Chapter 5

Context-aware Influential Nodes Tracking in Dynamic Social Networks

We focus on the third objective of the thesis, i.e., Context-Aware Influential Nodes Tracking in Dynamic Social Networks, in this chapter. We give an introduction, motivation, and contributions for the considered problem in section 5.1. Section 5.2 gives the preliminaries and problem statement. Section 5.3 explains the proposed framework as the solution to the defined problem. Experimental details are given in section 5.4, and their outcomes are discussed in section 5.5. Section 5.6 concludes the overall outcome of the chapter.

5.1 Introduction

In the last two decades, we have witnessed significant advancements in information sciences that have made online social networks (OSN) important interaction platforms for the interchange of ideas and information. Studies have modeled the process of information diffusion in social networks for application domains such as social media [60], epidemiology [61], viral marketing [62, 63, 64], political campaigning [65], fake news containment [66] and many more.

Several models have been proposed during the last two decades to formulate the information diffusion process. Independent Cascade (IC) model and Linear Threshold (LT) model are the two elementary diffusion models proposed by David Kempe et al. [59]. In the *IC model*, a node has a probability of convincing each of its neighbours. And in the *LT model,* a node accepts a new idea if the influence from all its neighbours has crossed a threshold. Initially, David Kempe et al. formally formulated the *IM problem* to find the seed set S of size k to maximize the influence spread in the network using the IC/LT diffusion model. They also proved that the IM problem is *NP-hard*, and the corresponding objective function is monotone and submodular. They also proposed a hill-climbing greedy algorithm to solve the IM problem, which is quite close to the optimal solution.

However, two major challenges still exist. The first challenge is that it is not time efficient; therefore, it is not suitable for large networks. The second challenge is the effectiveness of seeds; this is due to ignorance of many other important factors responsible for seed set selection. To tackle these challenges, several research works have been proposed, such as centrality based [67, 68, 69, 70], sub-modularity based [59, 71, 72, 73], and influence path based [74, 75, 76, 77] approaches. These approaches are significantly faster than the traditional greedy approach. However, these models have limitations in terms of the quality of seed nodes which is still an issue. To handle this, several context/topic-aware [78, 79, 80, 81, 82, 83, 84] IM techniques were introduced. These methods improve the quality of the seed set as compared with traditional structure-based approaches. However still, efficiency and scalability is an issue. To address the above challenges, we propose an efficient and effective IM algorithm for influence maximization in dynamic social networks.

Motivation: In real-world society, if we observe the social networks, we find that: "the users' posts and comments can appropriately determine their interests and activities. Users' interest in different topics can imply their interest in different products. A user might have different levels of interest in different topics, and a product can also be relevant to different topics. Thus, the similarity between the topics that interest a user and those relevant to the product can indicate the amount of the users' interest in the product." [181, 182]. We also observe the following phenomenon:

- Popular personalities of the society have more influence on the people connected to them.
- If two people share a common geographical location repeatedly, they can influence each other.

Motivated from the above observations, we propose a multi-feature based influential node tracking method named MINT algorithm for influence maximization in dynamic social networks. Here, we use the structure of the network, users' topic-of-interest, users location sharing information, and the popularity of nodes in the network to propose a context-aware independent cascade diffusion model for influence spread. We also propose an efficient topic-aware seed selection technique that uses the Topic-aware Influence sub-Graph for finding the seed set for topic-based information spread. The David Kempe et al. model [59] is a basic model that did not consider such attributes. Our proposed Influence Maximization model in Chapter 3 considered only the structure of the network. However, this work focus on topic-aware influence maximization. See Appendix A for our research paper supporting this work.

5.2 Problem Description

5.2.1 Data Model

We consider n_t number of users as a set of vertices denoted as $V^t = \{v_1, \ldots, v_{n_t}\}$ at timestamp t and the set of edges among these users as $E^t = e_{ij}$, where each edge e_{ij} indicates a link (e.g., friendship) between v_i and v_j at timestamp t. $N(u)$ denotes the set of neighbours of user/node u. The node's geographical location information at time t is given by $L^t = [l_1, l_2, \ldots, l_{n_t}]^t$, where $l_i \in \mathbb{R}^r$, denotes the checked-in information of user v_i

at r different locations.

Formally, at given timestamp $T = \{1, 2, 3, \ldots, t\}$, we define location-aware dynamic attributed networks as follows:

Location-aware dynamic social networks: At a particular time stamp t, the corresponding location-aware dynamic attributed network is represented as $G^t = (V^t, E^t, D^t, L^t)$, where vertices V^t denotes the set of users, $E^t \subseteq (V^t \times V^t)$ denotes the pairs of users having a friendship relationship at t, $D^t = [d^1, d^2, \ldots, d^{n_t}]^t$ denotes the text documents (messages/comments/tweets/retweets) exchanged between nodes and $L^t = [l_1, l_2, \ldots, l_{n_t}]^t$ denotes the nodes check-in information.

The ***Popularity Vector*** $\mathscr{P}^t = [\mathscr{P}_{v_1}, \mathscr{P}_{v_2}, \ldots, \mathscr{P}_{v_{n_t}}]$ can be computed using the definition of popularity given in Section 4.2 of Chapter 4.

5.2.2 Information Theory and Similarity Measure

To determine the relative closeness or similarity between the interest distributions, we use the Bhattacharyya distance [183]. For probability distributions p and q over the same domain X, the Bhattacharyya distance is defined as:

$$C(p, q) = -\log(BC(p, q)), \tag{5.1}$$

where $BC(p, q)$ is the Bhattacharyya coefficient for discrete probability distributions and it can be evaluated as:

$$BC(p, q) = \sum_{x \in X} \sqrt{p(x)q(x)}. \tag{5.2}$$

5.2.3 Influence Maximization

Given the seed set S, we describe the influence spread of S as the expected number of activated nodes when the diffusion procedure stops, represented by the influence function $\sigma(S)$.

Influence Maximization: The Influence Maximization process is to find a seed set $S \subseteq V$ of maximum size k to maximize the influence function $\sigma(S)$. Formally, the Influence Maximization method can be defined as the following optimization problem:

$$I^* = \arg\max_{|S| \leq k} \sigma(S). \tag{5.3}$$

The Influence Maximization is described in more detail in Section 2.3.2.2.

5.2.4 Influence Maximization in Dynamic Networks

To find the most influential seed set in online social networks, we need to track the dynamic behaviour of the networks. Here, we consider the sequence of the snapshot graphs $G_1, G_2, \ldots, G_t$ at time-stamp $T = 1, 2, \ldots, t$, respectively. In this chapter, we consider the dynamics of networks in terms of the addition or deletion of nodes and edges, geographical locations of the users, and shift/change in interest distribution of the users. For instance, a new user may join the social network, or an existing user may deactivate his/her profile on the social network. The user may visit/shift/share a new location. There may be a shift of views or interests with time, like change in political views or change of favorite tv-shows/actor/actress/sports-person/sports-team/news-channel/political-leader/topic-of-interest. Considering all these factors as the dynamics of the networks, we propose an effective model for influence maximization in dynamic social networks.

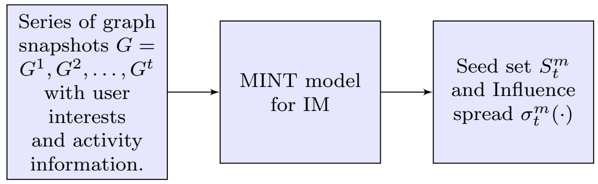


Figure 5.1: Block Diagram of the Proposed IM Framework

5.2.5 Problem Definition

Multifeature based Influential Nodes Tracking (MINT): *Given, a series of snapshots as* $G = \{G^1, G^2, \ldots, G^t\}$ *of the location-aware dynamic social network at different timestamps* $= \{1, 2, \ldots, t\}$*, respectively, the MINT problem's objective is to use the key factors responsible for the influence spread in the network and to compute the efficient and effective seed set* S_t^m *of size* k *for the spread of message* m *in current snapshot* G^t*. Formally, we have to optimize the following objective function:*

$$S_t^m = \arg\max_{S_t^m \in V, |S_t^m| \leq k} \sigma_t(S_t^m | G^t, m). \tag{5.4}$$

5.3 Proposed Framework

Figure 5.1 shows the block diagram for the proposed framework. We give the steps involved in the proposed MINT model in Figure 5.2. In this model, we assume that the users with similar interests have a stronger influence on each other. Firstly, we discover each users' interests and then find the similarity between the users' interests.

5.3.1 Discovering Users Interest

We consider the posts and comments exchanged by the user to find his/her interest. For instance, the content of the message posted by a user on Twitter "My heart was truly overjoyed yesterday when I received this gift. It is a new book called 'Women of Spirit Share Rituals Divine.' Included in the package was a very inspiring message Gift yourself some alone time to wallow in the Spirit and the Divine." This indicates the personal interest of the user in reading books related to women empowerment. This suggests that

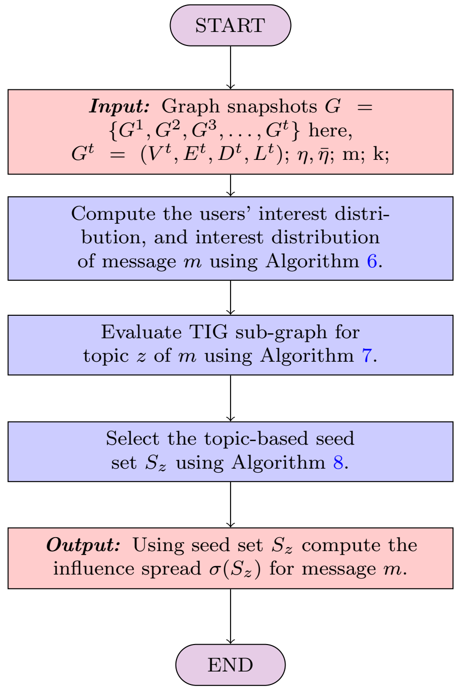


FIGURE 5.2: Flow of Steps Showing MINT Model

the user is interested in reading books on women empowerment and spirituality, which can be a subcategory of Women and Spirituality [184]. By considering the messages exchanged by the users, we can find the major topics related to them and can determine the extent of interest on each topic for all the users. In many social networks, the most usual topics of interest [185], including Politics, Education, Business and Finance, Health and Medical Conditions, Women Empowerment, Spirituality, Entertainment and Sports, and Disasters and Accidents. Various methods of topic modelling [186, 187, 181, 185, 188, 189] are available to quantify how much a user is interested in each of the considered topics, which gives the interest probability distribution for each user and also for the message to be spread in the network.

5.3.2 Computing Interest Distribution

We apply the Biterm Topic Model (BTM) [190] as a text-based topic discovery method for computing the topic-based interest distribution over the messages. BTM is specifically used for short texts. To make the inference effective with the rich corpus-level information, BTM learns topics by modelling the generation of word co-occurrences patterns (i.e., biterms) in the corpus. Here, the key idea is that "if two words co-occur more frequently, they are more likely to belong to the same topic".

Suppose a corpus is given with d documents that contain d_B biterms $B = \{b_j\}_{j=1}^{d_B}$ with $b_j = (w_{j,1}, w_{j,2})$, and $\mathscr{T}$ topics expressed over $\bar{w} = |\{w_1, w_2, \ldots, w_M\}|$ unique words in the vocabulary. Following the assumption of Latent Dirichlet Allocation (LDA) [186], the symmetric Dirichlet priors for θ and $\phi_{z_{d,i}}$ is used; here, $\theta = \{\theta_i\}_{i=1}^{\mathscr{T}}$ with $\theta_i = P(z = i)$ is the prevalence of topics in the corpus and $\sum_{i=1}^{\mathscr{T}} \theta_i = 1$ is the $\mathscr{T}$-dimensional multinomial distribution of interest (topic) and $z \in [1, \mathscr{T}]$ is the topic indicator variable. The word distribution for topics ($P(\bar{w}|z)$) is represented by a $\mathscr{T} \times M$ matrix Φ where the i^{th} row ϕ_i is an M-dimensional multinomial distribution with entry $\phi_{iM} = P(\bar{w}|z = i)$ and $\sum_{w=1}^{M} \phi_{jw} = 1$. Formally the BTM algorithm is given as Algorithm 6. Using this algorithm, we have computed the interest distribution of users, also the interest distribution of the message to be spread.

Algorithm 6 BTM Algorithm

Require: : Graph $G^t = (V^t, E^t)$, Text exchanged between nodes.
Output: : Topic distribution z_{v_i} for each node v_i.

```
for each user v do
    for each associated document d do
        for each word i ∈ d do
            Draw a topic z_{d,i} ~ Multinomial(θ) from the topic mixture of user v_{d,i};
            Draw words w_{d,i,1}, w_{d,i,2} ~ Multinomial(φ_{z_{d,i}});
        end for
    end for
end for
return z_{v_i} = {z_{v_1}, z_{v_2}, ..., z_{v_n}}.
```

Probability distribution of message m for $\mathscr{T}$ major topics is given as $z_m = \{z_m^1, z_m^2, \ldots, z_m^{\mathscr{T}}\}$ and the topic-based interest distribution for user v is represented as $z_v = \{z_v^1, z_v^2, \ldots, z_v^{\mathscr{T}}\}$, here z_v^i gives the interest probability of user v under topic i. We have $\sum_{j=1}^{\mathscr{T}} z_v^j = 1$, and $\sum_{f=1}^{\mathscr{T}} z_m^f = 1$. The major topics can be represented as a set $z = \{z_1, z_2, \ldots, z_{\mathscr{T}}\}$.

5.3.3 Computing Interest Similarity

The similarity between interest distribution z_m of message m and the interest distribution z_v of user v determines the extent by which user v is interested in message m. To determine the similarity between interest distribution z_v and z_m, we use Bhattacharyya distance between z_v and z_m by using the equations 5.1, and 5.2 as:

$$C_v(z_v, z_m) = -\log(BC(z_v, z_m)), \tag{5.5}$$

where $BC(z_v, z_m)$ is the Bhattacharyya coefficient for discrete probability distributions and it can be evaluated as:

$$BC(z_v, z_m) = \sum_{j=1}^{\mathscr{T}} \sqrt{z_v^j \cdot z_m^j}. \tag{5.6}$$

Similarly, we can also compute the interest similarity between each user.

5.3.4 Additional Factors for Influence Maximization

We also consider two additional factors responsible for the influence maximization gives as hypothesis defined as follows:

- *Location Hypothesis*: In a real-life scenario, we often see that if user u and v are friends on social networks and share a common location repeatedly then the chances of being influenced by each other increase if the interests of both the users are similar.
- *Popularity Hypothesis*: In a real-life scenario, we often see that if users u and v of social networks are related to each other and their interests are similar, then in most of the cases the more popular user influences, the less popular user.

We adopted both the hypothesis in our proposed model and integrated them with interest similarity to find the propagation probability for our proposed diffusion model.

5.3.5 Diffusion Model Used

Motivated from the real-life experience that individuals with similar interests in a particular topic have great probability to get influence with each other view over that topic. We use this phenomenon in our proposed diffusion model named the Compatible Independent Cascade (CIC) model. We propose a compatibility-based information diffusion model named Compatible Independent Cascade model for multifeature-based influence maximization in dynamic social networks. We use the content of the message to be spread to find the topic distribution $z_m = \{z_m^1, z_m^2, \ldots, z_m^{\mathscr{T}}\}$ of the message m for $\mathscr{T}$ number of topics, and then we use the topic wise compatibility between each user termed as user interest distribution $z_{uv} = \{z_{uv}^1, z_{uv}^2, \ldots, z_{uv}^{\mathscr{T}}\}$ for $\mathscr{T}$ number of topics; here, z_{uv}^j represents the influence probability of user u on user v under topic j. We also consider the popularity $\mathscr{P}_u$ of user u and location sharing information l_{uv} between user u and v as the factors responsible for the influence spread. The location sharing information l_{uv} is the count of the number of times user u meets with user v during the considered time period. Later this value is normalized to lie in $[0, 1]$. The formal definition of CIC is given as follows:

Compatible Independent Cascade Model: In this model, when a node u becomes active by message m in step t, it attempts to activate all of its out-going inactive neighbours $v \in N_{out}(u)$ in step $t+1$. For each neighbour v, it succeeds with the known probability cp_{uv} computed using equation 5.7:

$$cp_{uv}^m = \gamma \cdot \sum_{j=1}^{j=\mathscr{T}} z_{uv}^j \cdot z_m^j + (1-\gamma)(l_{uv} \parallel \mathscr{P}_u), \tag{5.7}$$

where operator "$\|$" represents the logical OR operator, and $\gamma \in [0, 1]$ represents a hyper-parameter controlling the significance of interest similarity and other considered factors.

We use the Compatible Independent Cascade model in our proposed influence maximization model for dynamic social networks. The edge weights for the diffusion process can be evaluated by using equation 5.7 for the CIC model. This equation integrates the additional factors with the interest similarity. Here, if either the location sharing or the popularity of the message sender node is higher than a threshold, then the edge weight increases by a factor $(1-\gamma)$ where γ is a hyperparameter.

5.3.6 Topic-aware Influence Sub-Graphs (TIG)

The TIG set $G_z = (V_{z_i}, E_{z_i})_{i=1}^{\mathscr{T}}$ is computed using Algorithm 7. Here, we evaluate the similarity between the interest distribution of the given message to be spread and the interest distribution of each user as C_v. We take the set of users with similarity $C_v > \eta$; where $\eta \in [0,1]$ is a threshold. Further, we compute the edge weights $W_{i,j}$ for the existing edges between users in V_z. Finally, we get TIG for each topic z by considering the above-selected users and corresponding edges with a significant value of edge weight, i.e., $W_{i,j} > \bar{\eta}$; where $\bar{\eta} \in [0,1]$ is a threshold.

Algorithm 7 TIG Algorithm

Require: : Graph $G^t = (V^t, E^t)$, Topic distribution for each user and message m.
Output: : Topic-aware influence sub-graph set $G_z = (V_{z_i}, E_{z_i})_{i=1}^{\mathscr{T}}$.

Input topics $z = \{z_1, z_2, \ldots, z_{\mathscr{T}}\}$.
for Each topic z_1 to $z_{\mathscr{T}}$ **do**
  for Each user v_1 to v_n **do**
    Compute similarity C_v with topic z using Equation 5.5;
  end for
  Find set V_z with users having $C_v > \eta$;
  Compute weight $W_{i,j}$ for each existing edge between users of set V_z using the *CIC model* (Equation 5.7) ;
  Remove edges with $W_{i,j} < \bar{\eta}$;
  TIG $G_z = (V_{z_i}, E_{z_i})$ for topic z_i is obtained;
end for
return $G_z = (V_{z_i}, E_{z_i})_{i=1}^{\mathscr{T}}$.

5.3.7 Topic-based Influential Nodes Tracking

We propose an efficient seed selection technique implemented on TIGs to get the desired topic-aware seed sets for topic-specific influence maximization in social networks. In this method, firstly, we take a TIG subgraph related to the topic of the message, m to be spread. Then we compute the *influence factor (IF)* for each node. The IF for a node v is defined as:

$$IF_v = \sum_{e_i \in E_v^2} W_{e_i}, \tag{5.8}$$

here, E_v is the set of edges within a $2-hop$ distance from node v, and W_{e_i} is the edge weight of edge e_i.

Further, we find the node v_h with the highest IF_v value and add it to the seed node-set. Then, we remove node v and its neighbours within $2-hop$ distance from the considered TIG subgraph. We repeat this process k times, and finally, we get the required seed set for message m. Algorithm 8 gives the steps involved in the seed selection process.

Algorithm 8 The MINT Algorithm

Require: : TIG $G_z = (V_{z_i}, E_{z_i})_{z=1}^{\mathscr{T}}$, Size of seed set k.
Output: : Topic-aware seed sets $S_z = \{S_1, S_2, \ldots, S_{\mathscr{T}}\}$.

Initialize: Set $S = \phi$ and $SC = \phi$
for $z = 1$ `to` $\mathscr{T}$ **do**
    `S=0`
    for $j = 1$ `to` k **do**
        `Find` $v_h \in V_z$ `with highest` IF_v `value using Equation` 5.8;
        `Add` v_h `to` S_z;
        `Remove` E_v^2 `and` v_h `from` G_z;
    end for
    return S_z.
end for
return $S = \{S_1, S_2, \ldots, S_z, \ldots, S_{\mathscr{T}}\}$.

5.4 Experiments

5.4.1 Datasets

We used six real-world network datasets for the performance evaluation of our proposed model. These datasets are from online social networks and coauthor networks. The considered datasets are Facebook, Epinions, Brightkite, DBLP, Gowalla, and Twitter. The description of these network datasets is given in section 2.5.2.

5.4.2 Baseline Methods

We compare our proposed model with five state-of-the-art methods using their published codes or our implementation. The considered baseline methods are introduced in section 2.6.3. Two of these methods are from basic techniques of influence maximization, and the rest of the other methods use topic-based influence maximization.

5.4.3 Evaluation Metrics

We have used two evaluation metrics to compare the performance of the influence maximization of our proposed model with other methods. The considered metrics are *Spread of Influence*, and *Speedup*. The formal definitions of these metrics are given in section 2.4.2. Better results have a greater Spread of Influence and Speedup values.

5.4.4 Experimental Settings

We have divided each dataset into a series of snapshots $G = \{G^1, G^2, G^3, \ldots, G^t\}$. The experiments are performed on each snapshot dataset. We use the standard parameter settings for all the baseline methods to implement them on our considered datasets. To evaluate the topic-of-interest and interest distribution of users and messages, we perform preprocessing of text data available as messages/comments of users. To improve the quality of the text; we processed the raw content by applying the following normalization steps: (a) removing non-Latin characters and stop words; (b) removing words with document frequency less than 10; (c) filtering out messages with length less than 3; (d) removing duplicate messages. For evaluating location information, we consider only M key locations based on the frequency of visits of the networks' users. For computing the popularity of nodes, we consider the window of size ten snapshots and by taking the value $t_y = 2$ for fresh links and links formed during the last ten snapshots as all links in Equation 2.2 and keeping all other variables fixed. The value of η and $\bar{\eta}$ in Algorithm 7 lies in $[0, 1]$.

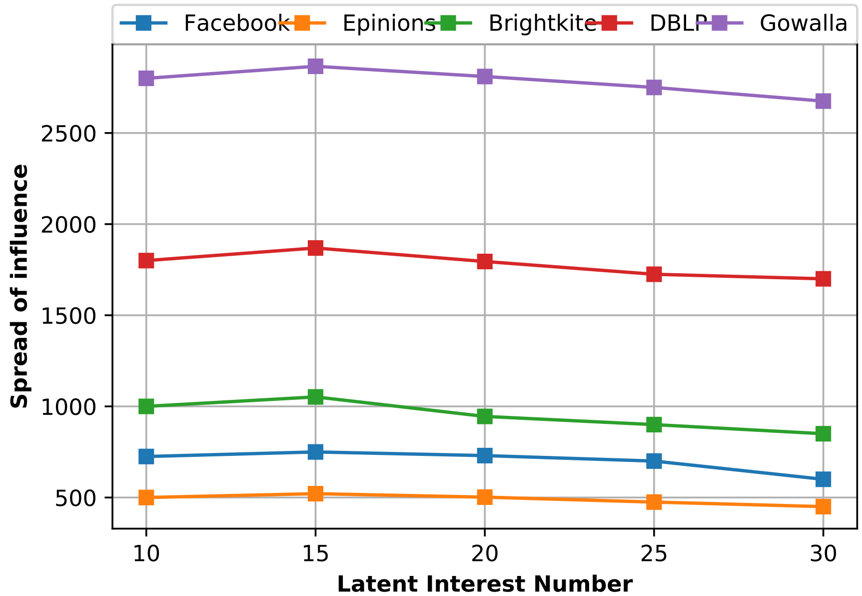
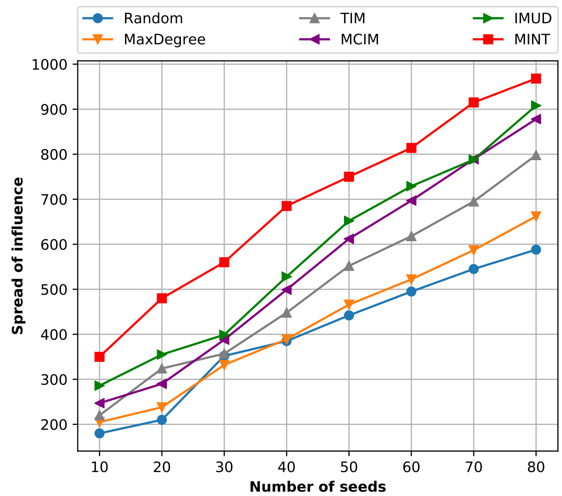


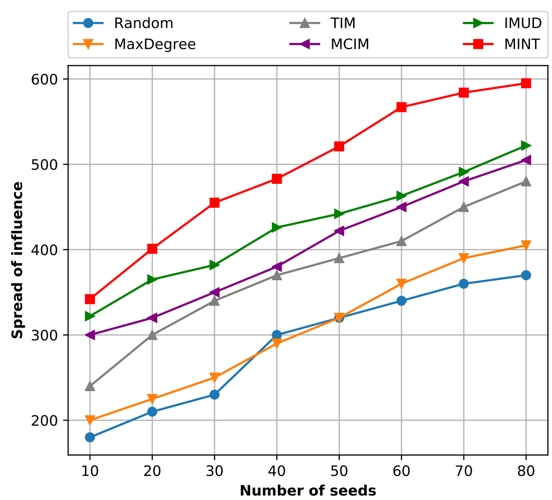
FIGURE 5.3: Spread of Influence versus Number of Topic-of-interest on Different Considered Datasets using Seed Set Size $k = 50$

5.5 Results and Discussions

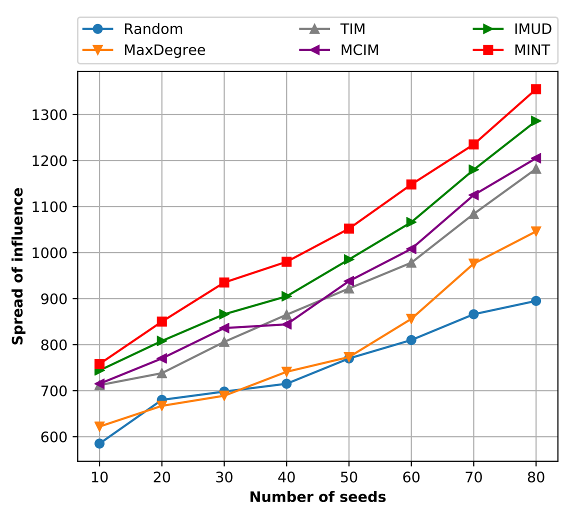
In this section, the results of the experiments performed are presented. Firstly, the result of user latent interest distribution is analyzed. We select the representative users from Facebook, DBLP, and Twitter datasets to show their latent interest distribution. In Fig. 4.4 (a)-(f), the graphs show the interest ID (latent interest number $\mathscr{T}$=10, and $\mathscr{T}$=20) on the x-axis and the probability of latent interest on the y-axis. As shown in Fig. **??** (a), (c), and (e), when $\mathscr{T}$=10, user u1 from the Facebook dataset have interest mainly concentrated in Interest ID=2, 6; for user u3 from DBLP have interest in Interest ID=8 and user u5 from the Twitter dataset has an interest in Interest ID=4, 7. We can observe that users u1 and u5 have prominent interests, and user u3 has some concentrated interests. Similarly, latent interest distribution for users u2, u4, and u6 are also shown (for $\mathscr{T}$=20). We can find that each user's latent interest preferences are different, and because of their differences in latent interests, the impact of this factor will affect the influence maximization task.



(a) Facebook

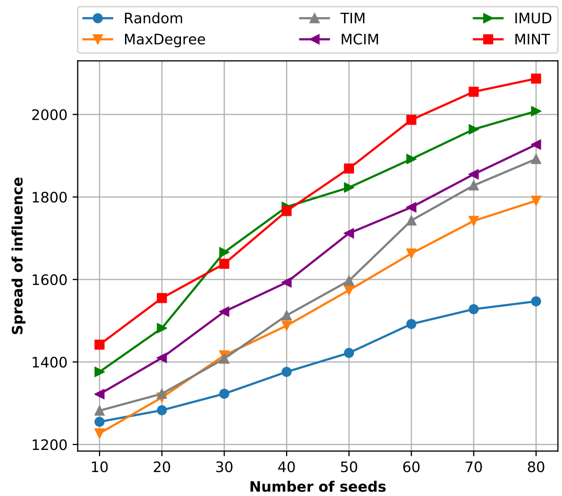


(b) Epinions

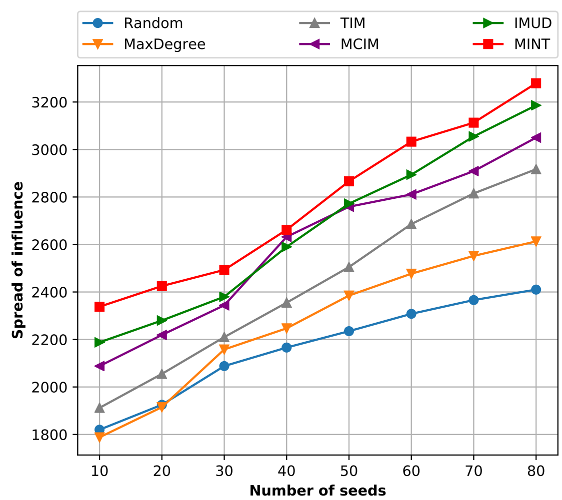


(c) Brightkite

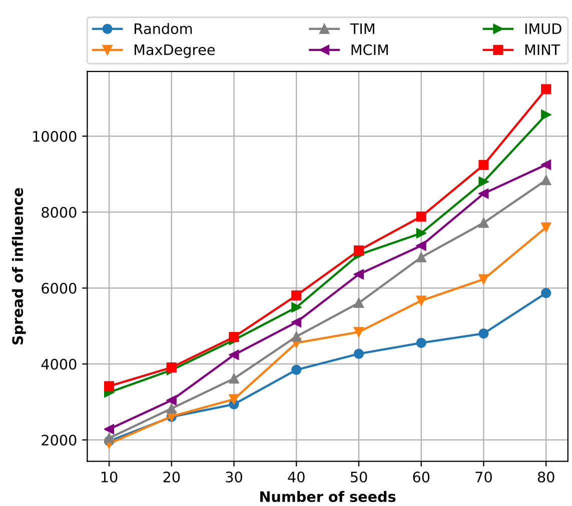
FIGURE 5.4: Spread of Influence versus Number of Seed Nodes for Facebook, Epinions, and Brightkite Datasets.



(a) DBLP



(b) Gowalla



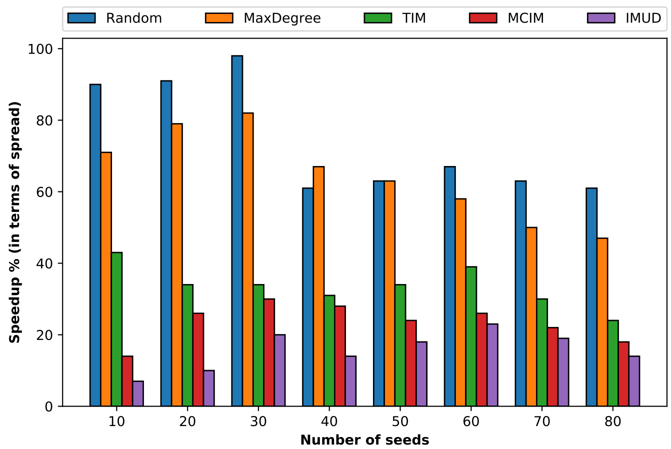
(c) Twitter

FIGURE 5.5: Spread of Influence versus Number of Seed Nodes for DBLP, Gowalla, and Twitter Datasets.

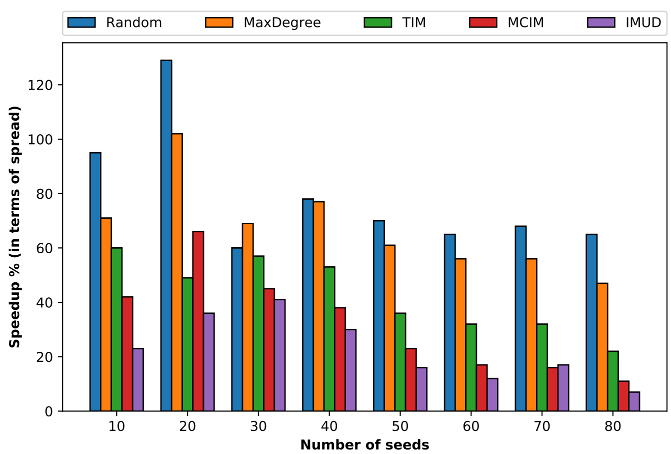
Figures 5.4 and 5.5 represents the graph showing the spread of influence in terms of the total number of nodes that get influenced after the diffusion process stops on the $y-axis$, and the number of seed nodes selected as initial influencers $|S| = k$ is represented on the $x-axis$. Here, seed size varies as $k = \{10, 20, 30, 40, 50, 60, 70, 80\}$ for all the considered baselines and proposed method on all the considered datasets. The results show that the proposed MINT algorithm outperforms in terms of the spread of influence as compared with the baseline methods. MINT performs far better than the traditional heuristic-based IM techniques such as Random and MaxDegree algorithms, and it also performs much better than the context-aware algorithms such as TIM, MCIM, and IMUD algorithms. Only the IMUD algorithm performs almost equally as the MINT algorithm for DBLP and Twitter datasets for seed size $k = \{30, 40\}$ and $k = \{30, 60\}$, respectively. The reason for better performance of the MINT algorithm is undoubtfully the considerations of multiple features for computation for influence spread and also due to the selection of effective seed set.

The results shown in Figure 5.3 represents that the number of topic-of-interest $\mathscr{T}$ affects the IM results. Here, the $x-axis$ represents the number $\mathscr{T}$ of latent interests z considered for topic-based IM, and the $y-axis$ represents the spread of influence in the considered datasets. We observe that the peak of the spread of influence reaches when $\mathscr{T} = 15$ in all the considered datasets. So the value of $\mathscr{T}$ in this model should select the small value preferably, in the range of 10-20. A too-large value of $\mathscr{T}$ may make the model more sensitive to noise information. A too-small value of $\mathscr{T}$ may overestimate user interest and increase the estimation error.

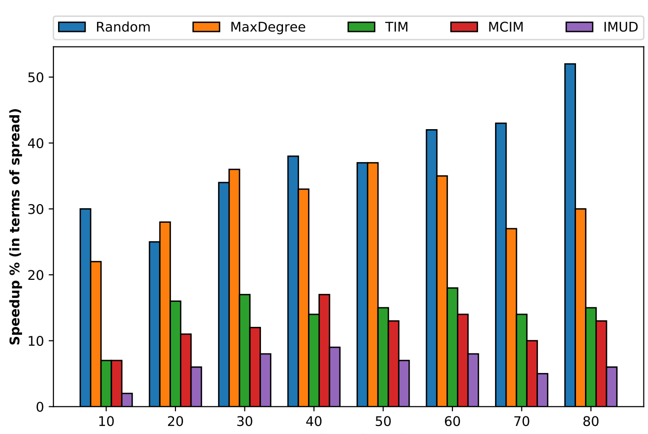
In Figures 5.6, and 5.7 the speedup % is shown on the $y-axis$ and the $x-axis$ represents the varying number of seed set size k. The result shows the proposed algorithms' efficiency compared with the baseline algorithms in terms of time taken to spread the influence with the selected seed set S in the different considered datasets. Higher values of speedup show that the seed set selected with the MINT algorithm can spread the influence in less time than the considered baseline algorithms. Here, we can observe that the influence spread by the MINT algorithm is around $30\% - 120\%$ more efficient than Random and MaxDegree



(a) Facebook

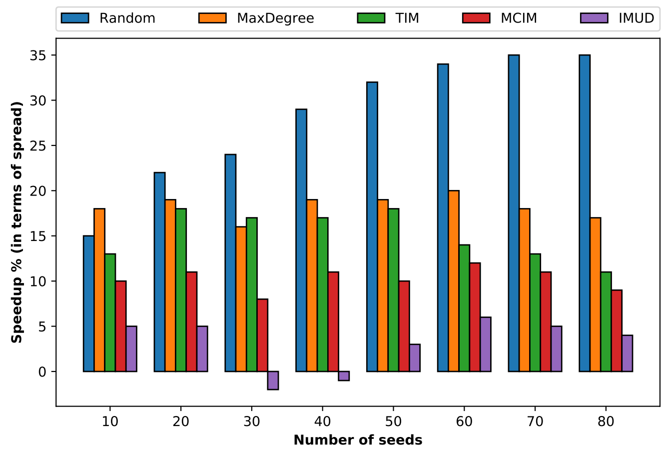


(b) Epinions

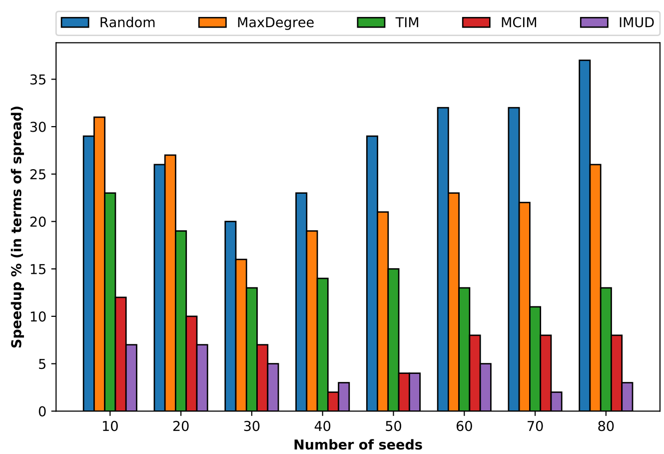


(c) Brightkite

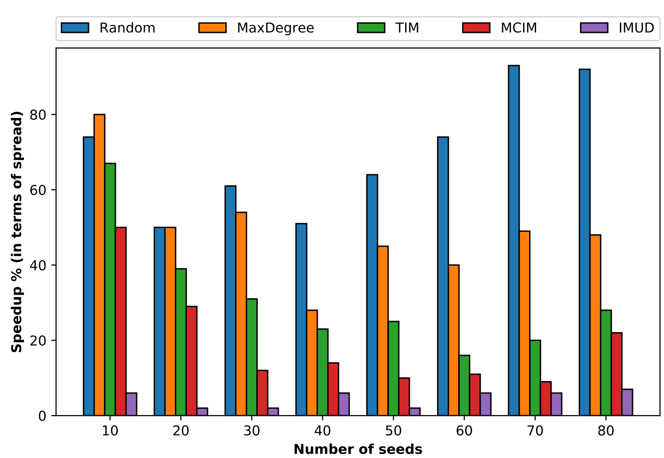
FIGURE 5.6: Speedup % (in terms of spread) Compared with Considered Baseline Methods versus Number of Seed Nodes for Facebook, Epinions, and Brightkite Datasets.



(a) DBLP



(b) Gowalla



(c) Twitter

FIGURE 5.7: Speedup % (in terms of spread) Compared with Considered Baseline Methods versus Number of Seed Nodes for DBLP, Gowalla, and Twitter Datasets.

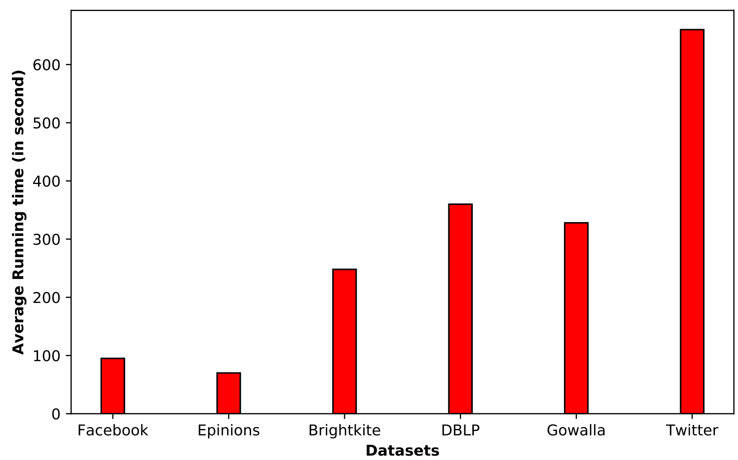


FIGURE 5.8: Average Running Time of MINT Algorithm over Different Networks

algorithms, and its' speedup varies between approximately 8% − 60% as compared with the considered context-aware seed selection algorithms. Results prove that the selected seed nodes are better influencers. The reason for getting better influencers through the MINT algorithm is the consideration of important factors such as the topic-of-interest of users, their location information and popularity in networks, and finding the topic-of-message (to be spread) based influencers. Figure 5.8 shows the average running time of seed selection for our proposed MINT algorithm on different considered datasets.

5.5.1 Insightful Discussion

The results of the experimental evaluation discussed above prove the efficiency and effectiveness of our proposed algorithm. We can observe that considering the interest similarity between users, location, and popularity based assumptions makes the model more effective because these features are based on the real-life scenario of our society. Using a Topic-aware influence graph for topic-based seed selection makes our model efficient and scalable because this model can be easily implemented on large graphs. Using TIG for seed selection reduces the time and space complexity for the seed selection process compared to the considered baselines. However, our proposed model can not give any theoretical guarantee for influence maximization, which is a limitation of our proposed MINT model.

5.6 Conclusions

This chapter presented a topic-aware influence maximization technique in dynamic social networks in which the influence spread depends on multiple features and seed nodes are discovered according to the topic-of-interest of users and message/product. We propose a novel multifeature based diffusion model, CIC, which is a modified version of the IC diffusion technique. The proposed diffusion model considers the similarity of topic-of-interest between users and also between users and messages. It also considers the popularity and location information of the users to perform the diffusion process. We also propose a novel topic-aware influence maximization algorithm based on the CIC diffusion model named the MINT algorithm for topic-aware seed set selection. Experimental results represent that the proposed MINT algorithm performs better in comparison to the considered baseline algorithms in terms of influence spread and speedup.