Intelligent Automation & Soft Computing
DOI: 10.32604/iasc.2023.034019
Article

Identifying Influential Communities Using IID for a Multilayer Networks

C. Suganthini* and R. Baskaran

Department of Computer Science and Engineering College of Engineering, Guindy, Anna University, Chennai, Tamilnadu, India
*Corresponding Author: C. Suganthini. Email: chinna246@gmail.com
Received: 04 July 2022; Accepted: 19 August 2022

Abstract: In online social networks (OSN), they generate several specific user activities daily, corresponding to the billions of data points shared. However, although users exhibit significant interest in social media, they are uninterested in the content, discussions, or opinions available on certain sites. Therefore, this study aims to identify influential communities and understand user behavior across networks in the information diffusion process. Social media platforms, such as Facebook and Twitter, extract data to analyze the information diffusion process, based on which they cascade information among the individuals in the network. Therefore, this study proposes an influential information diffusion model that identifies influential communities across these two social media sites. Moreover, it addresses site migration by visualizing a set of overlapping communities using hyper-edge detection. Thus, the overlapping community structure is used to identify similar communities with identical user interests. Furthermore, the community structure helps in determining the node activation and user influence from the information cascade model. Finally, the Fraction of Intra/Inter-Layer (FIL) diffusion score is used to evaluate the efficiency of the influential information diffusion model by analyzing the trending influential communities in a multilayer network. However, from the experimental result, it observes that the FIL diffusion score for the proposed method achieves better results in terms of accuracy, precision, recall and efficiency of community detection than the existing methods.

Keywords: Influential information diffusion model; community detection; influential communities; social network

1 Introduction

Social media sites such as Facebook, Twitter, and Instagram generate a large amount of data that requires analysis to identify user activities, such as sharing the most popular content, in the social network information diffusion process. However, online user activities are difficult to identify or predict, as they are unstructured and change dynamically daily. Moreover, the big data analytics component of social network analysis (SNA) is in the progressive research phase.

Generally, SNA depends on the properties of the network undergoing analysis. For instance, the analysis of undirected networks requires metrics that use symmetric edges between nodes. In other words, the paths through which information passes within communities can be identified; however, these directional paths cannot be managed. Thus, the analysis of undirected networks uses symmetric relationships between the users, to identify sub-communities and users that are important to those communities. The major contributions of the proposed method are as follows:

- Identifying the user activation in the information intervention phase and evaluating the social influence by assigning thresholds to determine the degree of social contagion.
- Identifying the influenced communities and their user behavior in the influential information diffusion model by predicting the interested communities in a multilayer network.

Section 1 provides a brief introduction to the benefits of analyzing online social networks (OSN) for complex multilayer networks. Section 2 deals with related work on information diffusion and community detection. Section 3 describes the proposed architecture in detail. Section 4 provides the experimental results and evaluation metrics of the proposed method. Finally, Section 5 concludes the studies and presents future research directions.

2 Related Work

Graph illustrates entities and the mutual relationships between these entities in social networks. For instance, a graph G (V, E) represents vertices and edges that connect in a network. The relationship between entities determines the values, such as the common node degree and the typical path range between nodes. Typically, studies in graph mining employ the conceptual information model instead of a mathematical entity [1]. Community structure modeling and its characteristics are also studied using graph mining concepts that focus on node and edge properties.

2.1 Activity Analysis in Social Media

Community structure models can be used to quantitatively determine the possible extent to which social media users are influenced by the opinion or decision of other users within the network [2,3]. A previous study on the influence exerted by network users demonstrated that social interactions occur more frequently among similar individuals than among dissimilar individuals [4]. Another study established that individual participants have the greatest influence on others in social networks in comparison with other sources of influence [5]. In addition, social influence has been defined as the phenomenon that induces interactive behavior based on interactions between two nodes [6,7]. In other words, influential user's exhibit high competence in discriminating between highly cited and less cited articles [8]. Active learning method provide more efficient learning of supervised relation extraction models [9].

2.2 Information Diffusion

Information diffusion predicts user interest based on several aspects, such as information sharing, ideas, and their interests in different OSN [10,11]. In this context, two-phase diffusion proposes an effective algorithm for identifying the individuals in the diffusion model [12]. Strong interaction links show which users were influenced and who motivated their activities based on social interaction. The model determines the source of influence by identifying users with whom the subjects of the analysis have shared more information in a network [13]. Another method of determining interdisciplinary influence is measure based on user connections with neighbors [14].

In the Independent Cascade model, each edge is interconnected with an influence probability that specifies the probability with which the source node influences the target node [15,16]. If the source node successfully influences the neighbor nodes, then the newly activated nodes remain activated in the information diffusion process [17]. Immediately following activation at a particular time step, each node gets exactly one attempt to activate inactive neighbor nodes with a different probability for each neighbor node in the information diffusion process [18,19]. The model represents the interaction link that connects multiple users across the network. Betweenness measures a set of optimal targets for spreading content or information throughout the connected social network [20].

2.3 Community Detection

Community detection is crucial in analyzing the concept of entropy, measures network information and unknown information, such as efficient modules and topological structures in complex networks [21,22]. Furthermore, community detection can be used to identify clusters with closely connected nodes, to ensure that nodes with higher similarity are partitioned into the same group [23,24]. The dynamic principle involved in this type of detection is that a community often comprises several weak cliques instead of cliques in complex networks, especially those networks that lack a clear community structure [25,26]. In general, computation models formalize the individual properties, interaction, and communication between individuals in a dynamic network [27,28].

In OSN, focus must be placed on particular cases of information diffusion, such as portions of information or popular topics that diffuse the most, the prospective path of information diffusion, and the primary members of networks involved in spreading information [29,30]. However, the edges in the inter-community are denser than those in the intra-community edges. Thus, identifying community structures without prior knowledge about the number of communities is difficult [31]. Furthermore, quantifying the strength of social ties and identifying the strong relationships between features and neighbor nodes is even more challenging [32,33].

Most existing studies deal with the neighbor node interactive activities that participate in the information diffusion process. In this diffusion process, information-sharing among the neighbor nodes through node interaction stops as soon as activation ends. Therefore, influential user activities must be identified across the network in the influential information diffusion model. In the information diffusion process, users interact with each other by gaining knowledge and sharing information among themselves, thereby leading to the formation of a multilayer network. In addition, misinformation in the diffusion model and its propagation, which is a deciding factor in content popularity, must be studied. Thus, the influential heterogeneous community structure is identified for a multilayer network.

3 Proposed System

The proposed model comprises four phases: in the first and second phases influential users and their corresponding community structures in Facebook and Twitter are identified. The third and fourth phases are the information intervention and information diffusion phase. Facebook and Twitter network construction includes node identification based on user activity and link creation based on the interactions between them. The hyper-edges that exist between inter-community and intra-community helps to identify influential users in the Facebook community structure by extracting multiple relationships between the users. In the third phase of the information cascade model, input is received from the Facebook community structure. Based on the input, the social contagion score and misinformation diffusivity in the extracted communities are determined, following which the communities are ranked.

The K-clan method identifies influential users in the Twitter community structure which is used as input to the information diffusion phase. In the fourth phase the output of social contagion helps to activate the influential users in the multilayer network and to construct a superimposition network. The influential information diffusion model is thus useful for determining the diffusion of innovation in a multilayer network and identifying the influential communities for a multilayer network. The architecture of the proposed system is shown in Fig. 1.

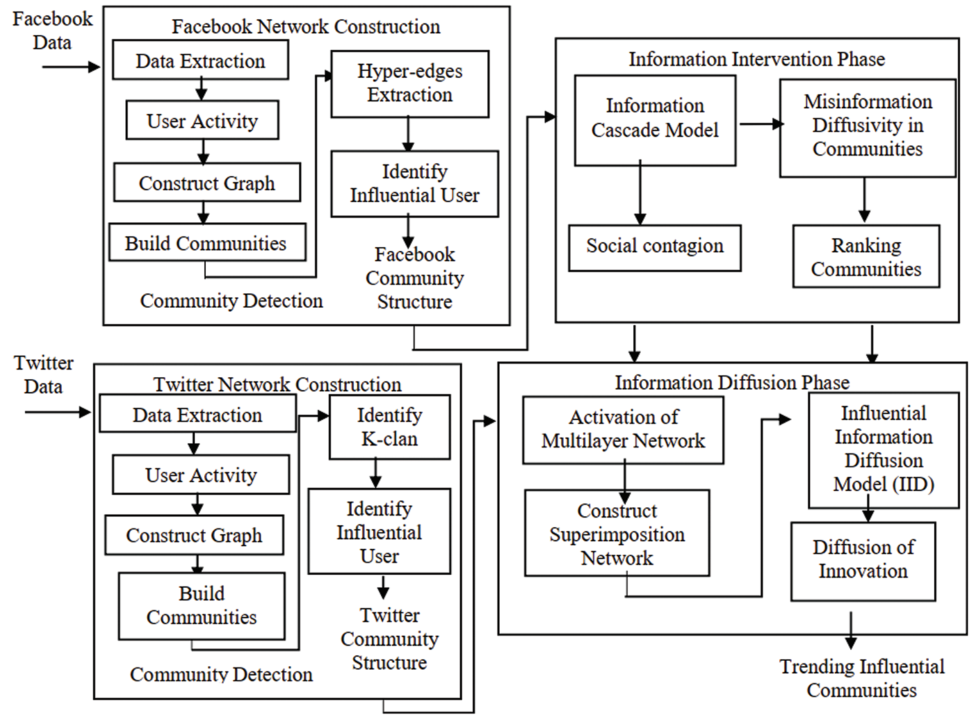

Figure 1: Architecture of the proposed system

3.1 Facebook Network Construction

Facebook data such as likes, shares, comments, and messages are extracted from Facebook in JSON format. Subsequently, the JSON file is converted to a CSV file. The user activity is built from the social graph where the nodes represent the users and the edges represent their activities [34]. Thus, a graph is constructed for the Facebook data and communities identified. The hyper-edge detected in the existing inter-communities and intra-communities is identified. The degree centrality of a user v is defined in terms of the number of incident edges it possesses.

$$C_D(v) = \sum_{i=1}^{n} e(u_i, v_i) \quad (1)$$

where $e(u_i, v_i) = 1$, if the users u_i and v_i are connected, i.e., an edge exists between them
$= 0$, otherwise

The degree centrality calculated for the Facebook network is based on the structural property from Eq. (1). Communities are identified on Facebook based on clusters. A group of users is identified using community detection extraction as shown in Algorithm 1.

Algorithm 1: Community Structure detection

Input: User activity graph of Facebook

Output: Community structures

1. begin community structure detection
2. max_com ← number of communities in fb
3. for each community in max_com
4. community_id ← c_num; // c_num begins from 0

(Continued)

Algorithm 1 (continued)

```
                for each edge in community c_num
                    if community id! = source_id
                        c_num → active users++
                    else
                        c_num → inactive users++
                    increase c_num by 1
                end
            end
            for each community in max_com
                return active users
            end
        end
```

User influence is determined using modularity or community membership. Influential users are identified from the Facebook community structure. Generally, each of these communities mutually overlap depending on the user similarities evaluated, which are described in Section 3.3.

3.2 Twitter Network Construction

Twitter data pre-processing includes the user activities like tweets and re-tweets in JSON format. Subsequently, the JSON file converts into a CSV file. The influential nodes identify by determining the most commonly searched keywords by Facebook users. Finally, identifying the user activities to construct graphs and extract Twitter communities based on similar user activities within the network. However, the outcomes of social interaction depend on all the users [35]. In this context, the number of influential individuals helps to determine shared user interests based on the social interaction between users [36]. Thus, social interaction measures using degree centrality based on the structural property from Eq. (1).

The influential user interactions identify from the Twitter community structure [37]. Social cohesiveness represents the interaction between user activities. To determine social cohesiveness, the maximum number of interactions in the connected networks is extracted using the k-clan method, which is described in Section 3.4.

3.3 Information Intervention Phase

Identifying the most influential spreaders within a social network is a critical task for ensuring the efficient information diffusion process. The purpose of studying information diffusion in dynamic social networks is to present a global view of familiar or popular topics in the future. Thus, the study of these networks considers the activation of multiple layered networks in the process of information diffusion. The Facebook community structure output is taken as the input of the third, or information intervention phase. By determining the probability of social contagion score for inter-community and intra-community structures can be more easily extracted. Finally, the misinformation is extracted and community ranking is determined.

3.3.1 Information Cascade Model

An OSN is assumed to be a closed world where the information spreads because of informational cascades, i.e., the path followed by a portion of information in the network (diffusion graph). In the information spreading process each node is designated as activated or inactivated. The process of propagation, which is observed as a continuous activation of nodes throughout the networks, is called the activation sequence.

Influence Probability $\underline{p_{uv}}$ denotes the probability with which an initial user u influences another user v. The diffusion begins at time step 0, in each time step the user is influenced by the previous time step while attempting to influence their neighbors and succeeds or fails based on the influence probabilities associated with the edges. Subsequently, the influenced neighbors successfully become recently activated users and stay activated for the rest of the diffusion. On activation at a particular time step, a user u has exactly one attempt to activate each of its inactive neighbors with a probability p_{uv} for each neighbor v. The diffusion ends when no further users are activated.

3.3.2 Social Contagion

Social influence and user similarity are the features that represent the user's interest and can be used to predict user behavior in the future. Thus, identifying the contagion in the individual user attitudes that affects social attitudes toward the intervention, can help to identify the social influence [38]. Furthermore, the influencing node is assumed to be the node that can adopt the cascade, which spreads to inactive nodes and activates them. The contagions are those groups of nodes that are not activated and are also closer to the epidemic of the influential nodes [39].

In addition, social contagion identifies inactive users in the network who become activated by re-sharing information and spreads the information within the community structure. This increase in information sharing leads to stronger interaction within a larger group. In other words, social contagion is higher when similar interests are shared in the intra-communities, and lower when dissimilar interests are shared in the inter-communities. Let us assume a complex network with N nodes (users), where each node i represents the activity $x_i(t)$ following the rate equation as

$$\frac{dx_i}{dt} = W(x_i(t)) + \sum_{j=1}^{N} A_{ij} Q\left(x_i(t),\ x_j(t)\right) \tag{2}$$

Eq. (2) represents the dynamical model and gives a general deterministic description for pair-wise interactions. $W(x_i(t))$ represents the node i self-dynamics. $\sum_{j=1}^{N} A_{ij} Q\left(x_i(t),\ x_j(t)\right)$ is the interaction between node i along with its neighbor node. A_{ij} is the adjacency matrix and $Q\ (x_i(t), x_j(t))$ represents the dynamic mechanism of the pair-wise interactions, when x_i provides the corresponding actual meaning. In epidemic processes, x_i represents the contagion probability.

3.3.3 Misinformation Diffusivity

In the case of misinformation diffusion, information does not spread properly across the network and results in the propagation of misinformation, which plays an important role in the diffusion process. Consequently, the absolute quantity of interactions with misinformation remains significant and may not fully capture the trending communities. Therefore, the least square method is used to fit the distribution of misinformation in the communities. It is a statistical procedure to determine the best fit for the set of users who belong to the intra-communities and inter-communities and predict user behavior depending on user activity [40]. To be precise, similar interests will fit in the intra-community, whereas dissimilar interests will fit in the inter-community.

3.3.4 Ranking Communities

The capability of nodes to spread information across the network is ranked based on the neighbor nodes that perform in the diffusion process [41]. The collected misinformation is used to rank them within the communities. Community ranking in the information cascade model is performed to identify the influential information spreader, which also helps to classify influential spreaders in the intervention process.

3.4 Information Diffusion Phase

The community structure output of the second phase is taken as the input for the fourth, or information diffusion phase. The social contagion output helps to activate the node in a multilayer network. Subsequently, community ranking helps to identify the changes in user behavior in the influential information diffusion process.

3.4.1 Activation of Multilayer Network

In multilayer networks, distinct entities are connected via different social interactions. Community detection identifies structurally similar pairs in multilayer networks by grouping them together. It is based on the interaction between distinct entities, which groups the Facebook and Twitter data using a multilayer network model that optimally captures the overlapping community structure in the network. Furthermore, considering that the users adopting information shared in the multilayer network play a significant role in activating the users in the network, the interactions and mutual impact of these unique connections must be faithfully captured. Finally, the Fraction of Intra/Inter-Layer diffusion score is used to evaluate the IID model for a multilayer network, which is discussed with the results in Section 4.3.1.

3.4.2 Construct Superimposition Network

Social influence is wielded by users through social interactions with other participants by sharing information, opinion, or decisions in the network [42]. The social distance value is less for a higher level of social influence and vice-versa. Accordingly, the social distance value will be lower in the intra-community structure, whereas it will be greater in the inter-community structure, which can be attributed to the presence of more influential users with less social distance value in the network. In this situation, a multilayer network is optimal for handling the dynamic network model.

3.4.3 Influential Information Diffusion Model (IID)

The similarity between the two networks is determined using the subgraph of overlapping community structures in Twitter and Facebook. Influential nodes are detected using fixed attributes that belong to user activities in the social network. The top-k influential node is typically the one that spreads the most information in each community [43]. Thus, the IID model visualizes the community structure for multiple networks. The proposed IID model spreads the influential information, which helps to identify the influential communities in a multilayer network. The diffusion degree centrality of a user v is defined as

$$C_{DD}(v) = \sum\nolimits_{u\epsilon\Gamma(v)} \left\{ \lambda_{u,v} + \left(\lambda_{u,v} x \sum\nolimits_{i\epsilon\Gamma(v)} \lambda_{i,v} \right) \right\} \tag{3}$$

where $\Gamma(v)$ denotes the neighbor set of *v* and $\lambda_{u,v}$ denotes the propagation probability of user v influencing user u. Eq. (3) incorporates the property of information diffusion along with structural information.

3.4.4 Diffusion of Innovation

The influence of social consensus information can change individual preferences with respect to mingling with minority group members, even beyond the intervention phase [44]. Furthermore, for an innovation to be adopted, it must have certain qualities [45]. Thus, social consensus information influences the individual's exact opinions through their attitudes [46].

The diffusion of innovation model identifies trending communities, thereby determining the changes in user interest across the networks. The rate depends on the performance of the spread of innovative information and affects the potential user spread that has not yet influenced the spreader. The rate at which the number of adopters changes with time in given in Eq. (4)

$$\frac{dA(t)}{dt} = i(t)[P - A(t)] \tag{4}$$

where A(t) is the total number of users that adopted the innovation until time t, i(t) is the coefficient of diffusion innovativeness of information spread, and P denotes the number of potential users.

The challenge in visualizing multilayer graphs is that multiple edges between two nodes may be plotted atop each other, thereby making them impossible to be discerned [47,48]. Initially, a community detection graph is constructed for the two social media sites. Thus, multilayer network visualization can be achieved for both the Facebook and Twitter data. The influential users are used to characterize the behavior of information spread within the network. Furthermore, the influential trending communities are identified from the shared information of the influential users in the multilayer network.

4 Results

The experiment was performed using the proposed method over two real-world datasets, which was subsequently compared with six existing methods: Social Influence Model (SI), Susceptible View Forward Removed Model (SVFR), Susceptible/Infective/Recovered (SIR) model, Fully Adaptive Cross Entropy Method (FACE), Hydrodynamic information diffusion prediction (Hydro-IDP) and Unknown-View-Share-Removed (UVSR) model. The results show that the proposed method exhibited higher accuracy compared with the existing approaches.

4.1 Experiment Setup

4.1.1 Dataset Description

In this section, the proposed system is evaluated and experimental results for the popular social sites are discussed. Social media users typically interact with each other by sharing or exchanging information. Thus, data from Facebook and Twitter are using extracted using graph API Streams. Not only are these API streams valid as common data for pre-processing, they can also be effectively interpreted to achieve graph data. User influence in the social network is analyzed in terms of participation, content sharing, popularity, and activity. The statistical properties of datasets are summarized in Table 1.

Table 1: Features of datasets

Property	Facebook	Twitter
Nodes	89527	108493
Edges	145772	1048576
Degree centrality $C_D(v)$	10.958	4.491
Diffusion degree centrality $C_{DD}(v)$	254.779	123.197
Avg. clustering coefficient	0.068	0.787
Modularity	0.7535	0.8639
No. of communities	38	45
Trending communities	25	36
Information cascade	0.678	0.7438

4.1.2 Comparison of Proposed Method with Existing Methods

The proposed IID model was compared with six existing methods: SI, SVFR, SIR, FACE, Hydro-IDP and UVSR. The descriptions of the implementation details of these methods are summarized below.

IID: The influential communities across the network are identified based on the information spread among the communities. Information diffusion plays the major role in identifying the influential interaction. The fraction of intra/inter layer (FIL) helps in measuring the efficiency of the model.

SI [6]**:** The interactive and non-interactive activities are identified using the top k-influence ranks based on node selection.

SVFR [7]**:** The three types of user reactions to a message are view, ignore, or forward. The view or forward probability is determined based on the content. However, this method requires a reduction in the time delay in information diffusion.

SIR [10]**:** An EM algorithm is used to predict the diffusion probabilities. However, this method requires an improvement of its propagation into the independent cascade model.

FACE [17]**:** The golden selection search algorithm is applied under moderate temporal constraints. This method requires a reduced time taken to spread influence information.

Hydro-IDP [26]**:** The characteristics of information diffusion are extracted to describe and predict the spreading process of information in OSN. However, the influence and diffusivity on social platforms should be improved.

UVSR [48]**:** A continuous-time, stochastic model is used in this method to characterize the information diffusion process and understand the topological features and temporal dynamics of information diffusion. In this method, the delay probability of diffusion and speed in viewing and sharing must be improved.

4.2 Experiment Results

This section presents the evaluation of the community detection, influence of information spread, and identification of the trending communities in the social network. The proposed method was applied to two real-world datasets to analyze its effectiveness from the aspects of influence information spread.

4.2.1 Community Detection

In a social graph, nodes or vertices are described as users or actors and the association between nodes are represented by the activity links or edges between them. Initially, Facebook and Twitter communities are identified in the social graph, and the overlap between communities is determined to identify the influence of persons or groups of persons. The user interaction reveals a friendly relationship and forms an overlapping community structure between Facebook and Twitter data.

Modularity metrics are used to evaluate the community detection capabilities of multilayer networks. The value of modularity Q is defined as shown in Eq. (5).

$$Q = \sum_{i}^{k}\left(e_{ii} - \left(\sum_{i}^{k} e_{ij}\right)^{2}\right) \tag{5}$$

where k denotes the number of communities, e_{ii} is the ratio of the number of edges within the community i to the total number of edges in the entire network and e_{ij} is the ratio of the number of edges between communities i and j to the total number of edges in the entire network [49,50]. The different users identified are potential users, adopted users, and influential users in the information diffusion process for a multilayer network. The modularity Q value is calculated for a multilayer network based on the different users in the communities, which is shown in Fig. 2.

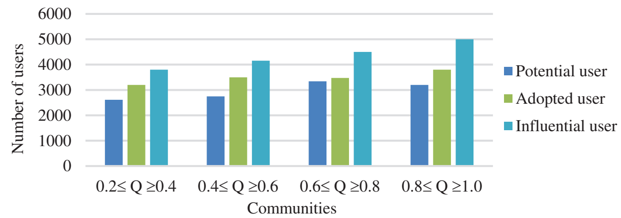

Figure 2: Communities comparison based on different networks

4.2.2 Influence Information Spread

In the IID model, active nodes are considered as senders, and the nodes being activated are considered as receivers. On activation, each node has one attempt at activating each of its neighbor nodes in the social graph. A node u has a random threshold $\theta_u \sim U$ [0, 1]. Each Neighbor node v influences node u according to a degree centrality $C_{(D)}u$, v as given by Eq. (6)

$$\text{Influence node} = \sum\nolimits_{\text{v neighbor of u}} C_{(D)}u,\ v \leq 1 \tag{6}$$

A node u becomes active when the fraction of its active neighbors is at least θ_u, as given by Eq. (7)

$$\text{Active node} = \sum\nolimits_{\text{v active neighbor of u}} C_{(D)}u,\ v \geq \theta_u \tag{7}$$

The spread of information begins with a collection of active nodes and continues until no further activation of nodes is possible. When a user u becomes activated at a particular time step, it has exactly one attempt to activate each of its inactive neighbors with the probability p_{uv} for each neighbor v. The diffusion ends when there are no further users that can be activated. Depending on the number of nodes interacting in the network, the degree centrality and active nodes increase, as shown in Table 2.

Table 2: Spreading the influence information

No. of nodes	Random threshold θ_u	Influenced node $C_{(D)}u$, v	Active node $C_{(D)}u$, v
100	0.5743	0.6503	0.6235
500	0.6543	0.7245	0.7054
700	0.7245	0.7525	0.7451
1000	0.7769	0.7914	0.7854
2500	0.8214	0.8547	0.8375
3000	0.8545	0.8853	0.8742
5000	0.8651	0.9176	0.8963

In social networks, an active node denotes that the node was selected to spread the behavior, innovation, or decision. Based on the social influence effect, information can spread across the network through the principles of herd behavior and informational cascade [51,52]. Some topics can become quite popular, spread worldwide, and contribute to new trends.

Finally, the components of an information diffusion method practiced in an OSN can be similar to a discussion of information carried by messages that spread along the edges of the network according to a particular mechanism. The interaction based on specific properties depends on the edges and nodes in the social network [53]. For instance, the most relevant recent activity as well as the weaknesses, strengths, and improvements for each feature must be analyzed, as shown in Table 3.

Table 3: Summary of information diffusion w.r.t information used for feature modeling

Features	Doo et al. [6]	Liu et al. [7]	Saito et al. [10]	Dhamal et al. [17]	Hu et al. [26]	Liu et al. [48]	Proposed
Diffusion	√	-	√	-	√	√	√
Network connection	√	√	√	-	√	√	√
User activities	-	√	-	√	-	√	√
Time delay	-	-	-	√	√	√	-
Hashtags, URL's mentions	√	-	-	-	√	-	√
Topic information	√	√	√	-	-	-	-
Content based similarities	√	-	√	√	-	√	√
Facebook dataset	√	√	√	√	-	√	√
Twitter dataset	-	√	-	-	√	-	√
Output Type (P/C) (probabilistic/Classifications)	C	P	P	C	C	P	P

4.2.3 Trending Communities in Multilayer Networks

Identifying the numerous influential spreaders in a social network is critical for ensuring the efficient diffusion of information. For instance, a social media campaign can improve efficiency by targeting the influential individuals who can initiate huge information cascades that will result in more adoptions. The output of user interested communities or the trending communities are shown in Table 1. The communities are ranked based on the extraction of misinformation in the communities, as shown in Table 4.

Table 4: Communities score based on the ranking

Communities	Category	No. of users in community	Avg. information cascade	Avg. size	Score	Rank
C12	Environment	1988	0.854	173.16	9.457	1
C7	News	1875	0.825	117.62	9.296	2
C5	Sports	1655	0.793	99.18	8.547	3
C15	Disease	1548	0.764	91.59	8.123	4
C1	Education	1384	0.748	79.34	7.747	5
C3	Politics	1158	0.721	67.63	7.475	6
C8	Product	1058	0.714	62.91	7.245	7
C11	Music	984	0.685	54.97	6.874	8
C13	Movie	824	0.667	51.81	6.425	9
C2	Entertainment	753	0.653	46.83	6.157	10

(Continued)

Table 4 (continued)

Communities	Category	No. of users in community	Avg. information cascade	Avg. size	Score	Rank
C9	Food	633	0.648	40.42	5.925	11
C14	Hotels	599	0.625	35.03	5.783	12
C6	Game	524	0.591	30.15	5.642	13
C4	Travel	453	0.573	25.94	5.124	14
C10	Photography	409	0.543	23.45	4.784	15

The top five ranked communities are C12, C7, C5, C15, and C1 and the categories they belong to are environment, news, sports, disease, and education respectively. These are the topics on which significant diffusion of information occurs. Based on the information cascade that occurs in each of these communities, the score is calculated and the communities are ranked.

The analysis of community interaction behavior shows that the users who join communities determine the factors that are shared among them. These new influential users form the larger group that was analyzed for influential community structure behavior. The trending communities for the multilayer networks of Facebook and Twitter are shown in Fig. 3.

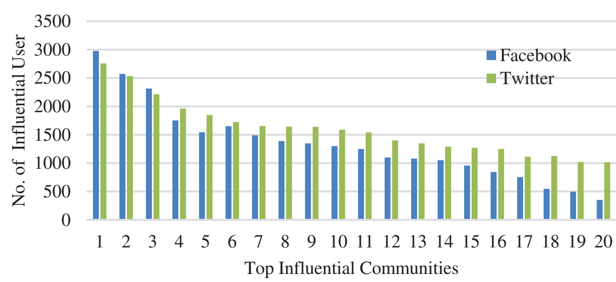

Figure 3: Trending communities for different networks

4.3 Evaluation Metrics

In this section, the ground truth of two real-world datasets is used to evaluate the quality of influence information spread using the evaluation metrics which are discussed below.

4.3.1 Fraction of Intra/Inter-Layer (FIL)

The FIL is used to measure the spreading of an information cascade within or between the layers of a multilayer network. The FIL score denotes the fraction of user interaction in the diffusion network and the average rate of information cascade over the different layers. Following this, the probability of information diffusion was applied to calculate the FIL diffusion score of information spread from one user to another in the multilayer network. Thus, the trending influential communities are identified using the IID model, and the FIL diffusion score is used to measure the efficiency of the IID model on a different social network. The precision, recall, accuracy, and F-measures evaluated based on the FIL diffusion score are shown in Table 5.

Table 5: Different possibilities for information spreading from one layer to another in a multilayer network

	Same user	Different user
Same layer	Information spreads to the same user on the same layer (**SLSU**)	Information spreads to a different user on the same layer (**SLDU**)
Different layer	Information spreads to the same user on a different layer (**DLSU**)	Information continues spreading to a different user on a different layer (**DLDU**)

Note: SLSU = Information spread in a similar layer with similar users leads to a correct positive prediction.
SLDU = Information spread in a similar layer with dissimilar users leads to an incorrect negative prediction.
DLSU = Information spread in different layers with similar users leads to an incorrect positive prediction.
DLDU = Information spread in different layers with dissimilar users leads to a correct negative prediction.

Precision. The ratio of the number of similar users in the same layer to the total number of users. It is also called a positive predictive value. The precision value is calculated using Eq. (8).

$$\text{Precision } = \frac{\sum \text{SLSU}}{\left(\sum \text{SLSU } + \sum \text{SLDU}\right)} \tag{8}$$

Recall. The ratio of the number of similar users in the same layer to the total number of users in the different layers. It is also called a true positive rate. The recall value is calculated using Eq. (9).

$$\text{Recall } = \frac{\sum \text{SLSU}}{\left(\sum \text{SLSU } + \sum \text{DLSU}\right)} \tag{9}$$

Accuracy. The ratio of correctly predicted observations referred to as users in a similar layer to the total observation referred to as the total number of users in different layers. The accuracy value is calculated using Eq. (10).

$$\text{Accuracy } = \frac{\sum \text{SLSU} + \ \sum \text{DLDU}}{\sum \text{SLSU} + \ \sum \text{SLDU} + \ \sum \text{DLSU} + \ \sum \text{DLDU}} \tag{10}$$

F measure. The F1 score is used to consolidate precision and recall into one measure; the F1 measure is calculated using Eq. (11).

$$\text{Fmeasure} = 2 \times \frac{\text{Precision} * \ \text{Recall}}{\text{Precision } + \text{ Recall}} \tag{11}$$

The fraction of Intra/Inter-Layer diffusion score is determined using Eqs. (8)–(11) as shown in Fig. 4.

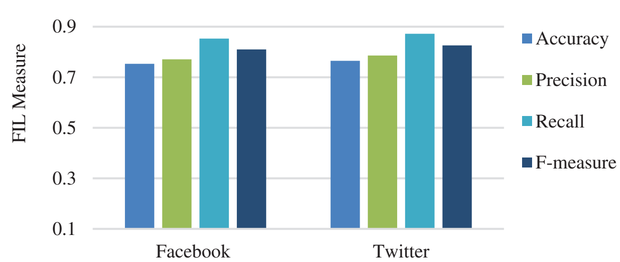

Figure 4: FIL diffusion score for a multilayer network

4.3.2 Overall Performance Metrics

The overall performance in terms of precision, recall, and F-measure for the proposed method compared with the six existing methods was evaluated based on the FIL diffusion score, as shown in Fig. 5.

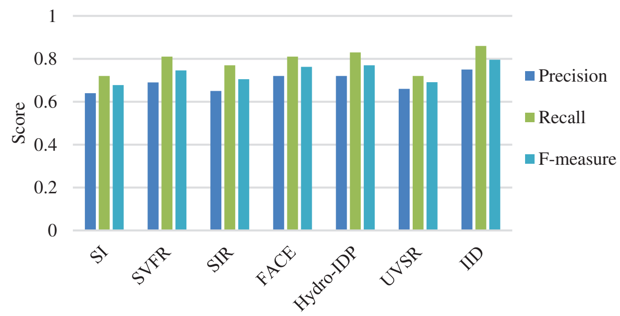

Figure 5: FIL diffusion score for different methods

5 Conclusion

This paper proposes an influential information diffusion model that can be used to analyze the activities of social media users and their influence on the user's timeline across the network. Typically, social contagion induces the spread of information within the community structure, thereby resulting in strong interactions within social groups. An important step in the information diffusion process involves predicting user behavior by identifying influential communities across several networks. Experimental results show that the proposed method effectively identifies the influencing community structure extracted for real-world data. Thus, the FIL can efficiently evaluate the IID model from one layer to another in the social network. Consequently, the influential community structure across the network can be achieved in multilayer networks as well. From this perspective, dynamic OSN are an interesting field of study that can reveal the trends involved in user interactions, which change over a period. In future work, temporal graph analysis can be used to achieve more efficient graph mining techniques. Geo location and time factor consideration is important such as in the case of the temporal graph analysis in multilayer networks. Therefore, future studies should ideally aim to identify the heterogeneous community structure based on user interest and predict future user behavior with the time factor and geo location for dynamic social networks.

Funding Statement: This publication is an outcome of the R & D work undertaken in the project under the Visvesvaraya Ph.D. Scheme of the Ministry of Electronics and Information Technology, Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia).

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

[1] D. F. Nettleton, "Data mining of social networks represented as graphs," *Computer Science Review*, vol. 7, no. 1, pp. 1–34, 2013.

[2] A. T. Stephen and O. Toubia, "Explaining the power-law degree distribution in a social commerce network," *Social Networks*, vol. 31, no. 4, pp. 262–270, 2009.

[3] W. Yang, G. Wang, M. Z. A. Bhuiyan and K. K. R. Choo, "Hypergraph partitioning for social networks based on information entropy modularity," *Journal of Network and Computer Applications*, vol. 86, pp. 59–71, 2017.

[4] T. Pozzoli and G. Gini, "Friend similarity in attitudes toward bullying and sense of responsibility to intervene," *Social Influence*, vol. 8, no. 2–3, pp. 161–176, 2013.

[5] A. Guille, H. Hacid, C. Favre and D. A. Zighed, "Information diffusion in online social networks," *ACM SIGMOD Record*, vol. 42, no. 2, pp. 17–28, 2013.

[6] M. Doo and L. Liu, "Extracting top-k most influential nodes by activity analysis," *Proceedings of the 2014 IEEE 15th International Conference on Information Reuse and Integration (IEEE IRI 2014)*, vol. 1, pp. 227–236, 2014.

[7] L. Liu, B. Qu, B. Chen, A. Hanjalic and H. Wang, "Modelling of information diffusion on social networks with applications to WeChat," *Physica A: Statistical Mechanics and Its Applications*, vol. 496, pp. 318–329, 2018.

[8] S. -U. Hassan, T. D. Bowman, M. Shabbir, A. Akhtar, M. Imran *et al.,* "Influential tweeters in relation to highly cited articles in altmetric big data," *Scientometrics*, vol. 119, no. 1, pp. 481–493, 2019.

[9] H. Sun and R. Grishman, "Employing lexicalized dependency paths for active learning of relation extraction," *Intelligent Automation & Soft Computing*, vol. 34, no. 3, pp. 1415–1423, 2022.

[10] K. Saito, R. Nakano and M. Kimura, "Prediction of information diffusion probabilities for independent cascade model," *Knowledge-Based Intelligent Information and Engineering Systems, KES 2008, Lecture Notes in Computer Science*, vol. 5179, pp. 67–75, 2008.

[11] D. Varshney, S. Kumar and V. Gupta, "Predicting information diffusion probabilities in social networks: A Bayesian networks based approach," *Knowledge-Based Systems*, vol. 133, pp. 66–76, 2017.

[12] S. Mondal, S. Dhamal and Y. Narahari, "Two-phase influence maximization in social networks with seed nodes and referral incentives," *Proceedings of the Eleventh International AAAI Conference on Web and Social Media (ICWSW, 2017)*, vol. 11, no. 1, pp. 620–623, 2017.

[13] E. Ferrara, "A large-scale community structure analysis in facebook," *EPJ Data Science*, vol. 1, no. 9, pp. 1–30, 2012.

[14] X. Wu and C. Zhang, "Finding high-impact interdisciplinary users based on friend discipline distribution in academic social networking sites," *Scientometrics*, vol. 119, no. 2, pp. 1017–1035, 2019.

[15] Y. Lin and J. C. S. Lui, "Analyzing competitive influence maximization problems with partial information: An approximation algorithmic framework," *Performance Evaluation*, vol. 91, pp. 187–204, 2015.

[16] S. Kundu and S. K. Pal, "Deprecation based greedy strategy for target set selection in large scale social networks," *Information Sciences*, vol. 316, pp. 107–122, 2015.

[17] S. Dhamal, K. J. Prabuchandran and Y. Narahari, "Information diffusion in social networks in two phases," *IEEE Transactions on Network Science and Engineering*, vol. 3, no. 4, pp. 197–210, 2016.

[18] J. Jankowski, P. Brodka, P. Kazienko, B. K. Szymanski, R. Michalski *et al.,* "Balancing speed and coverage by sequential seeding in complex networks," *Scientific Reports*, vol. 7, no. 891, pp. 1–13, 2017.

[19] G. D. Angelo, L. Severini and Y. Velaj, "Selecting nodes and buying links to maximize the information diffusion in a network," *42nd International Symposium on Mathematical Foundations of Computer Science (MFCS 2017)*, pp. 1–14, 2017.

[20] A. V. Mantzaris, "Uncovering nodes that spread information between communities in social networks," *EPJ Data Science*, vol. 3, no. 26, pp. 1–17, 2014.

[21] Y. Li, G. Zhang, Y. Feng and C. Wu, "An entropy-based social network community detecting method and its application to scientometrics," *Scientometrics*, vol. 102, no. 1, pp. 1003–1017, 2015.

[22] C. Suganthini and R. Sridhar, "A survey on community detection in social network analysis," *International Journal of Applied Engineering Research*, vol. 10, no. 75, pp. 1–6, 2015.

[23] Y. Gandica, A. Decuyper, C. Cloquet, I. Thomas and J. C. Delvenne, "Measuring the effect of node aggregation on community detection," *EPJ Data Science*, vol. 9, no. 6, pp. 1–16, 2020.

[24] Y. Li, X. Zhao and Z. Qu, "A dynamic programming framework for large-scale online clustering on graphs," *Neural Processing Letters*, vol. 52, no. 4, pp. 1613–1629, 2020.

[25] X. Zhang, C. Wang, Y. Su, L. Pan and H. -F. Zhang, "A fast overlapping community detection algorithm based on weak cliques for large-scale networks," *IEEE Transactions on Computational Social Systems*, vol. 4, no. 4, pp. 218–230, 2017.

[26] Y. Hu, R. J. Song and M. Chen, "Modeling for information diffusion in online social networks via hydrodynamic," *Special Section on Socially Enabled Networking and Computing, IEEE Access*, vol. 5, pp. 128–135, 2017.

[27] M. Pellert, S. Schweighofer and D. Garcia, "The individual dynamics of affective expression on social media," *EPJ Data Science*, vol. 9, no. 1, pp. 1–14, 2020.

[28] A. Said, T. D. Bowman, R. A. Abbasi, N. R. Aljohani, S. -U. Hassan *et al.,* "Mining network-level properties of twitter altmetrics data," *Scientometrics*, vol. 120, pp. 217–235, 2019.

[29] M. Farajtabar, Y. Wang, M. G. Rodriguez, S. Li, H. Zha *et al.,* "COEVOLVE: A joint point process model for information diffusion and network Co-evolution," *Journal of Machine Learning Research*, vol. 18, no. 1, pp. 1305–1353, 2017.

[30] A. Vogelsang and S. Fuhrmann, "Why feature dependencies challenge the requirements engineering of automotive systems: An empirical study," *2013 21st IEEE International Requirements Engineering Conference (RE)*, pp. 267–272, 2013.

[31] H. Lu, Q. Zhao, X. Sang and J. Lu, "Community detection in complex networks using nonnegative matrix factorization and density-based clustering algorithm," *Neural Processing Letters*, vol. 51, pp. 1731–1748, 2020.

[32] J. Urena-Carrion, J. Saramaki and M. Kivela, "Estimating tie strength in social networks using temporal communication data," *EPJ Data Science*, vol. 9, no. 37, pp. 1–20, 2020.

[33] H. Feng, J. Tian, H. J. Wang and M. Li, "Personalized recommendations based on time-weighted overlapping community detection," *Information and Management*, vol. 52, no. 7, pp. 789–800, 2015.

[34] R. Sridhar, A. Kumar, S. B. Roshini, R. K. Sundaresan and S. Chinnasamy, "Feature based community detection by extracting facebook profile details," *ICTACT Journal on Soft Computing*, vol. 8, no. 4, pp. 1706–1713, 2018.

[35] N. J. V. Doesum, J. C. Karremans, R. C. Fikke, M. A. de Lange and P. A. M. V. Lange, "Social mindfulness in the real world: The physical presence of others induces other-regarding motivation," *Social Influence*, vol. 13, no. 4, pp. 209–222, 2018.

[36] M. Huang, G. Zou, B. Zhang, Y. Gan, S. Jiang *et al.,* "Identifying influential individuals in microblogging networks using graph partitioning," *Expert Systems with Applications*, vol. 102, pp. 70–82, 2018.

[37] C. Suganthini and R. Baskaran, "Identifying the influential user based on user interaction model for twitter data," *Advances in Signal Processing and Intelligent Recognition Systems, SIRS 2019, Communications in Computer and Information Science*, Springer, vol. 1209, pp. 48–63, 2020.

[38] R. M. Bond, "Contagion in social attitudes about prejudice," *Social Influence*, vol. 13, no. 2, pp. 104–116, 2018.

[39] G. Berry, C. J. Cameron, P. Park and M. Macy, "The opacity problem in social contagion," *Social Networks*, vol. 56, pp. 93–101, 2019.

[40] A. Zareie and A. Sheikhahmadi, "A hierarchical approach for influential node ranking in complex social networks," *Expert Systems with Applications*, vol. 93, pp. 200–211, 2018.

[41] D. Gowsikhaa, S. Abirami and R. Baskaran, "Automated human behavior analysis from surveillance videos: A survey," *Artificial Intelligence Review*, vol. 42, no. 4, pp. 747–765, 2014.

[42] G. Song, Y. Li, X. Chen, X. He and J. Tang, "Influential node tracking on dynamic social network: An interchange greedy approach," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, no. 2, pp. 359–372, 2017.

[43] P. Kim and S. Kim, "Detecting community structure in complex networks using an interaction optimization process," *Physica A: Statistical Mechanics and Its Applications*, vol. 465, pp. 525–542, 2017.

[44] F. Jansson and M. Bursell, "Social consensus influences ethnic diversity preferences," *Social Influence*, vol. 13, no. 4, pp. 192–208, 2018.

[45] L. Lorincz, J. Koltai, A. F. Gyor and K. Takacs, "Collapse of an online social network: Burning social capital to create it?" *Social Networks*, vol. 57, pp. 43–53, 2019.

[46] K. Kobayashi, "Communicating highly divergent levels of scientific and social consensus: Its effects on people's scientific beliefs," *Social Influence*, vol. 14, no. 1, pp. 1–12, 2019.

[47] J. F. Sanchez-Rada and C. A. Iglesias, "Social context in sentiment analysis: Formal definition, overview of current trends and framework for comparison," *Information Fusion*, vol. 25, pp. 344–356, 2019.

[48] L. Liu, B. Chen, B. Qu, L. He and X. Qiu, "Data driven modeling of continuous time information diffusion in social networks," *2017 IEEE Second International Conference on Data Science in Cyberspace (DSC)*, vol. 2, pp. 655–660, 2017.

[49] P. Chopade and J. Zhan, "Structural and functional analytics for community detection in large-scale complex networks," *Journal of Big Data*, vol. 2, no. 11, pp. 1–28, 2015.

[50] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Physics Reports*, vol. 659, pp. 1–44, 2016.

[51] N. Meghanathan, A. Essien and R. Lawrence, "A two-hop neighbor preference-based random network graph model with high clustering coefficient for modeling real-world complex networks," *Egyptian Informatics Journal*, vol. 22, no. 3, pp. 389–400, 2021.

[52] F. Altunbey and B. Alatas, "Overlapping community detection in social network using parliamentary optimization algorithm," *International Journal of Computer Networks and Applications*, vol. 2, no. 1, pp. 12–19, 2015.

[53] G. K. Orman, V. Labatut and H. Cherifi, "Comparative evaluation of community detection algorithms: A topological approach," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 8, pp. 1–16, 2012.