Journal of Intelligent Information Systems (2021) 56:355–377
https://doi.org/10.1007/s10844-020-00623-8

Modeling information diffusion in online social networks using a modified forest-fire model

Sanjay Kumar[1,2] · Muskan Saini[3] · Muskan Goel[4] · B. S. Panda[2]

Received: 12 July 2020 / Revised: 22 September 2020 / Accepted: 22 September 2020 /
Published online: 12 October 2020

Abstract

Information dissemination has changed rapidly in recent years with the emergence of social media which provides online platforms for people worldwide to share their thoughts, activities, emotions, and build social relationships. Hence, modeling information diffusion has become an important area of research in the field of network analysis. It involves the mathematical modeling of the movement of information and study the information spread pattern. In this paper, we attempt to model information propagation in online social networks using a nature-inspired approach based on a modified forest-fire model. A slight spark can start a wildfire in a forest, and the spread of this fire depends on vegetation, weather, and topography, which may act as fuel. On similar lines, we labeled users who haven't joined the network yet as $Empty$, existing users as $Tree$, and information as $Fire$. The spread of information across online social networks depends upon users-followers relationships, the significance of the topic, and other such features. We introduce a novel $Burnt$ state to the traditional forest-fire model to represent non-spreaders in the network. We validate our method on six real-world data-sets extracted from Twitter and conclude that the proposed model performs reasonably well in predicting information diffusion.

Keywords Information diffusion · Forest-fire model · Nature-inspired algorithm · Online social networks · Twitter

1 Introduction

Many complex systems like biological, communication and social networks can be modeled as graphs (Newman 2010). These systems constitute a large number of nodes with a complicated interaction among its members (Wang et al. 2012; Barabási 2016). Social networks are online resources that link people and help in the spread of information (Bakshy et al. 2012). Due to the rapid advancements of the internet and mobile networks, online

✉ Sanjay Kumar
sanjay.kumar@dtu.ac.in

Extended author information available on the last page of the article.

social networks (OSNs) like Twitter, Facebook, Siena Weibo have become very popular and connects billions of people worldwide. These platforms excel as tools for people to share news, trending topics, ideas, and opinions. Hence, online social networks have brought people together, enhance communication speed and generate a massive amount of data in a few minutes. Some actions of a few numbers of people may lead to a large scale spreading of information. It takes only a few clicks for information to reach from one corner of the world to another. Twitter, a micro-blogging website is one such network. As compared to other networks, Twitter is more diverse and densely connected. These properties make Twitter an engrossing network for research. Social networks like Twitter play an essential role in the analysis of the spread of information. Modeling information diffusion on these networks has many applications like finding trending topics (Mashiach and Sharma 2020), finding influential users (Kumar and Panda 2020), devising marketing strategies, identifying opinion leaders (He et al. 2020; Kimura et al. 2013), and many more. In the past, information diffusion models like susceptible-infected-recovered (SIR), susceptible-exposed-infected-recovered (SEIR), and other similar epidemic models have been used Guille et al. (2013) and Cai et al. (2012). These models try to model the information diffusion on the network and allowing the estimation of the spread of the information with certain limitations. Recently, many researchers have made proposals in the field of information diffusion simulating natural processes like spring damper, bee colonies, etc Cai et al. (2012) and Sankar and Kumar (2016).

A common hazard in forests is forest-fire, which occurs due to human-made or natural reasons. Broad research has demonstrated that the fundamental variables influencing the spread of forest-fires are forest flammability, climate, and landscape (Hawley 1926). Many studies were conducted in past to model the spread of wildfire in forests in form of computer simulations and algorithms (D'Ambrogio et al. 2016; Kanga and Singh 2017). It is an important area of research aiming to help mangers to quickly anticipate the spread of fires, giving a decision premise to work out viable fire extinguishing plans in advance. This natural phenomenon can be used to model many real time problems efficiently. The traditional forest-fire model has numerous applications in applied mathematics like cell automata, and community detection (Pattanayak et al. 2019; Wang and Liu 2011).

In this paper, we assume information spreading in an online social network is analogous to the spread of fire in a deep forest. We explore the usage of the forest-fire algorithm for modeling the spread of information on real-life data-sets of Twitter based on user's tweets around a given hashtag. The contributions of our paper are as follows:

1. We successfully extend the usage of forest-fire model on complex networks to study information diffusion and predict influenced people in online social networks.
2. We introduce a $Burnt$ node status in the traditional forest-fire model to categorize the infected population into spreaders, users who are responsible for spreading the information, and non-spreaders, users who keep information to themselves and do not spread it further.
3. We model the information spread for multiple hashtags including the information regarding the global pandemic of COVID-19. The proposed algorithm performs reasonably well on real-world tweet data-sets of Twitter.

The rest of the paper is organized into the following sections: Section 2 explores the traditional models in the field of information diffusion. Section 3 discusses the methodology used and our proposed algorithm. Section 4 presents the description of the datasets used to validate the model. In Section 5, we visualize the proposed algorithm on parts of the Twitter network. In Section 6, we discussed the results obtained for various real-world datasets. Section 7 concludes the paper and discusses the contribution of the paper in the field of information diffusion.

2 Backgrounds and related works

The process of information diffusion involves the flow of information from one part of a network to another. Information Diffusion in online social networks is an interesting area of research (Bakshy et al. 2012; Chakraborty et al. 2018). Several attempts have been made in the past to study the dynamics of the information diffusion process (Guille et al. 2013). Modeling information propagation can assist in predicting the extent of information spread. The models can be broadly divided into two categories, as shown on Fig. 1: Explanatory models and Predictive Models (Guille et al. 2013). Explanatory Models try to reiterate the path of information. Predictive models, on the other hand, predict the diffusion process in any network. Predictive Models can be further categorized into graph-based and non-graph-based models. Graph-based models include threshold models like Linear Threshold (LT) and cascading models like Independent Cascades (IC). These models work on the assumption that a static graph exists for the network and, thus, try to model the diffusion process. Non-Graph based include epidemic models like SI, SIR, SIS, SIRS Model, and Partial Differential Equation (PDE) Model (Zhou et al. 2006). These approaches do not assume the presence of a static graph. Epidemicmodeling is a popular approach in the field of information diffusion modeling (Daley and Gani 2001). Such models represent the spread of information in a manner similar to the spread of a disease in a community. In the epidemic spread, infected people come in contact with susceptible people and pass on the infection.

Similarly, information spreads from one person to another Stai et al. (2018). Several attempts have been made to study the dissemination of information using traditional epidemic models like the Susceptible-Infected model, Susceptible-Infected-Recovered model.

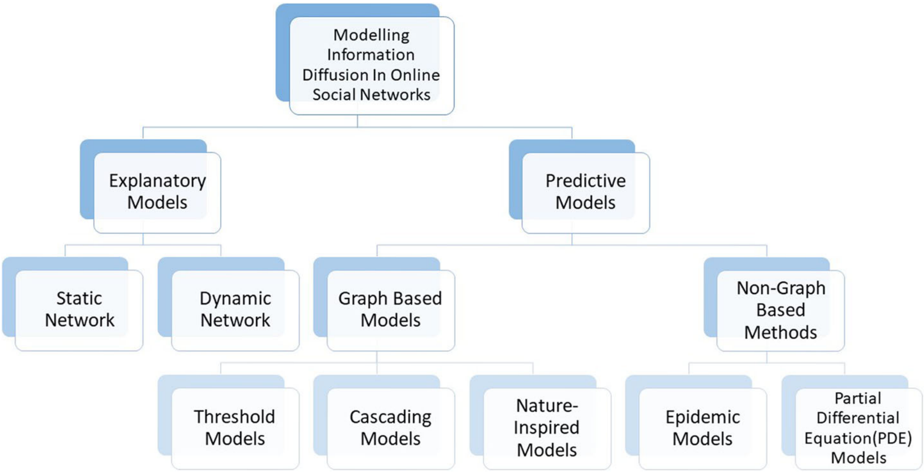


Fig. 1 Information diffusion models

Also, variants of these traditional models like SEIR (Biswas et al. 2014), SHIR (Liu et al. 2016), and others have been developed in the past. Recent work proposed a model named CISIR (competitive information susceptible infected recovered) for studying information dissemination in online social networks (He and Liu 2020). Some of the models were developed, especially for Twitter. Saito et al. (2015) observed information diffusion data and devised a generic algorithm to detect changes in parameter values of an information diffusion model. Hoang and Mothe (2018) proposed a predictive model for information diffusion on twitter based on user-based, content-based, and time-based features to evaluate if a post will be retweeted or not. Ding and Li (2018a) presented a topologically biased random walk for the diffusion process in multiplex networks displaying how topological properties affect the diffusion process. In the recent past, attempts have been made to model information diffusion using nature-inspired algorithms. Various nature-inspired algorithms like a firefly, cuckoo search algorithm, forest-fire, and similar have been used to solve problems related to social networks. Yang et al. (Ding and Li 2018b) discusses challenges and problems like scalability, parameter tuning, etc. that occur while dealing with nature-inspired algorithms. Forest-fire Model has been used to generate graphs (Rui et al. 2018) and investigate interactions in social networks (Fischer et al. 2013; Indu and Thampi 2019). We attempt to use the forest-fire algorithm to study information diffusion on Twitter and modify it to find the actual spreaders of information. Hu et al. (2018) gave an overview of different perspectives to view the information diffusion process. These perspectives are macroscopic, microscopic, and interaction between macroscopic and microscopic. The macroscopic perspective focuses on the extent to which information spreads in a network, whereas a microscopic perspective deals with individual nodes i.e., how a particular node will act upon a piece of information. The perspective of the interaction between macroscopic and microscopic levels deals with the relations between two levels. In this paper, we deal with both macroscopic and microscopic levels. We model the extent of information spread and also discover the nodes who may spread the information further and who may not.

3 Methodology

In this section, we describe the forest-fire algorithm and its relevance for modeling information diffusion. Firstly, we describe an algorithm to use forest-fire model online social networks to model the extent of information diffusion i.e., the total number of infected people. Secondly, we propose a modification in the traditional forest-fire model by adding a *Burnt* node to divide the infected population into spreaders and non-spreaders. This helps in identifying the people who actually spread the information. Further, we define the parameters on which the information spread depends.

3.1 Forest-fire algorithm

The traditional forest-fire model defined for a cell automata consists of three states: *Empty*, *Tree* and *Fire*. The model has two probabilities p and f where p is the probability with which new trees originate in a forest, and f is the probability with which any tree catches fire. Whenever a cell consisting of a tree catches fire, its adjacent cells also catch fire, and then they further spread the fire. We argue the same idea can be applied to graph datasets to study information diffusion in online social networks. Thus, in the case of graphs, whenever a node spreads information, its neighbors can also spread the information. The model can be applied to social network graphs like twitter as well. We present below Algorithm 1 based on the forest-fire model to simulate information diffusion in the Twitter network. In this algorithm, all the current Twitter users are considered as $Tree$. New users join the twitter network every day, leading to an $Empty$ to $Tree$ transformation. When a user tweets or retweets on a topic, he is given the status $Fire$. Also, all his followers are considered informed on the topic and are thus converted into $Fire$. The followers can further retweet a tweet spreading the information further. The fire keeps on spreading and soon covers the entire forest. Similarly, information unfurls over the whole network. In the algorithm, we calculate the probability f_u of a user posting a tweet on a topic considering various factors like user activity, and topic significance where f_0 is the threshold value for the probability f. The Algorithm 1 explains how forest-fire algorithm can be applied on graphs of social networks like twitter to study the extent of information diffusion. The input to the algorithm is the social network graph (G) and output is the list of infected or informed nodes (I) about the particular topic or subject.

Algorithm 1 Forest-fire algorithm for modeling information diffusion.

```
Input: Social Network, $G = (V, E)$
Output: List of informed nodes ($I$)
 1: procedure ForestFire($G$)
 2:     Initialize $I = \{\}$
 3:     Initialize status of each node
 4:     for each node $u$ in $G$ do
 5:         if status $[u]$ = Tree then
 6:             if $f_u \geq f_0$ then
 7:                 status $[u]$ = Fire
 8:                 $I = I \cup \{u\}$
 9:             end if
10:         end if
11:         if status$[u]$ = Fire then
12:             for each neighbour $v$ of $u$ do
13:                 status$[v]$ = Fire
14:                 $I = I \cup \{v\}$
15:             end for
16:         end if
17:     end for
18:     Return $I$
19: end procedure
```

The detailed explanation of the steps involved in the Algorithm 1 is as follows:

Step-1 The status of each node is initialized as per line 3 of Algorithm 1. The users who had posted a tweet on a topic are initialized as *fire*, and all others are labeled as *tree*.

Step-2 Each node of the graph is then iterated. If the status of node (u) is *tree*, then the tweet probability, f_u is calculated for node u using the Eq. No. (2). If $f_u \geq f_0$, where f_0 is the threshold value, then the user is considered to have posted a tweet on the topic and thus, is labeled as *fire*. The node (u) is then added to the list of informed nodes (I). This step explains lines 5-8 of Algorithm 1.

Step-3 If the status of the node is *Fire*, then all of its neighbors are also labeled as *fire* and added to the list of infected nodes. This is just like a forest-fire, when a tree is burning, all of its neighboring trees also catch fire. This step covers lines 9-12 of Algorithm 1.

3.2 Modified forest-fire (MFF) algorithm

In this section, we propose a modified forest-fire (MFF) algorithm for online social networks. In the forest-fire Algorithm 1, once a tree catches fire, it spreads it to its neighbors, which spread it further. In the case of social networks, all people do not spread the information further. Thus, infected or informed people can be classified into spreaders and non-spreaders. To model the non-spreaders, we introduce a new *Burnt* node. This can be correlated with the forest-fire phenomenon as some trees may get entirely burnt without spreading the fire further. Whether a tree may spread the fire further depends on the composition of the tree, climate, and vegetation of the area. Similarly, whether a user may pass on the information further depends on user characteristics and various other factors like user activity, user interests, topic significance. Thus, we calculate the retweet probability r_u, the probability with which a user retweets a tweet considering various characteristics of users and interests on tweets. The algorithm presented in 2 is written for the Twitter platform but can be easily extended to other online social networks like Facebook, Instagram, and others. We have created a correlation between forest-fire and information diffusion on the Twitter network. The algorithm partitions the entire population into 4 categories: *Empty*, *Tree*, *Fire*, *Burnt*

(i) *Empty* represents the users who have not yet joined the network. Twitter is a dynamic network, with new users joining the network everyday. This can be considered analogous to a forest where new trees originate every now and then.

(ii) *Tree* refers to users that have no knowledge about the topic.

(iii) *Fire* refers to those users that are informed and are actively involved in spreading the information further by either tweeting or retweeting on that topic.

(iv) *Burnt* refers to the nodes that are infected but do not spread the information further. A user v can become infected in two ways, either a user u whom v follows publishes a tweet on a particular topic or user v becomes aware of the topic from some outside source. It is important to note that *Burnt* nodes are not spreading the information further.

There are three types of probabilities that come into play in this algorithm p , f_u and r_u.

(i) p is the probability that a new user joins the network i.e. *Empty* to *Fire* transformation.

(ii) f_u is the probability that an existing user will post a tweet related to a topic i.e. transformation from *Tree* to *Fire* state.

(iii) r_u is the probability that a user v retweets a tweet that the person u he follows tweeted.

Algorithm 2 describes the modified forest-fire algorithm and explains its applicability on online social networks to model information diffusion. In the algorithm, f_0 and r_0 are the threshold probability values for the probabilities f_u and r_u, respectively. The input to the Algorithm 2 is the social network graph (G). The output is the list of spreader nodes (S) and non-spreader nodes (NS).

Algorithm 2 Modified forest-fire algorithm.

Input: Social Network, $G = (V, E)$
Output: List of spreaders (S) and non-spreaders (NS)

```
 1: procedure ModifiedForestFire( $G$ )
 2:     Initialize $S = \{\}$, $NS = \{\}$
 3:     Initialize status of each node
 4:     for each node $u$ in $G$ do
 5:         if status[$u$] = $Tree$ then
 6:             if $f_u \geq f_0$ then
 7:                 status [$u$] = $Fire$
 8:                 $S = S \cup \{u\}$
 9:             end if
10:         end if
11:         if status [$u$] = $Fire$ then
12:             for each neighbour $v$ of $u$ do
13:                 if status [$v$] = $Tree$ then
14:                     if $r_u \geq r_0$ then
15:                         status [$v$] = $Fire$
16:                         $S = S \cup \{v\}$
17:                     else
18:                         status [$v$] = $Burnt$
19:                         $NS = NS \cup \{v\}$
20:                     end if
21:                 end if
22:             end for
23:         end if
24:     end for
25:     Return $S$ and $NS$
26: end procedure
```

The detailed explanation of the steps involved in the Algorithm 2 is as follows:

Step-1 The status of each node is initialized as either *tree* or *fire*. The users who had posted a tweet on a topic are labeled as *fire*, and all others are labeled as *tree*. The list of spreaders (S) and non-spreaders (NS) is also initialized. This explains lines 2 and 3 of the Algorithm 2.

Step-2 It explains lines 4-8 of the algorithm. Each node of the graph is iterated. If the status of the node (u) is *tree*, then the tweet probability, f_u is calculated for that node using the Eq. No. (2). If $f_u \geq f_0$, where f_0 is the threshold value, then the user is considered to have posted a tweet on the topic and thus, is labeled as *fire*. The node (u) is added in the list of spreaders.

Step-3 If the node's status is *fire*, then it means that the user has posted a tweet or retweet on a topic, then each of its neighbor or follower (v) is considered. If the status of node v is *tree*, then the retweet probability, r_u is calculated for that follower using the Eq. No. (8). If $r_u \geq r_0$, where r_0 is the threshold value, the follower (v) is considered to have re-tweeted the person's tweet or retweet he follows and thus, v is labeled as *fire* and is added to the list of spreaders. If $r_u < r_0$, it means that the person (v) did not retweet the tweet and did not spread the information further.

Therefore, user v is labeled as *Burnt* and is added to the list of non-spreaders. This explains lines 9-17 of the Algorithm 2.

Step-4 After iterating all the nodes of the graph, the list of spreaders and non-spreaders is returned as output.

3.3 Defining probabilities

In this subsection, we study the probabilities p, f_u and r_u and enumerate the factors on which these probabilities depend. After scrutinizing the features of tweets and users, we model these probabilities and use them in the modified forest-fire model to study the rate of spread of information.

3.3.1 Probability of a new user joining the network (p)

Online social networks have become so popular and advanced in the last decade. They are still expanding and reaching people worldwide. Any new user joining the social network can act as a *Tree* to spread information. Hence, p represents the probability that state changes from *Empty* to *Tree*.

3.3.2 Probability of an existing user tweeting about the topic (f_u)

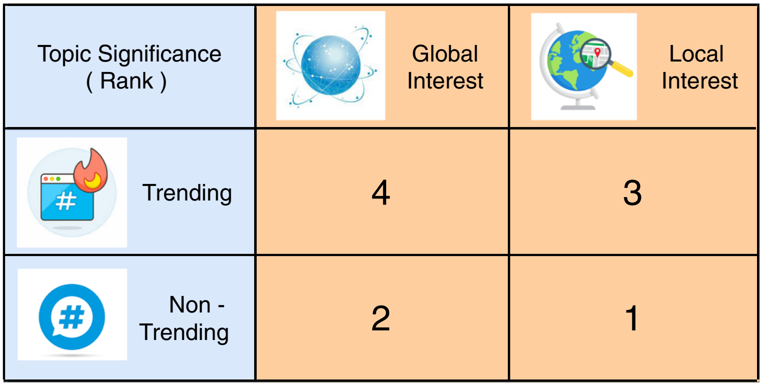
The entire forest area could burst into flames ignited by the slightest spark. Similarly, in online social networks, the users who tweet about the topic act as a spark in the system. These are the users who bring information to the network. After analyzing Twitter statistics, we found that whether or not a user tweets about a topic largely depends on the user's general activity and the importance of the topic.

(i) **User Activity** (UA): The general activity of a user can be determined by taking into consideration the total number of tweets or retweets posted by a user over a certain period of time. All the tweets/retweets posted by the user from the account registration period are considered. Its high value specifies that the user is highly active on Twitter, and thus, there is a high possibility that our user will tweet about a topic. Depending on the account activity, we assign a range of values to users. We compute the user's activity (UA) as follows:-

$$UA = \frac{\text{Total number of tweets or retweets made by the user}}{\text{Number of days from account registration period till date}} \tag{1}$$

(ii) **Topic Significance** (TS): In general, users tweet about the more popular and trending topics. Depending upon the significance of the topic of the tweet, we provide them ranks or scores (TS), as shown in Fig. 2. Higher the score, higher is the importance of topic and thus, higher is the probability that any user will tweet about it. We divide the topics into following categories and then give them scores according to their importance.

(a) *Global and Local Interest* : A topic can be of global interest like COVID-19 with multiple users from diverse corners of the world sharing and tweeting about the same or of local interest like elections in an area attracting the interests of only locals residing in or involved with the area. Topics with global affairs attract a lot of attention worldwide and hence, tend to get shared a lot as compared to topics of local interests.



Topic Significance (Rank)	Global Interest	Local Interest
Trending	4	3
Non - Trending	2	1

Fig. 2 Assigning ranking to tweets on the basis of their topic significance

(b) *Trending and Non-Trending* : If the content is related to hot news recently or say has a trending topic like #MeToo movement, COVID-19 Pandemic, etc. The users tend to get more captivated by it and may post their status on that particular topic. On the other hand, non-trending topics like #Travel, #Tuesday are just like regular contents that are less popular and may get less recognition.

Thus, global and trending topics are given highest score and local and non-trending topics are given the least score.

The probability f_u is calculated as the weighted summation of above discussed features like user activity, topic interest as shown in Fig. 3. Tweet probability (f_u) can be represented mathematically as:

$$f_u = w_1 \times UA + w_2 \times TS \tag{2}$$

Here, w_1 and w_2 are the constant weights associated with the above mentioned features namely User Activity and Topic Significance respectively.

$$w_1 + w_2 = 1 \tag{3}$$

3.3.3 Probability of an existing user retweeting a tweet (r_u)

To properly understand how information flows in online social networks, we first need to calculate the probability by which any user will retweet since retweeters (people who

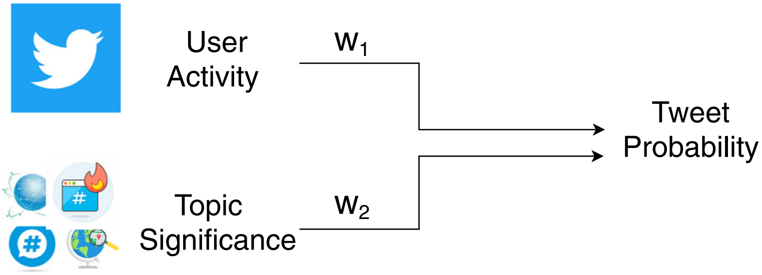


Fig. 3 Calculation of tweet probability (f_u)

re-tweet or re-post another person's tweet) play a significant role in disseminating the information over the entire network. Several works have been done in past as well to study the retweet phenomenon (Kuang et al. 2014; Nesi et al. 2018). We analyzed some important features of users listed below which can help us in answering that with what probability a particular user may retweet a particular tweet.

(i) **IsMentioned (IM)**:- One of the most popularly used functions on Twitter is the @mention function. The @mention function is used to 'tag' users in the status updates. Mentioning a user on Twitter is an effective way to grab a user's attention quickly and hence, there is a very high probability that the user will re-tweet or interact with it in some way or other. Recursive partitioning procedure models (RPART) also labelled mentions count as the most correlated feature with the action of retweeting (Nesi et al. 2018). Here, we take IsMentioned score as a binary variable.

$$\text{IsMentioned (IM)} = \begin{cases} 1, & \text{if user is mentioned in the tweet} \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

(ii) **Similarity Score (SS)** Relationships between the users are the backbone of social media services. One of the most significant features that describe a relationship is how similar the two users are in terms of their interests, activities and behaviours. More similar the two users/accounts are, higher is the possibility of one retweeting another's content and spreading it in the network. We consider some user-specific features to calculate similarity score. We use Jaccard similarity index to calculate these measures or features. The formula to find Jaccard similarity index for a given measure *m* for two user accounts can be re-written as:

$$A_m(X, Y) = \frac{X_m \cap Y_m}{X_m \cup Y_m} \tag{5}$$

Here, $\mathrm{A_m}$ $(\mathrm{X_m},\mathrm{Y_m})$ represents the Account similarity between two user accounts X and Y for a given feature or measure *m*. $\mathrm{X_m}$ and $\mathrm{Y_m}$ are the values of the measure under consideration. $X_m \cap Y_m$ represents the number of values of corresponding measure common in both accounts and $X_m \cup Y_m$ represents the total number of values of corresponding measure in either of the accounts. We consider five different features to find the similarity between two users X and Y using Jaccard similarity index. These features as follows :-

(a) *Followings List* (A_{Fg}): By intuition, users who share common interests must be following some common accounts. Hence, we find the "following" list of both the users and find the number of common accounts they are following.

(b) *Hashtags Used* (A_{Ht}): Hashtags have become the new fashion to grab more crowd's attention towards any topic on social media platforms. Users use hashtags to highlight some relevant keywords or phrases while posting content (or tweet). Clicking on these words containing hashtags take you to the other tweets labeled with the same hashtags. We can clearly say that hashtags define the interests and ideas of our users. Evidently, if two users have common hashtags, we can infer that they are interested in similar topics. We extracted hashtags of both the users, more the common hashtags and keywords in user's tweets, similar their interests, tend to be.

(c) *Languages used* (A_{Lg}): The online social networks constitute people from diverse backgrounds and cultures spread out geographically. Only those users who can understand each other's languages will take interest in another's content (tweet) and may retweet it. Hence, we found the common languages used by both the users by extracting and analysing languages used in tweeting/re-tweeting by the user on the platform.

(d) *Location of Account* (A_{Lc}): Location can be also considered in calculating similarity score as people from same geographical boundaries may share some common interests, say related to local activities, although not necessarily. Providing your profile location is an optional feature in Twitter. So, we took this factor in account only for those users who have provided their location.

(e) *Followers List* (A_{Fw}): Let us consider a user who tweets about a topic. We can assume that the followers of our user can share some common interests with our user and are similar since similar people tend to follow each other. Furthermore, when these followers of our user are also following some other user, we can infer that they too may appear similar in some aspects. Hence, the two users are similar and hence, we can use followers list of both the users to find the common followers as one of the similarity characteristics.

Finally, similarity score is calculated as the weighted summation of above five measures as also shown in Fig. 4. The similarity score (SS) can be represented mathematically as:

$$SS = w_{Fg} \times A_{Fg} + w_{Ht} \times A_{Ht} + w_{Lg} \times A_{Lg} + w_{Lc} \times A_{Lc} + w_{Fw} \times A_{Fw} \quad (6)$$

Here, w_{Fg}, w_{Ht}, w_{Lg}, w_{Lc} and w_{Fw} are constant weights associated with above-discussed features, namely the Followings List, Hashtags Used, Languages Used,

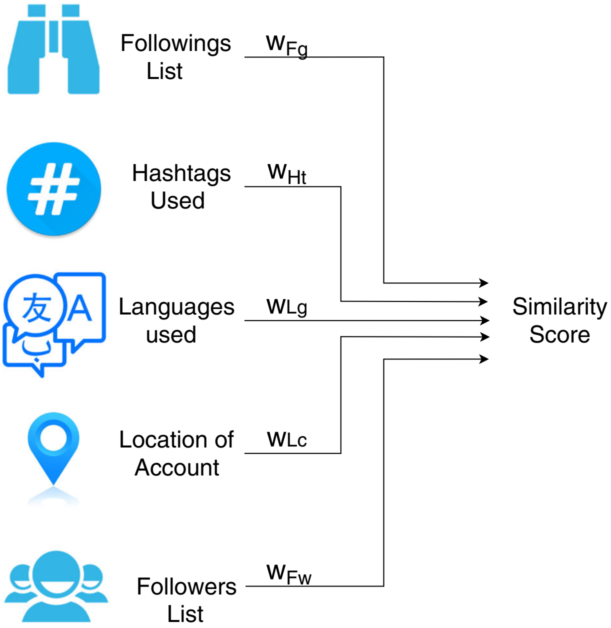


Fig. 4 Calculation of similarity score between users

Location of the Account, and Followers List features respectively. The sum of all these weight constants is one.

$$w_{Fg} + w_{Ht} + w_{Lg} + w_{Lc} + w_{Fw} = 1 \tag{7}$$

(iii) **User Activity (UA)**: The general activity of the user gives us the idea of how much active a user is on Twitter. A high value of user activity indicates that our user is much engaged with the platform and thus, there is a high possibility that our user will be retweeting another person's tweet. We use Eq. No. (1) to calculate the user activity.

(iv) **Topic Significance (TS)** The huge amount of retweeting happens just when the content is appealing, which we name as the topic significance. Users will in general give more consideration to those tweets with attractive and popular contents, that is, with high evaluation of significance of topic. Depending upon the significance of the topic of the tweet, we gave them scores (TS) using Fig. 2. Higher the score, higher is its probability to get retweeted or dispersed further in the online network.

Finally, retweet probability (r_u) of any user is just a weighted summation of the scores of four features listed above as shown in Fig. 5. It can be represented mathematically as:

$$r_u = w_{IM} \times IM + w_{SS} \times SS + w_{UA} \times UA + w_{TS} \times TS \tag{8}$$

, where w_{IM}, w_{SS}, w_{UA}, and w_{TS} are constants weight associated with IsMentioned, similarity score, user activity, and topic significance, respectively. The sum of all these weight constants is one.

$$w_{IM} + w_{SS} + w_{UA} + w_{TS} = 1 \tag{9}$$

4 Datasets description

We test our proposed model for the Twitter network. We perform the results validated for six real-time Twitter datasets on the topics: Corona Virus (Smith 2020), FIFA World Cup

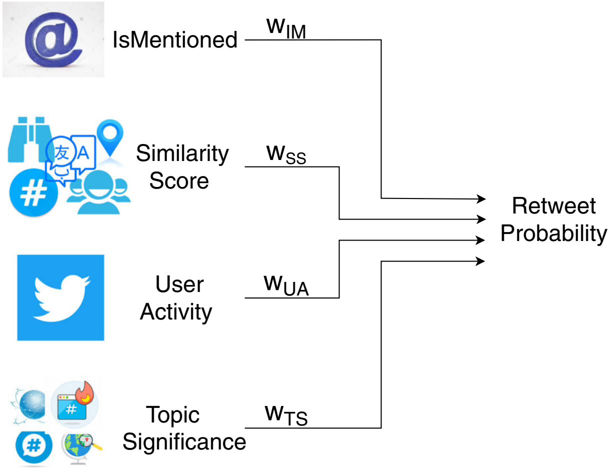


Fig. 5 Factors on which retweet probability (r_u) depends

(Rituparna 2018), NBA Finals (Pesic 2018), Game of Thrones S8 (de Abreu 2019), MeToo movement (Ramirez 2018) and Coachella Festival (Pesic 2019). “Coronavirus” or “COVID-19” refers to the global pandemic caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). It started in December 2019 in Wuhan, China, and has been declared a public health emergency of international concern by the World health organization (WHO). As of April 2020, around 210 countries in the world have reported cases of COVID-19.

“FIFA World Cup” refers to the international football tournament in which men’s national teams of 32 countries competed with each other. It was hosted by Russia from 14 June 2018 to 15 July 2018.

“NBAFinals” refers to the championship series of the National Basketball Association’s 2017–18 season and conclusion of the season’s playoffs. The dataset contains the tweets captured during the 3rd game of the 2018 NBA Finals between Cleveland Cavaliers and Golden State Warriors.

“Me Too” movement was a global campaign against sexual assault. It was first started by Tarana Burke in 2006 to empower girls who had faced sexual harassment. This movement gained popularity when a hashtag #metoo started creating a surge on social media platforms and women started opening up regarding their issues. The movement took hold in India around 2018 when women began sharing stories about harassment at the workplace and various aspects of their life. From actors to politicians to professionals to writers, everyone started calling out their experiences. In the past, women had mostly remained silent on such issues. This movement gave them the courage to open up and speak for themselves.

“Game of Thrones” refers to the American drama television series created by David Benioff and D. B. Weiss based on George R. R. Martin’s series of novels. The series was launched in April 2011. Eight seasons have been broadcasted till now, with the latest one being launched in April 2019. The series has a huge fan following and active viewership.

“Coachella 2019” refers to the Coachella Valley Music and Arts Festival, an annual music and arts festival organized at the Empire Polo Club in Indio, California, in the Coachella Valley in the Colorado Desert. It was Coachella’s 20th anniversary in 2019.

Table 1 presents the hashtags used to collect tweets of a topic and the duration for which the tweets were collected for each dataset.

We adopt the following pre-processing steps on the collected data:

i) The datasets that we utilize have many fields or tokens like user_id, status_id, created_at, screen_name, text_of_tweet, hashtags, language, account_created_at, location, retweet_count, followers_count, url, etc. We extracted required columns like followers count, re-tweet count, timestamp, etc. from the collected data based on the hashtag.

ii) Based on each topic or hashtag, tweets are segregated into time intervals using the created_at or timestamp attribute. We split tweets according to observed time intervals, like on an hourly basis or daily basis.

iii) For each time interval, we calculate the total tweets count, re-tweet count, and followers count using cumulative sum. Then, we compute the number of spreaders as the number of people who tweeted or re-tweeted and the number of non-spreaders as the total number of followers of people who tweeted or re-tweeted minus the number of followers who re-tweeted.

Table 2 presents the total number of tweets, retweets, and followers count for each dataset. Here, followers count refers to the total sum of followers of the users who tweeted on a topic.

Table 1 Datasets description

Topic	Hashtags used	Duration
Coronavirus Pandemic	#covid19 #coronavirus #coronavirusoutbreak #coronavirusPandemic	March 13, 2020 00:00 am to 23:59 pm IST
FIFA World Cup 2018	#FIFA #WorldCup #FIFA2018	July 2, 2018 to July 7, 2018
NBA Finals 2018	#NBAFinals	June 7, 2018 01:13 am to 01:58 am UTC
MeToo Movement	#metoo #SexualAssault #MarchForOurLives	February 20, 2018 19:05 pm to 19:55 pm IST
Game of thrones S8	#GameOfThrones #WinterIsComing #GoTS8 #GoT #TheMadKing	April 14, 2019 to May 14, 2019
Coachella festival 2019	#coachella #coachella19 #coachella2019	April 7, 2019 to April 23, 2019

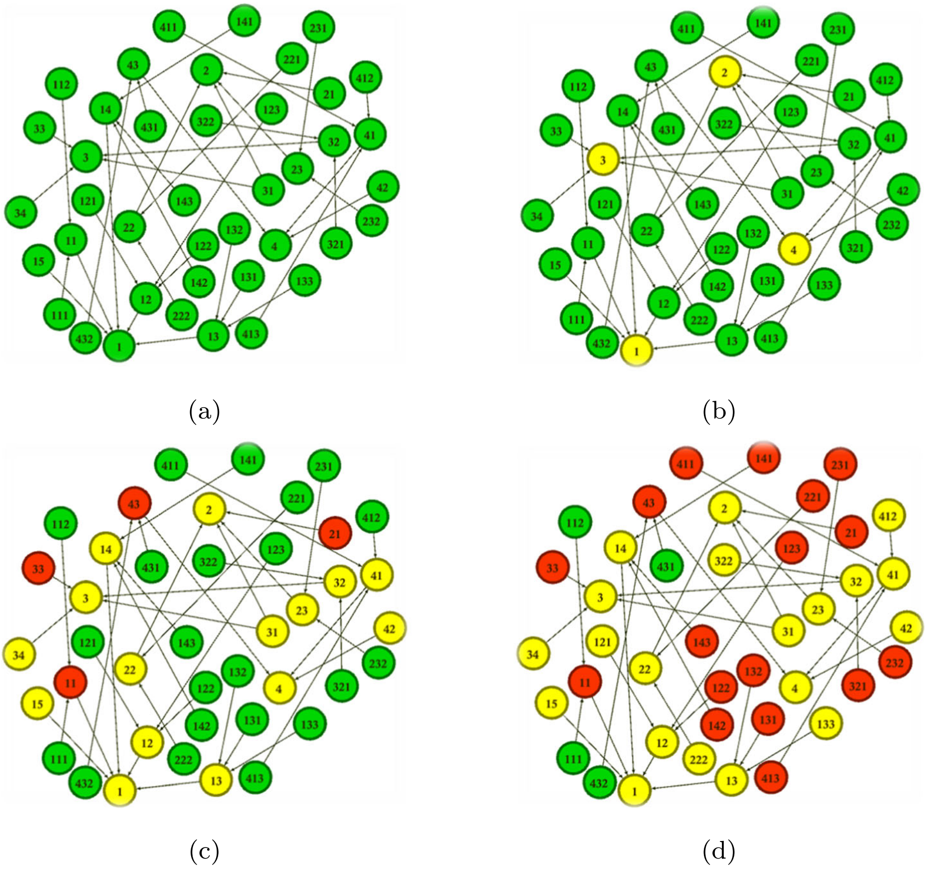
5 Experimental setup and visualization

We generated graphs using user-follower data mined using Twitter API and then applied modified forest-fire (MFF) on the graphs. We perform the experimental results on a personnel system with configuration : 8 GB RAM and Intel(R) Core(TM) i5-8250U CPU @1.60 GHz (8 CPUs), 1.8GHz processor.

We visualized the obtained graphs using the Gephi Tool, which is an open-source network analysis, exploration, and visualization software package written in Java on the NetBeans platform. Figure 6 gives a visual understanding of how information is being propagated in a toy network. It represents a step-by-step illustration of our proposed model on

Table 2 Number of tweets, retweets and followers count

Topic	Number of tweets	Number of retweets	Total followers count
Coronavirus Pandemic	300273	694338	42819166666
FIFA world Cup 2018	242876	41927	2441853742
NBA finals 2018	19986	31439	748874447
MeToo Movement	1000	12	7620496
Game of Thrones S8	504726	12499	186646420
Coachella Festival 2019	390276	412949	22794218542



(a)

(b)

(c)

(d)

Fig. 6 Modified Forest-fire (MFF): **a** All trees are initially green means all the nodes have no knowledge about the topic or information **b** Some people have posted a tweet on a particular topic and are colored in yellow, **c** Some of the followers choose to spread the information and are colored in yellow. However, some followers decide not to spread the information further and hence, are considered burnt and colored red, and **d** The final stage of the simulation. As the fire runs over in the entire forest, the majority of the nodes become either red or yellow

the part of the Twitter network graph generated. Here, the green nodes represent *Tree*, the yellow nodes represent *Fire* and the red nodes represent *Burnt*. The red nodes are the users who are infected but choose not to spread the information further and thus, are kept in the category of *Burnt*, whereas yellow nodes are the users who are infected and are actively spreading the information further like a fire in a forest and are thus, labeled as *Fire*. It is evident from the figure that with time most of the nodes turn either yellow or red i.e., information reaches a large group of users worldwide. We present the simulation of the proposed algorithm on the toy network in the following four steps:-

Step 1: Initially, all nodes in the network are green representing trees in the forest. These nodes represent the set of Twitter users i.e. the entire network. These nodes have no knowledge about the topic.

Step 2: Then some of the nodes catch fire and are represented by yellow color. This means that some nodes have posted a tweet on that topic, thus starting the information diffusion process. These nodes are the spreaders of information on that topic.

Step 3: Now, the neighbours or followers of the spreader nodes also learn about the information. Some of them further spread the information by retweeting the tweet. These nodes have now become spreaders and hence, are colored yellow. Some of the followers choose not to spread the information further and hence, are considered burnt and colored red. These are the non-spreaders.

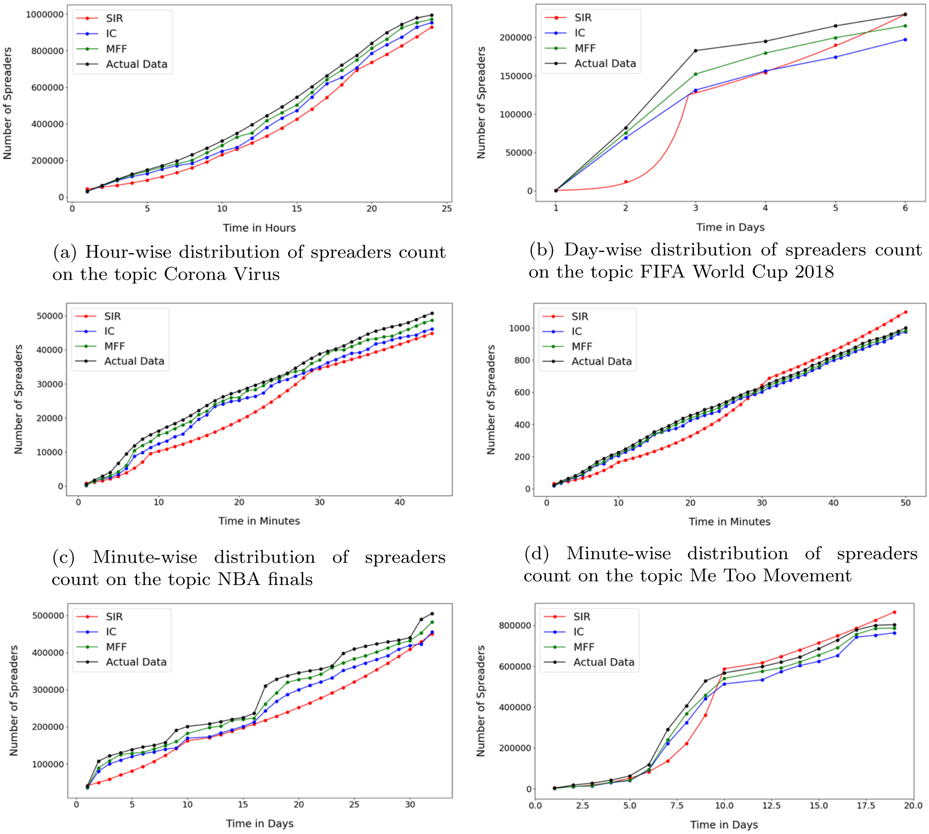
Step 4: With the passage of time, this fire keeps on spreading further with green nodes turning yellow or red i.e. conversion of most of the users from *Tree* to *Fire* or *Burnt*. This implies that information has reached to most of the users with time in the network. It is due to the highly inter-connected nodes that make fire spread swiftly in such a dense network of users worldwide. The extent of information spread depends on the rate at which users post tweets and retweets.

6 Results and discussion

We analyzed information spread for six real-time datasets extracted from Twitter. The tweets were clustered into several groups with respect to time (minutes, hours, days) depending on the datasets. A graph is generated based on the user-followers data extracted from Twitter. The modified forest-fire (MFF) model is applied to the generated graph. Since the data collected ranges over a few days, the number of new users joining the network is very low as compared to the twitter population of more than 300 million. Thus, the population can be considered constant during our time of research. Therefore, the probability p with which new users join the network can be considered zero. The probability f_u of any user posting a tweet on a topic is calculated as per Eq. No.(2). The factors user activity and topic significance were assigned equal weights. Thus, w_1 = 0.5 and $w_2 = 0.5$. The weights assigned to various factors involved in calculating similarity score in the (6) are $w_{Fg} = 0.3$, $w_{Ht} = 0.25$, $w_{Lg} = 0.25$, $w_{Lc} = 0.1$, $w_{Fw} = 0.1$. The retweet probability r_u with which a user retweets a tweet is calculated using the Eq. No. (8). Here, different measures are assigned different weights, $w_{IM} = 0.25$, $w_{SS} = 0.3$, $w_{UA} = 0.2$, and $w_{TS} = 0.25$. These weight constants are obtained by performing experiments on the Twitter data which gives optimal results.

The modified forest-fire (MFF) algorithm is applied on the Twitter graph to distinguish between spreaders and non-spreaders users. At $t = 0$, those users who had posted a tweet about the given topic were labeled as *Fire*. There can be some users who can get information from outside sources and post a tweet with probability f_u greater than or equal f_0. Here, we consider the value of threshold, i.e, f_0 = 0.5. Thus, we also considered the case of self-infection. The followers of an infected user are segregated into spreaders and non-spreaders based on the retweet probability (r_u). For each follower of the user who posted a tweet or retweet, we calculated the retweet probability using the formula mentioned in (8). We consider those followers for whom the retweet probability is greater than a minimum threshold (r_0) of 0.5 as spreaders nodes; otherwise, nodes are labeled as non-spreaders. The overall informed or infected count is the sum of spreaders and non-spreaders. The count of spreader nodes, non-spreader nodes and total informed nodes is plotted against time. The experimental values are found to be very close to actual values. For each topic or hashtag, the actual data is the cumulative sum of the number of tweets and re-tweets over time. Spreaders are the people who actively take part in information diffusion process by spreading the information. In the case of twitter, these are the people who post tweets on a particular topic or retweet the tweets of users whom they follow. We compare the spreaders count predicted by our model with other classical models in literature. The models used for comparison are Independent Cascade (IC) Model and Susceptible-Infected-Recovered (SIR) Model. Both of these are prominent models used to study information propagation and are known to produce satisfactory results. We utilize these two models because both of these predict the active spreaders in a network.

Figure 7 illustrates the spreaders count predicted by the modified forest-fire (MFF) model, independent cascades (IC) model, and susceptible-infected-recovered (SIR) model along with the actual spreaders count as calculated from the tweets data. The actual data comprises of all the tweets and retweets posted by users. For each topic or hashtag, the actual data is the cumulative sum of the number of tweets and re-tweets over time. We considered people who tweeted or re-tweeted on a topic or hashtag as spreaders of information. Thus for spreaders, the actual data comprises of the number of tweets and retweets.



(a) Hour-wise distribution of spreaders count on the topic Corona Virus

(b) Day-wise distribution of spreaders count on the topic FIFA World Cup 2018

(c) Minute-wise distribution of spreaders count on the topic NBA finals

(d) Minute-wise distribution of spreaders count on the topic Me Too Movement

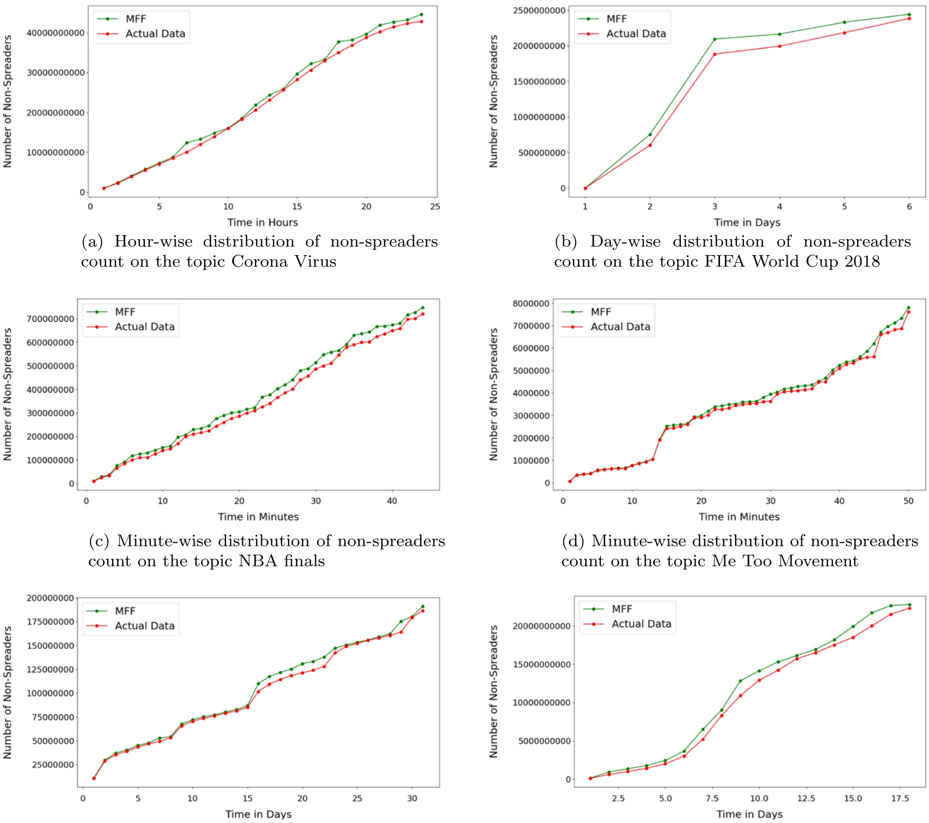
(e) Day-wise distribution of spreaders count on the topic Game of Thrones

(f) Day-wise distribution of spreaders count on the topic Coachella 2019

Fig. 7 Comparison of distribution of spreader nodes with respect to time according to various algorithms for various topics

Non-spreaders are the people who have knowledge of a topic but choose not to spread the information further. These are the followers of people who tweeted on a topic. These people chose not to retweet further. Thus, for non-spreaders, actual data is the difference of followers count and retweet count of the users tweeted on a topic. The total infected count is the sum of spreaders and non-spreaders. For the total infected count, we took the sum of tweets, retweets, and followers count. The model has been validated for six datasets as described in Table 1. Figure 7a illustrates the hour-wise spreaders count for the topic Corona Virus Pandemic. It can be seen that our MFF model predicts values close to the actual data followed by IC Model and SIR Model. Figure 7b shows the day-wise spreaders count for the topic FIFA World Cup 2018. It can be observed that SIR Model predicts very low spreaders count initially but later on, the count increased rapidly reaching close to the actual data. The MFF model predicts better results throughout the duration. The FIFA World Cup data was collected from July 2, 2018 to July 7, 2018. There were no football matches on July 4 and 5. The frequency of tweets was high during the match or on the day of the match. Thus, after day 3, there is a drop in tweet count. As the infected count is a cumulative sum of tweets over days, there is a plateau in the graph after day 3. Figure 7c represents the minute-wise spreaders count for the topic NBA Finals. Here, SIR Model underestimates the spreaders count. Our MFF Model gives better results than IC Model and SIR Model. Figure 7d illustrates the minute-wise spreaders count for the topic Me Too Movement. The SIR model underestimates the spreaders count initially and then overestimates it. The IC model and Modified Forest-fire model predicted results close to the actual data. Figure 7e shows the day-wise spreaders count for the topic Game of Thrones Season 08. It can be observed that MFF Model was able to fit the actual data better than IC Model and SIR Model. Figure 7f illustrates the day-wise spreaders count for the topic Coachella Music and Arts Festival. The SIR model initially underestimated the spreaders count but in later stages, gave optimal results. The MFF Model gave satisfactory results throughout. It can be observed from Fig. 7, that our model (MFF) predicted results which are close to the actual data. The SIR model produces varying results, sometimes overestimating the spreaders population and sometimes underestimating it. It can be seen from the graphs, that our model performed better than IC and SIR model as the values of spreaders count predicted were close to the actual values. Thus, it can be concluded that our model predicts active spreaders efficiently and can be used for modeling for information diffusion in real-life scenario.

Non-spreaders are people who have knowledge of the information but do not pass it on to others. The classical information diffusion models like IC and SIR model do not consider non-spreaders as infected. They only take in to account the users who spread the infection. However, our proposed MFF model considers non-spreaders in the information diffusion process. The number of non-spreaders with respect to time are also plotted for the six datasets as depicted in Table 1. Figure 8 shows the plot of number of non-spreaders vs. time duration either in minutes, hours or days for various topics. It can be seen from the results obtained on all the topics mentioned, that the non-spreaders count predicted by our model is in harmony with the actual count calculated from the tweets data. Here, the actual data is the difference of followers count and retweet count of the users who tweeted on a topic. It is observed from the results that the non-spreaders count is much higher than the spreaders count, indicating that only a few people spread the information further as compared to a large number of people who have knowledge about the information. It can be deduced that the actions of a few people can lead to large-scale dissemination of information. Thus, segregating spreaders and non-spreaders is useful to predict information diffusion.



(a) Hour-wise distribution of non-spreaders count on the topic Corona Virus

(b) Day-wise distribution of non-spreaders count on the topic FIFA World Cup 2018

(c) Minute-wise distribution of non-spreaders count on the topic NBA finals

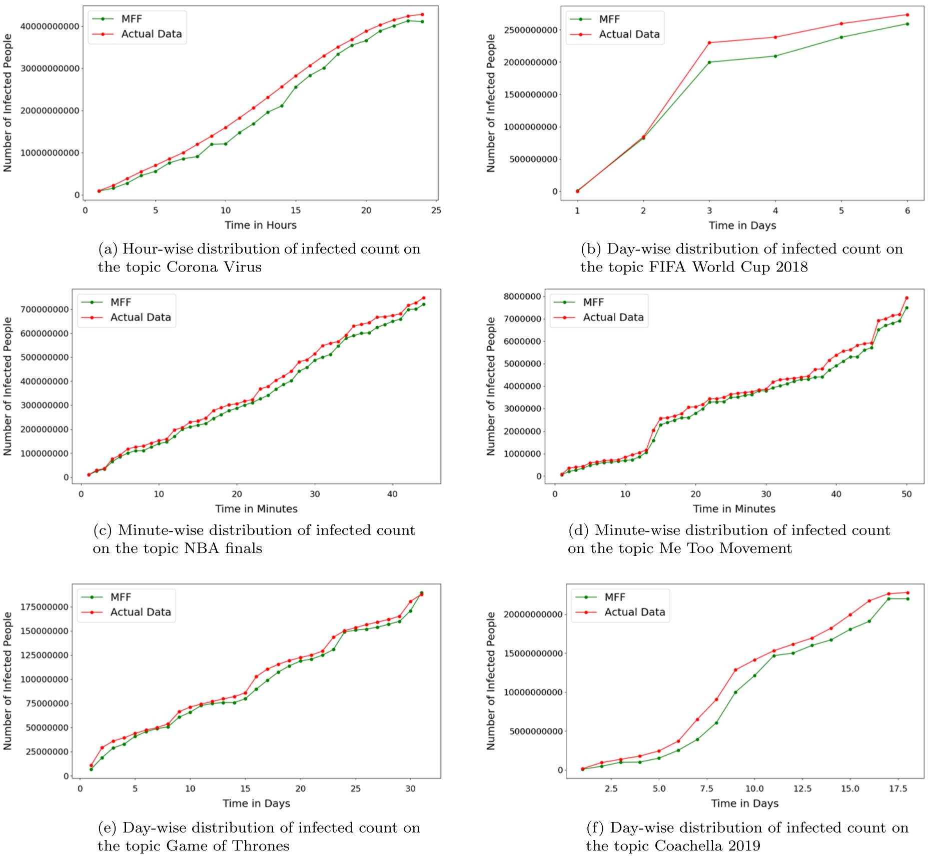
(d) Minute-wise distribution of non-spreaders count on the topic Me Too Movement

(e) Day-wise distribution of non-spreaders count on the topic Game of Thrones

(f) Day-wise distribution of non-spreaders count on the topic Coachella 2019

Fig. 8 Distribution of non-spreader nodes with respect to time for various topics

Figure 9 represent the plot of the total infected or informed count with respect to time for various topics. Here, the infected or informed count includes the number of spreaders and non-spreaders. These figures display the extent of information diffusion as predicted by our proposed model (MFF) in comparison to actual data. The actual data comprises of a number of users who posted a tweet/re-tweet on a particular topic and their respective followers count. Figure 9a shows the plot of infected users with respect to time in hours for the topic Corona Virus or COVID-19. COVID-19 is a global topic affecting the entire world. Thus, the infected count increased rapidly with time. Figure 9b represents the distribution of infected nodes related to the topic FIFA world Cup 2018 with respect to time in days. FIFA is a global topic bringing together football enthusiasts all over the world, thus attracting many tweets. Figure 9c illustrates the distribution of infected users against time in minutes for the topic of the NBA Finals. The infected count increased rapidly with each minute during the match. Figure 9d represents the plot of infected users related to the topic of Me Too Movement with respect to time in minutes. The infected count was low initially but then increased gradually. Figure 9e represents the plot of infected users related to the topic of Game of Thrones Season 8 with respect to time in days. The infected count kept on increasing with time. Figure 9f represents the plot of infected users related to the topic of



(a) Hour-wise distribution of infected count on the topic Corona Virus

(b) Day-wise distribution of infected count on the topic FIFA World Cup 2018

(c) Minute-wise distribution of infected count on the topic NBA finals

(d) Minute-wise distribution of infected count on the topic Me Too Movement

(e) Day-wise distribution of infected count on the topic Game of Thrones

(f) Day-wise distribution of infected count on the topic Coachella 2019

Fig. 9 Distribution of infected or informed nodes with respect to time for various topics

Coachella music and arts festival with respect to time in days. The infected count increased with a low rate initially, that is, a week before the festival but increased rapidly as the festival inched close. The number of tweets was high during the festival. It can be observed from the plots that our proposed MFF algorithm models the process of information diffusion efficiently. The predicted results very well fit the actual data. Also, it can be deduced from the plots that total infected count is much higher than the actual spreaders, thus, validating the fact that even a few people can lead to massive dissemination of information. Our proposed model considers the infected or informed people as the sum of spreaders and non-spreaders. This is because, in the case of online social networks, people can be infected i.e., aware of information even if they are not spreading further. Hence, our model intends to efficiently predict information diffusion in the case of online social networks like Twitter.

7 Conclusion

In this paper, we modeled information diffusion in online social networks using a nature-inspired model named as modified forest-fire model. We exhibited the flow of information in a network using the modified forest-fire model. We divided the entire population into four categories: *Empty*, *Tree*, *Fire*, and *Burnt*. We also segregated the informed population into spreaders and non-spreaders. The study on the spreading of information in real-world data-sets is similar to as illustrated by our model, and thus, information spreads like fire in a network. We performed analysis on a densely connected and popular network of Twitter by using different hashtags. The modified forest-fire model resembles the real-world spreading of information diffusion in highly complex networks like Twitter.

References

Bakshy, E., Rosenn, I., Marlow, C., Adamic, L. (2012). The role of social networks in information diffusion.

Barabási, A.L. (2016). Network science. Cambridge University Press.

Biswas, M.H.A., Paiva, L.T., De Pinho, M.D.R. (2014). A SEIR model for control of infectious diseases with constraints. *Mathematical Biosciences & Engineering*, *11*(4), 761.

Cai, G., Wang, R., Qiang, B. (2012). Online social network evolving model based on damping factor. *Procedia Computer Science*, *9*, 1338–1344.

Chakraborty, A., Dutta, T., Mondal, S., Nath, A. (2018). Application of graph theory in social media. *International Journal of Computer Sciences and Engineering*, *6*, 722–729.

D'Ambrogio, A., Gaudio, P., Gelfusa, M., Luglio, M., Malizia, A., Roseti, C., Zampognaro, F., Giglio, A., Pieroni, A., Marsella, S. (2016). Use of integrated technologies for fire monitoring and first alert. In *2016 IEEE 10th International Conference on Application of Information and Communication Technologies (AICT)* (pp. 1–5): IEEE.

Daley, D.J., & Gani, J. (2001). Epidemic modelling: an introduction (Vol. 15). Cambridge University Press.

de Abreu, L.F. (2019). Game of Thrones S8 (Twitter) 7th April 2019 - 28th May 2019, US, Version 1. https://www.kaggle.com/monogenea/game-of-thrones-twitter.

Ding, C., & Li, K. (2018). Topologically biased random walk for diffusions on multiplex networks. *Journal of Computational Science*, *28*, 343–356.

Ding, C., & Li, K. (2018). Topologically biased random walk for diffusions on multiplex networks. *Journal of Computational Science*, *28*, 343–356.

Fischer, A., Korejwa, A., Koch, J., Spies, T., Olsen, C., White, E., Jacobs, D. (2013). Using the forest, people, fire agent-based social network model to investigate interactions in social-ecological systems. *Practicing Anthropology*, *35*(1), 8–13.

Guille, A., Hacid, H., Favre, C. (2013). Predicting the temporal dynamics of information diffusion in social networks. arXiv:1302.5235.

Guille, A., Hacid, H., Favre, C., Zighed, D.A. (2013). Information diffusion in online social networks: a survey. *ACM Sigmod Record*, *42*(2), 17–28.

Hawley, L.F. (1926). Theoretical considerations regarding factors which influence forest fires. *Journal of Forestry*, *24*(7), 756–763.

He, D., & Liu, X. (2020). Novel competitive information propagation macro mathematical model in online social network. *Journal of Computational Science*, *41*, 101089.

He, Q., Wang, X., Mao, F., Lv, J., Cai, Y., Huang, M., Xu, Q. (2020). CAOM: A community-based approach to tackle opinion maximization for social networks. *Information Sciences*, *513*, 252–269.

Hoang, T.B.N., & Mothe, J. (2018). Predicting information diffusion on Twitter–Analysis of predictive features. *Journal of Computational Science*, *28*, 257–264.

Hu, Y., Aiello, M., Hu, C. (2018). Information diffusion in online social networks: a compilation. *Journal of Computational Science*, *28*, 204–205.

Indu, V., & Thampi, S.M. (2019). A nature-inspired approach based on Forest Fire model for modeling rumor propagation in social networks. *Journal of Network and Computer Applications*, *125*, 28–41.

Kanga, S., & Singh, S.K. (2017). Forest fire simulation modeling using remote sensing & GIS.International Journal of Advanced Research in Computer Science 8 (5).

Kimura, M., Saito, K., Ohara, K., Motoda, H. (2013). Learning to predict opinion share and detect anti-majority opinionists in social networks. *Journal of Intelligent Information Systems*, *41*(1), 5–37.

Kuang, L., Tang, X., Guo, K. (2014). Predicting the times of retweeting in microblogs. Mathematical Problems in Engineering 2014.

Kumar, S., & Panda, B.S. (2020). Identifying influential nodes in Social Networks: Neighborhood Coreness based voting approach. Physica A: Statistical Mechanics and its Applications, pp. 124215.

Liu, Y., Diao, S.M., Zhu, Y.X., Liu, Q. (2016). SHIR Competitive information diffusion model for online social media. *Physica A: Statistical Mechanics and its Applications*, *461*, 543–553.

Mashiach, L.T., & Sharma, A. (2020). Selecting user posts related to trending topics on online social networks. U.S. Patent 10,,535,106.

Nesi, P., Pantaleo, G., Paoli, I., Zaza, I. (2018). Assessing the reTweet proneness of tweets: predictive models for retweeting. *Multimedia Tools and Applications*, *77*(20), 26371–26396.

Newman, M.E.J. (2010). *Networks: An Introduction*, (p. 18). New York: Oxford University Press.

Pattanayak, H.S., Sangal, A.L., Verma, H.K. (2019). Community detection in social networks based on fire propagation. *Swarm and Evolutionary Computation*, *44*, 31–48.

Pesic, P. (2018). Tweets during Cavaliers vs Warriors 3rd game of the 2018 NBA Finals #NBAFInals, Version 24. https://www.kaggle.com/xvivancos/tweets-during-cavaliers-vs-warriors.

Pesic, P. (2019). Coachella 2019 Tweets, Version 2. https://www.kaggle.com/pdp2600/coachella-2019-tweets.

Ramirez, V. (2018). MeToo Dataset. Retrieved from https://data.world/bikthor/metoo.

Rituparna (2018). FIFA World Cup 2018 Tweets, Version 4. https://www.kaggle.com/rgupta09/world-cup-2018-tweets.

Rui, X., Hui, S., Yu, X., Zhang, G., Wu, B. (2018). Forest fire spread simulation algorithm based on cellular automata. *Natural Hazards*, *91*(1), 309–319.

Saito, K., Ohara, K., Kimura, M., Motoda, H. (2015). Change point detection for burst analysis from an observed information diffusion sequence of tweets. *Journal of Intelligent Information Systems*, *44*(2), 243–269.

Sankar, C.P., & Kumar, K.S. (2016). Learning from bees: an approach for influence maximization on viral campaigns. *PloS One*, *11*(12), e0168125.

Smith, S. (2020). Coronavirus (covid19) Tweets-Tweets using hashtags associated with Coronavirus, Version 13. https://www.kaggle.com/smid80/coronavirus-covid19-tweets.

Stai, E., Milaiou, E., Karyotis, V., Papavassiliou, S. (2018). Temporal dynamics of information diffusion in twitter: Modeling and experimentation. *IEEE Transactions on Computational Social Systems*, *5*(1), 256–264.

Wang, X.F., Li, X., Chen, G.R. (2012). *Network Science: an Introduction*, (pp. 10–15). Beijing: Higher Education Press.

Wang, J., & Liu, X. (2011). The improvement of computer algorithm for forest fire model based on cellular automata. In *2011 2nd International Conference on Artificial Intelligence, Management Science and Electronic Commerce (AIMSEC)* (pp. 2567–2570): IEEE.

Zhou, T., Fu, Z., Wang, B.H. (2006). Epidemic dynamics on complex networks. *Prog Natural Science*, *16*, 452–457.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Sanjay Kumar[1,2] · **Muskan Saini**[3] · **Muskan Goel**[4] · **B. S. Panda**[2]

Muskan Saini
muskansaini257@gmail.com

Muskan Goel
muskangoel210499@gmail.coms

B. S. Panda
bspanda@maths.iitd.ac.in

[1] Department of Computer Science and Engineering, Delhi Technological University, Main Bawana Road, New Delhi, 110042, India

[2] Computer Science and Application Group, Department of Mathematics, Indian Institute of Technology Delhi, Hauz Khas, New Delhi, 110016, India

[3] Microsoft India, Hyderabad, Telangana, 500032, India

[4] Microsoft India, Banglore, Karnataka, 560025, India