BIG DATA MINING AND ANALYTICS
ISSN 2096-0654 15/15 pp217–243
DOI: 10.26599/BDMA.2023.9020021
Volume 7, Number 1, March 2024

A Survey on Event Tracking in Social Media Data Streams

Zixuan Han, Leilei Shi*, Lu Liu*, Liang Jiang, Jiawei Fang, Fanyuan Lin,
Jinjuan Zhang, John Panneerselvam, and Nick Antonopoulos

Abstract: Social networks are inevitable parts of our daily life, where an unprecedented amount of complex data corresponding to a diverse range of applications are generated. As such, it is imperative to conduct research on social events and patterns from the perspectives of conventional sociology to optimize services that originate from social networks. Event tracking in social networks finds various applications, such as network security and societal governance, which involves analyzing data generated by user groups on social networks in real time. Moreover, as deep learning techniques continue to advance and make important breakthroughs in various fields, researchers are using this technology to progressively optimize the effectiveness of Event Detection (ED) and tracking algorithms. In this regard, this paper presents an in-depth comprehensive review of the concept and methods involved in ED and tracking in social networks. We introduce mainstream event tracking methods, which involve three primary technical steps: ED, event propagation, and event evolution. Finally, we introduce benchmark datasets and evaluation metrics for ED and tracking, which allow comparative analysis on the performance of mainstream methods. Finally, we present a comprehensive analysis of the main research findings and existing limitations in this field, as well as future research prospects and challenges.

Key words: Event Detection (ED); event propagation; event evolution; social networks

1 Introduction

With the rapid development of Internet technology, social media has become one of the major tools for communication, information acquisition, marketing, and social interaction in today's society. This emerging communication medium has not only changed the way people access and share information, but has also had a broad impact on business and society. Online Social Networks (OSNs) are now densely networked and fast-spreading, and characterize high commercial value[1–4]. Social media has become an "amplifier" for public opinion events[5], especially for the rampant spread of false information. The unprecedented spread of false information in social media brings complex challenges

• Zixuan Han, Leilei Shi, Jiawei Fang, Fanyuan Lin, and Jinjuan Zhang are with School of Computer Science and Communication Engineering and Jiangsu Key Laboratory of Security Technology for Industrial Cyberspace, Jiangsu University, Zhenjiang 212013, China. E-mail: hanshu8880@gmail.com; 937109472@qq.com; 1336823278@qq.com; 2822174280@qq.com; 1737901499@qq.com.
• Lu Liu and John Panneerselvam are with School of Computing and Mathematical Sciences, University of Leicester, Leicester, LE1 7RH, UK. E-mail: l.liu@leicester.ac.uk; jkp15@leicester.ac.uk.
• Liang Jiang is with Ocean College, Jiangsu University of Science and Technology, Zhenjiang 212100, China. E-mail: jiangliang@just.edu.cn.
• Nick Antonopoulos is with University Executive Office, Edinburgh Napier University, Edinburgh, EH11 4BN, UK. E-mail: N.Antonopoulos@napier.ac.uk.
* To whom correspondence should be addressed.
Manuscript received: 2022-11-03; revised: 2023-04-24; accepted: 2023-08-11

in the management of public opinions whilst restoring social harmony, security, and stability. Event tracking has been regarded as an efficient means of regulating opinions that are posted in social media aided with accurate tracking of events in real time[6]. Moreover, event tracking facilitates accurate predictions of future events in social media, with which a timely discovery of major events with a potential to affect social security can be achieved[7−9].

As of 2022, the number of social network users worldwide has exceeded 4.62 billion, which accounts for 58.4% of the global population[10]. This trend demonstrates how social networks have become deeply integrated into people's daily lives, becoming an indispensable and important part of it. Its rapid popularity has greatly improved interpersonal connections and communication, as well as the dissemination of information and knowledge. For example, Twitter accounts for around 320 million active users every month, while Facebook has over 2.9 billion registered users, and these numbers are continuously growing. Online social networks not only provide a platform for people to interact socially, but also offer a wide range of applications in the fields of business, media, politics, and so on. By leveraging the data generated by users in social networks, data-based methods can be analyzed and developed to monitor and control the spread of disinformation. This data provides a valuable resource to help develop more accurate and reliable algorithms to further improve the efficiency and accuracy of disinformation detection.

Social networks can be viewed as virtual representations of real-world user relationships. By monitoring the development of popular events on social media, it is possible to better identify social phenomena and potential social problems that may be difficult to detect in real-world networks. By analyzing the data generated by social network users, we can better understand the nature of social phenomena and take appropriate measures to address related social issues[11]. Event tracking in social networks driven by detection, propagation, and evolution based models may offer novel methodologies for the identification and monitoring of public opinion events[12−14]. The efficiency and performance of these three models are inextricably linked to the dependability of event tracking in social networks. Therefore, it becomes especially crucial to deeply investigate efficient algorithms for Event Detection (ED), propagation, and evolution in social networks. This will not only help reveal the fundamentals and behavioral patterns of social networks, but also enable the development of more effective countermeasures to protect users from disinformation and inappropriate content[15−17].

A graph neural network model is a deep learning model for processing graph-structured data. In event tracking, graph neural networks can help us better understand the relationships among events, and can extract meaningful information from large amounts of data. They are highly adaptive, so they can detect, propagate, and evolve different types of events, including natural disasters, political events, or trending topics on social media. To this end, events tracking techniques with a combination of deep learning algorithms[16] and graph neural network models[17] have emerged recently.

However, the continuous, rapid, and variable nature of social networks leads to the user interest shift phenomenon, which brings several challenges in public opinion detection[18], propagation[19], and evolution[20] models, thereby restraining the accuracy and efficiency of existing models for many real-life scenarios. Existing models of event tracking in social media encounter the below issues amid the user interest shift phenomena:

(1) Existing public opinion detection methods characterize high requirements of pre-processing to mitigate data noise, and largely ignore the multi-dimensional situational features of public opinion events, such as their syntactic structure and potential dependence relationships. It is difficult to analyze the deep semantic expression mechanism of major public opinion events under data noise, resulting in insufficient detection and low model accuracy.

(2) Existing public opinion propagation methods ignore the impact of user interest shift phenomenon on the continuous propagation range of public opinion under multiple communities, making it difficult to summarize the true propagation rules of public opinion events. Thus, existing models cannot analyze the adaptive propagation mechanism during continuous propagation of public opinion events.

(3) The majority of existing public opinion evolution approaches are solely focused on identifying and tracking multiple static graph structures, and ignore the user interest evolution characteristics across networks. This limits existing models from accurately identifying new and old public opinion events evolution

relationships under the existence of user interest shift phenomenon, thereby resulting in low efficiency of constructing chains of public opinion events evolution tracking graphs.

With this in mind, this paper provides a comprehensive analysis of various techniques that are used to track events by exploiting social media data. We thoroughly examine the strategies and performance of existing models, along with reviewing their data collection methods. Important contributions of this paper are as follows:

(1) This paper systematically reviews various techniques of machine learning and deep learning in relevance to their applications in tracking events from social media.

(2) A large selection of existing ED methods is critically reviewed with a detailed analysis of their methodology and performance.

(3) We identify current research gaps in the field of ED in social networks, and derive strategies for future research directions.

The remainder of this work is structured as follows: In Section 2, we conduct a detailed survey of event tracking techniques, by analyzing and comparing a large selection of relevant techniques. Section 3 discusses related existing works of event tracking methods. Section 4 presents a summary of our analysis along with highlighting the challenges encountered in our analysis. This paper's conclusion and a discussion of potential future study topics are provided in Section 5.

2 Event Detection

In the digital age, social networks have built a vast network of connections among Internet users through the convenience and diversity of online platforms[21–23]. These connections often revolve around shared interests or user interactions, creating an extension of the actual social relationships among friends and family. Popular social network platforms, such as instant messaging and microblogging, leverage these online social connections to offer users options for posting, discovering, and sharing information[24]. The availability, accessibility, and effective communication channels of diverse information on social networking platforms continue to attract more users, thus opening up new avenues for research on event tracking techniques in social networks[25–27]. Importantly, social network platforms, such as the Sina API and Twitter API, are open and available for researchers[28–31], which facilitates identifying and evaluating hot events in social media and lays the foundation for further research into event tracking methodologies[32–34].

In this regard, we start by reviewing the definition of event tracking and highlighting the primary challenges associated with it[35–38]. Next, we introduce and critically review the mainstream event tracking methods, which involve three primary technical steps: ED, event propagation, and event evolution. Finally, we introduce benchmark datasets and evaluation metrics for ED and tracking, which facilitate fellow researchers with empirical analysis on mainstream ED methods. Finally, we summarize the main research findings in this area in recent years and potential research directions for future work for the reference of subsequent researchers.

In social networks, the optimization and improvement of ED models are important core issues in opinion tracking technology. An effective ED model should consider multiple aspects of public opinion events to improve the effectiveness of the algorithm, such as event scale, timeliness, locality, and other factors, and be able to predict the future development trend of events while improving the accuracy and efficiency of public opinion events detection, thus providing a scientific basis for relevant decisions. ED in social networks refers to identifying and analyzing the topics, opinions, and emotional tendencies related to the current hot events in social media platforms by monitoring the Shanghai volume of text, images, videos, and other information using advanced natural language processing technology and machine learning algorithms. With the continuous development of topic clustering techniques, ED has been significantly improved in semantic analysis. By automatically classifying and labeling large amounts of text data, topic clustering techniques can help us better understand and summarize event information, and generate relevant models. The continuous optimization of this technology allows ED algorithms to more accurately identify and extract useful information from massive and complex data, providing more accurate and efficient decision support and services for various industries. By classifying and clustering text data on social media platforms, researchers can describe popular events in social networks by grouping posts on similar topics into the same category. This quantitative analysis approach allows us to more objectively study and assess the impact and trends of epidemic events. At the same time, with the wide application of artificial intelligence and machine learning technologies, the efficiency and accuracy of ED and analysis continue to improve, providing strong support for us to grasp the current situation and trends of events. As a result, social network based ED research has taken popularity in recent years, with many scholars contributing to the study and design. Social networks provide users with a wide range of ways to engage in online social activities, as users can extend and create their own social networks by following, retweeting, and commenting on content on social media. This type of interaction not only enhances the connection and interaction among users, but also facilitates information sharing and knowledge dissemination among users. Moreover, users of social networks can obtain information provided by other connected users, and can continue to spread information they obtain.

ED plays a crucial role in the information extraction process. Its main purpose is to extract structured and event-related information from the natural language of social media posts. This information can include aspects, such as the time and place of the event, the participants, and the nature of the event. Through ED, we can quickly and accurately understand the situation and background of current hot events, monitor and analyze public opinion, and take timely measures to respond to possible problems. ED finds applications in various business domains, such as knowledge atlases and financial analysis. In Automatic Content Extraction (ACE) evaluation, event extraction is a process that includes two subtasks: ED and event element extraction. Among them, ED is mainly responsible for discovering posts related to current hot events from the massive text data; while event element extraction further analyzes these posts to extract information, such as the time, place, participants, and event attributes of the events[39]. ED serves as the foundation for event extraction in online social networks, whereby facilitating the extraction and classification of events from text. The results obtained from ED directly impact the efficiency of event element extraction[40]. Figure 1 provides an overview of the ED process and its associated phases[41].

The core goal of ED is to extract event trigger words from social media text data and categorize them into different event types[42]. Using natural language processing techniques and machine learning algorithms, ED identifies content related to current hot events in a fast and accurate way, and classifies and analyzes it. Traditional ED methods typically involve using a set of Natural Language Processing (NLP) tools to create features, which are subsequently classified using a statistical classification model[43]. Features extracted from texts can be categorized into two groups, such as lexical and sentence-level characteristics. The syntax and context dependence of the whole phrase are known as sentence-level features, whereas the semantic or prior knowledge of vocabulary is referred to as vocabulary-level features. However, the use of advanced NLP tools in these approaches can lead to error propagation, where manual feature engineering construction is a time-consuming and arduous process.

Compared with traditional machine learning algorithms, neural networks can automatically learn richer and more useful features from large-scale natural language text, and also have better performance in event recognition. The development of this technique provides new ideas and methods for processing and

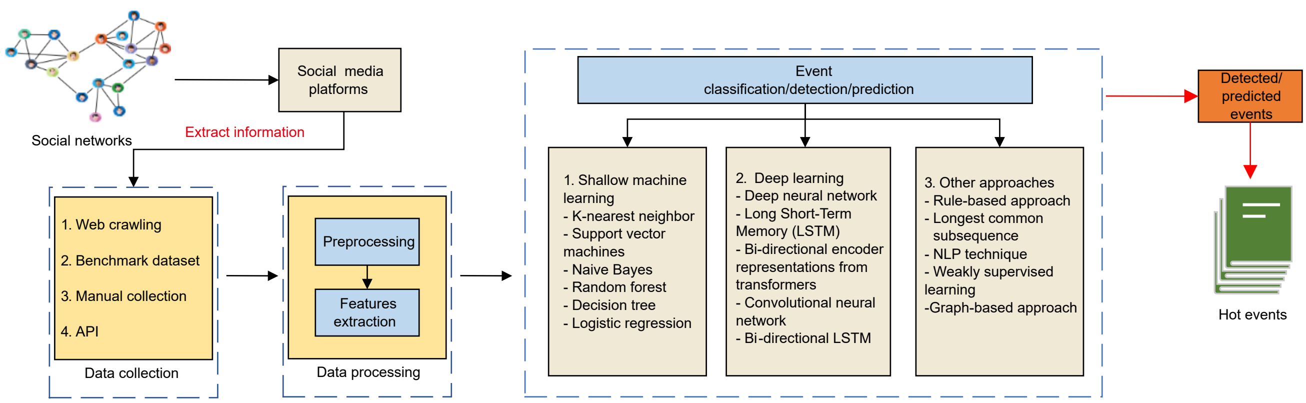


Fig. 1 Main phases of the ED model.

analyzing social media data. Recently, a specific type of Recurrent Neural Network (RNN) has been proposed that has been used to jointly extract event trigger words and event elements in social media texts[44]. This approach allows the processing of multiple related events simultaneously and enables the fast and accurate extraction of key information related to hot events from the data. In addition to this, a new approach is proposed in research, which takes advantage of the Bi-directional Long and Short-Term Memory (Bi-LSTM) model and Convolutional Neural Networks (CNN), thus enabling more accurate recognition of events in social media texts[45]. This approach collects both sequential and structural semantic information from the context and integrates them, thus improving the accuracy and efficiency of event recognition. The model proposed by Jin et al. in their research combines CNN with Bidirectional Gated Recurrent Units (Bi-GRU) for trigger word extraction. However, these models did not adequately consider the contextual information of the text, and relying solely on separate feature extraction may be insufficient[46].

Attention mechanisms have gained popularity in ED due to their ability to collect global information and focus on key characteristics[47, 48]. A mixed neural network model based on attention mechanisms has been proposed, which utilizes CNNs to extract data of varying granularity to generate sentences[49]. This approach extracts contextual semantic information features from social media texts by combining sentence feature representations to improve the performance of sentiment classification. In this process, the attention mechanism is applied to calculate the attention score for each word and assign different learning weights to different features. This technique can help us better understand and characterize the sentiment tendencies of social media users and provide more accurate opinion analysis and prediction for organizations such as enterprises and governments. In Chen et al.'s study[50], they placed a Bi-GRU layer between the convolutional and pooling layers of the CNN to extract global features from the local features obtained from the CNN. This approach enables us to more effectively capture contextual information in social media data and extract additional valuable features. In addition, attention mechanism is also applied in this method to assign dynamic learning weights to the extracted features.

However, despite their effectiveness, these models have notable flaws as they only examine sentence-level text characteristics, and ignore word-level features and vocabulary location information. Additionally, the pooling operation in CNN leads to a loss of location information, whereby rendering the ED task ineffective. ED is crucial for popular social network ED and tracking systems, where an efficient detection model should significantly boost accuracy and efficiency of ED.

Research approaches for ED can be summarized into two types: feature engineering approaches and neural network based approaches. The former mainly relies on manual extraction of key features to achieve ED, while the latter automatically learns features from data with the help of deep learning techniques in machine learning. This approach not only alleviates the reliance on manual feature engineering, but also enables a more comprehensive mining of feature information in the data to improve the accuracy and efficiency of ED. Scholars have attempted to address the challenge of ED using token level features and structured features using deep learning techniques. Recent progress has been made in neural network approaches that embed contextual semantic information into low-dimensional spaces, whereby treating ED as a word-by-word classification task. Bidirectional Encoder Representations from Transformers (BERT) has been frequently employed in event extraction tasks, especially since the introduction of pre-training language models[51].

Although the fully supervised ED algorithms described above have made significant progress, their restricted data size precludes them from obtaining better performance[52]. Furthermore, the overfitting problem restrains the efficiency of fully supervised deep learning algorithms, particularly in real-world applications. A novel weak supervision pair based on data enhancement Anti training technique has been proposed, which exploits a hybrid text anti training approach based on Bert-based Mix-text A Dversarial (BMAD) training[53]. In this method, unsupervised data is initially built from the original representation, then the unsupervised data is upgraded. Then, to build virtual training data, the BMAD method concentrates on a novel data augmentation approach called mix text. This method attempts to enhance efficiency by training the model's generalization ability along with avoiding overfitting issues resulting from erroneous and noisy data. Finally, to improve the model's resilience, a technique based on mix text's countermeasure training approach has been developed.

As shown in Fig. 2, this article comprehensively considers and categorizes the basic elements of social ED from multiple aspects, such as data sources, challenges in data collection, key features, evaluation methods, visualization, application, and information fusion. The source and quality of data have a crucial impact on the efficiency of social ED. Key features are essential elements for identifying and tracking the development process of events and are crucial for improving the detection accuracy. Visualization technology can visually present event-related features and trends in social media data, thus helping us to better understand the development process and impact of hot events. Information fusion, on the other hand, is one of the important ways to improve the accuracy and efficiency of social media ED. It enables the combination of different data sources, algorithms or models to obtain more accurate and comprehensive analysis results. By combining multiple technical approaches, we can more comprehensively describe and analyze the events occurring in today's society and provide more accurate and useful reference information for decision makers. Therefore, in social ED, it is necessary to use multiple methods and technologies comprehensively to achieve better results.

Traditional approaches of ED activity rely heavily on hand-engineered features, which can achieve excellent results in specific domains but face significant challenges when migrating to a different language or changing annotation standards. Liu et al.[54] presented a supervised attention strategy for encoding argument information in ED. To achieve ED, Yan et al.[55] proposed a new approach using an aggregated attention graph convolutional network model based on dependency trees. This model can model and analyze social media text data and extract key information related to hot events from them. This approach enables more accurate identification of hot events on social media and quick access to key features and trends of the events. Wang et al.[56] proposed a unique multilayer residual and gating-based convolutional neural network architecture to capture more scale contextual information by increasing the perceptual field. This approach can effectively capture multi-level features in social media data and provide more refined and accurate analysis results for ED.

Currently, feature-based ED models[57] and document-based ED models are the most widely used forms of ED models. Feature-based ED algorithms primarily investigate and evaluate features in data streams[58, 59], and then identify or cluster certain representative features to detect new hot events or unknown events[60]. Sakaki et al.'s research[61] utilizes real-time social media data streams to identify various characteristics of earthquakes, including the earthquake's geographical distribution, the geographic locations of affected users, the intensity of the earthquake's epicenter, and scenarios related to emergency situations. They transformed the ED problem into a classification problem, abstracted the real situation of the earthquake's impacted users from findings informed by cluster analysis, and determined the affected users' geographical position and distribution range. Long et al.[62] proposed a comprehensive ED model that incorporates several additional ED features, such as the ability to select keywords from microblog data in social networks for ED. In Weng and Lee's study[63], they transformed

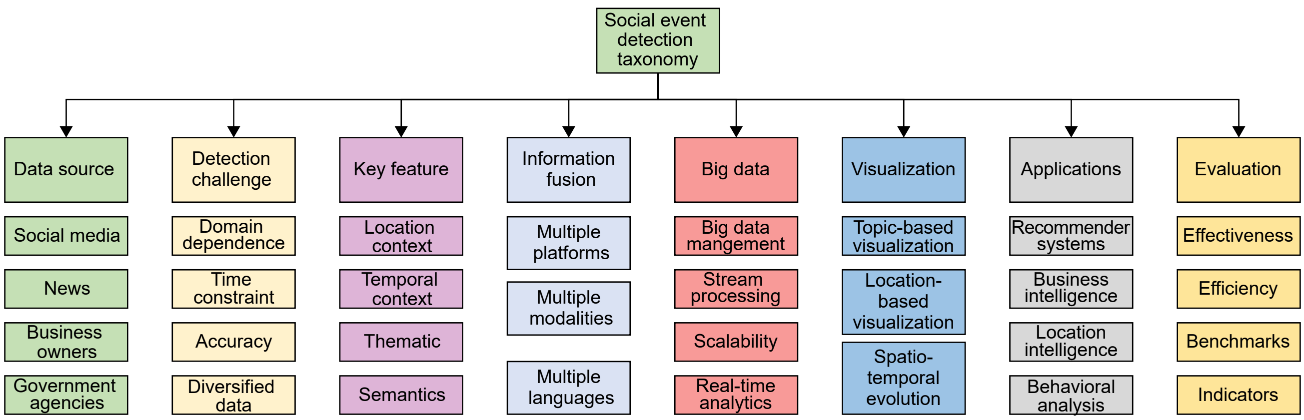


Fig. 2 Social ED taxonomy.

Twitter data into signals, applied wavelet analysis, used autocorrelation functions to filter out unnecessary lexical information, denoised the data stream text, and calculated text similarity to construct a network. Ultimately, they organized the discovered events into a list of terms with similar features.

Matuszka et al.[64] demonstrated that ED can be achieved by introducing the concept of the lifespan of each feature keyword and utilizing the changing frequency of predefined feature keywords. He has proposed a real-time ED model based on Twitter data, which takes into account the lifecycle of each feature keyword. The average behavior over time (such as average frequency change) is represented in such a way to enable the ED model to detect new feature keywords in real-time. This technique can more accurately identify and analyze hot events on social media and obtain key information about the events in a timely manner. Allan[65] introduced a real-time ED algorithm based on convolutional neural networks, with the aim of detecting unexpected events in tweets. It uses these tweets as input data for ED, and employs a convolutional neural network to detect the characteristic keywords related to earthquake in real time and with high accuracy. To summarize, while the feature-based ED model considers the characteristics of popular events as well as content information such as topic keywords, it disregards the potential impact of a large amount of noisy data in social networks. As a result, some high-quality event subject keywords are not detected in a timely and accurate manner.

The document-based ED model detects events primarily by clustering similar texts based on the contents of document data. This class of models has long been studied in the field of topic detection and tracking. Topic detection and tracking is a data processing technique that is used for automatically identifying and tracking known topics in news media streams[66]. Since the topic detection and tracking data mostly comprise long text documents, their data scale is usually small. On the contrary, the scale of data in social networks is usually large, irregular, and relatively colloquial. As a result, the direct application of this approach to the ED process on social networks does present some difficulties due to some of the limitations mentioned above. However, researchers are continuously improving existing algorithms and models to overcome these limitations and extend its applicability to a wider range of domains.

Meanwhile, the document-based ED model is a standard clustering process that analyzes and processes text information using a clustering algorithm to identify new events and/or merge existing events[67]. As a result, traditional clustering algorithms like divide-and-conquer and hierarchical clustering[68], K-means[69], and other clustering methods, can be used to detect events[70, 71]. Soucy and Mineau[72] introduced a data-weighted representation method based on Term Frequency-Inverse Document Frequency (TF-IDF) and performed clustering analysis by constructing a network relationship graph among texts. Finally, each popular event topic is represented by a series of network graphs in the temporal dimension. Kaleel and Abhari[73] proposed a new ED method based on clustering textual information to detect popular events. This method constructs a feature vector of microblog data using a high-dimensional incremental TF-IDF algorithm, and detects popular events using the approximate nearest neighbor fast lookup technique based on high-dimensional data. Giovanetti and Lancieri[74] proposed and implemented a novel online data processing system based on the Twitter platform, which is capable of real-time clustering and analysis of information on the Twitter platform. Unlike traditional ED methods, this online data processing system based on the Twitter platform has higher real-time performance and better scalability, and it can quickly identify and track popular events and provide real-time event analysis results. However, these studies cannot automatically determine the number of topics contained in the dataset, due to their diminishing detection ability under relatively large dataset.

Topic model based ED is a subset of document-based ED models[41], and has been an important research hot spot in recent years[75–77]. Topic models are suitable for detecting implicit topics in texts that generate semantic representations of documents, by mining and modeling the underlying semantic correlations among documents. Probabilistic topic models are a commonly used text analysis method for discovering hidden topics and semantic structures in text data. Two of the popular probabilistic topic models include Probabilistic Latent Semantic Analysis (PLSA)[78] and Latent Dirichlet Allocation (LDA)[79], as shown in Fig. 3. Both PLSA and LDA are probabilistic distribution-based models that can represent documents as a combination of topic and word distributions and learn model parameters by maximizing the likelihood function. Especially in the

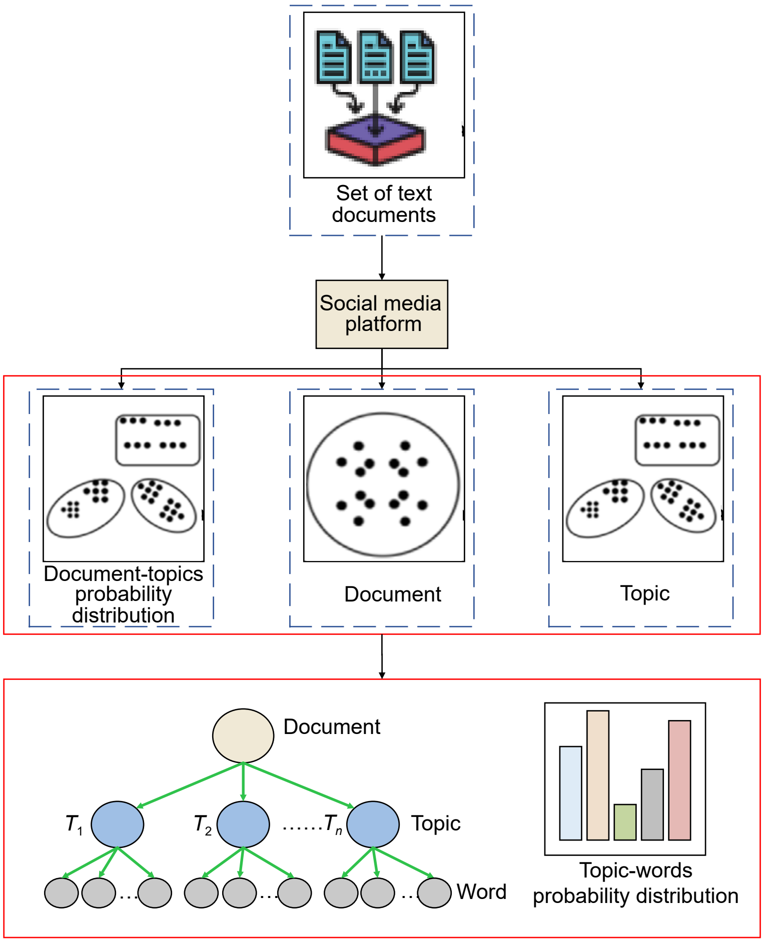


Fig. 3 Topic modeling of LDA.

processing of large-scale text data, probabilistic topic models have become a widely used tool by researchers. Besides, there are many improved and extended probabilistic topic models, such as Pachinko Allocation Model (PAM), Correlated Topic Model (CTM), etc., which have also achieved good results in different application scenarios. With the rapid growth of text data and the increasing demand of applications, the research and application of probabilistic topic models are also evolving. In the future, we can expect more efficient, accurate and flexible probabilistic topic models to appear, bringing new ideas and methods for text analysis and information processing.

The Bilateral Topic Model (BTM)[80] has been introduced to directly simulate the co-occurrence model of words throughout a document, with the motivation of solving the feature sparse problem of short texts. The BTM model has been improved using a Dirichlet process by exploiting keyword co-occurrence to mine topics in short texts for detecting popular events[81]. However, the efficiency of ED has not been significantly improved, and the ability to analyze the evolutionary relationship among popular event topics has not been fully explored. To address this issue, Li et al.[82] introduced BEE model for ED on the Sina Weibo platform. The BEE model employs the PLSA probabilistic topic model to model microblog text information and message events, as well as hashtag information. The BEE topic model uses a sliding window mechanism to build the event relationship in adjacent time periods and analyzes the evolution process among event topics to detect popular events, thereby improving the efficiency and accuracy of ED.

The Efficient eVent dEtection (EVE) model[77] uses the PLSA probabilistic topic model and the Expectation-Maximization (EM) parameter estimation method to detect events in microblog text information. Furthermore, data preprocessing operations are added prior to the PLSA topic model detection, which significantly improved the ED efficiency. Chen et al.'s research[83] conducted a thorough analysis of popular ED methods within the context of microblogging platforms. They observed that existing methods often significantly overlooked untagged microblogs, rendering them unable to detect events in such cases. To address this issue, they introduced a Time-Hashed label Latent Dirichlet Allocation (TH-LDA) ED model. This model combines microblog tags and posting times, employing the LDA topic modeling technique to enhance the accuracy of ED in Chinese microblogs. Similarly, Yu and Qiu[84] proposed the User Latent Word Dirichlet Multinomial Mixture (ULW-DMM) event topic detection model in combination with user latent features for topic detection and classification of short text microblogs, by extending the DMM topic model based on the user-LDA model. However, the topic distribution of users on social media is often ignored, which makes it more difficult to accurately identify users' interests and topics[85]. To solve this problem, Gunawan et al.[86] proposed the Twitter LDA topic model, which can accurately identify users' topics and reveal their interests and preferences more comprehensively by modeling various aspects of information such as user behavior.

However, this model relies on external resources to supplement short text contents to model topics, which adversely affects the quality of topic identification[87]. To make better use of the real-time data provided by social media users, Nolasco and Oliveira[88] proposed the concept of sub-events. Considering social media users as human sensors, real-time data about entities and events can be obtained by analyzing the information they post, such as tweets. At the same time, the detection of sub-events enriches our understanding of the main event, not only improves the quality of the subjects, but also helps us to understand all aspects of the event more comprehensively and solves the problem of scarcity of data on short textual tweets. By identifying and analyzing sub-events, we can capture the details and changes of the event more accurately, and thus understand the development process and influencing factors of the event more comprehensively. Similarly, Troudi et al.[66] investigated the extraction and tracking of popular microblog event topics using LDA models, one of which is in the literature. Researchers presented a better topic extraction technique, Microblog Feature Latent Dirichlet Assignment (MF-LDA), for extracting well-liked themes from microblog posts in order to more reliably extract event subjects from microblog posts. The model provides a better model based on the LDA topic model to determine the joint probability distribution of all terms and themes. The model incorporates five characteristics, including praise, comments, retweets, postings, and user authority. By modeling the multidimensional information of microblog content, the MF-LDA model can reflect the topics and hotspots of microblog content more comprehensively and improve the accuracy and reliability of topic extraction. At the same time, the model can also track the evolution of microblog events and reveal the change patterns and influencing factors of the events.

Furthermore, Gibbs sampling method has been proposed to obtain an optimal model parameter to considerably improve ED accuracy. Moreover, we gathered the most recent ED algorithms from various social media data streams and elaborated on the benefits of different algorithms by comparatively analyzing their respective optimization strategies with an emphasis on the data sources. Table 1 summarizes our analysis of ED methods in social networks. In addition, Fig. 4 presents a classification of various mainstream topic modeling methods addressed in this article.

In conclusion, as shown in Table 1, despite yielding some results, existing methods suffer several issues that affect the overall performance of their ED and tracking technology.

Important observations from our analysis of existing ED methods are listed as follows:

(1) The social network text data streams encompass a plethora of cluttered textual information, including low-quality microblogs, content generated by ordinary users, and topics that are both irrelevant to potential popular events and unpopular. In the topic model based ED, the prior parameters of the topic model must be manually configured in advance, resulting in a significant reduction in ED efficiency and accuracy. Traditional topic model-based ED methods require manual configuration of the topic model's a priori parameters in advance when performing topic modeling, which requires specialized domain knowledge and experience, and requires constant adjustment and optimization of the model to adapt to new data and contexts. Such a process is very time-consuming and labor-intensive, and also prone to bias and misclassification of detection results. To solve the

Table 1 Summary of ED on various social media.

Index	Reference	Media	Method	Feature	Limitation	Merit
1	Shi et al.[26]	Twitter	Bag-of-words based approach: BOWED model and heuristics EAAS	Unique words	● Slow in progress ● Larger dataset is required for validation	● Proposing a three-phase based incremental clustering algorithm
2	He and Duan[40]	News article	Joint annotation model based on CRF multi-task learning	Multimodality data	● Multi-tab situations for event elements cannot be completely eliminated	● Effectively alleviating the problems of small data scale and data imbalance
3	Zhan et al.[43]	Multimedia	Chinese event extraction method based on HMM and multi-stage approach	Multimodality data	● Effect of segmentation is limited ● Algorithms cannot process long-term information	● Alleviating class imbalance problem ● Improving matching degree
4	Feng et al.[45]	News article	Hybrid neural network	Multimodality data	● Limitation with event structure module	● Can be easily applied to any languages
5	Yan et al.[49]	Weibo	CNN-BiGRU-AT model	Weibo time and location information	● Limitation with implicit expression analysis	● Effectively improving the accuracy of student text sentiment analysis
6	Cao et al.[51]	News article	Knowledge Consolidation Network (KCN)	Multimodality data	● Slow in progress ● Limitation with other corpora	● Alleviating semantic ambiguity
7	Liu et al.[54]	News article	ED via supervised attention mechanisms	Multimodality data	● Loss of algorithm accuracy when additional event samples are automatically extracted	● Modeling argumentative information for ED ● Using a supervised attention mechanism.
8	Yan et al.[55]	News article	ED based on graph convolution network	Multimodality data	● Recall capabilities of algorithms are constrained when dealing with lengthy sentences	● Better captureing dependency contextual information for ED
9	Wang et al.[56]	News article	RG-ACNN framework	Multimodality data	● Precision needs to be tested on more datasets ● Limitation with other corpora	● Highly efficient and effective compared to the benchmark method
10	Hasan et al.[59]	Twitter	Using specialized inverted indices and incremental clustering approach.	Unique words	● Random indexing may compromise algorithm's stability	● Reducing computational cost ● With a variety of more accurate filters
11	Troudi et al.[66]	Multimedia	Collecting data from various social media; based on the concept of mashup; based on the hadoop framework	Multimodality data	● Asynchronous data collection across platforms results in time loss ● Noise handling has limitations	● Handling different data sources to achieve more accurate ED ● overcoming the problem of existing algorithms dealing with big data
12	Wang et al.[75]	Multimedia	Sparse topic model based semi-supervised method	Multimodality data	● Stability of the algorithm needs to be optimized when processing short text	● Having clear advantages in fields of ED
13	Tan et al.[89]	Multimedia	Character-word fusion gate mechanism	Multimodality data	● Topics of events cannot be extracted in an unsupervised manner	● Correcting the mismatch between Chinese words and event triggers

(To be continued)

Table 1 Summary of ED on various social media. (Continued)

Index	Reference	Media	Method	Feature	Limitation	Merit
14	Yin et al.[90]	News article	Method based on CNN-BiGRU model	Multimodality data	● Dimensional limitation of word vectors ● Limitation with convolutional layer	● Improving effect of trigger words ● Reducing model training time
15	Araki and Mitamura et al.[91]	WordNet and Wikipedia	Distant supervision	WordNet data and an additional dataset	● No method is provided for the extraction of event types and cognitive states	● Generating high-quality training data automatically
16	Huang and Ji[92]	News article	Semi-supervised vector quantized approach	Multimodality data	● Handling multiple-type events is not performing as expected	● Popular events are extracted with a high degree of accuracy
17	Hong et al.[93]	News article	Self-regulated learning approach	Multimodality data	● No evaluation of the performance of discriminators in adversarial networks	● Discriminator performs well by fierce competition
18	Wang et al.[94]	Supervised data	Effective method for weakly supervised ED	Labeled instances and event type	● Computational cost of event-related candidates is high	● Automatically constructing more diverse and accurate training data
19	Rossi et al.[95]	Twitter	Automated set of services that start from the avail-ability of weather forecasts	Time and location information	● Accuracy of ED is heavily influenced by noise ● Limitation with other corpora	● Machine learning based methods can mine events for in-depth features
20	Xie et al.[96]	Twitter	New approach for identifying leadership in group learning	Unique words	● Influence of lack of multi-type event data on results	● Highlighting dynamic nature of leadership behavior
21	Dabiri and Heaslip[97]	Twitter	Traffic ED model using deep-learning architectures	Labeled traffic feature words	● Lack of an effective geocoder to locate traffic incidents accurately from relevant tweet text	● Giving good results in traffic information and condition detect.
22	Cui et al.[98]	Weibo	SVM classifier, and novel method to identify the location for the foodborne disease event	Multimodality data	● Detection algorithm's keyword selection needs to be improved	● Selecting more relevant tweets within the context dynamically to obtain a larger corpus of candidates
23	Abdelhaq et al.[99]	Twitter	Novel spatial regularization technique and hierarchical space-partitioning index structure	Time and location information	● Limitations exist when handling spatial outliers ● Noise is not handled when processing documents	● Conducting real-time extraction and tracking of local keywords with fine-grained spatio-temporal resolution
24	Alsaedi et al.[100]	Twitter	Integrated framework based on temporal TF-IDF	Namely temporal, spatial, and textual content	● Background characteristics of event discovery lack thorough analysis	● Ability to aggregate events corresponding to Weibo posts without prior knowledge
25	Hu et al.[101]	News article	Link prediction based method for ED	Time and content information	● In the case of short texts, generalizability cannot be confirmed	● Automatic describing unknown network evolution

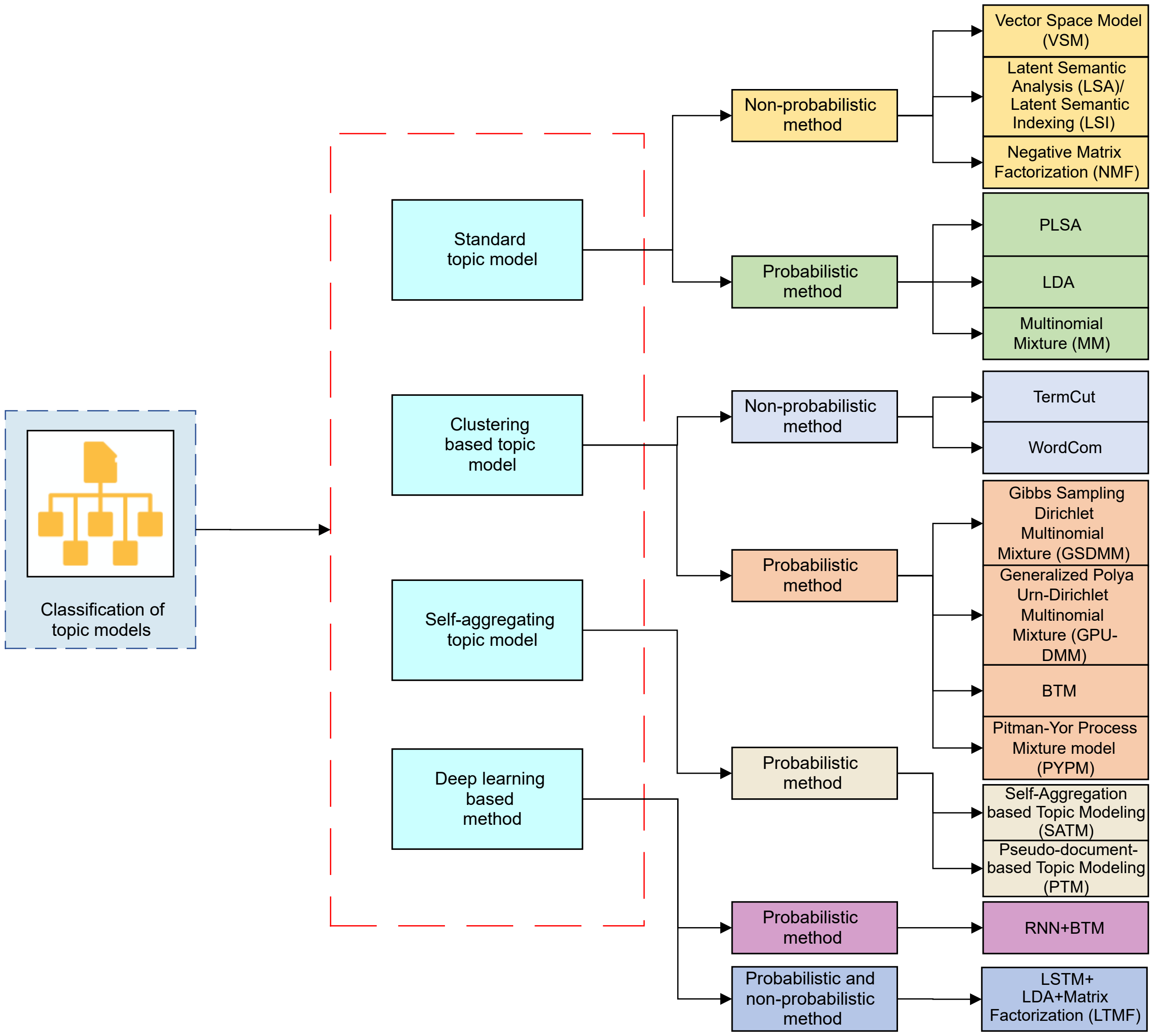


Fig. 4 Classification of various topic modelling methods.

above problems, researchers need to propose a series of improvement solutions. For example, deep learning techniques can be introduced to automatically learn and generate topic models, better handle massive amounts of social network text data, etc., to improve the efficiency and accuracy of ED. In addition, network analysis techniques can be used to filter out influential users and key tweets to reduce noise interference and improve the credibility of detection results. In the future, with the continuous development and innovation of technology, we can expect more new solutions for these problems to emerge, bringing more accurate, efficient and practical means for the analysis and application of social media data.

(2) The existing topic-based ED methods in social networks rely heavily on high-frequency topic keywords, which makes the topic description of events relatively vague. However, due to the short characteristics of social network texts and the diversity of user expressions, the subject keywords of a given event are difficult to be accurately described and captured. At the same time, the uneven distribution and sparsity problems of social network data in time and space also lead to the problem that it becomes increasingly challenging to represent the original state of events completely and accurately. To address these problems, researchers need to consider improvement solutions from multiple perspectives. For example, more advanced natural language processing techniques, such as word vector modeling and semantic analysis, can be used to analyze and interpret social network texts from different perspectives in order to obtain

more comprehensive and accurate information about events. In addition, other data sources, such as traditional media reports, government announcements, etc., can be combined to consider multiple dimensions and aspects of the event in order to improve the effectiveness and credibility of ED.

(3) When analyzing social media data, topic-based ED methods usually enable the detection and tracking of events by identifying topics and keywords associated with specific events. However, the approach also has some limitations. One of them is that automatic identification of key tweets and influencers is still difficult, and the final detection results are still only a list of topic keywords, lacking in-depth understanding and insight into the events. Researchers can introduce network analysis and machine learning techniques to build more comprehensive and accurate models to analyze and process social media data. These methods can better identify key tweets and influencers, and evaluate and predict events based on their contribution and level of influence. Also, techniques such as text sentiment analysis and time series analysis, can be combined to improve the accuracy and usefulness of ED.

ED models on social networks are one of the cores of popular ED and tracking techniques, whose main task is to identify hot events in the data of tweets posted or retweeted by users on social networks. Topic models are one of the core components of ED technology, which can automatically identify and extract topic information by analyzing and modeling text data.

An effective topic model should be able to improve the accuracy and efficiency of hot events identification, and significantly improve the performance of ED techniques. This can be achieved by introducing new techniques and methods, such as deep learning, incremental learning, graph neural networks, etc. We also need to explore and study the application of topic models in other fields, such as recommendation systems, ad placement, opinion monitoring, etc.

As mentioned above, traditional topic models (PLSA or LDA) exhibit unreliable ED performance in social network microblog text data. This is attributed to the short length characteristics of text information in social networks. When modeling such short text, the problem of sparse features arises, where text features cannot be extracted accurately. This problem can be addressed from two different perspectives[102]: firstly, several external resources can be used to supplement the expression of short text, and secondly, the short text can be integrated into long text before modeling the subject. Although former method can solve the problem of sparse data, it heavily relies on external resources, which can change the original semantics of the text, resulting in unreliable ED results. The latter solution of transforming short text into long text for topic modeling simply constructs the long text according to some common characteristics identified in the short text. Although it solves the problem of sparse features to a certain extent, it does not improve the quality of topic mining, thereby imposing only a little impact on improving the efficiency of ED. In addition, users are important participants in social networks. Therefore, a user level analysis is pivotal to better understand the themes of social networks. In order to improve the quality of topic mining and optimize the performance of ED models in social networks, both of these approaches need further in-depth research and practice.

From the above analysis, although research and applications of public opinion detection have achieved certain results, key issues still exist that need immediate attention to optimize the performance of existing public opinion detection models. We draw important findings from our analysis as follows: (1) Feature-based public opinion detection methods ignore the multidimensional characteristics of public opinion events resulting from the user's interest migration phenomenon and do not consider the syntactic structure and potential dependencies, making it difficult to explore the deep semantic expression mechanisms of major public opinion events under high noise; (2) Model-based public opinion detection methods do not require the number of optimal categories to be predetermined, which result in their reduced computational complexity. However, their fragmented clustering methods deliver the detection results as a series of topic keywords with vague descriptions, thereby failing to reveal the semantic generation rules of public opinion events, which leads to insufficient detection efficiency and low accuracy.

3 Event Tracking

3.1 Event propagation

Opinion tracking technology relies on information posted or shared by users on social media in order to achieve the tracking of event dissemination. The efficiency of the information dissemination model, on the other hand, is a key factor in determining the accuracy and effectiveness of public opinion dissemination, which ultimately affects the performance of the overall public opinion dissemination technology. An efficient information dissemination model can accurately predict the trend, scale, and speed of event dissemination, thus providing important support for public opinion analysis and early warning. For this reason, we need to continuously optimize and improve the algorithms and implementations of information dissemination models to enhance their accuracy, efficiency and robustness.

The independent cascade model and linear threshold model are examples of widely used traditional information propagation models[103]. Various other event dissemination models have been built upon the two traditional propagation models[104, 105]. With the motivation of improving the event dissemination efficiency, various algorithms have been developed recently[106–110], where mostly the time cost is optimized[90]. These models and algorithms provide valuable insights on online information diffusion behaviors, which helps improving the event dissemination efficiency and accuracy.

In social networks like Weibo, user tweets are propagated to connected users via event propagation as soon as they are published. The published tweets are further propagated by a network of users, which allows organizations to quickly disseminate information through user relationship networks. Propagation models have brought various sophistication into people's life. However, correctly, and effectively identifying hot events and tracking user's opinions amid the proliferation of user-generated content and real-time events remain challenging[20, 21].

The transmission of popular events in social networks is similar to the fission process, where microblogs posted by users propagate automatically to neighbors, who may then share the events among their network. Such a process helps with quick spread of information in social networks, whereby local discussions produce group effects[111, 112]. The Communityzer algorithm outlines the stages of birth, growth, shrinkage, merging, splitting, and dying of group communities in a social network by detecting overlapping communities along with performing user community evolution operations during information dissemination[98]. This two-stage community mining process defines how local discussions generate group effects in social network information dissemination to accelerate information spreading.

Existing studies have mainly focused on either the information dissemination model or the user community process, but largely ignore the user's interested topic information and event content information[113, 114]. User's participation behavior on social networking platforms varies greatly in terms of their topic interests and activity, which contribute to topic heterogeneity. Moreover, microblog contents have a considerable impact on user behaviors. Most topics quickly disappear from the list of discussed topics, with only a few standing out and attract more attention in the competitive environment[115, 116].

Kwak et al.[117] noted that the short text feature in social networks shortens the time required for users to access and post information, thus accelerating the dissemination of information, which is confirmed in a statistical analysis of the entire Twitter network. To improve the effectiveness and accuracy of the dissemination of popular events in social networks, a large-scale statistical analysis of social network information is being conducted.

The content of social network information dissemination emerges from the discussion and sharing of current events from our daily life[118]. Local events have greater communication capability, but are frequently ignored by traditional media[119–122] (e.g., emergencies and disaster events). Regional differences exist in the dissemination of information when keywords are used to extract microblogs[122]. Moreover, users are interested in events around their locality. Common users pay more attention to a given event's general information rather than the finer details, and they typically retweet news headlines or links to express their opinions and thoughts. On contrary, information posted or forwarded by users belonging to the locality of the events are more specific and detailed.

The issue of event propagation[123, 124] has been studied from the perspectives of information propagation models and user topics. A two-stage algorithm[123] was proposed for disseminating popular events for a specific topic, which improves event dissemination efficiency and accuracy to some extent. Furthermore, this method aids the spread of popular events based on exploiting a set of influential users under a specific topic. Karthika and Geetha's research[111], by analyzing users' interest distribution, estimated users' activation probabilities at the topic level, and proposed a novel event propagation method. This method utilizes users' activation probabilities to rapidly disseminate popular events related to specific topics, which improves the event's propagation speed, along with extending their propagation range. The significance of individuals in online social networks and the topical relevance of information dissemination have been discussed[125]. Firstly, based on the PageRank algorithm, a new information dissemination model was proposed. Subsequently, existing algorithms were employed to approximate the optimal solution for the problem of influence maximization based on this model. Xuan et al.[126] introduced a new framework designed to explore the concept of reverse influence effects for studying information propagation during the process of social network reconstruction. First, a new information diffusion model for social networks was introduced, which considers two types of people as smart and normal, as well as two types of information as true and false. As social networks consist of individuals who learn from their experiences, receiving a true (or false) message from a neighbor can increase (or decrease) the sender's trust in their neighborhood. By utilizing self-learning mechanisms, a social network can become more intelligent and improve its ability to distinguish between genuine and fake information. In the recently identified model based social stratification phenomenon, authentic information that is initially posted by a person in a closer proximity to a smart person could be retweeted by more people using self-learning mechanisms. Through the validation of simulation-based experiments, researchers can confirm their theories and explore deeper questions about the underlying mechanisms of social networks and information dissemination, providing readers with a deeper understanding.

From the above analysis, existing public opinion propagation models still face critical issues that need to be resolved to improve the overall performance of public opinion tracking technology. Identifying these issues from two perspectives: (1) current public opinion propagation methods ignore the impact of user interest migration on the continuous propagation of public opinion across multiple communities. Existing models are unable to capture the rich structural and semantic features that are embedded in the public opinion propagation graph, which is pivotal to learn and model the dynamic process of information propagation. As a result, it is difficult to summarize the real dissemination patterns of public opinion events. (2) There is a lack of research on the selective propagation rules of public opinion events based on user interest migration. It is difficult to explore the inherent correlation between information and its spatiotemporal features, such as the evolving characteristics of user interests, semantic attributes, and other information in the public opinion propagation graph, whereby hampering the discovery of adaptive propagation mechanisms for continuous public opinion events.

3.2 Event evolution

Nowadays, opinion tracking technology is very necessary. In addition, the evolution of public opinion is one of the crucial factors for public opinion tracking technology. Only by deeply understanding the changing trend of public opinion and the new and old public opinion events that emerge in the process of evolution, can we better conduct effective opinion tracking. To achieve this goal, it is especially important to use efficient and accurate models to track the evolution of public opinion. Such models can continuously collect, organize, analyze, and summarize various forms of public opinion information, and turn them into graphs or reports that can be visually presented to provide reliable reference bases for decision makers. Meanwhile, if the correlation between old and new public opinion events can be grasped in time during the process of public opinion changes, the chain of public opinion events can be greatly accelerated, thus enhancing the performance of public opinion tracking technology and better serving the public and decision makers. Therefore, we should realize the importance of the evolution of public opinion for public opinion tracking technology, and actively promote the development and application of efficient and accurate opinion tracking models.

In social networks, event evolution refers to the process of spreading and diffusion of opinion events on the network. This process usually consists of some users sharing or posting information about a topic and drawing the attention of other users by liking, commenting, or retweeting, which leads to the formation and spread of opinion events. In social networks, the speed and scale of event evolution can be rapidly scaled up, because social networks are characterized by fast dissemination, wide coverage, and high interactivity, and also because users in social networks are highly connected and influential. Therefore, once an important event or topic emerges, it will quickly permeate the entire social network, and may trigger public discussion and attention to the issue. The evolution of public opinion events in social media can be presented in various forms. We summarize both the scope and period of public opinion dissemination. To measure the dynamic spread of public opinion from a spatial perspective, we focus on the general trend of public opinion spread to assess its coverage and influence. In addition, we need to consider the period of public opinion dissemination, i.e., the development process of public opinion events in time, to understand its evolution pattern and the impact it has on society. The comprehensive analysis of these aspects helps us understand and manage public opinion events in a more comprehensive way. The public opinion dissemination period measures the dynamic spread of public opinion from a temporal perspective, with emphasis on the speed and duration of public opinion spread in the network. We need to pay attention to the evolution of public opinion events in social networks to understand the temporal trajectory, speed, and trend of their spread. This helps us better assess and respond to the trends of public opinion events, and thus manage and control public opinion risks more effectively.

Exploring the evolutionary mechanisms of popular event information in social networks is a very important area of scientific and practical research. In social networks, information can be disseminated and diffused in different ways, and the laws and mechanisms of this dissemination phenomenon need to be studied and explored in depth. The techniques for detecting and tracking popular events in social networks rely heavily on event evolution models. A good event evolution model should improve the accuracy and efficiency of identifying new and old popular events to achieve accurate monitoring and management of popular events. Event evolution models play a crucial role in the evolution of epidemic events. A good event evolution model needs to consider not only the propagation process of an event, but also multiple aspects, such as the scope of the event, its duration, and its possible consequences. Only through in-depth research and analysis of event evolution models can we better improve the performance of popular ED and tracking technologies, and provide scientific support for social opinion management. However, this is a difficult procedure that includes huge data processing, text content analysis and processing, network structure analysis, and information retrieval. At the same time, the recent proliferation of user-generated content has resulted in a slew of issues around information overload. Therefore, the study of the evolution of popular events in social media has attracted widespread attention. In the age of social media, the speed and scale of information dissemination have been greatly enhanced, which makes the evolution of popular events more complex and diverse. Therefore, we need to explore in depth the dissemination patterns and mechanisms of popular events in social media in order to better understand the formation and development trends of popular events. This research involves various fields, such as data mining, network analysis, and machine learning, which not only helps improve the efficiency of epidemic ED and management, but also provides a scientific basis for us to better understand and grasp social opinion. Furthermore, government agencies collect and analyze network security information by tracking the evolution of popular events, as well as the evolution of user opinions and attitudes, allowing for timely control and governance[127, 128].

Event evolution refers to the development process of popular events spreading and diffusion in social networks, which is closely related to the form and dissemination method of popular events. In social networks, event evolution usually shows dynamic changes, diversity, and complexity, and is influenced by various factors, such as user engagement, information quality and diversity, and social network structure. In video sharing websites, the evolution of video events is typically depicted by the number of views and shares[129]. The dynamic dissemination of events in news platforms is represented by the number of news comments[130]. The evolution of event information in social networking can be presented in a variety of ways, but two important perspectives include propagation range[131] and propagation period[131]. The dissemination range is a spatial measure of dynamic dissemination of information that focuses on the overall trend of information dissemination. The propagation cycle is a time-based measurement of dynamic propagation of information, which is measured based on the information propagation rate in the network and its duration. Figure 5 illustrates the process of event evolution.

Traditional event evolution research is primarily based on high-frequency keywords of emergencies or the topic model based method for detection and clustering events[132–135]. Majority of existing methods employ high-frequency keywords as abstracts of sudden or popular events. However, popular events in real life evolve over time, thus it is critical to track their evolution and development. The evolution of popular events require emphasis, which refers to the changes in interests of influential communicators in the event evolution process. As a result, we must concentrate on hot events and the evolution of influential user’s interests in relevance to those events, rather than solely focusing on the hot events.

Various studies have focused on event evolution recently. In classical topic detection and tracking research, the event evolution model (Story Link Detection (SLD) is an essential method, which can discover the associations among documents. Qian et al.[21] constructed a multi-vector event model to describe events, where the relationship between events is described by calculating vector similarity. This method, however, treats different vectors equally and cannot distinguish between different weights of words, geographic locations, names, and background information. Yang et al.[130] proposed an event evolution ranking method to evaluate the relationship among events during the process of event evolution. Although this method integrates event contents, background information, and similarity of document distribution, it does not consider other important event attributes, such as the geographic location, participants etc. Various combinations of feature information[136] has been considered to compute the average content similarity and optimal temporal features. However, these methods are applied to long news text corpus, and cannot be directly applied to short text of Weibo.

To address the issue of the evolution of short-text events in social networks, Zhang et al.[123] proposed an approach in his research to build a context-aware

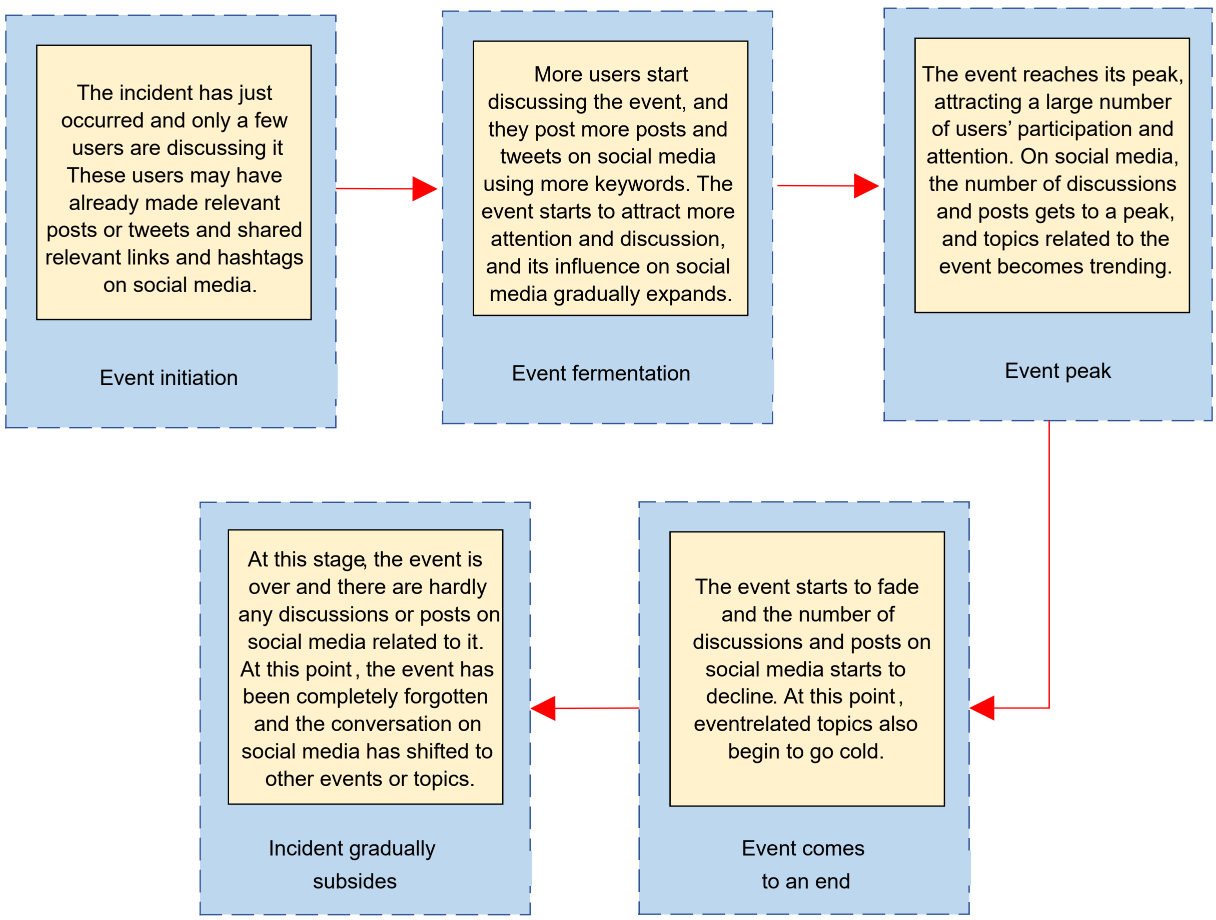


Fig. 5 The evolution of popular events on social media.

storyteller model. This method initially builds a microblog relationship graph, and uses a time sliding window to dynamically detect the event's core subgraph, and finally builds the event's evolution relationship by monitoring the subgraph's evolution process. This method can determine the interdependencies between events. Lin et al.[137] proposed a graph optimization model, which employs dynamic pseudo-relevant feedback to determine the evolution of related microblogs and events. This method, however, requires users to query keywords, and the model's effectiveness is dependent on the ability of users' query words to identify relevant popular events.

As a result, the graph optimization method is deemed unsuitable for the event evolution problem of short texts without prior knowledge. To this end, a microblog time and tag-based evolution tracking model (e.g., TH-LDA)[60] was proposed. To improve performance, the TH-LDA model incorporates tags and time factors. The TH-LDA model can detect popular events along with identifying unlabeled microblogs that belong to the same popular event as the label text, which is used to detect the popular event's sub-events and track the popular event's evolution. A Hot Topic LifeCycle Model (HTLCM) was created by academics to follow the evolution trend of hot event subjects[66]. The model divides the evolution of hot topics into five stages, which are birth, growth, maturity and death. Through an in-depth study of the propagation pattern of hot topics in social networks, participants' behaviors, and topic contents, the HTLCM model is expected to reveal the development trajectory and pattern of hot topics, which is of great significance for public opinion management and prediction. At the same time, the HTLCM model can determine whether the current topic is a potential hot event topic and can track the evolution of hot topics in real time, thus significantly improving the ability to identify and track the evolution of events. The model combines social network analysis, text mining and machine learning to effectively identify and track the evolution of hot event topics by collecting, processing and analyzing data.

While social networks provide us with services, such as finding friends, sharing resources, and creating virtual communities, information overload has become a growing issue amid the evolution of popular events in social networks, making it increasingly difficult to find useful information about popular events during specific time periods[138]. As a result, identifying information quickly related to popular events in unstructured, diverse, and dynamic social networks, has become a hot research topic. In social networks, event evolution based on user interest evolution primarily refers to the search for people and information related to specific user interests or specific events, as well as their evolution process. Due to the issues of social network information overload, existing solutions are primarily divided into two categories, methods based on collaborative filtering and methods based on interest community discovery[139].

The method based on collaborative filtering primarily finds users similar to target user by analyzing the user and social information score matrix, and predicts a given user's interest score by calculating the interests of similar users. The predicted score is used to determine the relevance of discovered events to the users. While collaborative filtering recommendation methods are widely popular, they may face challenges when dealing with data sparsity and cold start issues.

The method of interest-based community discovery is used to discover people or information related to users through communities, which is derived from the idea of "things gather together, and people are divided into groups". People have multiple social identities in social networks, for example, users may have interest in multiple communities. As a result, in the recommendation method, the target user's social network information can be used to improve recommendation quality. Therefore, Lin et al.'s research[140] has reached the following conclusion: the discovery of resource information within networks of friends with shared interests is highly effective and accurate. This finding underscores the significance of sharing interests and resources within social networks, further reinforcing the pivotal role of social networks as vital channels for information and resource dissemination. Additionally, this conclusion offers a fresh perspective for social network research, enabling deeper exploration into the formation of interest-based communities and information propagation mechanisms, thus enhancing our understanding of the information ecosystem within social networks. These insights may also contribute to guiding the development of social media platforms and the formulation of social network management strategies. However, the nature of social networks is dynamic, and users' interests change on a regular basis. As a result, majority of existing community interest discovery algorithms are incapable of accurately reflecting the real network.

From the above analysis, evolution model still needs improvement in various perspectives: (1) Existing event evolution models are not capable of identifying user interest communities in popular events. Furthermore, their inefficiency in identifying influential change process of key microblogs and key figures make them less capable of tracking and identifying the influence of dynamic microblogs and user. (2) Tracking the evolution of popular events based on user interest migration is difficult in existing models, thereby efficiently and accurately clustering and tracking popular events is less feasible. (3) The low recognition rate of new and old popular events in the process of event evolution leads to a low event evolution control ability. Overall, efficiently and accurately tracking the evolution of popular events along with identifying the evolution of interests from influential users still need improvement.

4 Discussion

Research on tracking popular events in social networks is still in its early stages. As social networks have evolved as popular and important communication medium, research on next-generation social networks is urgently needed to improve user experience. Furthermore, by analyzing the research gaps in existing detection and tracking technology of popular events, this paper presents important inferences for researchers on potential future directions.

The main issues in existing ED models, event propagation models, and event evolution models in social networks are summarized below:

(1) Social network data are prone to noise, such as low-quality microblogs, low-influence users, and unpopular topics, which is less useful for ED. At the same time, the number of a priori parameters in existing topic model-based ED methods must be set manually, which significantly reduces ED efficiency and accuracy. Furthermore, existing topic model based social network ED methods rely on high-frequency topic keywords, thereby necessitating manual discrimination. The description of event topic keywords is relatively vague due to short text characteristics of social networks. Furthermore, due to data scarcity, it is difficult to comprehensively represent the original state of a given event, which further reduces the ED accuracy and efficiency.

Existing public opinion detection methods commonly use topic models or methods to cluster high-frequency keywords, but ignore the multi-dimensional and complex characteristics of public opinion events under the user interest shift phenomena. Moreover, these methods do not effectively utilize the semantic information of public opinion texts under high noise and fail to consider the syntactic structure and potential dependency relationships. Furthermore, during the detection process, all topic keywords are treated as independent features, without in-depth consideration of the correlations among multidimensional and complex characteristics. This results in their failure to effectively explore the deep semantic expression mechanism of significant public opinion events, which is not conducive to track subsequent evolution of public opinion events, and leads to insufficient detection and low accuracy.

(2) Existing event dissemination models in social networks cannot selectively disseminate events based on the popularity of user interests, cannot intelligently identify and update influencers, and cannot learn from previous dissemination processes. These constraints reduce their efficiency. At the same time, existing studies do not consider the significance of event content and user preferences. However, in social networks, event propagation is related to users and event topics. Moreover, a given user may characterize different event propagation capabilities for different events. There is currently no research on the impact of changes in user preferences on information dissemination in social networks. Furthermore, experience or knowledge from previous propagation process cannot be learned, thereby resulting in low accuracy and efficiency. Past knowledge is crucial in continuous event propagation within a small range, which significantly affects the accuracy and efficiency of event propagation model.

The public opinion propagation method based on information propagation models is an effective propagation technique, but requires a prior determination of the propagation probability among users to disseminate public opinion events through the network. Moreover, this method ignores the impact of user interest shift phenomena on multi-source information propagation, and cannot learn and model the dynamic process of information propagation by capturing the rich structural features and semantic characteristics embedded in the public opinion propagation graph. This leads to problems, such as poor flexibility of the public opinion propagation model under multiple communities, and over-dependence on propagation probability for continuous propagation range.

In addition, existing public opinion propagation methods require retraining of model parameters during continuous propagation, which not only incurs a considerable amount of time to construct the model, but also causes the new model to lose the feature information of past propagation probabilities, thus affecting the performance of the model.

(3) A large number of replies, retweets, and other operations has led to a continuous expansion of social networks. Only a portion of microblog data can be analyzed in any given time frame under incomplete network topology. Thus, the accuracy of event evolution model is impacted by the evolving nature of social networks. Many existing methods of user influence analysis are carried out using static network analysis, thus incapable of exploiting the internal mechanism of a dynamic network. Tracking the evolution of popular events based on user interest migration is often challenging. Moreover, it is impossible to automatically cluster and track popular events in an efficient and accurate manner. The low recognition rate of new and old popular events in the process of event evolution leads to a low event evolution control ability, making it difficult to track the evolution of popular events and the evolution of influential communicators’ interests efficiently and accurately.

Most of the existing public opinion evolution methods only focus on identifying and tracking multiple static graph structures, but ignore user interest evolution characteristics across networks. This makes it difficult to accurately identify evolution relationship of the new and old public opinion events under the user interest shift phenomenon, thereby losing efficiency whilst constructing chains of new and old public opinion events evolution relationship graphs. At the same time, existing public opinion evolution models assign same weights to new and old public opinion events during the training process, and ignore the impact of user interest shift phenomena during the process of recognizing the evolution relationship between new and old public opinion events. This makes it difficult to explore the generation mechanism of public opinion events evolution relationship graph chain.

5 Conclusion

In summary, research on public opinion tracking technology has been studied from different perspectives in recent years.

Public opinion events characterize complex internal and external relationships and evolutionary characteristics. Overcoming the shortcomings of low detection accuracy under high noise resulting from user interest shifts is an immediate requirement. Furthermore, the narrow range of continuous public opinion dissemination under multiple communities and the difficulty in constructing chains of new and old public opinion events amid the evolution of public opinion events across networks affect the detection efficiency. Therefore, in-depth research on rapid and accurate public opinion tracking technology, especially under user interest shift phenomena, has significant practical value.

We derive future research directions as below:

(1) Improve classification accuracy: Classification accuracy of existing public opinion tracking algorithms still need significant improvement. Introducing more advanced text analysis techniques, optimizing feature selection methods, and establishing more complete classifiers can be used to improve classification accuracy.

(2) Introduce deep learning technology: Traditional machine learning methods are still used for feature extraction and classification methods in public opinion tracking algorithms. However, these methods cannot efficiently process complex semantics and facilitate training of larger datasets. Therefore, the introduction of deep learning techniques, such as graph neural networks and Generative Pre-trained Transformer 4 (GPT-4) should be considered for higher-level analysis in the future when performing text feature extraction and classification analysis. To provide more accurate and comprehensive data support for opinion analysis and prediction.

(3) Optimize sentiment analysis model: Sentiment analysis is one of the core modules of public opinion tracking algorithms. In the future, sentiment analysis models should be optimized using more training data, smarter feature selection, and classifiers to improve the accuracy of sentiment analysis.

(4) Expand dataset: The classification effect of public opinion tracking algorithms is limited by smaller datasets, and traditional datasets are usually limited to specific fields and languages. Therefore, in the future, more languages and real-world datasets should be used in the training process to better deal with challenges of diversity and complexity.

(5) Combine with other algorithms: In addition to text-based public opinion tracking algorithms, other algorithms, such as machine vision and social network analysis, should be incorporated into public opinion tracking models. This requires improving the interoperability between algorithms and promoting cross-development of algorithms.

Acknowledgment

The work was supported by the National Natural Science Foundation of China (No. 62302199), the China Postdoctoral Science Foundation (No. 2023M731368), the Natural Science Foundation of the Jiangsu Higher Education Institutions (No. 22KJB520016), the Jiangsu University Innovative Research Project (No. KYCX22_3671), the Youth Foundation Project of Humanities and Social Sciences of Ministry of Education in China (No. 22YJC870007), the Jiangsu University Undergraduate Student English Teaching Excellence Program, and the Ministry of Education's Industry-Education Cooperation Collaborative Education Project (No. 202102306005).

References

[1] Y. Peng, Y. Zhao, and J. Hu, On the role of community structure in evolution of opinion formation: A new bounded confidence opinion dynamics, *Inf. Sci.*, vol. 621, pp. 672–690, 2023.

[2] H. Hassani, R. Razavi-Far, M. Saif, and E. Herrera-Viedma, Blockchain-enabled trust building for managing consensus in linguistic opinion dynamics, *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 8, pp. 2722–2733, 2023.

[3] I. V. Kozitsin, Opinion dynamics of online social network users: A micro-level analysis, *J. Math. Sociol.*, vol. 47, no. 1, pp. 1–41, 2023.

[4] P. Liu, Y. Li, and P. Wang, Opinion dynamics and minimum adjustment-driven consensus model for multi-criteria large-scale group decision making under a novel social trust propagation mechanism, *IEEE Trans. Fuzzy Syst.*, vol. 31, no. 1, pp. 307–321, 2023.

[5] L. Weng, Q. Zhang, Z. Lin, L. Wu, and J. H. Zhang, Integrating interactions between target users and opinion leaders for better recommendations: An opinion dynamics approach, *Comput. Commun.*, vol. 198, pp. 98–107, 2023.

[6] Y. Liu, J. Liu, and K. Wu, Cost-effective competition on social networks: A multi-objective optimization perspective, *Inf. Sci.*, vol. 620, pp. 31–46, 2023.

[7] Z. Wu, Q. Zhou, Y. Dong, J. Xu, A. H. Altalhi, and F. Herrera, Mixed opinion dynamics based on DeGroot model and Hegselmann-Krause model in social networks, *IEEE Trans. Syst. Man, Cybern.*: *Syst.*, vol. 53, no. 1, pp. 296–308, 2022.

[8] A. Saravanou, N. Panagiotou, and D. Gunopulos, News monitor: A framework for querying news in real time, in *Proc. 43rd European Conf. on Information Retrieval*, Virtual Event, 2021, pp. 543–548.

[9] Z. K. Abbas and A. A. Al-Ani, An adaptive algorithm based on principal component analysis-deep learning for anomalous events detection, *Indones. J. Electr. Eng. Comput. Sci.*, vol. 29, no. 1, pp. 421–430, 2023.

[10] X. Chen, H. Wang, L. Ke, Z. Lu, H. Su, and X. Chen, Identifying Cantonese rumors with discriminative feature integration in online social networks, *Expert Syst. Appl.*, vol. 215, p. 119347, 2023.

[11] 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, 2019.

[12] L. L. Shi, L. Liu, Y. Wu, L. Jiang, J. Panneerselvam, and R. Crole, A social sensing model for event detection and user influence discovering in social media data streams, *IEEE Trans. Comput. Soc. Syst.*, vol. 7, no. 1, pp. 141–150, 2020.

[13] Y. Xiao, C. Song, and Y. Liu, Social hotspot propagation dynamics model based on multidimensional attributes and evolutionary games, *Commun. Nonlinear Sci. Numer. Simul.*, vol. 67, pp. 13–25, 2019.

[14] O. Loyola-González, A. López-Cuevas, M. A. Medina-Pérez, B. Camiña, J. E. Ramírez-Márquez, and R. Monroy, Fusing pattern discovery and visual analytics approaches in tweet propagation, *Inform. Fusion*, vol. 46, pp. 91–101, 2019.

[15] D. Diallo and T. Hecking, Gradual network sparsification and georeferencing for location-aware event detection in microblogging services, in *Proc. 11th Int. Conf. on Complex Networks and Their Applications XI*, Province of Palermo, Italy, 2023, pp. 108–120.

[16] M. Liu, X. Wang, S. Liang, X. Sheng, and S. Lou, Single and composite disturbance event recognition based on the DBN-GRU network in φ-OTDR, *Appl. Opt.*, vol. 62, no. 1, pp. 133–141, 2023.

[17] P. H. Khotimah, A. Arisal, A. F. Rozie, E. Nugraheni, D. Riswantini, W. Suwarningsih, D. Munandar, and A. Purwarianti, Monitoring Indonesian online news for COVID-19 event detection using deep learning, *Int. J. Electr. Comput. Eng.*, vol. 13, no. 1, pp. 957–971, 2023.

[18] Y. Jiang, R. Liang, and J. Zhang, Network public opinion detection during the coronavirus pandemic: a short-text relational topic model, *ACM T. Knowl. Discov. D.*, vol. 16, no. 3, pp. 1–27, 2021.

[19] L. Zhang, T. Wang, and Z. Jin, The research on social networks public opinion propagation influence models and its controllability, *China Commun.*, vol. 15, no. 7, pp.

98–110, 2018.

[20] M. Zhuang, Y. Li, and X. Tan, Analysis of public opinion evolution of COVID-19 based on LDA-ARMA hybrid model, *Complex Intell. Syst.*, vol. 7, pp. 3165–3178, 2021.

[21] C. Qian, N. Mathur, N. H. Zakaria, R. Arora, V. Gupta, and M. Ali, Understanding public opinions on social media for financial sentiment analysis using AI-based techniques, *Inf. Process. Manage.*, vol. 59, no. 6, p. 103098, 2022.

[22] P. Wu, J. Liu, and F. Shen, A deep one-class neural network for anomalous event detection in complex scenes, *IEEE Trans. Neural Netw. Learn.*, vol. 31, no. 7, pp. 2609–2622, 2019.

[23] J. Kim and M. Hastak, Social network analysis: Characteristics of online social networks after a disaster, *Int. J. Inf. Manag.*,, vol. 38, no. 1, pp. 86–96, 2018.

[24] I. Afyouni, Z. Al Aghbari, and R. A. Razack, Multi-feature, multi-modal, and multi-source social event detection: A comprehensive survey, *Inf. Fusion*, vol. 79, pp. 279–308, 2022.

[25] A. Bondielli and F. Marcelloni, A survey on fake news and rumour detection techniques, *Inf. Sci.*, vol. 497, pp. 38–55, 2019.

[26] 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.

[27] W. Hu, H. Wang, Z. Qiu, C. Nie, L. Yan, and B. Du, An event detection method for social networks based on hybrid link prediction and quantum swarm intelligent, *World Wide Web*, vol. 20, no. 4, pp. 775–795, 2017.

[28] D. Mumin, L. L. Shi, L. Liu, and J. Panneerselvam, Data-driven diffusion recommendation in online social networks for the internet of people, *IEEE Trans. Syst. Man, Cybern.*: *Syst.*, vol. 52, no. 1, pp. 166–178, 2022.

[29] J. Tong, L. Shi, L. Liu, J. Panneerselvam, and Z. Han, A novel influence maximization algorithm for a competitive environment based on social media data analytics, *Big Data Mining and Analytics*, vol. 5, no. 2, pp. 130–139, 2022.

[30] L. L. Shi, L. Liu, Y. Wu, L. Jiang, M. Kazim, H. Ali, and J. Panneerselvam, Human-centric cyber social computing model for hot-event detection and propagation, *IEEE Trans. Comput. Soc. Syst.*, vol. 6, no. 5, pp. 1042–1050, 2019.

[31] D. M. Kingsbury, M. P. Bhatta, B. Castellani, A. Khanal, E. Jefferis, and J. S. Hallam, The personal social networks of resettled Bhutanese refugees during pregnancy in the United States: A social network analysis, *J. Commun. Health*, vol. 43, no. 6, pp. 1028–1036, 2018.

[32] L. Jiang, L. Liu, J. Yao, and L. Shi, A user interest community evolution model based on subgraph matching for social networking in mobile edge computing environments, *J. Cloud Comput.*, vol. 9, no. 1, p. 69, 2020.

[33] J. Teevan, D. Ramage, and M. R. Morris, # TwitterSearch: A comparison of microblog search and web search, in *Proc. 4th ACM Int. Conf. on Web Search and Data Mining*, Hong Kong, China, 2011, pp. 35–44.

[34] H. Sun, Y. Geng, L. Hu, L. Shi, and T. Xu, Measuring China's new energy vehicle patents: A social network analysis approach, *Energy*, vol. 153, pp. 685–693, 2018.

[35] B. Liu, Q. Zhou, R. X. Ding, I. Palomares, and F. Herrera, Large-scale group decision making model based on social network analysis: Trust relationship-based conflict detection and elimination, *Eur. J. Oper. Res.*, vol. 275, pp. 737–754, 2019.

[36] R. Bridgstock, Employability and career development learning through social media: Exploring the potential of LinkedIn, in *Challenging Future Practice Possibilities*, J. Higgs, S. Cork, and D. Horsfall, eds. Rotterdam, The Netherlands: Sense-Brill, 2019, pp. 143–152.

[37] B. Palani, S. Elango, and K. V. Viswanathan, CB-Fake: A multimodal deep learning framework for automatic fake news detection using capsule neural network and BERT, *Multimed. Tools Appl.*, vol. 81, no. 4, pp. 5587–5620, 2022.

[38] L. L. Shi, L. Liu, L. Jiang, R. Zhu, and J. Panneerselvam, QoS prediction for smart service management and recommendation based on the location of mobile users, *Neurocomputing*, vol. 471, pp. 12–20, 2022.

[39] G. Doddington, A. Mitchell, M. Przybocki, L. Ramshaw, S. Strassel, and R. Weischedel, The automatic content extraction (ACE) program-tasks, data, and evaluation, in *Proc. 4th Int. Conf. on Language Resources and Evaluation*, Lisbon, Portugal, 2004, pp. 837–840.

[40] R. F. He and S. Y. Duan, Joint Chinese event extraction based multi-task learning, (in Chinese), *J. Softw.*, vol. 30, no. 4, pp. 1015–1030, 2019.

[41] W. Z. Al-Dyani, F. K. Ahmad, and S. S. Kamaruddin, A survey on event detection models for text data streams, *J. Comput. Sci.*, vol. 16, no. 7, pp. 916–935, 2020.

[42] B. Jang, M. Kim, G. Harerimana, S. U. Kang, and J. W. Kim, Bi-LSTM model to increase accuracy in text classification: Combining Word2vec CNN and attention mechanism, *Appl. Sci.*, vol. 10, no. 17, p. 5841, 2020.

[43] L. Zhan, X. Jiang, and Q. Liu, Research on Chinese event extraction method based on HMM and multi-stage method, *J. Phys.*: *Conf. Ser.*, vol. 1732, no. 1, p. 012024, 2021.

[44] H. Fei, Y. Ren, and D. Ji, A tree-based neural network model for biomedical event trigger detection, *Inf. Sci.*, vol. 512, pp. 175–185, 2020.

[45] X. Feng, B. Qin, and T. Liu, A language-independent neural network for event detection, *Sci. China Inf. Sci.*, vol. 61, no. 9, p. 092106, 2018.

[46] J. Jin, H. Guo, J. Xu, X. Wang, and F. Y. Wang, An end-to-end recommendation system for urban traffic controls and management under a parallel learning framework, *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 3, pp. 1616–1626, 2021.

[47] J. Liu, Y. Chen, K. Liu, and J. Zhao, Event detection via gated multilingual attention mechanism, in *Proc. 32nd AAAI Conf. on Artificial Intelligence, the 30th Innovative*

Applications of Artificial Intelligence, and the 8th AAAI Symp. on Educational Advances in Artificial Intelligence, New Orleans, LA, USA, 2018, pp. 4865–4872.

[48] A. Brown, A. Tuor, B. Hutchinson, and N. Nichols, Recurrent neural network attention mechanisms for interpretable system log anomaly detection, in *Proc. 1st Workshop on Machine Learning for Computing Systems*, Tempe, AZ, USA, 2018, p. 1.

[49] W. Yan, L. Zhou, Z. Qian, L. Xiao, and H. Zhu, Sentiment analysis of student texts using the CNN-BiGRU-AT model, *Sci. Program*, vol. 2021, p. 8405623, 2021.

[50] N. Chen, J. Li, and Y. Man, Research on aviation unsafe information text analysis based on improved CNN-BiGRU-att model, in *Proc. IEEE 3rd Int. Conf. on Civil Aviation Safety and Information Technology*, Changsha, China, 2021, pp. 322–327.

[51] P. Cao, Y. Chen, J. Zhao, and T. Wang, Incremental event detection via knowledge consolidation networks, in *Proc. 2020 Conf. on Empirical Methods in Natural Language Processing*, doi: 10.18653/v1/2020.emnlp-main.52.

[52] L. Li, Y. Lin, and B. Du, Real-time traffic incident detection based on a hybrid deep learning model, *Transp. A*: *Transp. Sci*, vol. 18, no. 1, pp. 78–98, 2022.

[53] P. Luo, L. Ding, X. Yang, and Y. Xiang, Chinese event detection based on data augmentation and weakly supervised adversarial training, (in Chinese), *J. Comput. Appl.*, vol. 42, no. 10, pp. 102–114, 2022.

[54] S. Liu, Y. Chen, K. Liu, and J. Zhao, Exploiting argument information to improve event detection via supervised attention mechanisms, in *Proc. 55th Annu. Meeting of the Association for Computational Linguistics*, Vancouver, Canada, 2017, pp. 1789–1798.

[55] H. Yan, X. Jin, X. Meng, J. Guo, and X. Cheng, Event detection with multi-order graph convolution and aggregated attention, in *Proc. 2019 Conf. on Empirical Methods in Natural Language Processing and the 9th Int. Joint Conf. on Natural Language Processing*, Hong Kong, China, 2019, pp. 5766–5770.

[56] Z. Wang, Y. Guo, and J. Wang, Empower Chinese event detection with improved atrous convolution neural networks, *Neural Comput. Appl.*, vol. 33, no. 11, pp. 5805–5820, 2021.

[57] D. Shah, M. Hurley, J. Liu, and M. P. Daggett, Unsupervised content-based characterization and anomaly detection of online community dynamics, in *Proc. 52nd Hawaii Int. Conf. on System Sciences*, Grand Wailea, HA, USA, 2019, pp. 2264–2273.

[58] A. Bhuvaneswari and C. Valliyammai, Social IoT-enabled emergency event detection framework using geo-tagged microblogs and crowdsourced photographs, in *Emerging Technologies in Data Mining and Information Security*, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, eds. Singapore: Springer, 2019, pp. 151–162.

[59] M. Hasan, M. A. Orgun, and R. Schwitter, Real-time event detection from the Twitter data stream using the TwitterNews+ framework, *Inf. Process. Manage.*, vol. 56, no. 3, pp. 1146–1165, 2019.

[60] O. Kolchyna, T. T. P. Souza, P. C. Treleaven, and T. Aste, A framework for Twitter events detection, differentiation and its application for retail brands, in *Proc. 2016 Future Technologies Conf.*, San Francisco, CA, USA, 2016, pp. 323–331.

[61] T. Sakaki, M. Okazaki, and Y. Matsuo, Earthquake shakes Twitter users: Real-time event detection by social sensors, in *Proc. 19th Int. Conf. on World Wide Web*, Raleigh, NC, USA, 2010, pp. 851–860.

[62] R. Long, H. Wang, Y. Chen, O. Jin, and Y. Yu, Towards effective event detection, tracking and summarization on microblog data, in *Proc. 12th Int. Conf. on Web-Age Information Management*, Wuhan, China, 2011, pp. 652–663.

[63] J. Weng and B. S. Lee, Event detection in Twitter, in *Proc. 5th Int. AAAI Conf. on Weblogs and Social Media*, Barcelona, Spain, 2011, pp. 401–408.

[64] T. Matuszka, Z. Vincellér, and S. Laki, On a keyword-lifecycle model for real-time event detection in social network data, in *Proc. IEEE 4th Int. Conf. on Cognitive Infocommunications*, Budapest, Hungary, 2013, pp. 453–458.

[65] J. Allan, Introduction to topic detection and tracking, in *Topic Detection and Tracking*: *Event-based Information Organization*, J. Allan, ed. Boston, MA, USA: Springer, 2002, pp. 1–16.

[66] A. Troudi, C. A. Zayani, S. Jamoussi, and I. A. B. Amor, A new mashup based method for event detection from social media, *Inf. Syst. Front.*, vol. 20, no. 5, pp. 981–992, 2018.

[67] F. Atefeh and W. Khreich, A survey of techniques for event detection in Twitter, *Comput. Intell.*, vol. 31, no. 1, pp. 132–164, 2015.

[68] T. Zhang, R. Ramakrishnan, and M. Livny, BIRCH: An efficient data clustering method for very large databases, *ACM SIGMOD Rec.*, vol. 25, no. 2, pp. 103–114, 1996.

[69] Z. Lv, T. Liu, C. Shi, J. A. Benediktsson, and H. Du, Novel land cover change detection method based on K-means clustering and adaptive majority voting using bitemporal remote sensing images, *IEEE Access*, vol. 7, pp. 34425–34437, 2019.

[70] M. Mathioudakis and N. Koudas, TwitterMonitor: Trend detection over the Twitter stream, in *Proc. 2010 ACM SIGMOD Int. Conf. on Management of Data*, Indianapolis, IN, USA, 2010, pp. 1155–1158.

[71] A. Angel, N. Koudas, N. Sarkas, D. Srivastava, M. Svendsen, and S. Tirthapura, Dense subgraph maintenance under streaming edge weight updates for real-time story identification, *VLDB J.*, vol. 23, no. 2, pp. 175–199, 2014.

[72] P. Soucy and G. W. Mineau, Beyond TFIDF weighting for text categorization in the vector space model, in *Proc. 19th Int. Joint Conf. on Artificial Intelligence*, Edinburgh, UK, 2005, pp. 1130–1135.

[73] S. B. Kaleel and A. Abhari, Cluster-discovery of Twitter messages for event detection and trending, *J. Comput.*

Sci., vol. 6, pp. 47–57, 2015.
[74] R. Giovanetti and L. Lancieri, Model of computer architecture for online social networks flexible data analysis: The case of Twitter data, in *Proc. 2016 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, San Francisco, CA, USA, 2016, pp. 677–684.
[75] J. Wang, L. Xia, X. Hu, and Y. Xiao, Abnormal event detection with semi-supervised sparse topic model, *Neural Comput. Appl.*, vol. 31, no. 5, pp. 1607–1617, 2019.
[76] S. Liu and P. Jansson, City event detection from social media with neural embeddings and topic model visualization, in *Proc. 2017 IEEE Int. Conf. on Big Data*, Boston, MA, USA, 2017, pp. 4111–4116.
[77] X. Sun, Y. Wu, L. Liu, and J. Panneerselvam, Efficient event detection in social media data streams, in *Proc. 2015 IEEE Int. Conf. on Computer and Information Technology; Ubiquitous Computing and Communications; Dependable, Autonomic and Secure Computing; Pervasive Intelligence and Computing*, Liverpool, UK, 2015, pp. 1711–1717.
[78] T. Hofmann, Unsupervised learning by probabilistic latent semantic analysis, *Mach. Learn.*, vol. 42, nos. 1&2, pp. 177–196, 2001.
[79] D. M. Blei, A. Y. Ng, and M. I. Jordan, Latent dirichlet allocation, *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003.
[80] X. Yan, J. Guo, Y. Lan, and X. Cheng, A biterm topic model for short texts, in *Proc. 22nd Int. Conf. on World Wide Web*, Rio de Janeiro, Brazil, 2013, pp. 1445–1456.
[81] Y. Pan, J. Yin, S. Liu, and J. Li, A biterm-based dirichlet process topic model for short texts, in *Proc. 3rd Int. Conf. on Computer Science and Service System*, doi: 10.2991/csss-14.2014.71.
[82] J. Li, Z. Tai, R. Zhang, W. Yu, and L. Liu, Online bursty event detection from microblog, in *Proc. IEEE/ACM 7th Int. Conf. on Utility and Cloud Computing*, London, UK, 2014, pp. 865–870.
[83] J. Chen, Q. Shang, and H. Xiong, Hot events detection for Chinese microblogs based on the TH-LDA model, in *Proc. 2018 Int. Conf. on Transportation & Logistics, Information & Communication, Smart City*, Chengdu, China, 2018, pp. 157–166.
[84] J. Yu and L. Qiu, ULW-DMM: An effective topic modeling method for microblog short text, *IEEE Access*, vol. 7, pp. 884–893, 2019.
[85] S. Jiang, X. Qian, J. Shen, Y. Fu, and T. Mei, Author topic model-based collaborative filtering for personalized POI recommendations, *IEEE Trans. Multimed.*, vol. 17, no. 6, pp. 907–918, 2015.
[86] D. Gunawan, R. F. Rahmat, A. Putra, and M. F. Pasha, Filtering spam text messages by using Twitter-LDA algorithm, in *Proc. 2018 IEEE Int. Conf. on Communication, Networks and Satellite*, Medan, Indonesia, 2018, pp. 1–6.
[87] M. Gupta and P. Gupta, Research and implementation of event extraction from Twitter using LDA and scoring function, *Int. J. Inf. Technol.*, vol. 11, no. 2, pp. 365–371, 2019.
[88] D. Nolasco and J. Oliveira, Subevents detection through topic modeling in social media posts, *Future Gener. Comput. Syst.*, vol. 93, pp. 290–303, 2019.
[89] X. Tan, G. Deng, and X. Hu, Multi-granularity context semantic fusion model for Chinese event detection, in *Proc. 10th Int. Conf. on Internet Computing for Science and Engineering*, Guilin, China, 2021, pp. 1–7.
[90] H. Yin, J. Cao, Y. Du, G. Wang, L. Cao, X. Wang, and H. Zhang, Chinese abrupt event recognition based on CBiGRU-ATT model, *IOP Conf. Ser.: Mater. Sci. Eng.*, vol. 782, no. 5, p. 052043, 2020.
[91] J. Araki and T. Mitamura, Open-domain event detection using distant supervision, in *Proc. 27th Int. Conf. on Computational Linguistics*, Santa Fe, NM, USA, 2018, pp. 878–891.
[92] L. Huang and H. Ji, Semi-supervised new event type induction and event detection, in *Proc. 2020 Conf. on Empirical Methods in Natural Language Processing*, doi: 10.18653/v1/2020.emnlp-main.53.
[93] Y. Hong, W. Zhou, J. Zhang, G. Zhou, and Q. Zhu, Self-regulation: Employing a generative adversarial network to improve event detection, in *Proc. 56th Annu. Meeting of the Association for Computational Linguistics*, Melbourne, Australia, 2018, pp. 515–526.
[94] X. Wang, X. Han, Z. Liu, M. Sun, and P. Li, Adversarial training for weakly supervised event detection, in *Proc. 2019 Conf. of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, Minneapolis, MN, USA, 2019, pp. 998–1008.
[95] C. Rossi, F. S. Acerbo, K. Ylinen, I. Juga, P. Nurmi, A. Bosca, F. Tarasconi, M. Cristoforetti, and A. Alikadic, Early detection and information extraction for weather-induced floods using social media streams, *Int. J. Disast. Risk Reduct.*, vol. 30, pp. 145–157, 2018.
[96] K. Xie, G. Di Tosto, L. Lu, and Y. S. Cho, Detecting leadership in peer-moderated online collaborative learning through text mining and social network analysis, *Internet High. Educ.*, vol. 38, pp. 9–17, 2018.
[97] S. Dabiri and K. Heaslip, Developing a Twitter-based traffic event detection model using deep learning architectures, *Expert Syst. Appl.*, vol. 118, pp. 425–439, 2019.
[98] W. Cui, P. Wang, Y. Du, X. Chen, D. Guo, J. Li, and Z. Zhou, An algorithm for event detection based on social media data, *Neurocomputing*, vol. 254, pp. 53–58, 2017.
[99] H. Abdelhaq, M. Gertz, and A. Armiti, Efficient online extraction of keywords for localized events in Twitter, *GeoInformatica*, vol. 21, no. 2, pp. 365–388, 2017.
[100] N. Alsaedi, P. Burnap, and O. Rana, Can we predict a riot? Disruptive event detection using Twitter, *ACM Trans. Internet Technol.*, vol. 17, no. 2, p. 18, 2017.
[101] W. Hu, H. Wang, C. Peng, H. Liang, and B. Du, RETRACTED: An event detection method for social networks based on link prediction, *Inf. Syst.*, vol. 71, pp. 16–26, 2017.
[102] 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 and Analytics*., vol. 1, no. 1, pp. 34–46, 2018.

[103] D. Kempe, J. Kleinberg, and É. Tardos, Maximizing the spread of influence through a social network, in *Proc. 9th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Washington, DC, USA, 2003, pp. 137–146.

[104] H. Liang and K. W. Fu, Network redundancy and information diffusion: The impacts of information redundancy, similarity, and tie strength, *Commun. Res.*, vol. 46, no. 2, pp. 250–272, 2019.

[105] J. Wu, J. Chang, Q. Cao, and C. Liang, A trust propagation and collaborative filtering based method for incomplete information in social network group decision making with type-2 linguistic trust, *Comput. Ind. Eng.*, vol. 127, pp. 853–864, 2019.

[106] M. Richardson and P. Domingos, Mining knowledge-sharing sites for viral marketing, in *Proc. 8th ACM SIGKDD Int. Conf. on Knowledge Discovery and Data Mining*, Edmonton, Canada, 2002, pp. 61–70.

[107] J. Yang and J. Leskovec, Modeling information diffusion in implicit networks, in *Proc. 2010 IEEE Int. Conf. on Data Mining*, Sydney, Australia, 2010, pp. 599–608.

[108] M. Wu, J. Guo, C. Zhang, and J. Xie, Social media communication model research bases on Sina-weibo, in *Proc. of the Sixth International Conference on Intelligent Systems and Knowledge Engineering,* Shanghai, China, 2011, pp. 445–454.

[109] F. Xiong, Y. Liu, Z. J. Zhang, J. Zhu, and Y. Zhang, An information diffusion model based on retweeting mechanism for online social media, *Phys. Lett. A*, vol. 376, nos. 30&31, pp. 2103–2108, 2012.

[110] P. Yang, G. Yang, J. Liu, J. Qi, Y. Yang, X. Wang, and T. Wang, DUAPM: An effective dynamic micro-blogging user activity prediction model towards cyber-physical-social systems, *IEEE Trans. Ind. Inf.*, vol. 16, no. 8, pp. 5317–5326, 2020.

[111] S. Karthika and R. Geetha, Communalyzer—understanding life cycle of community in social networks, in *Innovations in Computer Science and Engineering*, H. S. Saini, R. Sayal, A. Govardhan, and R. Buyya, eds. Singapore: Springer, 2019, pp. 197–204.

[112] M. J. Kim, C. K. Lee, and N. S. Contractor, Seniors' usage of mobile social network sites: Applying theories of innovation diffusion and uses and gratifications, *Comput. Hum. Behav.*, vol. 90, pp. 60–73, 2019.

[113] D. Bhattacharya and S. Ram, Sharing news articles using 140 characters: A diffusion analysis on Twitter, in *Proc. 2012 IEEE/ACM Int. Conf. on Advances in Social Networks Analysis and Mining*, Istanbul, Turkey, 2012, pp. 966–971.

[114] E. Bakshy, I. Rosenn, C. Marlow, and L. Adamic, The role of social networks in information diffusion, in *Proc. 21st Int. Conf. on World Wide Web*, Lyon, France, 2012, pp. 519–528.

[115] J. S. More and C. Lingam, A gradient-based methodology for optimizing time for influence diffusion in social networks, *Soc. Netw. Anal. Min.*, vol. 9, no. 1, p. 5, 2019.

[116] L. Yang, J. Wang, C. Gao, and T. Li, A crisis information propagation model based on a competitive relation, *J. Ambient Intell. Hum. Comput.*, vol. 10, pp. 2999–3009, 2019.

[117] H. Kwak, C. Lee, H. Park, and S. Moon, What is Twitter, a social network or a news media? in *Proc. 19th Int. Conf. on World Wide Web*, Raleigh, NC, USA, 2010, pp. 591–600.

[118] R. Ghebleh, A comparative classification of information dissemination approaches in vehicular ad hoc networks from distinctive viewpoints: A survey, *Comput. Netw.*, vol. 131, pp. 15–37, 2018.

[119] Z. Tan, D. Wu, T. Gao, I. You, and V. Sharma, AIM: Activation increment minimization strategy for preventing bad information diffusion in OSNs, *Future Gener. Comput. Syst.*, vol. 94, pp. 293–301, 2019.

[120] A. Bovet and H. A. Makse, Influence of fake news in Twitter during the 2016 US presidential election, *Nat. Commun.*, vol. 10, no. 1, p. 7, 2019.

[121] S. Vieweg, A. L. Hughes, K. Starbird, and L. Palen, Microblogging during two natural hazards events: What Twitter may contribute to situational awareness, in *Proc. SIGCHI Conf. on Human Factors in Computing Systems*, Atlanta, GA, USA, 2010, pp. 1079–1088.

[122] Y. Qu, C. Huang, P. Zhang, and J. Zhang, Microblogging after a major disaster in China: A case study of the 2010 Yushu earthquake, in *Proc. ACM 2011 Conf. on Computer Supported Cooperative Work*, Hangzhou, China, 2011, pp. 25–34.

[123] Y. Zhang, J. Zhou, and J. Cheng, Preference-based top-K influential nodes mining in social networks, in *Proc. IEEE 10th Int. Conf. on Trust, Security and Privacy in Computing and Communications*, Changsha, China, 2011, pp. 1512–1518.

[124] D. Varshney, S. Kumar, and V. Gupta, Modeling information diffusion in social networks using latent topic information, in *Proc. 10th Int. Conf. on Intelligent Computing*, Taiyuan, China, 2014, pp. 137–148.

[125] S. Pei, L. Muchnik, J. S. Andrade Jr, Z. Zheng, and H. A. Makse, Searching for superspreaders of information in real-world social media, *Sci. Rep.*, vol. 4, no. 1, p. 5547, 2014.

[126] Q. Xuan, X. Shu, Z. Ruan, J. Wang, C. Fu, and G. Chen, A self-learning information diffusion model for smart social networks, *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 3, pp. 1466–1480, 2020.

[127] N. Dakiche, F. B. S. Tayeb, Y. Slimani, and K. Benatchba, Tracking community evolution in social networks: A survey, *Inf. Process. Manag.*, vol. 56, no. 3, pp. 1084–1102, 2019.

[128] Z. Wang and Y. Guo, Rumor events detection enhanced by encoding sentimental information into time series division and word representations, *Neurocomputing*, vol. 397, pp. 224–243, 2020.

[129] F. Figueiredo, F. Benevenuto, and J. M. Almeida, The tube over time: Characterizing popularity growth of youtube videos, in *Proc. 4th ACM Int. Conf. on Web*

Search and Data Mining, Hong Kong, China 2011, pp. 745–754.

[130] C. C. Yang, X. Shi, and C. P. Wei, Discovering event evolution graphs from news corpora, *IEEE Trans. Syst., Man, Cybern.-Part A*: *Syst. Hum.*, vol. 39, no. 4, pp. 850–863, 2009.

[131] B. Suh, L. Hong, P. Pirolli, and E. H. Chi, Want to be retweeted? Large scale analytics on factors impacting retweet in twitter network, in *Proc. IEEE 2nd Int. Conf. on Social Computing*, Minneapolis, MN, USA, 2010, pp. 177–184.

[132] Y. Artzi, P. Pantel, and M. Gamon, Predicting responses to microblog posts, in *Proc. 2012 Conf. of the North American Chapter of the Association for Computational Linguistics*: *Human Language Technologies*, Montréal, Canada, 2012, pp. 602–606.

[133] M. Hasan, M. A. Orgun, and R. Schwitter, A survey on real-time event detection from the Twitter data stream, *J. Inf. Sci.*, vol. 44, no. 4, pp. 443–463, 2018.

[134] S. Dutta, A. K. Das, G. Dutta, and M. Gupta, A comparative study on cluster analysis of microblogging data, in *Emerging Technologies in Data Mining and Information Security*, A. Abraham, P. Dutta, J. K. Mandal, A. Bhattacharya, and S. Dutta, eds. Singapore: Springer, 2019, pp. 873–881.

[135] H. J. Choi and C. H. Park, Emerging topic detection in twitter stream based on high utility pattern mining, *Expert Syst. Appl.*, vol. 115, pp. 27–36, 2019.

[136] R. Nallapati, A. Feng, F. Peng, and J. Allan, Event threading within news topics, in *Proc. 13th ACM Int. Conf. on Information and Knowledge Management*, Washington, DC, USA, 2004, pp. 446–453.

[137] C. Lin, C. Lin, J. Li, D. Wang, Y. Chen, and T. Li, Generating event storylines from microblogs, in *Proc. 21st ACM Int. Conf. on Information and Knowledge Management*, Maui, HA, USA, 2012, pp. 175–184.

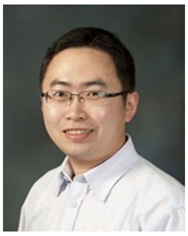
[138] S. Mishra, Bridging models for popularity prediction on social media, in *Proc. 12th ACM Int. Conf. on Web Search and Data Mining*, Melbourne, Australia, 2019, pp. 810–811.

[139] D. R. Liu, H. Omar, C. H. Liou, H. C. Chi, and C. H. Hsu, Recommending blog articles based on popular event trend analysis, *Inf. Sci.*, vol. 305, pp. 302–319, 2015.

[140] D. Lin, B. Ma, M. Jiang, N. Xiong, K. Lin, and D. Cao, Social network rumor diffusion predication based on equal responsibility game model, *IEEE Access*, vol. 7, pp. 4478–4486, 2018.

Zixuan Han received the BEng degree from Jilin Normal University, China in 2018, and the MEng degree from Jilin Normal University, China in 2021. He is currently a PhD candidate at School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include data mining, social networks, and cloud computing.

Leilei Shi received the BEng degree from Nantong University, China in 2012, the MEng degree from Jiangsu University, China in 2015, and the PhD degree from Jiangsu University, China in 2020. He is currently a lecturer at School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include event detection, data mining, social computing, and cloud computing.



Lu Liu received the BEng degree from South Central Minzu University, China in 2002, the MEng degree in data communication system from Brunel University, UK in 2003, and the PhD degree in electronic engineering from University of Surrey, UK in 2007. He is a professor at School of Computing and Mathematical Sciences, University of Leicester, UK, and an adjunct professor at Jiangsu University, China. He is a fellow of British Computer Society (BCS). His research interests are in areas of cloud computing, social computing, service-oriented computing, and peer-to-peer computing.

Liang Jiang received the MEng and the PhD degrees from Jiangsu University, China in 2011 and 2021, respectively. He is currently a lecturer at Ocean College, Jiangsu University of Science and Technology, China. His research interests include Online Social Networks (OSNs), data mining, social computing, and network security.

Jiawei Fang is an undergraduate student at School of Computer Science and Communication Engineering, Jiangsu University, China. His research interests include data mining, social networks, and cloud computing.

Fanyuan Lin is an undergraduate student at School of Computer Science and Communication Engineering, Jiangsu University, China. Her research interests include data mining and social networks.

Jinjuan Zhang is an undergraduate student at School of Computer Science and Communication Engineering, Jiangsu University, China. Her research interests include data mining and social networks.

John Panneerselvam is a lecturer in computing at University of Leicester, UK. He received the MEng degree in advanced computer networks and the PhD degree in computing from University of Derby, UK in 2013 and 2018, respectively. He is an active member of IEEE and British Computer Society. His research interests include cloud computing, fog computing, internet of things, big data analytics, opportunistic networking, and P2P computing. He has won the best paper award in IEEE International Conference on Data Science and Systems in 2018.

Nick Antonopoulos received the PhD degree in computer science from University of Surrey, UK in 2000. He is currently the vice principal of Research and Innovation at Edinburgh Napier University, UK. His research interests include cloud computing, P2P computing, and software agent architectures and security. He has more than 18 years of academic experience and has published more than 150 articles in fully refereed journals and international conferences. He is a member of the IEEE.