



# Identifying the Influential User Based on User Interaction Model for Twitter Data

C. Suganthini and R. Baskaran<sup>(✉)</sup>

Department of Computer Science and Engineering, College of Engineering, Anna University,  
Chennai, Tamil Nadu, India

chinna246@gmail.com, baaski@annauniv.edu

**Abstract.** In the social network, the user Interaction model is used to measure the interest depending on the interaction behavior between different entities. An entity or node can be people, groups or organizations and the links or edges are shown to represent the relationship between them. Edges are used to identify whether there is communication or social interaction between different users. Based on the user's interest or activities made in a social group, it is used to identify whether the interaction in the network is active or inactive. Social media is one of the fast-growing, dynamic and unpredictable in detecting future influencers which becomes harder in the social group. The features identified from the user interaction model and thus the potential influencers and current influencers are identified. Incidence matrix is used to identify the interaction behavior between different users in a social group. Thus, the incidence edges play an essential role in the vertex-edge incidence matrix interaction to identify the active and inactive user interactions in a social network. The Fraction of Strongly Influential (FSI) users score is used to evaluate the user interaction model by analyzing the influential user's interest and thus the top trending user interests are achieved.

**Keywords:** Feature extraction · Influential user interaction and Edge Incidence matrix interaction · Fraction of Strongly Influential users

## 1 Introduction

In a social network sites where one connects with those sharing personal or professional interest in it. Some of the social network sites such as Facebook, Twitter, and LinkedIn which allow users to create or join groups of common interest where they can do some of the activities such as post, comments, photos and videos can be shared and discussion can take place. Social media has become a key media for sharing their opinions, interest and personal information such as messages, photos, videos, news, etc. Users spend several hours on social media daily. Similar to influence analysis, researchers also evaluate information diffusion models on social media to identify key users who increase the diffusion of information. In this work, social influence defines as the effect of users on others that results in sharing information, which is the most common definition of social influence. The users tend to be influential experts on specific topics such as sports, economy, politics, rather than being global experts. This result leads us to explore

more on user's activities such as sharing information on their interests. In the proposed approach, the features are related to user activities such as ideas, opinions, interest, and thoughts. In this paper, the influential user interaction model of user-related features is introduced. These user's features are classified into core interest and marginal interest of users in the user interaction model. The proposed model is applied in a distributed manner to efficiently analyze the influence of the user's interest.

One of the most important aspects of social media is that it is so dynamic and fast, users and relations may appear and disappear very frequently. It is important to make predictions about who may be influential in the future rather than only identifying current influencers. In this proposed method, the influential users are identified, who would spread the information more, on specific topics. To believe that these users tend to be topic experts and those members are observed as experts on a few numbers of topics.

The remaining part of this paper is organized as follows. In Sect. 2 the existing literature related to the research topic is reviewed. The proposed user influential model based on the vertex-edge incidence matrix for identifying influential users in the social group is presented in Sect. 3. Section 4 describes the experimental data and evaluation results are discussed. Finally, Sect. 5 exposes the conclusion by giving future research challenges.

## 2 Related Work

### 2.1 Influential User

David et al. [1] introduce the two basic diffusion models is adopted for spreading an idea or innovation and influence among its members through a social network. Understanding the behavior and comparing its performance for identifying influential individuals. Wei et al. [2] identify the problem to select the initial user who influences the large number of people in the network, i.e. finding the influential individuals in a social network. Michael et al. [3] developed an approach to determine which users have significant effects on the activities of others using the longitudinal records of member's log-in activity. This approach identifies the specific users who most influence the site activity of others. Amit et al. [4] propose the propagation model to maximize the expected influence spread nodes that eventually get activated.

### 2.2 User Interaction

Francesco [5] proposed a framework to analyze the interaction between social influence and homophily. The features analyzed from social relations and similarity is more important to predict future behavior than either the social influence or similarity feature. Fangshuang et al. [6] Diversified influence maximization which is performed to help the task such as recommending a set of movies with different user's interest and finding a team of experts (answerers) with a comprehensive scope of knowledge. Qindong et al. [7] propose the user interaction process which is one of the most important features in analyzing user behavior, information spreading model, etc. Li et al. [8] propose a framework based on frequent pattern mining to find the influence users as well as the proper

time to spread information. To identify influential bloggers to set collective statistics and find users.

Fredrik et al. [9] proposed work analyzes the influential users and their behavior to predict user participation in the social network. Wang et al. [10] propose a dynamic regional interaction model is to evaluate the influential user identification in understanding the information dissemination process according to the interactions between adjacent and non-adjacent users in online social networks. Jiyoung et al. [11] propose topic diffusion in web forms using the epidemic model where the user in online communities spread the diffusion information. Amrita et al. [12] propose the influence of users through the information spreading and diffusion region. Thus finds the number of active users is influenced by the top influential user with higher network reachability holds the larger diffusion in the network region.

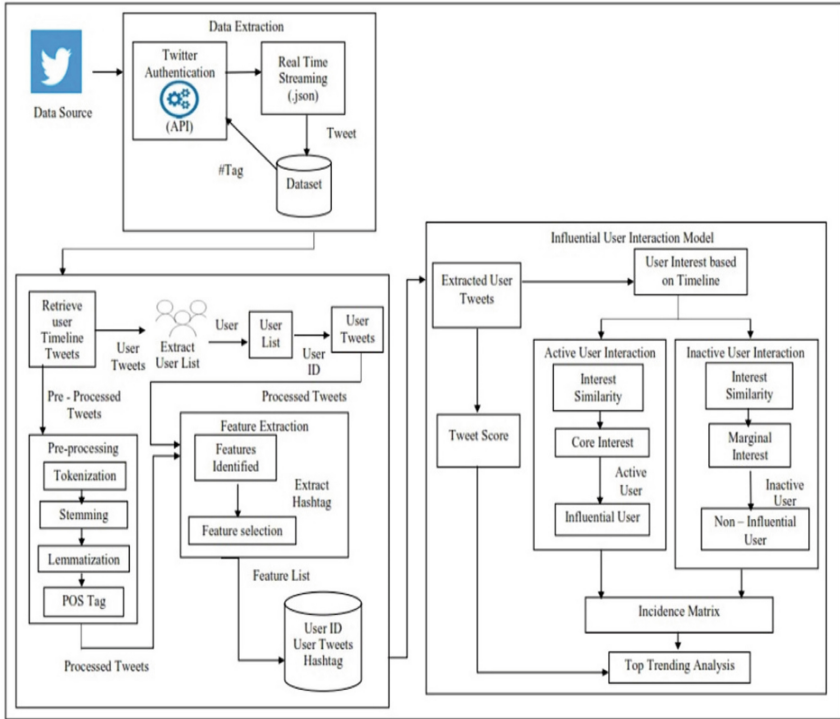
Wenzheng et al. [13] propose an influence propagation process the social decision of a user depends more on the network structure and the social structures of different groups of its neighbors rather than the number of its neighbors. Madhura et al. [14] propose a technique to identify the behavior and opinion of influential users in a social network. Yuchen et al. [15] propose the diffusion model and the influence spread under the model. The information diffusion process identifies the  $k$  users to maximize the expected number of influenced users in the social network. Mohammed et al. [16] propose an algorithm for the top  $k$ -influential users by the selection process. Identification of such influential users helps to identify, understand and discover the underlying interactions of interesting users in the network. Bo et al. [17] propose the most influential node discovery method for discovering influential nodes in the social network to identify the maximum influential node. Zeynep et al. [18] proposed approach include identification authorities on network is used to maximize the spread of information and to recommend other users.

Most of the existing works does not handle the dynamic interaction of users in social network sites. Therefore it is essential to adopt the user interaction model approach which helps in identifying the influential individuals interacted during their time interval in the social network. Incidence matrix interaction is an efficient method which plays an important task in analyzing the influential and non-influential individuals. Thus the influential user interests and influential topics help to identify top trending topics from the user interaction model.

### 3 Proposed System

The proposed work includes the data extraction, pre-processing, feature extraction and the influential user interaction model is shown in Fig. 1. The first level includes data retrieved from the user timeline tweets. The second level includes the pre-processing and feature extraction. Tweets are pre-processed and the user list is generated to store the user information, both produce the processed tweets. From the processed tweets the hashtag is extracted in the feature extraction level. The third level includes the influential user interaction model, in which the user interest is classified into active and inactive user interaction. Based on the parameters of core interest and marginal interest, the influential and non-influential user interaction is identified respectively. Then the incidence matrix is used to identify the structure of user interaction from the user's timeline. Tweet score

is considered to find influential topics from the extracted tweets. Based on the influential topics and the influential users identified the top trending interested topics are analyzed.



**Fig. 1.** Overall Architecture of the Proposed System

### 3.1 Data Extraction

Social site information is not publicly available, a password is required to collect data from the user. The Twitter platform provides various API, SDK for developing an application that is used to access the data and data retrieved from the social site's Oauth account of Twitter API. Social Interaction data includes more active users among social media sites. Users retrieve the data which are connected or related to that particular user's post. The information in the retrieved user data includes tweets, followers, Likes, retweets, etc. Twitter users can control photo tagging and sharing of their friend list with the public user can also share the status with specific people [22]. Collecting and analyzing such network sites data allows us to study the structure of the relationship between users in a social network.

### 3.2 Feature Extraction

The first step of feature extraction is pre-processing. Tweets are tokenized, stemmer and lemmatization applied and analyzed based on part of speech tag. Irrelevant content and

URLs in the tweets are removed. From the extracted tweets the names of users who have shared the message on a specific topic of interest extracted. The extracted information is stored using a user's list which contains the user's names and their hashtag. The processed tweets retrieved from the user's timeline via his/her hashtag by using the Twitter API. Features extracted from the user's activities such as the sharing of information. The hashtag is extracted based on the user who shares their interest from the processed tweets.

### 3.3 User Interaction Model

Social members with similar features often connected than with more dissimilar ones. If there is any relationship between the members of the social group, there must be some common characteristics between those members which form a social relation. This basic idea leads to introduce the structure of the user interaction model. These characteristics are referred to as an influence. For example, in online social sites such as Twitter, Facebook, LinkedIn, etc., the certain influence is shown as the frequency with which friends tag each other based on common interests (movies, songs, books, etc.), in their posts. These influences suggest us how similar are the members of the social network [23]. Users are more willing to make friend with the one who has the same interest than the person who holds different attitudes [24].

The user interaction model is classified into two types such as active and Inactive user interaction based on the user's interest which further grouped into core interest and marginal interest respectively. In this paper, the two main problems by identifying how user interests change over time and whether user interests have a preference. Thus to define user interest based on core and marginal interest and the incidence edge matrix is used to capture user interests change over time and user interest structure is also identified. The user Interaction model generates the interesting topics of users from the tweets retrieved from the user timeline. This module first identifies entities that are can be directly extracted from the user's tweets and then scores them to reflect the influence of each topic of interested users. The time-based dynamics of user interest and how the topics of individual user's interest change over time. The proposed work deals with the user's interest from user tweets as a set of influenced concepts where concept may refer to an arbitrary entity and the influence indicates how important the concept is for the user's interests.

#### User Interest

The topics of interest of a user  $u$  is a set of influenced concepts where a concept  $c$  is represented via the entity.

$$UI(u, t) = \{(c_u, I(c_u, t, tweet_u)) \mid c_u \in C_{UE}\} \quad (1)$$

Where  $I(c_u, time, tweet_u)$  is an influence which is estimated for the concept  $c$  by the given user  $u$  based on user tweets denoted as  $tweet_u$  posted by  $u$  and based on the given time  $t$ .  $C_{UE}$  is consists of a set of entities.

### Tweet Score

The topic score of each tweet  $T$ , posted by given a user  $u$ , can be measured by frequency and confidence for an entity  $e$ , where a tweet  $t$  is represented via a set of entities  $E$ .

$$TS(u, T) = \sum_{i=1}^n f_{ei} c_{ei} \text{ where } f_{ei} \in F_E \text{ and } c_{ei} \in C_E \quad (2)$$

### Interest Similarity

The tweets retrieved from the user timeline is considered with a vector. Assume a user  $u$  interests is the vector  $S(u, t_1)$  at time  $t_1$  and  $S(u, t_2)$  at time  $t_2$ , where  $t_1 < t_2$ , which are represented via topics sharing of interests using the same vector representation that is used for a given user interests and the similarity score is used to identify the users interest change degree according to their cosine similarity.

$$\text{SimI}(S(u, t_1), S(u, t_2)) = \frac{(S(u, t_1) \cdot S(u, t_2))}{\|S(u, t_1)\| \|S(u, t_2)\|} \quad (3)$$

### Active User Interaction

Identifying the core interest among like, share and comment with a similar interest in a group. User's interest among the shared information in social media may not change depends on the trending information in the network. Active user interaction in the social group is formed due to frequent interaction among the user who shares the common interest leads to strong tie strength between them. This type of interaction is also referred to as a strong interaction. The user interest score is calculated from the given Eq. (1), if the score is high then they are identified as an active user, then the interest similarity score is calculated from the given Eq. (3), if the value of SimI is high, it indicates the interests of the same user have less change over time i.e. their interest is stable for a long period thus more amount of users has interacted.

#### *Core Interest (CI)*

The number of times that the user interacts with each other and thus they tend to have a closer relationship between them. The individual core interests will not change in a short period will stay for the long-term. The User's core interest is calculated which depends on the interaction among them and they are stable. The active users will hold the same core interest based on the user's interaction which belongs to a similar interest. The core interest score is calculated from the given Algorithm 1.

### Inactive User Interaction

Inactive user interaction in the social network is formed by fewer activities or dissimilar interests among them which leads to a weak tie strength between them. This type of interaction is also referred to as a weak interaction. Users may not share the related information or most likely topic in social media. Inactive user interaction in the social network is formed by fewer activities or dissimilar interests among them which leads to a weak tie strength between them. This type of interaction is also referred to as a

weak interaction. Users may not share the related information or most likely topic in social media. User's interest among the shared information in social media may change depends on the shared information in the network. Interaction between the users is an irrelevant topic. Weak ties between the users indicated the interaction between them is weak/less and inactive. The user interest score is calculated from the given Eq. (1), if the score is low then they are identified as an inactive user, then the interest similarity score is calculated from the given Eq. (3), if the value of SimI is low, it indicates the interests of the same user has a large change over time.

---

**Algorithm 1: User Interaction Algorithm**


---

**Input:** User Interest

**Output:** Active and inactive user identification

S: Active and Inactive User Interaction

M: CI of each Category in Aa

R: MI of each Category in IAa

Aa: Active interested user who shares the common interest

IAa: Inactive interested user who shares the dissimilar interest

---

1.  $S := \emptyset$
  2.  $M := \emptyset, R := \emptyset$
  3. for  $i = 1, 2, 3, \dots, k$  users do
  4.     Evaluate UI according to Equation 1
  5.     Evaluate SimI according to Equation 3
  6.     If UI and SimI value is high, then update to Aa
  7.     If UI and SimI value is less, then update to IAa
  8.      $P_{a_i}$  = Total number of users in the interaction
  9.     Threshold value = 0.5
  10.  for  $j = 1, 2, 3, \dots, m$  in Aa
  11.     Core Interest  $CI = |A_{a_j}| / |P_{a_i}|$
  12.      $CI >$  Threshold value
  13.     Active user is obtained by CI
  14.     append M
  15.  for  $j = 1, 2, 3, \dots, r$  in IAa
  16.     Marginal Interest  $MI = |IA_{a_j}| / |P_{a_i}|$
  17.      $MI <$  Threshold value
  18.     Inactive user is obtained by MI
  19.     append R
  20.  $S = \{ M, R \}$
  21. return S
- 

### *Marginal Interest (MI)*

The temporary or marginal interest is the interaction among the user's interest will change over a short period and thus the marginal interest is unstable. Inactive users will have dissimilar or different interests which change in a short period. Thus the interaction among them is distinguished with dissimilar interest. The marginal interest score is calculated from the given Algorithm 1.

### 3.4 Incidence Matrix Interaction

A graph with no parallel edges,  $m$  vertices represented as  $v_1, v_2, v_3, v_4, \dots, v_m$ , and  $t$  edges denoted as  $e_1, e_2, e_3, e_4, \dots, e_t$ . An Incidence matrix of an undirected graph shows the interaction between two nodes. The representation of matrix which includes nodes or vertex for each 'm' row and edges is the relation between them for each 't' column. The incidence matrix of 'm x t' order is denoted by  $[m_{ij}]$  is shown in Eq. (4).

$$M(G) \quad m_{ij} = \begin{cases} 1 & \text{if } e_i \text{ edge is incident to } v_j \text{ vertex} \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

Incidence matrix is also called a Vertex-edge Incidence matrix. If a vertex  $v$  incident upon edge  $e$ , then the pair of  $(v,e)$  will be equivalent to one otherwise it is zero. Thus the incidence graph  $GU_1$  and  $GU_2$  for user  $u_1$  and user  $u_2$ . The vertices denote the users and the edges represent the relations among the users. The incidence graph  $U_1$  and  $U_2$  as  $GU_1 (VU_1, EU_1)$  and  $GU_2 (VU_2, EU_2)$  where  $VU_1$  and  $VU_2$  are the vertices corresponding to the users  $u_1$  and  $u_2$  and  $EU_1$  and  $EU_2$  are the edges corresponding to social interaction among the users  $u_1$  and  $u_2$ .

The user interaction model shows the interaction between Users and their tweets and the incidence matrix represents the structure of the user interaction. Users are represented as vertex and the interaction between them is represented as edges. Multiple users post or share the information for certain hashtags. The interested user who tweets for the particular hashtag and their relationship between them exist through the interaction. The incidence matrix interaction is used to represent the structure of active and inactive user interaction. The topmost frequency distribution of hashtag is measured to identify the active user interaction by using a core interest method. The least frequency distribution of hashtag is measured to identify the inactive user interaction by the marginal interest method.

The incidence matrix interaction is a more efficient way to detect the user interest change over time and identify the user interest structure. Incidence matrix for an active user shows the number of users who frequently interacted with other users in the social group, whereas the inactive user shows fewer users who interact with other users over a long duration of the period. Incidence matrix is an efficient method to identify the influential user interest on the user's timeline from the given Eq. (4). From the given Eq. (2), the tweet score is calculated and identifies the influential topics. Thus based on the influential user interest and influential topics identified the Top trending interested topics is achieved by the fraction of strongly influential users score.

## 4 Performance Evaluation

In this section, the evaluation of the system and experimental results are discussed for the real-world datasets. Mainly, users in social networks will interact by sharing opinions or exchanging messages with each other. Everyone is interested in understanding user behavior and comparing their performance for identifying influential individuals. Social influence and user's similarity are both shows the users own interest and are more



predictive for finding future user behavior. Tweets are extracted based on the user's timeline interaction for 10 different categories. The interaction among the users is dependent on how they share their interests among them. The number of tweets extracted in each category and the hashtag frequency distribution for 10 different categories is evaluated based on their extracted hashtag is shown in Table 1.

**Table 1.** Tweets extracted and hashtag frequency distribution

Category	No. of Tweets	Frequency
Game	5347	1001
Education	5879	1007
Entertainment	9200	2022
Environment	13578	5079
Food	11303	3130
Music	12428	4454
News	10829	4975
Sports	14588	9868
Politics	7409	1464
Travel & tourism	8030	2462

#### 4.1 User Interaction Evaluation

From the results of the frequency distribution, the most influential users identified using the user interaction method. For each category, the top 10 influential users are identified and compared to the number of influenced users at a different timeline. The top frequency count is calculated to identify influential users. According to each user participating in the tweeting and retweeting process, the user interaction model divides the user influence into two different types as the active and inactive user influence according to the user interaction among them. The core interest and marginal interest score are evaluated to find the highest and least frequency count respectively for each category from the extracted tweets. The user interest (UI) and Interest Similarity (SimI) score is calculated from the Eqs. 1 and 3 for active and inactive user as shown in the Table 2.

The most influential user leads to active user interaction which is predicted from the top frequency calculated from core interest for the frequently interacted users. The non-influential user leads to inactive user interaction is predicted from the least frequency calculated from marginal interest for not frequently interacted users. The active and inactive user interaction is shown in Table 3.

Consider any three categories from the extracted 10 different categories of tweets to explain the difference between active and inactive user participation. Consider some of the different scenarios such as environment, music and news categories respectively.

**Table 2.** UI and SimI score for Active and Inactive User

S. No.	Category	Active User		Inactive User	
		UI Score	SimI	UI Score	SimI
1.	Game	0.667864	0.781384	0.1069	0.0692
2.	Education	0.561072	0.575867	0.1609	0.1246
3.	Entertainment	0.540059	0.627811	0.0772	0.0925
4.	Environment	0.888758	0.849076	0.0382	0.0549
5.	Food	0.623323	0.671565	0.0834	0.1035
6.	Music	0.598788	0.524465	0.0903	0.1166
7.	News	0.636181	0.516462	0.0482	0.0512
8.	Sports	0.760438	0.61923	0.0636	0.0764
9.	Politics	0.588362	0.732885	0.1034	0.0905
10.	Travel & tourism	0.706742	0.675854	0.0646	0.0942

### Active User Interaction

To assume that if user u1 tagged in most of the tweets where user u2 is tagged as well, there is a high chance of user u1 will also have more interest in a new topic where u2 is already active because the interest of u2 influences u1 to be active and they both have similar interests. For active user interaction for different categories as shown in Fig. 2.

Case 1: In the environment category the number of tweets extracted as 13578 among them the frequency distribution is calculated as 5079. The topmost topics preferred by the core interest of active user interaction with their top frequency score is 4675.

Case 2: In the Music category the no. of tweets extracted as 12428 among them the frequency distribution is calculated as 4975. The topmost topics preferred by the core interest of active user interaction with their top frequency score is 2667.

Case 3: In the News category the no. of tweets extracted as 10829 among them the frequency distribution is calculated as 4454. The topmost topics preferred by the core interest of active user interaction with their top frequency score is 3165.

### Inactive User Interaction

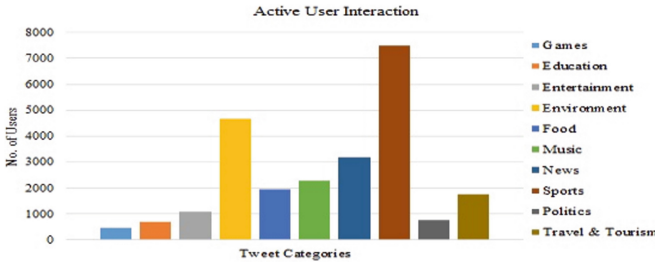
If user u1 does not tag the tweets where user u2 is tagged, then there is less chance of user u1 is interested in a new topic where u2 is active because the interest of u2 does not influence u1 to be active and they both have a dissimilar interest. For inactive user interaction for different categories as shown in Fig. 3.

Case 1: In the environment category, the least topics preferred by the marginal interest of inactive users interaction with their least frequency score is 322.

Case 2: In the music category, the least topics preferred by the marginal interest of inactive users interaction with their least frequency score is 402.

**Table 3.** Active and Inactive User Interaction

S. No.	Category	Top	Least
1.	Game	451	107
2.	Education	690	162
3.	Entertainment	1092	156
4.	Environment	4675	322
5.	Food	1951	261
6.	Music	2667	402
7.	News	3165	240
8.	Sports	7504	628
9.	Politics	768	118
10.	Travel & tourism	1740	159



**Fig. 2.** Active User Interaction based on different categories

Case 3: In the news category, the least topics preferred by the marginal interest of inactive users interaction with their least frequency score is 240.

The timeline interaction graph is considered for a better understanding of the user interaction shown in the different scenarios. Timeline interaction graph which is used to measure the evolution changes of the user interaction in the social network is achieved by the incidence matrix interaction. The active and inactive user interaction is measured based on the interaction with each other is shown over the period.

Strong interaction deals with the common interest shared among the users who frequently interact for a long period. From the results, the interaction among the users is frequent and the user’s distribution over the period is very dense. Weak interaction shows the deviation of user interest over a period. The user does not interact frequently and they are inactive users. From the results, the interaction among the users is not frequent and the distribution looks sparse for a certain period is very sparse. The user interaction for the different scenarios shown in Figs. 4 and 5.

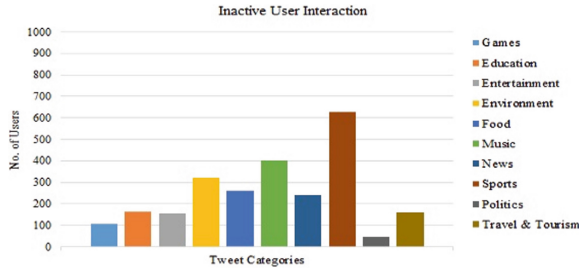


Fig. 3. Inactive User Interaction based on different categories

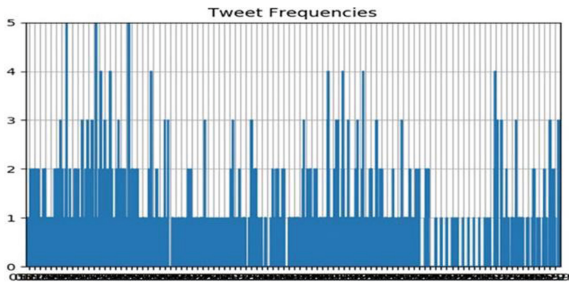


Fig. 4. Active User Interaction for Environment Category

### 4.2 Evaluation Metrics

Adamic/Adar measure is the inverted sum of degrees of common neighbors among the interaction of two users. A value of 0 indicates that two users are not close to each other, while higher values indicate users are closer.

$$A(x, y) = \sum_{u \in N(x) \cap N(y)} \left( \frac{1}{\log |N(u)|} \right) \tag{5}$$

where  $N(u)$  is the number of users adjacent to  $u$ .

This measure is used to identify the neighbor’s user in the social group. The incidence matrix interaction is used to represent the structure of the user interaction and the values

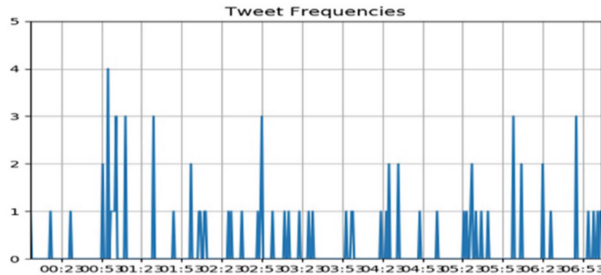


Fig. 5. Inactive User Interaction for Environment Category

are evaluated from the given Eq. (5). This measure is used to know the strong relationships among the user’s interaction. If they share a similar interest, then the users are related to each other. If they share the dissimilar interest, then the users are not closer to each other. If two users have a significant number of common interest than the number of their total interest. The measure is higher and the interaction between them is stronger refers to influential user interaction [21].

**4.3 Fraction of Strongly Influential (FSI)**

The Fraction of Strongly Influential is the value of the number of users interacted among other users over a period. A user is strongly influenced by an active user who frequently shares the common interest during the long time interval. A user is weakly influenced with inactive user among which they have less interaction based on the dissimilar interest with long duration. The fraction of Strongly Influential (FSI) users score will lie between 0 and 1, with 1 being the frequently strongly influenced whereas 0 leads to weakly influenced users. The precision, recall, accuracy, and F-measures evaluated based on the Fraction of Strongly influential (FSI) users score as shown in Table 4.

**Table 4.** User Interaction based on FSI

	Similar interest	Dissimilar interest
Active User	Most Influential user with similar interest (MIUSI)	Influential user with a dissimilar interest (IUDSI)
Inactive User	Non-influential user with similar interest (NIUSI)	Non-Influential user with a dissimilar interest (NIUDSI)

MIUSI = MIUSI is active with frequent interaction. This is a correct positive prediction.  
 IUDSI = IUDSI is active with frequent interaction. This is an incorrect negative prediction.  
 NIUSI = NIUSI is inactive with non-frequent interaction. This is an incorrect positive prediction.  
 NIUDSI = NIUDSI is inactive with non-frequent interaction. This is a correct negative prediction.

**Precision**

Precision is calculated as the number of Influential similar user predicated divided by the total number of similar user predictions. Thus it is also called a positive predictive value. Precision is calculated from Eq. 6.

$$\text{Precision} = \frac{\sum \text{MIUSI}}{(\sum \text{MIUSI} + \sum \text{IUDSI})} \tag{6}$$

**Recall**

The recall is known as true positive value or sensitivity. It is defined as the number of correct results divided by the number of relevant results. It is also called a true positive rate. The Recall is calculated from Eq. 7.

$$\text{Recall} = \frac{\sum \text{MIUSI}}{(\sum \text{MIUSI} + \sum \text{NIUSI})} \tag{7}$$

**Accuracy**

Accuracy measures a ratio of correctly predicted observation referred to as influential users to the total observations referred to as the total number of users. Accuracy is calculated from Eq. 8.

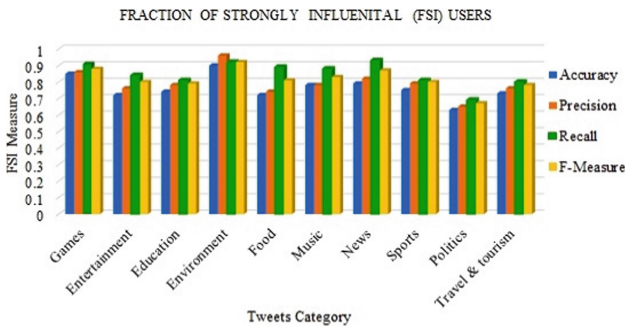
$$\text{Accuracy} = \frac{(\sum \text{MIUSI} + \sum \text{NIUDSI})}{(\sum \text{MIUSI} + \sum \text{IUDSI} + \sum \text{NIUSI} + \sum \text{NIUDSI})} \tag{8}$$

**F measure**

F1 score is used to consolidate precision and recall into one measure, the F1 measure is calculated from Eq. 9.

$$F = 2X \frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \tag{9}$$

The fraction of strongly influential users score is evaluated from the given Eq. (6), Eq. (7), Eq. (8) & Eq. (9) and shown in Fig. 6.



**Fig. 6.** User Interaction based on FSI Score

## 5 Conclusion

In many Social network sites such as Facebook, Twitter or LinkedIn the users find new friends or imitate a real-life relationship with friends and establish new friendship relations that cannot exist in real life due to distance or other factors. This study represents the social phenomenon of how users influence the establishment of a new friend's relationship in social network sites. In this proposed work, the user interaction model is used to identify an active user or inactive user interaction in a social group. To characterize a particular relationship between the users by using incidence matrix interaction i.e. an edge plays the important role in identifying the interaction, by the strength of relationship based on common interest among the users in a social group. Thus it is used to identify the influential user based on the vertex–edge incidence matrix interaction. Thus the Fraction of strongly influential Users measure is an efficient way to evaluate the Influential user interaction. As a part of future work, it would be interesting to see whether these model helps to detect the communities in the social network.

**Acknowledgement.** This publication is an outcome of the R&D work undertaken in the project under the Visvesvaraya Ph.D. Scheme of Ministry of electronics of Info. Technology, Government of India, being implemented by Digital India Corporation (formerly media Lab Asia).

## References

1. Kempe, D., Kleinberg, J., Tardos, E.: Maximizing the spread of influence through a social network. In: Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining – KDD 2003 (2003)
2. Chen, W., Wang, Y., Yang, S.: Efficient influence maximization in social networks. In: Proceedings of the 15th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 2009 (2009)
3. Trusov, M., Bodapati, A.V., Bucklin, R.E.: Determining influential users in internet social networks. *J. Market. Res.* **47**(4), 643–658 (2010). ISSN: 0022-2437
4. Goyal, A., Bonchi, F., Lakshmanan, L.V.S.: A data-based approach to social influence maximization. *Proc. VLDB Endowment* **5**(1), 73–84 (2011)
5. Bonchi, F.: Influence propagation in social networks: a data mining perspective. In: 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology, vol. 12, no. 1, pp. 8–16 (2011)
6. Tang, F., Liu, Q., Zhu, H., Chen, E., Zhu, F.: Diversified social influence maximization. In: 2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014), pp. 455–459 (2014)
7. Sun, Q., Wang, N., Zhou, Y., Wang, H., Sui, L.: Modeling for user interaction by influence transfer effect in online social networks. In: 39th Annual IEEE Conference on Local Computer Networks, pp. 486–489 (2014)
8. Kao, L.-J., Huang, Y.-P.: Mining influential users in social network. In: 2015 IEEE International Conference on Systems, Man, and Cybernetics, pp. 1209–1214 (2015)
9. Erlandsson, F., Brodka, P., Borg, A., Johnson, H.: Finding influential users in social media using association rule learning. *Entropy* **18**(5), 164 (2016)
10. Nan, W., Sun, Q., Zhou, Y., Shen, S.: A study on influential user identification in online social networks. *Chin. J. Electron.* **25**(3), 467–473 (2016)

11. Woo, J., Chen, H.: Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. Springer Plus **5**(1), 1–19 (2016)
12. Namtirtha, A., Gupta, S., Dutta, A., Dutta, B., Coenen, F.: Algorithm for finding influential user: base on user's information diffusion region. In: 2016 IEEE Region 10 Conference (TENCON) - International Conference, pp. 2734–2738 (2016)
13. Xu, W., Liang, W., Lin, X., Yu, J.X.: Finding top-k influential users in social networks under the structural diversity model. Inf. Sci. **355–356**, 110–126 (2016)
14. Kaple, M., Kulkarni, K., Potika, K.: Viral marketing for smart cities: influencers in social network communities. In: 2017 IEEE Third International Conference on Big Data Computing Service and Applications (Big Data Service), pp. 110–126 (2017)
15. Li, Y., Fan, J., Wang, Y., Tan, K.-L.: Influence maximization on social graphs: a survey. IEEE Trans. Knowl. Data Eng. **30**(10), 1852–1872 (2018)
16. Alshahrani, M., Zhu, F., Zheng, L., Mekouar, S., Huang, S.: Selection of top-K influential users based on radius-neighborhood degree, multi-hops distance and selection threshold. J. Big Data **5**(1), 1–20 (2018). Springer
17. Zhang, B., Zhang, L., Mu, C., Zhao, Q., Song, Q., Hong, X.: A most influential node group discovery method for influence maximization in social networks: a trust-based perspective. Data Knowl. Eng. **121**, 71–87 (2019)
18. Zengin, Z., Şule, A., Oguducu, G.: Influence factorization for identifying authorities in Twitter. Knowl. Based Syst. **163**, 944–954 (2019)
19. Gowsikhaa, D., Abirami, S., Baskaran, R.: Automated human behavior analysis from surveillance videos: a survey. Artif. Intell. Rev. **42**(4), 747–765 (2012)
20. Suganthini, C., Sridhar, R.: A survey on community detection in social network analysis. Int. J. Appl. Eng. Res. **10**(75), 1–6 (2015). ISSN 0973-4562
21. Sridhar, R., Kumar, A., Roshini, S.B., Sundaresan, R.K., Chinnasamy, S.: Feature based community detection by extracting Facebook profile details. ICTACT J. Soft Comput. **8**(4), 1706–1713 (2018)
22. Montreal, D.S.: All graphs in which each pair of distinct vertices has exactly two common neighbors. Mathematica Bohemica **1**, 101–105 (2005)
23. Verma, A.K., Pal, M.: Evolving social networks via friend recommendations. In: 11th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), IEEE Computer Society, pp. 379–383 (2015)
24. Malliaros, F.D., Vazirgiannis, M.: Clustering and community detection in directed networks: a survey. Phys. Rep. **533**(4), 95–142 (2013)