

Received August 21, 2020, accepted September 3, 2020, date of publication September 7, 2020, date of current version September 17, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3022096

Human-Driven Dynamic Community Influence Maximization in Social Media Data Streams

JUN GE^{1,2}, LEI-LEI SHI², YAN WU², AND JIE LIU¹

¹School of Information Engineering, Suqian University, Suqian 223800, China

²School of Computer Science and Telecommunication Engineering, Jiangsu University, Zhenjiang 212013, China

Corresponding author: Lei-Lei Shi (937109472@qq.com)

This work was supported in part by the National Natural Science of Foundation of China Program under Grant 61502209 and Grant 61502207.

ABSTRACT Microblogging—a popular social media service platform—has become a new information channel for users to receive and exchange the most up-to-date information on current events. Consequently, it is a crucial platform for maximizing community influence which has broad application prospects in recommendation system, advertising and other fields. With the rapid development of the mobile Internet, online social networks are gradually infiltrating into our daily lives, in which the communities are an important part of social networks. The combination of social networking and edge computing technology has important application value and is the development trend of influence maximization in future networks. However, traditional influence maximization models look for the most influential seed nodes while ignoring the fact that the selected seed nodes are various for different event topics, which significantly reduces the efficiency and accuracy of event propagation. In addition, most existing methods focus only on event propagation and neglect multiple topics in event propagation. At the same time, the interests of users in the network are not always single and the user's interest and the topic of the event will change over time, thus making it challenging to track momentous events in a timely manner. To address these issues, this paper proposes a Multi-Topic Learning-based Independent Cascade model (MTL-IC), and a Similarity Priority Mechanism-based Event Evolution model (SPM-EE). MTL-IC incorporates multi-topic factors and considers the authority and hub in interests of user, which makes the results more efficient and more accurate. SPM-EE can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution. The experimental results, using Twitter datasets, demonstrate the effectiveness of our proposed methods for both dynamic community influence maximization and event evolution.

INDEX TERMS Dynamic community influence maximization, microblogging, IC, event evolution.

I. INTRODUCTION

With the rapid development of Internet, the proliferation of Internet of things (IoT) and the burgeoning of 5G networks will generate a larger volume of data than has been previously possible. This advance in technology will see hundreds of applications deployed at the edge to consume this data. Thus social network has become an important place for people to communicate [1]–[3], [38]–[44]. Edge computing emerged as a new paradigm application, pushing the frontiers of computing applications, data and services from centralized nodes to the edge of the network, forming a useful supplement to cloud computing, and obtaining a better user experience through resource collaboration [48].

The associate editor coordinating the review of this manuscript and approving it for publication was Maode Ma¹.

The combination of social networking and edge computing technology has important application value and is the development trend of future networks [49]. In the real world, user groups with various connections are mapped into a directed network in the virtual social network, and information will be transferred among the nodes in the directed network [4], [45], [46]. In the process of transmission, messages are either passed down one by one through the link relationship between nodes, or are interrupted for some reasons. If we want to maximize the dissemination of information, then we need to study the dynamic dissemination mechanism, so maximizing the impact has become an important research direction in the field of data mining [5].

In the process of message propagation, when a node receives information from its neighbor, it will judge whether to accept the information according to various conditions.

If accepted, the node will continue to disseminate information to its outgoing neighbors. This process of communication is called word-of-mouth effect [6]. The real situation shows that people are more inclined to accept “friendship recommendation” from their closely related social groups than the dazzling advertisements on the Internet. With the increasing dependence on the Internet, large social network platforms such as micro-blog, Twitter, Facebook and so on have become the best application sites of “viral marketing” strategy.

Because of the difference of preferences between people, the results of different commodities or events in the same network become different. However, people’s preferences cannot be obtained directly. They can only be obtained through the analysis of text information left by users in the network. Therefore, text information becomes an important factor in the process of information dissemination. The paper cited network and Twitter network in this paper are both text-rich social networks. The research on maximizing the influence of the paper cited network can help the new works to be better and faster known by the majority of relevant scholars. The analysis of maximizing the influence of Twitter network can help the promotion of new products or events. Accordingly, this paper integrates multi-topic information into the traditional problem of maximizing influence in order to achieve more effective and practical results.

Nowadays, the use of online social networks for product promotion and marketing is still a hot research direction of data mining, and maximizing impact is one of the key researches. The promotion of a product has a fixed budget, so only one user set can be selected to make use of “viral transmission” so that the final affected users are the most, and more users are willing to buy the product. In this process, how to use the existing conditions to select the initial user set to promote products or events, and ultimately achieve the greatest impact is the key to the problem.

There are many application scenarios for the maximization of influence, which not only plays an important role in marketing, but also has great research prospects in other directions such as academic exchanges. At present, due to the prosperity and growth of social platforms such as Facebook, Twitter and Weibo, the problem of maximizing influence has a new starting point. How to find the most influential users accurately and efficiently in large-scale networks is a very challenging problem.

The purpose of maximizing influence is to find a set of seed nodes in the network that can maximize information dissemination. The whole selection process mainly involves two parts: the propagation model and the influence maximization algorithm. The propagation model is mainly responsible for the propagation of the impact and the activation of the nodes, while the influence maximization algorithm is responsible for finding the seed nodes that meet the requirements. At present, most of the related studies are not topic-sensitive, and the seed nodes under different topics are often different. Shi *et al.* [39] gave a detailed description of the impact of topic distribution on information dissemination, which can reflect the important

role of topic factors in maximizing influence. Even though a few communication models consider topic factors, they ignore the fact that in the real world, any product or event that needs to be promoted urgently must contain multiple topics, considering only a single topic is one-sided and inaccurate. At the same time, the interests of users in the network are not always single, but it is the user’s behavior preferences that directly determine the results of communication in the social network. In addition, they also do not consider the geographical location, authority and other node factors, and cannot solve the problem of large-scale social networks.

To address this failing, this paper proposes a Multi-Topic Learning-based Independent Cascade model (MTL-IC), and a Similarity Priority Mechanism-based Event Evolution model (SPM-EE). MTL-IC incorporates multi-topic factors and considers the authority and hub in interests of user, which makes the results more efficient and more accurate. SPM-EE can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution. Using a Twitter dataset for our study, our experimental results demonstrate the effectiveness of our proposed methods for both community influence maximization and multi- topics event evolution.

The main contributions of this paper are as follows:

- 1) A new Multi-Topic Learning-based Independent Cascade model (MTL-IC) is designed by adding multi-topic, learning ability, geographic location and other information based on the classical independent cascade communication model.
- 2) Improve the greedy algorithm in the classical maximization of influence problem; make it pay more attention to multi-topic factors, more suitable for the proposed new model.
- 3) A Similarity Priority Mechanism-based Event Evolution model (SPM-EE) is proposed, which takes multi-topic factors into account and highlights the importance of multi-topic to seed node selection during event evolution.
- 4) We conducted experiments to evaluate the performance of our proposed models. The experimental results on a Twitter dataset demonstrate the efficiency and accuracy of our proposed models in both dynamic community influence maximization and multi- topics event evolution.

The rest of this paper is organized as follows. In section 2, we introduce previous studies of event detection. In section 3, we describe our proposed MTL-IC method. We introduce the SPM-EE model in section 4. We discuss our experimental analysis and the obtained results in section 5 and in Section 6, we draw our conclusions.

II. RELATED WORK

Influences maximization has always been a hot research direction in the field of data mining and social network analysis [44]. It mainly includes two important parts: communication model and influence maximization algorithm. As topic factors play an important role in the problem of maximizing influence, topic modeling has become an indispensable tool

to study the problem of maximizing influence. This paper will briefly introduce the current situation of topic model, traditional maximization of influence and topic-sensitive maximization of influence at home and abroad.

A. TOPIC MODEL

Topic modeling has been studied for more than ten years [44]. The earliest topic modeling technology is Probabilistic Latent Semantic Analysis (PLSA) [5]. PLSA is essentially a probabilistic generation model, which models the generation process of documents by introducing topic layer between documents and words. Blei's Latent Dirichlet Allocation (LDA) [6] proposed in 2003 is undoubtedly the most classical method. It can obtain the topic distribution and the word distribution of each topic through the analysis and calculation of each document. Among them, the topic distribution uses several probability values to show the user's interest in the corresponding topic. So far, many new models have improved the LDA model [7]–[9], and the Correlated Topic Model (CTM) proposed by Blei *et al.* [10] has improved the LDA model, which can obtain the correlation between topics and better interpret the relevance of topics in real situations. In addition, the dynamic Topic Model (DTM) proposed by Blei *et al.* [11] can divide time into several discrete time segments, express topic as polynomial distribution, and use Gauss distribution to model topics of adjacent time segments. Kalman Filtering and Nonparametric W-regression are conceived. The variational reasoning algorithm of avelet Regression solves the parameters in DTM.

In recent years, there are many advanced topic models. McCallum *et al.* [12] proposed Author-Recipient-Topic Model (ART). At the same time, the author and the receiver of the text were modeled, and the role-Author-Recipient-Topic Model (RART) was designed. It believed that users were associated with some roles in the process of sending and receiving information. There are also many excellent topic models, which are not listed here.

B. INFLUENCE MAXIMIZATION

Influences maximization is a classical problem in the field of data mining, which mainly includes two parts: influence maximization algorithm and propagation model. The propagation model is responsible for abstracting and simulating the transmission of messages and activation of nodes in the real network, while the algorithm is responsible for finding a set of neutron nodes that meet the requirements and can maximize the impact of propagation. At present, the related research on the maximization of influence has been fruitful. The specific introduction is as follows. The problem of maximizing influence was first proposed by Domingos *et al.* [13]. Its fundamental goal is to find a set of seed nodes that can maximize the spread of influence in the context of viral marketing. Subsequently, Kempe *et al.* [14] modeled the problem as a discrete optimization problem. Two classical propagation models, Independent Cascade Model (ICM) and Linear Threshold Model (LTM), were proposed and introduced in

detail. Greedy approximate KKT algorithm was proposed on these two models. Because the principle of KKT algorithm is to select the optimal solution every time, it is destined that the seed nodes found by KKT algorithm are optimal, but at the same time, it is also destined that KKT algorithm is highly complex and inefficient, so it is not very suitable for large-scale networks. Much work has been done to improve the efficiency of the influence algorithm. The direction of improvement is mainly focused on the following two aspects: reducing computational load by heuristic or pruning, and saving time by parallel computing.

At present, many improved influence maximization algorithms have good results; some of them can even be very close to the effect of KKT algorithm. Leskovec *et al.* [15] proposed the CELF (Cost-Effective Lazy forward) algorithm to reduce the time used in Monte Carlo simulation to improve efficiency, which has nearly 700 times the efficiency of KKT algorithm. Goyal *et al.* [16] improved the CELF algorithm and proposed the CELF++ algorithm, which can calculate the marginal benefit at the same time to better reduce the operation time. Chen *et al.* [17] proposed the New Green algorithm to prune the useless edges in the propagation network. At the same time, MixGreen algorithm was proposed, which combines CELF algorithm with New Green algorithm to improve efficiency. In addition, Liu *et al.* [18] improved KKT algorithm and proposed Bottom-up algorithm, which made the marginal impact statistics between nodes independent and in line with the requirements of parallelization. Therefore, IMGPU algorithm was proposed to speed up the algorithm by using GPU to realize parallel computing to alleviate the influence maximization analysis in large-scale networks.

Most of the algorithms mentioned above are aimed at the improvement of KKT algorithm. Although the efficiency has been improved, it is still unable to achieve high efficiency, so it is still difficult to apply to large-scale networks. In this case, some heuristic algorithms based on classical models are gradually designed. MIA algorithm proposed by Chen *et al.* [19] based on IC model, shortest-path algorithm proposed by Kimura *et al.* [20] based on IC model, local directed acyclic graph algorithm proposed by Chen *et al.* [21] based on LT model, LDAG algorithm and Simple Path-based impact algorithm proposed by Goyal *et al.* [22] are all very efficient heuristic algorithms. It can even be compared with KKT algorithm.

None of the studies mentioned above takes into account the differences in the ability of user nodes to influence propagation. The traditional method of estimating the propagation ability of user nodes is simulation. The most commonly used method is Monte-Carlo simulation (MC), but the efficiency of this simulation method is very low. Kimura *et al.* [23] apply it to KKT algorithm or other algorithms according to the method of link penetration and graph theory, so that it can effectively estimate the marginal revenue. Compared with Monte Carlo simulation, it can greatly reduce the amount of calculation. Because the selection process of seed nodes is a # P-hard problem, Kim *et al.* [24] proposed the parallel

algorithm IPA (Independent Path Algorithms), which uses OpenMP to greatly improve the processing speed.

With the development of the problem of maximizing influence, the related research on the clustering of social networks [25]–[31] has been gradually considered and concerned. The problem of maximizing influence based on group is mainly to improve operation efficiency by dividing the network into smaller groups. However, the problem of maximizing the influence based on groups is only some partitioning work to improve efficiency. Although the network is superficially partitioned, it still finds the most influential nodes in the whole.

The related studies mentioned above have neglected one of the important factors in information dissemination - text topic information. According to the research, the authority and conformity of users are also related to the topic, so the following article will introduce the related research of topic-sensitive impact analysis.

C. TOPIC-SENSITIVE INFLUENCE MAXIMIZATION

Topic factors are not taken into account in the previous section. However, in the real world, topic factors often play an important role in the process of influencing communication. In real life, the influence of user nodes is not in any way, but only in one or several fields, and each field can be considered as a topic, so the authority and influence are related to the topic. At the same time, ordinary user nodes will not focus on all areas, but only on some areas, so different topics, the choice of seed nodes should also be changed [42]–[46].

Liu *et al.* [32] designed a probabilistic derivation model, by which users' topic distribution and interaction based on topic can be obtained simultaneously. Zhang *et al.* [33] took the behavior pattern of user nodes as one of the factors affecting the propagation, and accordingly proposed the Extended Independent Cascade (EIC) model. According to the characteristics of the model, the GAUP algorithm was proposed, and the experimental results proved that the interest of user nodes can affect the propagation of a vital role. The AIR model proposed by Barbieri [34] *et al.* (Authoritativeness-Interest-Relevance) is one of the more advanced topic-sensitive models at present. It first learns parameters based on user's previous consumption information and adds time dynamics to it. Zhang *et al.* [35] proposed a topic-sensitive solution that can analyze influence in microblog networks. Li *et al.* [36] proposed a keyword-based target influence maximization problem (KB-TIM), which aims to select a set of seed nodes to maximize the impact among users associated with a given advertisement.

Although the above-mentioned achievements take into account the topic factors, they ignore the fact that any commodity or event involves multiple topics, so it is one-sided and inaccurate to consider a single topic. Not only those, the above studies have not considered the situation of large-scale social networks, but the development of reality is quite different. Therefore, it is urgent to study the influence maximization analysis of large-scale social networks. In view

of the shortcomings of the existing research, this topic integrates multi-topic information into the problem of maximizing influence, and considers geographic location information, user authority and influence to improve accuracy and efficiency.

However, traditional influence maximization models look for the most influential seed nodes while ignoring the fact that the selected seed nodes are various for different event topics, which significantly reduces the efficiency and accuracy of event propagation. In addition, most existing methods focus only on event propagation and neglect multiple topics in event propagation. At the same time, the interests of users in the network are not always single and the user's interest and the topic of the event will change over time, thus making it challenging to track momentous events in a timely manner.

To tackle the problems outlined above, this paper proposes a Multi-Topic Learning-based Independent Cascade model (MTL-IC), and a Similarity Priority Mechanism-based Event Evolution model (SPM-EE). MTL-IC incorporates multi-topic factors and considers the authority and hub in interests of user, which makes the results more efficient and more accurate. SPM-EE can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution. Finally, our proposed methods exhibit better efficiency and accuracy in both dynamic community influence maximization and multi-topics event evolution by addressing the above-noted drawbacks of existing methods [8]–[11], [36]–[38], [44].

III. MTL-IC METHOD

This section introduces in detail the LDA model which is very important to this paper and the principle and reasoning process of Topical HITS algorithm [37]. It also gives the concept of classical IC propagation model. On the basis of classical propagation model, it introduces such factors as multi-topic, authority and centrality, and puts forward multi-topic MTL-IC model.

Information propagation model is an extremely important part of the problem of maximizing influence. Its main responsibility is to simulate the propagation process of influence in the network. When products or events are promoted in a particular network, it is usually assumed that each user node will have two states - "activated" (accepting the goods or events) and "inactive" (not accepting the goods or events). When the "activated" entry neighbors of an "inactive" user node increase gradually, the possibility of activation of the node will also increase gradually, but the state of the user node may only change from "inactive" to "activated" and not be reversible. If the user node also becomes "active", it will continue to try to activate its "inactive" outgoing neighbor node. If you want to promote goods or events on the network, then in general, the node that can accept goods or events and continue to spread must be interested in the goods or events. The best way to judge whether a node is interested in commodities or events is to analyze their topic distribution. As a key indicator of the node's communication ability, topic

distribution can well describe the process of communication. In addition, geographic location information, node authority and centrality also play a key role in the dissemination process, so they are also integrated into the new model. This section will briefly introduce the classical communication model, and on the basis of it, propose a new communication model which integrates multi-topic, geographical location and other elements.

A. LDA TOPIC MODEL

LDA topic model is a Bayesian probability model with three layers of variable parameters proposed by Blei *et al.* [6], [37], [38] in 2003. It is called Potential Dirichlet Distribution Model. The three layers of variable parameters are words, topics and documents. LDA involves many theories such as Bayes theory, Dirichlet distribution and so on. It belongs to unsupervised machine learning technology and is used to infer potential topics contained in document set or corpus.

LDA treats each document as a word vector, which is used to perform complex mathematical calculations, thus transforming text information into digital information that is easy to model. A document should contain several topics, and words are obtained by calculating the probability distribution of topics. The polynomial distribution of words is used to represent a topic distribution. Similarly, the polynomial distribution of topics is used to represent a document.

More specifically, each document in the document set can be considered as a polynomial distribution of T topics. Each topic can be considered as a polynomial distribution of V words in the vocabulary. The vocabulary contains all the non-repetitive words in the document set, but in order to achieve better results, some commonly used stop words will be removed. There is a Prior Distribution of Dirichlet with Superparametric Condyle and Condyle, respectively. A document D must contain more than one word, and for each word, it should have multiple distributions from the document.

B. TOPICAL HITS ALGORITHM

It has been proved that besides text, documents contain some properties that can represent the characteristics of nodes. Sun [37] believes that documents have two potential attributes: Authority and Hub.

Jon Kleinberg believes that if a page has a high degree of authority, then the page will be linked by many centrality nodes; at the same time, if a page has a high degree of centrality, then the page will also be linked by many authoritative pages. Accordingly, Jon Kleinberg proposed Hyperlink Induced Topic.

Since topic factors were not taken seriously as an important factor when HITS was first proposed, HITS performed well in most text-based search engines. However, the topic factor has been paid more and more attention, and the HITS algorithm is no longer applicable when it plays a decisive role in determining the effect of the algorithm. In view of this situation, Shi *et al.* [38] integrated topic factors into HITS algorithm and proposed Topical HITS algorithm.

In Topical HITS algorithm, Authority vector and Hub vector of authority degree are considered instead of single authority degree and centrality degree. Each dimension of the Authority vector and Hub vector maps a topic, and the dimension is the number of topics contained in the current document. Topical HITS algorithm uses multi-surfer semantic model as random-access model. According to the behavior of surfer A , authority A can be obtained, and centrality H can be obtained according to the behavior of surfer H . Surfer A has two different decisions in each action:

The final scores can be iteratively calculated using Equations (1) and (2) for each post and user, respectively:

$$n.h = \Sigma d.a \quad (1)$$

$$d.a = \Sigma n.h, \quad (2)$$

where $d.a$ denotes post d 's authority score and $n.h$ denotes user n 's hub score [13]. The iterative processes for generating the final results are as follows:

$$A_n = M^T \cdot M \cdot A_{n-1} \quad (3)$$

$$H_n = M^T \cdot M \cdot H_{n-1} \quad (4)$$

where A_n and H_n denote the authority and hub scores at the n^{th} iteration, respectively, and M denotes the user-post matrix [38].

However, not all network links are the same as those mentioned above, so we need to make corresponding changes when calculating authority and centrality. For example, in a DBLP network, if Paper X refers to Paper Y , then the direction of influence propagation will be from x to y . The same situation exists in Twitter network, If user x pays attention to user y , then the influence will spread from user x to user y . The direction of transmission of the influence is opposite to that of links in Twitter network. In view of this, when Topical HITS algorithm is used to maximize impact, it becomes very important to distinguish the network link structure from the impact propagation structure, and it also needs to be adjusted and processed according to the different data sets. In this paper, the authoritativeness and centrality obtained by the Topical HITS algorithm mentioned above will be applied to the proposed MTL-IC model as parameters.

C. MULTI-TOPIC INFLUENCE PROPAGATION MODEL

When the weights between user nodes are the same, the impact propagation probability is equal. Therefore, when different products or events are promoted in the same network, the propagation probability between user nodes is equal, which is obviously not consistent with the real situation. In reality, even though the weight of links between user nodes is the same, the probability of impact propagation will be different due to the different interest of user nodes in the promoted products. When the user node is extremely interested in the promotion of goods, the probability of its activation will increase. If the commodities promoted change, the probability of their impact on transmission should also change. For example, if a manufacturer needs to use Twitter

to promote two items A and B , only a few users will have the opportunity to try them for free because of the limited budget. If the IC model is used to simulate the experiments that affect the propagation, the results of commodity A and B will be the same. This is because the IC model does not consider the topic factor, the simulated active users may not be interested in these two products, if it is promoted in the real world, its effect may be far less than expected. As the saying goes, “Skills have expertise”, so for different commodities and different users, we should consider the matching between commodities and users, that is, the degree of user’s interest in commodities. According to the above introduction, the role of multi-topic factors in maximizing influence is evident. Therefore, a multi-topic Learning-based Independent Cascade (MTL-IC) model is proposed below.

D. MULTI-TOPIC LEARNING-BASED MTL-IC MODEL

MTL-IC model is based on five very intuitive observations. It is observation 1 to 5 that makes MTL-IC model more realistic and more practical experimental results.

OB 1 If a user publishes a commodity or event related content in a large amount in the social network, he is more likely to accept the commodity or event.

OB 2 If a user is extremely concerned about a particular topic; he will be more inclined to accept goods or events that contain that topic. At the same time, the stronger the topic is in the topic distribution, the easier it will be accepted by users.

OB 3 stars or celebrities are more likely to influence their network followers, so these groups are also known as authoritarians, and their ability to influence others is also called authoritativeness. If the authority has more followers, his influence will be stronger.

OB 4 users are often influenced by their friends or stars in their network to drive them to accept some goods or events. Whether or not the user is interested in the goods or events, this ability to be influenced by others is called conformity. The more friends or stars the user has in the network, the more likely he is to be influenced by conformity.

OB 5 When geographical location restrictions exist; users tend to focus only on goods or events related to their geographical location and conduct corresponding behavioral operations.

In the above observations, there are several key factors: First, the user’s interest and the content of the product basically determine whether the product will be accepted by the user. Secondly, users have two related capabilities: authority and conformity. Based on the above observations, the principles and definitions of MTL-IC model are given below.

The whole activation process of MTL-IC model includes two stages: multi-topic activation and neighbor interaction activation, while multi-topic activation includes two stages: similarity activation and most prominent topic activation. In fact, the three activation stages of similarity activation, most prominent topic activation and neighbor interaction activation are carried out simultaneously. As long as any

activation stage is successful, the activation of the node is considered successful.

1) MULTI-TOPIC ACTIVATION

a: SIMILARITY ACTIVATION

The first activation stage is called similarity activation. When the user node u is activated, if it has outgoing neighbor v , then u will try to propagate the influence to v with probability p_{uv} . If v receives the influence from u , then v will judge whether the similarity activation stage is successful or not according to its own conditions. According to observation 1, the more similar the topic distribution of user node is to that of commodity, the more likely the user will accept the commodity, so it is necessary to calculate the similarity. Since topic distribution is essentially a probability vector and the sum of all dimensions is 1, the most direct cosine distance can be used to represent similarity. After the similarity of topic distribution between users and commodities is obtained, the similarity is compared with the set threshold. Once the similarity is greater than the threshold, the user nodes corresponding to the similarity will be activated immediately. The threshold is a decimal smaller than 1 and may vary under different circumstances. According to the experimental data and results, 0.85 is used as the threshold of this experiment. When the similarity is less than the threshold, the program generates a random decimal to compare with the similarity. If the random decimal is less than the similarity, the similarity activation stage is still considered successful. Otherwise, the similarity activation stage is considered to have failed and the most prominent topic activation stage is ready to start.

b: ACTIVATION OF THE MOST PROMINENT TOPIC

Similar to similarity activation, when v receives the influence from u , it judges whether the most prominent topic activation stage is successful or not according to its own conditions. From observation 2, it can be concluded that a single topic may also determine whether the whole activation process is successful, which is the reason why the most prominent topic activation stage exists. In the process of the activation stage, the first thing to do is to calculate the weight of each sub product of user nodes in the distribution of topic and topic distribution corresponding to the commodity neutron topic, and then select the largest product sub topic as the most prominent topic. The results of the phase weights due to activation depends directly on the user node topic distribution weights and commodity topic distribution corresponding topics, and each topic in the topic distribution component of the value is no more than 1, so this will be a topic of simple weight multiplication as the most prominent topic activation probability. As the similarity activation stage do, the program will generate a random number is compared, and the most prominent issue is that if the probability is less than, the activation was successful, otherwise, the stage of activation failure and began to prepare the neighbor interactive activation stage.

In extreme cases, if the user only cares about the only topic, the weight of the corresponding topic component in the topic distribution of the node will be close to 1. When the recommended commodity also only involves this topic, the most prominent topic activation probability will be close to 1. Then the user node will be very likely to be activated, which is consistent with the real world.

It is worth noting that, in reality, when and only when the user first received a specific item or event message, according to itself will interest in the goods or the contents of the event and whether to accept the goods or events to determine. If the user to receive the goods or event messages, as has been made to judge the goods or event preferences, thus not to judge whether this message is sent by the who, so many topics for the same commodity and activate the same user only make a judgment. However, the above situation can be divided into two kinds, that is, whether the user has accepted the goods or events at the time of the first evaluation. If the user first has to accept the goods, so it can be said that any penetration of the user as long as the neighbors sends messages to the user, and the user must accept and activation. Otherwise, you can only rely on neighbors to try to activate the user interactive activation.

2) NEIGHBOR INTERACTION ACTIVATION

The final activation stage of neighborhood interaction includes three stages: recommendation stage, communication stage and acceptance stage. From observation 3 and 4, it can be seen that even if the user is not too interested in the promoted product for the time being, the user may still be affected by his neighbor and accept the product. As shown in previous, a successful activation is related to three parameters, so the activation probabilities of these three parameters need to be obtained. For the propagation stage, for nodes u and v , the calculation method of propagation probability is the same as the classical IC model. For the recommendation stage, if only the similarity of topic distribution is taken as the recommendation probability, it will be different from the reality. For example, if a celebrity u has the same topic distribution as an ordinary person v , and if similarity is used as the recommendation probability, then the celebrity and the ordinary person will have the same influence on their outgoing neighbors. The fact is that the celebrity has more influence and can influence the people who pay attention to him. Therefore, celebrity u is more likely to succeed in the recommendation phase than common person v . In view of the above situation, it is unreasonable to regard similarity only as recommendation probability. Similarly, it is inappropriate to regard similarity only as acceptance probability. In fact, each user's recommendation probability and acceptance probability are different, so we need to find another way to calculate both the recommendation probability and acceptance probability.

In a given network G with influence relationship, if the node u has a directed edge pointing to the node v , then the node u has the opportunity to influence the node v .

When node u has a strong willingness to recommend a commodity to node v , node v will easily be affected and accept the commodity. At the same time, when the recommendation probability of node u is very high, the acceptance probability of node v also increases, and vice versa. The above results show that there is a mutually reinforcing relationship between the recommendation probability of a node and the acceptance probability of its outgoing neighbors. This relationship is very similar to the Authority and Hub obtained by Topical HITS algorithm. Therefore, Topical HITS algorithm is used to calculate the recommended probability vector R and the accepted probability vector A for each node.

In MTL-IC model, the authoritative obtained by Topical HITS algorithm cannot be directly used as recommendation probability, because the authoritative obtained by Topical HITS algorithm is topic-specific. However, it is noticed that the essence of topic distribution of commodities is actually a probability vector describing the intensity of each topic of commodities. Therefore, this paper uses topic distribution of commodities and authoritative degree obtained by Topical HITS algorithm to calculate the recommendation probability of user nodes.

IV. SPM-EE MODEL

Event evolution is an important issue in the study of influence maximization. Evolution is the basic characteristic of real networks. The event in the network will change with time. The evolution of the event is the result of the interaction between the network's own structure and the interaction process that frequently occurs on it. Event evolution analysis mainly studies the establishment of event evolution model according to the historical characteristics of the event in the network, and predicts the changes that may occur in the future. Discovering and analyzing the evolution of the user interest is important for analyzing changes in user interest, predicting trends in hotspots, or behaviors of users in the future.

A. AN AUTOMATIC EVENT CLUSTERING ALGORITHM

1) THE POST WEIGHTS

Suppose $E = \{E_1, E_2, \dots, E_K\}$ is a event clusters of a graph $G(V, E)$, where V is the set of posts and E is the set of edges between two posts. The N posts in the graph can be denoted by $\{P_1, P_2, \dots, P_N\}$. The matrix $V_{K \times N}$ denotes the post weights of N posts related to all the K events. As analysed previously, the authority of post can be used to express the opinion the post plays the centre role under its topic. Hence, the weight of post j 's influential degree in topic cluster E_r can be described as:

$$\text{if } P_j \in E_r V_{rj} = \frac{\text{Authority}_r(j)}{\sum_{(h:P_h \in E_r)} \text{Authority}_r(h)}$$

$$\text{else } V_{rj} = 0 \quad (r = 1, 2, \dots, k, j = 1, 2, \dots, N) \quad (5)$$

Therefore, for a given post P_i , the similarity between P_i and event E_j , described as, S_{ij} can be calculated as

$$\bar{S}_{ij} = \sum_{h=1}^N V_{jh} \times S_{jh} \quad (6)$$

where s_{jh} is the similarity between posts P_i and P_h . As we can see from function (5) and (6), S_{ij} is a sum of the similarity between post P_i and other posts in event E_j , thus the weights mainly rely on the contribution of the posts to the event.

2) THE EVENT AUTOMATIC CLUSTERING ALGORITHM

The algorithm for clustering posts into events in Microblogging networks is described as Algorithm 1 [38].

Algorithm 1 The Event Automatic Clustering Algorithm

Input: K , the number of events; A , the link matrix; $Nmax$, the maximum number of iterations.

Initialization:

(1) Select the top K posts with the highest authority values for the initial K events.

(2) Calculate the similarity matrix between any two posts in the graph.

(3) Extract the similarity matrix between the posts and the events. Partition the post into the event to which its nearest event belongs, and get the initial K classes of the graph: E_1, E_2, \dots, E_K .

Repeat

(4) Update the matrices $V_{K \times N}$ recording post weights of N post with respect to all the K events based on the current partitions using function (5).

(5) Calculate the similarity between post P_i and event E_j , \bar{S}_{ij} , using function (6), and then cluster the vertices into K events with every post being in the event it is most similar to.

Until: All the clustered events remain unchanged or the number of iterations comes to $Nmax$.

Output: All the members in each event.

B. THE COSINE MEASURE AND USER INTEREST DISCOVERING

After the completion of the LDA training process, the estimated $P(z)$ parameter is used to find an important event under a topic z . Posts related to the topic z are sorted according to $P(d|z)$ in descending order, and words related to the topic z are sorted according to $P(w|z)$ in descending order. But it is difficult for each topic to judge whether they belong to the same real-life event. Existing methods identify new real-life events manually and subjectively. In the HEE model, a new method based on cosine measure is presented to judge whether a new hot event is emerging and to identify whether some topics are belonging to one event automatically.

As time goes by in a social network, the user's behavior data will continue to accumulate, which may reflect changes

in user interest, which in turn leads to changes in the community structure. For example, based on previous community structure descriptions, if users in a community make friends with other strangers in the same community, or follow and comment on a friend's posts, it will lead to a closer relationship between users in the community. In contrast, if the user has many behaviors such as following and commenting posts published by other communities, it is a possibility that the user may withdraw from the current community and fall into other communities. In this paper, we propose a SPM-EE model, which can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution.

V. EXPERIMENTS

In this section, we detail the experiments we conducted on real-world short-text collections to demonstrate the effectiveness of our proposed MTL-IC method and SPM-EE model. We consider two typical topic models as benchmark methods, namely IC and HEE.

In the rest of this section, we describe our collection of the dataset, experimental setup and analysis, the baseline approaches, and model evaluation.

A. DATASET

The data set used in this experiment is Twitter platform data set. The Twitter data set is the Tweet blog data of some tweets composed of 126995 posts and 6589 users from December 28, 2015 to January 05, 2016.

Since the authoritative and centrality of users need to be obtained in MTL-IC model, after the topic distribution of user nodes is obtained through topic model, the topic distribution and the link relationship between users are taken as parameters, and Topical HITS algorithm is run to obtain the authoritative and centrality of each user node.

B. EXPERIMENTAL SETTINGS

We conducted the experiments on a computer with an Intel I7 3.4 GHz CPU and 16 G memory.

We tuned the parameters via a grid search. For LDA, $\alpha = 0.5$ and $\beta = 0.1$. In all the experiments, we used Gibbs sampling for 1,000 iterations. The results reported here is the average of five runs. In the process of filtering high-quality posts, we set all of the initial authority scores $d.a$ and hub scores $u.h$ to 1.

When the influence maximization problem was first proposed, the evaluation measure used to evaluate the advantages and disadvantages of the propagation model and the influence maximization algorithm was the last activated node IS (S). However, as topic factors play an increasingly important role in maximizing impact, the number of active nodes IS (S) becomes unable to fully reflect the results. For example, if a manufacturer wants to promote a product through viral marketing in the network, his ultimate goal is to hope that users on the network will be interested in his product or even buy it, rather than just passing by. Therefore, the group that

the producer finds should not only have enough audience, but also have enough interest in the product. Accordingly, according to the characteristics of multi-topic factors, the new evaluation measures should pay attention to two parts at the same time: the number of the last activated nodes IS (S) and the cumulative similarity of the last activated nodes. Therefore, this paper proposes a new evaluation measure called Similarity Impact Sum (SIS). As the evaluation measure of the experiment, we can judge the quality of the propagation model and the algorithm of maximizing the influence.

SIS (S, c) denotes the product of the cumulative similarity between the number of the last activated node and the last activated node and the commodity C when the seed node set is S, where V_C denotes the topic distribution similarity between the node V and commodity C. When SIS (S, c) is larger, it shows that the final set of activated nodes is more interested in commodities while the number of nodes is larger. Through this metric, we can accurately evaluate the advantages and disadvantages of each communication model and influence maximization algorithm under multi-topic factors.

Before the experiment is carried out, the initial parameters needed in the experiment should be set. For the propagation probability p, in the classical IC model, the value is 0.01 [14]. Therefore, the IC model is also used in this experiment. Because of the need of the experiment, the Monte-Carlo simulation (MC) is needed, and the simulation times are set. According to the existing experience [14], when the number of simulations exceeds 10,000, the experimental results change very little, so the number of simulations of model Carlo is set to 10,000 in this experiment. In MTL-IC model, a threshold should be set at the similarity activation stage to judge whether a node can be activated at that stage. The setting of this threshold needs careful consideration. If the setting is too large, it will be difficult for users to have such a high degree of agreement with the topic distribution of goods, and the activation effect will become worse. If the setting is too small, a large number of users will be activated at this stage, which may be different from the reality.

Therefore, the MTL-IC model takes the damping coefficient of 0.85 as the threshold value, which has been taken into account by a lot of experiments in the PageRank algorithm. Because the existing research has not considered the multi-topic factors, it is difficult to compare the experiment with them. Therefore, this paper will only compare with the most classical KKT algorithm and the AHP algorithm proposed in [47].

C. BASELINE APPROACHES

We validated the improved efficiency and effectiveness of the proposed MTL-IC and SPM-EE by evaluating our model against IC model and hot event evolution (HEE) [38], which are classic latent semantic analysis algorithms.

TABLE 1. The results of MTL-IC model combined with ANS algorithm.

The number of topic	Final activated nodes	Similarity
5	125	0.627
10	86	0.494
20	29	0.461

TABLE 2. The results of IC model combined with KKT algorithm.

The number of topic	Final activated nodes	Similarity
5	206	0.177
10	108	0.056
20	46	0.038

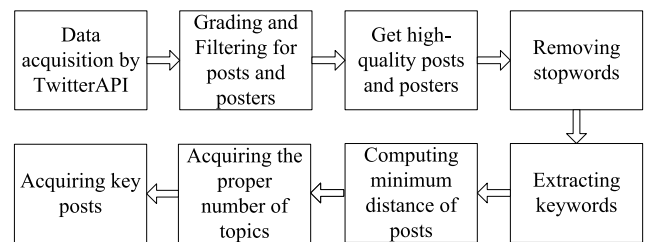


FIGURE 1. TD-HITS procedure.

D. EVALUATION METHODS

1) NUMBER OF TOPICS

Table 1 shows the results of MTL-IC model combined with ANS algorithm when the number of topics on Twitter datasets is 5, 10, 20, and the number of seed nodes is 5.

Table 2 shows the results of IC model combined with KKT algorithm when the number of topics on Twitter datasets is 5, 10, 20, and the number of seed nodes is 5.

According to Table 3 and 4, considering that KKT algorithm is the best and can not be surpassed in the index of the number of final activated nodes, the difference between MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm in this index is not large; in the index of cumulative similarity, the effect of MTL-IC model combined with ANS algorithm is far better than that of IC model combined with KT algorithm, and the average similarity is also higher. The same result shows that the activation nodes found by MTL-IC combined with ANS algorithm are very interested in the promoted goods or events, and in reality, these user nodes will be more receptive to the goods or events.

Figure. 3 gives the experimental results of MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm under SIS measurement. It can be seen that the results of MTL-IC model combined with ANS algorithm under SIS metric are significantly better than those of IC model combined with KKT algorithm, especially when

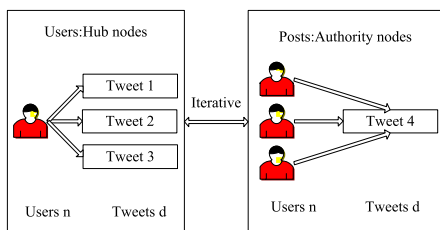


FIGURE 2. Iterative model for determining the authority score of posts and the Hub score of users.

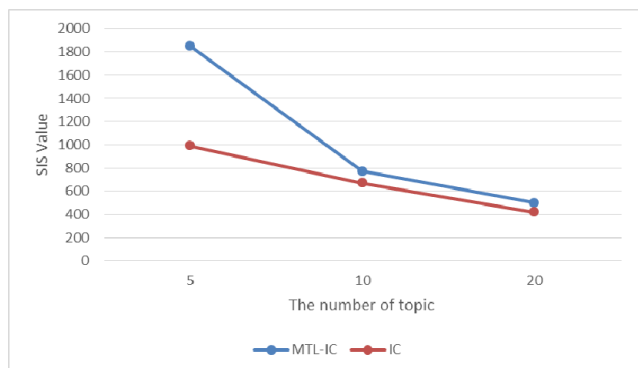


FIGURE 3. The experimental results of MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm under SIS measurement.

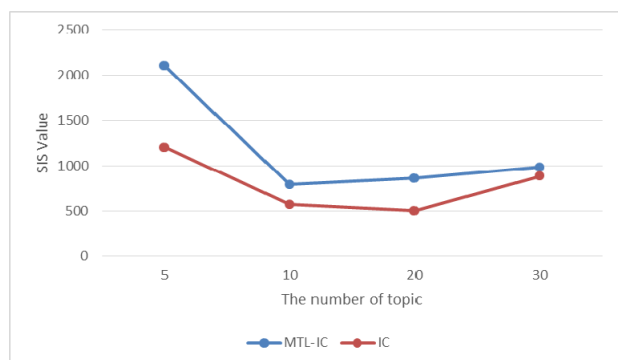


FIGURE 4. The trend of activation of MTL-IC model under different topics on Twitter dataset.

the number of topics is small. When the number of topics increases, the similarity of topic distribution between user nodes and commodities or events decreases due to the increase of topic number, which makes the activation effect of MTL-IC model worse.

However, considering from other aspects, the more topics, the more interested and targeted the final activation nodes are in reality.

Figure. 4 shows the experimental results of MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm under SIS measurement on Twitter data set. It can be seen that under SIS measurement, MTL-IC model combined with ANS algorithm has better effect, but the results are still affected by the topic distribution of the promoted goods or events.

TABLE 3. The activation of MTL-IC model in three activation stages under different topics.

The number of topic	Final activated nodes	Similarity
5	221	0.812
10	167	0.556
20	181	0.486
30	202	0.631

TABLE 4. The results of MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm under different topics.

The number of topic	Final activated nodes	Similarity
5	259	0.261
10	248	0.178
20	256	0.127
30	262	0.175

2) EFFECTIVENESS OF MTL-IC MODEL

this section shows the validity of MTL-IC model combined with ANS algorithm from the whole, but it cannot reflect the activation of MTL-IC model at each stage in detail, and it cannot reflect whether each activation stage of MTL-IC model has an indispensable role. Therefore, it is necessary to count the activation of MTL-IC model at three stages. Table 3 shows the activation of MTL-IC model in three activation stages under different topics. Figure. 4 shows the trend of activation of MTL-IC model under different topics.

Similarly, Table 3 and 4 show the results of MTL-IC model combined with ANS algorithm and IC model combined with KKT algorithm under different topics on Twitter datasets. It can be seen that when the number of topics is 5 or 10, the experimental results are similar to those on Twitter datasets. But when the number of topics is 10 and 20, whether MTL-IC model combines ANS algorithm or IC model combines KKT algorithm, their results are abnormal, that is, the effect is better when the number of topics increases. This is because the results of each experiment depend not only on the links and attributes of the user nodes in the network, but also on the topic distribution of the promoted goods or events. When the number of topics is 20 and 30, the result is better because, compared with topic 5 and 10, the selected promotion commodities or events when the number of topics is 20 and 30 are more interesting to users in the network. Therefore, the number of final activation nodes and cumulative similarity are increased, so there is an anomaly.

It should be noted that the number of active nodes in the table does not mean the number of active nodes, but the number of active edges. According to the results shown in Tables 5 and 6, it can be seen that, whether on Twitter

datasets, the number of edges activated by similarity activation process accounts for the vast majority of the total activated edges, more than 80%, while the number of neighbor interaction activation process accounts for a small part of the total activated edges, about 5% - 15%, while the most prominent topic activation process. The marginal proportion of total activation is the smallest, probably between 5% and 10%. The above may be the result of the following reasons: firstly, in the selected network, most user nodes are interested in the selected products or events, so the similarity activation process becomes very smooth; secondly, only a small number of user nodes are interested in a single topic, which happens to have a higher topic distribution in the promoted goods or events.

Weights are not activated in the process of similarity activation, so a small number of user nodes are finally activated through the most prominent topic activation process; finally, because some existing user nodes are activated through the similarity activation process or the most prominent topic activation process, these nodes will not be included in the activation results of neighbor interaction activation process, and the selected network can be.

The authoritativeness and centrality of the vast majority of users are not high, so the proportion of edges activated in the process of neighborhood interaction activation is not very significant, but still occupies a certain proportion. If the average authority and centrality of the user nodes in the selected network are relatively high, that is, the users in the network trust each other and have abundant links among users, and then the neighbor interaction activation process will have better effect.

The above experimental results show that the whole multi-topic activation process plays an important role in MTL-IC model, especially in similarity activation process, which also proves the validity of MTL-IC model from the side.

3) ANS ALGORITHM EFFECTIVENESS

Since the existing influence maximization algorithm does not consider multi-topic factors, and is not suitable for MTL-IC model which integrates multi-topic factors, this paper proposes an influence maximization algorithm based on multi-topic ANS. But whether the ANS algorithm can really find the desired seed nodes according to the needs, which requires the validity of the ANS algorithm to be proved. By changing the parameters in ANSM metric, the last set of activated nodes found by ANS algorithm under different cumulative similarity ratios is recorded, and compared with the last set of activated nodes found by KKT algorithm, to see if the last activated node found by ANS algorithm is different from that found by KKT algorithm, so as to prove that the ANS algorithm is different from that found by KKT algorithm. KKT algorithm can better select the required seed nodes, that is, the effectiveness of ANS algorithm.

Table 5 shows the average similarity between ANS algorithm and KKT algorithm under different topics on Twitter datasets.

TABLE 5. The average similarity between ANS algorithm and KKT algorithm under different topics.

Algorithm	The number of topic		
	5	10	20
ANS	0.175	0.155	0.134
KKT	0.171	0.143	0.112

TABLE 6. The average similarity between ANS and KKT algorithms.

Algorithm	The number of topic			
	5	10	20	30
ANS	0.314	0.221	0.158	0.179
KKT	0.286	0.166	0.103	0.118

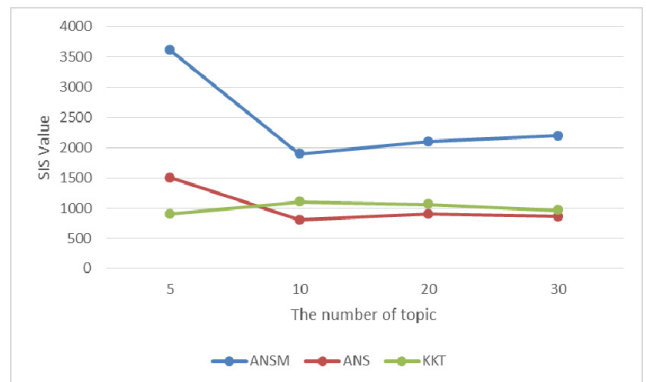


FIGURE 5. The result of activation of influence maximization under different algorithms.

Table 6 shows the average similarity between ANS and KKT algorithms for the last activated node which is not mutually included under different topics on Twitter datasets. The above results are obtained when the value is 0.5. As can be seen from Table 6, although the last activated node selected by ANS algorithm is partially the same as that of KKT algorithm, there are still some nodes selected by ANS algorithm that are not included in the results of KKT algorithm. From Table 5 and Table 6, it can be seen that the average similarity of nodes selected by ANS algorithm is higher than those selected by KKT algorithm, regardless of the number of topics. This proves that the user nodes selected by ANS algorithm are more interested in the promoted goods, and the more easily they accept the promoted goods or events in the real world.

As can be seen from Figure 5, when the number of topics increases gradually, the coincidence degree of the last activated node selected by ANS algorithm and KKT algorithm becomes smaller and smaller, which indicates that the difference between the selected nodes is also becoming larger and larger. This is because when the number of topics increases, the interests of user nodes are divided more carefully. While the KKT algorithm only focuses on the number of final activated nodes, the ANS algorithm also needs to pay attention to the similarity of the topic distribution between the activated

TABLE 7. The results of key users’ interests detection based on HEE model.

Time	Hot Event	Event evolution relationship
Sun Dec 27 2015	E1	
Mon Dec 28 2015	E2	
	E3	E2→E3
	E4	E2→E3→E4
	E5	E2→E3→E4→E5
Tue Dec 29 2015	E7	
	E8	
	E9	E8→E9
	E10	E8→E9→E10
	E11	E8→E9→E10→E11
	E12	E8→E9→E10→E11→E12
	E13	E8→E9→E10→E11→E12→E13

TABLE 8. The results of key users’ interests detection based on SPM-EE model.

Time	Hot Event	Event evolution relationship
Sun Dec 27 2015	E1	
Mon Dec 28 2015	E2	
	E3	E2→E3→E4
	E4	E2→E3→E4→E5
	E5	E2→E3→E4→E5
Tue Dec 29 2015	E7	
	E8	E7→E8
	E9	E7→E8→E9→E10
	E10	E7→E8→E9→E10→E11
	E11	E7→E8→E9→E10→E11
	E12	E7→E8→E9→E10→E11→E12
	E13	E7→E8→E9→E10→E11→E12→E13

nodes and the promoted commodities or events, which results in the difference between the ANS algorithm and the KKT algorithm in the final set of activated nodes.

As can be seen from Figure 5, the difference between the final set of activated nodes selected by ANS algorithm and KKT algorithm becomes larger as the number of active nodes increases gradually. This is because the value represents the proportion of cumulative similarity in ANSM metrics. As the number of seed nodes increases gradually, the ANS algorithm will tend to search for user nodes with higher cumulative similarity as seed nodes. At this time, the selected seed nodes are likely to be different from the seed nodes selected by KKT algorithm. Therefore, the difference between the final set of activated nodes selected by ANS algorithm and KKT algorithm will gradually increase. This also proves that ANS algorithm is more suitable to find user nodes interested in the promotion of goods or events, rather than focusing only

on the number of last activated nodes, and also proves the effectiveness of ANS algorithm.

4) THE EVENT EVOLUTION ANALYSIS

As is shown in Table 7 and 8, we can easily discover the more similar events during event evolution based on the SPM-EE model compared with HEE model. This is because a Similarity Priority Mechanism-based Event Evolution model, named SPM-EE model is proposed to judge correlation between events, which can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution. At the same time, we can also find which hot events have a long event evolution chain, which indicates the popularity of hot events compared with the HEE model [34].

VI. CONCLUSION AND FUTURE WORK

In this paper, the research on multi-topic sensitive influence maximization is carried out. MTL-IC model is proposed. The model incorporates multi-topic factors and considers the authority and centrality of user nodes, which makes the results more realistic. Since the existing influence maximization algorithm cannot be applied to MTL-IC model based on multi-topic, an ANSM metric is proposed by combining multi-topic factors with classical greedy algorithm, and an ANS algorithm is given according to the metric. A comparative experiment was conducted on Twitter real data sets with different propagation models and influence maximization algorithms. The performance advantages of the new model MTL-IC and ANS were analyzed by using the number of last activated nodes IS (S) and the new metric SIS (S) proposed in this paper. At the same time, a Similarity Priority Mechanism-based Event Evolution model, named SPM-EE model is proposed to judge correlation between events, which can update the seed users according to their changeable interest in time, which largely improve the precision of event evolution. At the same time, we can also find which hot events have a long event evolution chain, which indicates the popularity of hot events compared with the HEE model.

As an emerging intelligence event evolution computing paradigm, SPM-EE mainly involves satisfying the intelligent service requirements to adapt machine learning and natural language processing. Thus SPM-EE can accelerate the content deliveries and improve the quality of influence maximization and applications, which is attracting more and more interest from academia and industry because of its advantages in throughput, influence scope, network scalability and intelligence.

Although the model proposed in this paper incorporates many factors, such as multi-topic, authority and centrality, it is still static. In fact, with the passage of time, the interest and network structure of user nodes are also changing, so considering dynamic factors becomes a research direction. This paper uses Topical HITS algorithm to calculate the authoritativeness and centrality of user nodes, but there are many other alternative methods for parametric learning of

models. Meanwhile, SPM-EE also brings us new challenges, such as data moving and management, intelligent analysis and decisions. Thus in the future, we can study other more effective ways to estimate these parameters, so that the model is more accurate and closer to reality.

REFERENCES

- [1] Y. Lin, X. Wang, F. Hao, Y. Jiang, Y. Wu, G. Min, D. He, S. Zhu, and W. Zhao, "Dynamic control of fraud information spreading in mobile social networks," *IEEE Trans. Syst., Man, Cybern. Syst.*, early access, Aug. 9, 2014, doi: [10.1109/TSMC.2019.2930908](https://doi.org/10.1109/TSMC.2019.2930908).
- [2] X. Cheng, Y. Wu, G. Min, and A. Y. Zomaya, "Network function virtualization in dynamic networks: A stochastic perspective," *IEEE J. Sel. Areas Commun.*, vol. 36, no. 10, pp. 2218–2232, Oct. 2018.
- [3] W. Miao, G. Min, Y. Wu, H. Huang, Z. Zhao, H. Wang, and C. Luo, "Stochastic performance analysis of network function virtualization in future Internet," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 3, pp. 613–626, Mar. 2019.
- [4] H. Wang, Y. Wu, G. Min, J. Xu, and P. Tang, "Data-driven dynamic resource scheduling for network slicing: A deep reinforcement learning approach," *Inf. Sci.*, vol. 498, pp. 106–116, Sep. 2019.
- [5] T. Hofmann, "Probabilistic latent semantic indexing," in *Proc. 22nd Annu. Int. ACM SIGIR Conf. Res. Develop. Inf. Retr. (SIGIR)*, 1999, pp. 50–57, doi: [10.1145/312624.312649](https://doi.org/10.1145/312624.312649).
- [6] D. M. Blei, A. Y. Ng, and M. I. Jordan, "Latent Dirichlet allocation," *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, Mar. 2003.
- [7] Y. Papanikolaou and G. Tsoumakas, "Subset labeled LDA: A topic model for extreme multi-label classification," in *Proc. Int. Conf. Big Data Anal. Knowl. Discovery*, 2018, pp. 152–162.
- [8] D. Backenroth, Z. He, K. Kiryluk, V. Boeva, L. Pethukova, E. Khurana, A. Christiano, J. D. Buxbaum, and I. Ionita-Laza, "FUN-LDA: A latent Dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation: Methods and applications," *Amer. J. Hum. Genet.*, vol. 102, no. 5, pp. 920–942, May 2018.
- [9] C. Huang, Q. Wang, D. Yang, and F. Xu, "Topic mining of tourist attractions based on a seasonal context aware LDA model," *Intell. Data Anal.*, vol. 22, no. 2, pp. 383–405, Mar. 2018.
- [10] D. M. Blei and J. D. Lafferty, "Correlated topic models," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 18, no. 2, pp. 147–156, 2005.
- [11] D. M. Blei and J. D. Lafferty, "Dynamic topic models," in *Proc. 23th Int. Conf. Mach. Learn.*, 2006, pp. 113–120.
- [12] A. McCallum, X. Wang, and A. Corrada-Emmanuel, "Topic and role discovery in social networks with experiments on enron and academic email," *J. Artif. Intell. Res.*, vol. 30, pp. 249–272, Oct. 2007.
- [13] P. Domingos and M. Richardson, "Mining the network value of customers," in *Proc. 7th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2001, pp. 57–66.
- [14] D. Kempe, J. Kleinberg, and É. Tardos, "Maximizing the spread of influence through a social network," in *Proc. 9th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2003, pp. 137–146.
- [15] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance, "Cost-effective outbreak detection in networks," in *Proc. 13th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2007, pp. 420–429.
- [16] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "CELFF+: Optimizing the greedy algorithm for influence maximization in social networks," in *Proc. 20th Int. Conf. Companion World Wide Web (WWW)*, 2011, pp. 47–48.
- [17] W. Chen, Y. Wang, and S. Yang, "Efficient influence maximization in social networks," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, 2009, pp. 199–208.
- [18] X. Liu, M. Li, S. Li, S. Peng, X. Liao, and X. Lu, "IMGPU: GPU-accelerated influence maximization in large-scale social networks," *IEEE Trans. Parallel Distrib. Syst.*, vol. 25, no. 1, pp. 136–145, Jan. 2014.
- [19] W. Chen, C. Wang, and Y. Wang, "Scalable influence maximization for prevalent viral marketing in large-scale social networks," in *Proc. 16th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD)*, 2010, pp. 1029–1038.
- [20] M. Kimura and K. Saito, "Tractable models for information diffusion in social networks," in *Proc. 10th Eur. Conf. Princ. Pract. Knowl. Discovery Databases*, 2006, pp. 259–271.
- [21] W. Chen, Y. Yuan, and L. Zhang, "Scalable influence maximization in social networks under the linear threshold model," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2010, pp. 88–97.
- [22] A. Goyal, W. Lu, and L. V. S. Lakshmanan, "SIMPACT: An efficient algorithm for influence maximization under the linear threshold model," in *Proc. IEEE 11th Int. Conf. Data Mining*, Dec. 2011, pp. 211–220.
- [23] M. Kimura, K. Saito, and R. Nakano, "Extracting influential nodes for information diffusion on a social network," in *Proc. 22th Nat. Conf. Artif. Intell.*, 2007, pp. 1371–1376.
- [24] J. Kim, S.-K. Kim, and H. Yu, "Scalable and parallelizable processing of influence maximization for large-scale social networks?" in *Proc. IEEE 29th Int. Conf. Data Eng. (ICDE)*, Apr. 2013, pp. 266–277.
- [25] C. Zhou, P. Zhang, J. Guo, and L. Guo, "An upper bound based greedy algorithm for mining top-k influential nodes in social networks," in *Proc. 23rd Int. Conf. World Wide Web (WWW) Companion*, 2014, pp. 1039–1048.
- [26] Q. Yan, S. Guo, and D. Yang, "Influence maximizing and local influenced community detection based on multiple spread model," *Adv. Data Mining Appl.*, vol. 3, no. 2, pp. 82–95, 2011.
- [27] A. Galstyan, V. Musoyan, and P. Cohen, "Maximizing influence propagation in networks with community structure," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 79, no. 5, May 2009, p. 102.
- [28] J. Hu, K. Meng, X. Chen, C. Lin, and J. Huang, "Analysis of influence maximization in large-scale social networks," *ACM SIGMETRICS Perform. Eval. Rev.*, vol. 41, no. 4, pp. 78–81, Apr. 2014.
- [29] E. Cohen, D. Delling, T. Pajor, and R. F. Werneck, "Sketch-based influence maximization and computation: Scaling up with guarantees," in *Proc. 23rd ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2014, pp. 629–638.
- [30] G. Li, S. Chen, J. Feng, K.-L. Tan, and W.-S. Li, "Efficient location-aware influence maximization," in *Proc. ACM SIGMOD Int. Conf. Manage. Data (SIGMOD)*, 2014, pp. 87–98.
- [31] T. Zhou, J. Cao, B. Liu, S. Xu, Z. Zhu, and J. Luo, "Location-based influence maximization in social networks," in *Proc. 24th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2015, pp. 1211–1220.
- [32] L. Liu, J. Tang, J. Han, M. Jiang, and S. Yang, "Mining topic-level influence in heterogeneous networks," in *Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM)*, 2010, pp. 199–208.
- [33] Y. Zhang, J. Zhou, and J. Cheng, "Preference-based Top-K influential nodes mining in social networks," in *Proc. IEEE 10th Int. Conf. Trust, Secur. Privacy Comput. Commun.*, Nov. 2011, pp. 1512–1518.
- [34] N. Barbieri, F. Bonchi, and G. Manco, "Topic-aware social influence propagation models," *Knowl. Inf. Syst.*, vol. 37, no. 3, pp. 555–584, Dec. 2013.
- [35] C. Zhang, J. Sun, and K. Wang, "Information propagation in microblog networks," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, 2013, pp. 190–196.
- [36] Y. Li, D. Zhang, and K.-L. Tan, "Real-time targeted influence maximization for online advertisements," *Proc. VLDB Endowment*, vol. 8, no. 10, pp. 1070–1081, Jun. 2015.
- [37] X. Sun, Y. Wu, L. Liu, and J. Panneerselvam, "Efficient event detection in social media data streams," in *Proc. IEEE Int. Conf. Comput. Inf. Technol., Ubiquitous Comput. Commun., Dependable, Autonomous Secure Comput., Pervas. Intell. Comput.*, Oct. 2015, pp. 1711–1717.
- [38] L.-L. Shi, L. Liu, Y. Wu, L. Jiang, and J. Hardy, "Event detection and user interest discovering in social media data streams," *IEEE Access*, vol. 5, pp. 20953–20964, 2017.
- [39] L. Shi, Y. Wu, L. Liu, X. Sun, and L. Jiang, "Event detection and identification of influential spreaders in social media data streams," *Big Data Mining Anal.*, vol. 1, no. 1, pp. 34–46, Mar. 2018.
- [40] Y. Guo, L. Liu, Y. Wu, and J. Hardy, "Interest-aware content discovery in Peer-to-Peer social networks," *ACM Trans. Internet Technol.*, vol. 18, no. 3, pp. 1–21, May 2018.
- [41] L. Liu, N. Antonopoulos, M. H. Zheng, Y. Z. Zhan, and Z. J. Ding, "A socioecological model for advanced service discovery in machine-to-machine communication networks," *ACM Trans. Embedded Comput. Syst.*, vol. 15, no. 2, pp. 38:1–38:26, 2016.
- [42] L. Jiang, L. Shi, L. Liu, J. Yao, B. Yuan, and Y. Zheng, "An efficient evolutionary user interest community discovery model in dynamic social networks for Internet of people," *IEEE Internet Things J.*, vol. 6, no. 6, pp. 9226–9236, Dec. 2019.
- [43] L. Jiang, L. Shi, L. Liu, J. Yao, and M. A. Yousuf, "User interest community detection on social media using collaborative filtering," in *Wireless Networks*. Cham, Switzerland: Springer, Feb. 2019.

[44] L.-L. Shi, L. Liu, Y. Wu, L. Jiang, and A. Ayorinde, "Event detection and multi-source propagation for online social network management," *J. Netw. Syst. Manage.*, vol. 28, no. 1, pp. 1–20, Jan. 2020.

[45] Y. Wu, C. G. Yan, L. Liu, Z. J. Ding, and C. J. Jiang, "An adaptive multilevel indexing method for disaster service discovery," *IEEE Trans. Comput.*, vol. 64, no. 9, pp. 2447–2459, Sep. 2015.

[46] L.-L. Shi, Y. Wu, L. Liu, X. Sun, and L. Jiang, "Event detection and key posts discovering in social media data streams," in *Proc. IEEE Int. Conf. Internet Things (iThings) IEEE Green Comput. Commun. (Green-Com) IEEE Cyber, Phys. Social Comput. (CPSCom) IEEE Smart Data (SmartData)*, Jun. 2017, pp. 1046–1052.

[47] D. Wang, Z. Wen, H. Tong, C.-Y. Lin, C. Song, and A.-L. Barabási, "Information spreading in context," in *Proc. 20th Int. Conf. World Wide Web (WWW)*, 2011, pp. 735–744.

[48] J. Ren, D. Zhang, S. He, Y. Zhang, and T. Li, "A survey on end-edge-cloud orchestrated network computing paradigms: Transparent computing, mobile edge computing, fog computing, and cloudlet," *ACM Comput. Surv.*, vol. 52, no. 6, pp. 1–36, Jan. 2020, Art. no. 125.

[49] W. Z. Khan, E. Ahmed, S. Hakak, I. Yaqoob, and A. Ahmed, "Edge computing: A survey," *Future Gener. Comput. Syst.*, vol. 97, pp. 219–235, Aug. 2019.



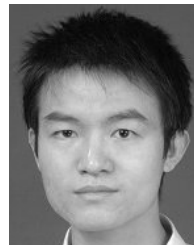
LEI-LEI SHI received the B.S. degree from Nantong University, Nantong, China, in 2012, and the M.S. and Ph.D. degrees from Jiangsu University, Zhenjiang, China, in 2015 and 2020, respectively. He is currently a Lecturer with the School of Computer Science and Telecommunication Engineering, Jiangsu University. His research interests include event detection, data mining, social computing, and cloud computing.



YAN WU received the M.S. degree from the Shandong University of Science and Technology, Qingdao, China, in 2009, and the Ph.D. degree from Tongji University, Shanghai, China, in 2014. He is currently a Lecturer with the School of Computer Science and Telecommunication Engineering, Jiangsu University, China. His research interests include formal methods, service-oriented computing, and cloud computing.



JUN GE received the B.S. degree from the University of South China, China, in 2001, and the M.S. degree from Jiangsu University, China, in 2010, where he is currently pursuing the Ph.D. degree with the School of Computer Science and Telecommunication Engineering. He is also an Associate Professor with the School of Institute of Information Engineering, Suqian University, China. His research interests include social networks and cloud computing.



JIE LIU received the B.S. degree from the Nanjing University of Aeronautics and Astronautics, China, in 2005, and the M.S. degree from Southeast University, China, in 2012. He is currently a Lecturer with the School of Institute of Information Engineering, Suqian University, China. His research interests include social networks, cloud computing, and the Internet of Things.

...