour and take proper steps to minimize the influence in social media.
7.1 Taxonomy of Rumour

Depending on the nature of rumours, it can be classified into following sections:

- **Misinformation**: Post that provides false information. It also includes posts that misrepresents the facts or represent information out of context. This type of information may cause harm to the society. For example, the propaganda against polio vaccination in India restricts the Indian government policies of removing polio from India.
- **Controversial information**: A post that can create dispute in another post, article, video or image unsubstantial information: A post that contains information that is either unsubstantiated or made some inference depending on the viewpoint of user.
- **Reporting misinformation**: A source that reports some rumour and supplies a secondary source such as newspaper reporting or some hyperlink.
- **Linked dispute**: A post that oppose some social network posts.
- **User belief**: A post that reflects opinion of that user.

7.2 Identification of Rumour

We can identify rumours or misinformation as unverified statement which starts from some sources without authentication and truthfulness and spreads in the network. A rumor can result in three directions: true (factual), false (nonfactual) or without conclusion. For the detection of rumours following analysis can be used.

7.2.1 Lexical Analyzer

Both semantic and syntactic analyses are required for lexical analysis of the data nodes involved in rumour propagation:

- **Semantic Analysis**:

  Semantic features include uses of opinion word/vulgar word/emoticons. It is common practice to identify commonly used opinion word and is a part of classification of sentiments which is used for identification of words reflecting strong negative or positive opinion (like trustworthy, terrible, traitor etc.). These types of words signify strong emotions and reflect personal view and indicate rumour. Similarly, the vulgar words tend to signal personal viewpoint. Emoticon, often used in online social network is also a good indicator of rumour.

  Speech Act word, such as expressions and recommendations are also a good indicator of user view point. It comprises of speech act verb like ask, promise, report and speech act phrase like I belief, we demand etc.
The uses of strong adjectives to qualify some incidents, events, persons are also an indicator of rumour.

- **Syntactic Analysis:**

Punctuations like ?(note of interrogation) and !(note of exclamation) often used for expressing the emotions and thus reflect personal viewpoint.

Part-of-speech tags, generated from the parser tree of dependency, can be used to identify the use of adjectives and interjections. In the sentence, where Interjections are used, mostly reflects emotion and thus can indicate personal expressions in the information. In the same way adjectives are used to indicate recommendations or personal expression of the user. Both of them are indicator of personal belief and thus indicator of misinformation.

### 7.2.2 Bursty Term Analysis

A keyword with bursty frequency generally appears only during a small time span. Generally it suddenly breaks out and end within a very short time span and show a tendency of rapid increase and decrease within the social network. The frequency of a general keyword generally continued as low steady level whereas frequent term occurrences always remain at high. But a sudden up and down of information frequency generally indicates a rumour. Periodic terms always occur once in a while, which implies it is less probable that periodic terms appears as bursty terms. Periodicity score of the topic can be scored and measured to find out the characteristics of the nature of occurrence.

### 7.2.3 Skewness of Information Distribution

For a periodic event, generally information generates a Gaussian distribution. For example, when there is football world cup, the information about the event grows in a regular pattern, reaches its peak at the time of event and then gradually goes down. But for a rumour, distribution of the keyword is generally show skewness (asymmetric distribution). Skewness score is a good measure for rumour detection in real-time.

### 7.2.4 User Identities

User’s identity is used for identification of the rumour. It includes following criterion for identification:

- **Controversiality**: it indicates the follower of the user
- **Originality**: it indicates whether the statement is original or not
- **Credibility**: Indicates whether online social network account is verified or not through the appropriate authority
- **Influence**: the information shared by the persons can create impact (positive or negative) or not
- **Role**: the kind of roles played by that user like active or stiffler user
- **Engagement**: It measures how active the user in online social network till joining

### 7.2.5 Source of Information

Source identification is an important criterion for rumour detection. Generally, rumours are generated from fewer numbers of sources and spread in an exponential manner. But for general information the sources are evenly distributed.

### 7.2.6 Sentence Clustering

After the mostly used terms have been extracted from the network through knowledge community detection, clustering of sentence is the next iteration to be followed. Sentence weight is defined as the average of all the weights of all the terms it comprises of. Sentences rank are determined and top k (number of k can be predefined or learned) sentences are considered as the centers of initial clusters and clusters are generated through clustering algorithm like K-means clustering.

### 7.3 Source Identification of Rumour

Source node can be identified from a social graph through verification of timestamp. As mentioned in earlier section, the two-phase algorithm, developed by Google, can convert a social graph into the union of stars where each star represents the connected components. Each vertex is assigned with a unique id. Searching and connecting all connected neighbours of the graph, it generates the star which is the union of all large and small of the network which includes that user as the vertex. From that star node with lowest time stamp can be considered as the vertex/apex node and signifies the source node of information.

### 7.4 Community Detection and Location Identification

For mapping the contagion of a rumour, it is important to find communities and where the nodes are affected by the information.
7.4.1 Community Detection

Communities are the structural property of the network indicating users of same community are interacting with each other as compared to other users of the network. Individuals of same community share more common and similar properties.

Community detection algorithms play crucial roles because characterizations of some clusters help us for the characterization of the network as a whole. This feature is very useful for scale free real network like online social network. For example we can identify the group of people who are interested in some particular topic. Similarly identification of group of users generating misinformation may results in proper measures against them. Identification of cyber-crime, terrorism can lead to resistance against the crime. Community detection has several application areas in smart governance.

7.4.2 Case Study

In the case study, Twitter account data are extracted and analyzed.

Using Twitter login the user should enter into a Twitter account and after authentication with Node XL, it will import Twitter data in form of adjacency matrix. For the data processing the downloaded data is stored as CSV file. Now this exported CSV file is used to plot the graph and apply different community detection algorithm to it. I-graph is used for graph visualization and R embedded in Hadoop platform is used for the analysis of community. For our case study we are using Edge betweenness algorithm which is a hierarchical algorithm. Edge betweenness score can be expressed as the ratio of shortest paths that include that particular edge and all possible shortest paths. In edge betweenness algorithm based on the edge betweenness value edges are removed in the decreasing order until the graph breaks into all disjoint components as shown in Fig. 9.

7.4.3 Location Identification

After the detection of community we can filter the information based on the location of data. In online spatial information are available: the location of the user from where it is registered and the location of the incidence, which is the topic of the rumour. For a smart city governance data will be collected based on this location of smart city.
7.5 Influence Minimization

Propagation of rumour can be controlled or restricted when there is an adverse effect in the society due to that propagation. Different rumour controlled strategy for rumour controlling can be adopted for that purposes.

7.5.1 Blocking the Source Node

From the previous subsection it is observed that source nodes of the rumour are generally small in number and can be identified from the network. Therefore it is possible that source node can be identified and blocked.

7.5.2 Blocking Centrality Nodes and Edges

For online social network, identification of rumour is associated with some time delay. At the time when information is recognized as rumour, information is already
spread into the medium. So it is required to restrict some influential nodes or edges from the network. Centrality nodes are somehow representing most influential node in the social graph. Blocking most influential node in the network, it is possible to restrict inoculation of the rumour in the network. It is observed that instead of blocking the centrality nodes, which costs the restriction of all connected edges of that node, it is less costly and more effective to block important edges. Centrality edges can be derived using some graph topological properties like edge betweenness, graph cut etc.

7.5.3 Beacon Model

In this model, Beacon nodes start anti-rumour campaigning after detection of rumour in the network and influence the user. In [10], random nodes are used as beacon node but centrality nodes will be more efficient for that purpose as centrality nodes are more efficient for spreading information [23]. The efficiency of Beacon model will increase with less delay time for the detection of rumour.

8 Conclusion and Future Research Scope

In this chapter, a framework for smart city governance based on online social network is proposed. For governance in a smart city, the prime criterion is that the governance should be citizen centric. As citizen has no means to communicate their view with the government directly, it will be the responsibility of the government to interact with the people proactively. Lot of research work is going on in that area, especially in information propagation and rumour detection. The fields that involve in this area are network analysis, big data analysis, natural language processing, machine learning and other interrelated disciplines. The main research challenges lies to its large volume of data and real time analysis. Lots of research scopes are present in this direction.

References

1. Alawadhi S, Scholl HJ (2016) Smart governance: a cross-case analysis of smart city initiatives. In: 49th Hawaii international conference on system sciences (HICSS), Koloa, HI, pp 2953–2963
2. Chowdhury C, Roy S, Song H, Srinivasan R, Sookoor T, Jeschke S (2017) Mobile crowdsensing for smart cities. In: Smart cities foundations principles and applications. Wiley, Hoboken, NJ, USA
3. Wang Q, Jin Y, Lin Z, Cheng S, Yang T (2016) Influence maximization in social networks under an independent cascade-based model. Physica A: Statistical Mech Appl 444:20–34. ISSN 0378-4371
4. Dey P, Chatterjee A, Roy S (2019) Influence maximization in online social network using different centrality measures as seed node of information propagation. Sadhana, Springer
5. Guzman JD, Deckro RF, Robbins MJ, Morris JF, Ballester NA (2014) An analytical comparison of social network measures. IEEE Trans Comput Soc Syst 1(1):35–45
6. Rodrigues FA (2019) Network centrality: an introduction. In: From nonlinear dynamics to complex systems: A Mathematical modeling approach, Macau E. (eds), Chapter-10, Springer
7. Fortunato S (2010) Community detection in graphs. Phys Rep 486:75–174
8. Raghavan UN, Albert R, Kumara S (2007) Near linear time algorithm to detect community structures in large-scale networks. Phys Rev E 76, 036106
9. Zhang Q, Zhang S, Dong J, Xiong J, Cheng X (2015) Automatic detection of rumor on social network. In: Li J., Ji H., Zhao D., Feng Y. (eds) Natural language processing and chinese computing. Lecture Notes in Computer Science, vol 9362. Springer, Cham
10. Rudra A, Tripathy M, Bagchi A, Mehta S (2010) A study of rumor control strategies on social networks. In: Proceedings of the 19th ACM international conference on Information and knowledge management, Toronto, ON, Canada, pp 1817–1820
11. Diao Q, Jiang J, Zhu F, Lim EP (2012) Finding bursty topics from micro blogs. In: Proceedings of the association for computational linguistics (ACL). pp 536–544
12. Zhao WX, Jiang J, Weng J, He J, Lim EP, Yan H, Li X (2011) Comparing twitter and traditional media using topic models. In: Proceedings of the 33rd European conference on advances in information retrieval, pp 338–349
13. Kentaro S, Tomohiro Y, Takeshi F (2014) Online topic model for twitter considering dynamics of user interests and topic trends. In: Proceedings of the 2014 conference on empirical methods in natural language processing, pp 1977–1985
14. Guille A, Hacid H, Favre C, Zighed DA (2013) Information diffusion in online social networks: a survey. SIGMOD Records 42(2):17–28
15. Polykarpos M, Giannis N, Francois R, Yannis S, Michalis V (2015) Degeneracy-based real-time sub-event detection in twitter stream. In: International AAAI conference on weblogs and social media
16. Diao Q, Jiang J (2014) Recurrent chinese restaurant process with a duration-based discount for event identification from twitter. In: SIAM international conference on data mining, pp. 388–397
17. Dey P, Roy S (2016) A comparative analysis of different social network parameters derived from Facebook profiles. In: Satapathy SC et al. (eds) Proceedings of the second international conference on computer and communication technologies, Advances in Intelligent Systems and Computing, vol 379
18. Dey P, Roy S (2016) Social network analysis. In: Meghanathan N (eds) Advanced methods for complex network analysis. Hershey, PA: IGI Global
19. Sarkar A, Chattopadhyay S, Dey P, Roy S (2017) The importance of seed nodes in spreading information in social networks: a case study. In: Communication systems and networks (COMSNETS), IEEE Xplore Digital Library
20. Arruda GFd, Barbieri AL, Rodríguez PM, Moreno Y, Costa LDdF, Rodrigues FA (2014) The role of centrality for the identification of influential spreaders in complex networks. In: Social and information networks, physics and society
21. Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking: bringing order to the web. In: Technical Report, Stanford InfoLab, pp 1–17
22. Dey P (2019) Characterization of online social network: a case study on Twitter data. In: Enabling technologies and architectures for next-generation networking capabilities, pp 272–299. IGI Global, USA
23. Dey P, Pyne S, Roy S (2017) Information spreading in online social networks: a case study on Twitter network. In: Proceedings of the 18th ACM international symposium on mobile Ad Hoc networking and computing