Check for updates

A survey of location-based social networks: problems, methods, and future research directions

Xuemei Wei[1] · Yang Qian[1,2] · Chunhua Sun[1,2] · Jianshan Sun[1,2] · Yezheng Liu[1,3]

Abstract

The development of mobile devices and positioning technology has facilitated the rapid growth of location-based social networks (LBSNs). Users in these networks can share geo-related information in real-time, including locations, trajectories, geo-tagged pictures, and tweets. LBSNs record massive amounts of spatiotemporal data and offer a great opportunity to analyze human and location-specific spatiotemporal characteristics. It plays an important role in various applications, such as marketing, recommendations, and urban planning. In this study, we collect relevant literature about LBSNs research in the past 10 years and use a topic model, latent Dirichlet allocation (LDA), to uncover the highly heterogeneous area of research related to LBSNs. Then, we conduct a systematic review of those works. In doing so, we organize identified literature into eight fine-grained directions. For each direction, we sum up the major research focus and contributions. We also systematize future research into four main themes concerning data simulation and fusion, privacy-aware methods, new applications and services, and technological innovations.

Keywords LDA · Data mining · Spatio-temporal · Recommendation · Mobility · Privacy

1 Introduction

Location-based social networks (LBSNs) or geographic social networks, such as Foursquare and Swarm, have rapidly developed in recent years. In contrast to the traditional social network. LBSNs combine social networks and location-based services, thus enabling the sharing of locations and location-related information amongst users on online platforms. For example, users can share their current locations with friends in Foursquare or send tweets with location tags to record their activities in specific places. Meanwhile,

✉ Yang Qian
   soberqian@hfut.edu.cn

1   School of Management, Hefei University of Technology, Hefei, China

2   Key Laboratory of Process Optimization and Intelligent Decision Making, Ministry of Education, Hefei, China

3   National Engineering Laboratory for Big Data Distribution and Exchange Technologies, Hefei, China

The location dimension in LBSNs bridge the gap between the physical world and online social networks, therefore, enables new interactions between users and locations. According to different interaction objectives, a network can be further divided into user relationship network, location relationship network, and user-location network, with each layer containing rich information. For users, the addition of location information enriches their profiles, which helps analyze user characteristics and user relationships from a spatiotemporal perspective. For locations, geo-related information enriches location semantics; simultaneously, user's movement patterns in different locations provide a new resource to study location characteristics and relationships. Numerous studies have analyzed the complex interactions in LBSNs, resulting in the eight main research directions discussed in this paper. Specifically, location recommendation is based on a user-location network, using user interaction and location relationship network to enrich their information, while route recommendation is an extension of location recommendation in the application. Friendship prediction and community detection are both based on user relationship network and use user-location interaction information to further explore the similarity between users in the offline world. The difference is that friendship prediction focuses on the similarity between user-pairs while community detection emphasizes more on the similarity between groups, and community discovery in LBSNs emphasizes the importance of geographic closeness. The influence maximization study in LBSNs measures the spatio-temporal influence of different users or locations based on the user-location network. Urban mobility analysis is a macro analysis of user behavior and emphasizes the application of LBSNs research to urban governance, urban security, etc. Privacy issues are widely concerned in LBSNs research, which affects the data quality and service quality of LBSNs. When users share geo-tagged information in LBSNs, service providers can capture and record those data as a quadruple $< user, time, location, content >$. Location is recorded by location ID or coordinates, and content is the text/figure shared by users. The data shared by a user at different moments constitutes his/her trajectory information. However, due to privacy constraints, the LBSNs dataset is a scarce resource, and publicly available data are very limited. Table 1 shows widely used public LBSNs datasets and their statistics, most of the existing works are based on those data. Also, many works have explored users' spatiotemporal information by collecting user-generated content with location tags from online social networks, or by obtaining users' mobile data through city sensors. However, these data are mostly non-public.

The generation and development of LBSNs have attracted the attention of researchers in different fields, such as psychology, economics, and geography. However, this paper focuses on applications of LBSN data and the main research topics or methods of LBSN dataset mining. The vast quantities of works published by numerous publishers with online presence and are available through the Internet pose a challenging task for conducting a proper literature analysis. Thus, the automated textual analysis approach is urgently needed, as it can discover and retrieve high-quality semantic information patterns and trends in the recurrence of text within the literature [1]. To conduct an efficient review of the literature,

Table 1 Public LBSB datasets

Dataset	Users	Links	Check-ins	Period
Gowalla[2]	196,591	950,327	6,442,890	2009.02–2010.10
Brightkite[2]	58,228	214,078	4,491,143	2008.04–2010.10
Foursquare[3]	114,324	607,333	22,809,624	2012.04–2014.01

we combine the application of text mining techniques with topic modeling procedures in the research. The detailed analysis process and results will be introduced in Sect. 2.

Many researchers have analyzed the research about LBSNs and published classic articles. Bao et al. [4] summarized existing studies on recommender systems for LBSNs from three perspectives: data source, recommendation objectives, and recommendation methods. Riaz et al. [5] analyzed location privacy and utility in geo-social networks, revealing that most existing mechanisms either ignored social platform-related user-privacy concerns or disregarded location data quality-related demands of platform providers. Silva et al. [6] discussed fundamental concepts of urban computing leveraging LBSNs data and surveyed recent urban computing studies that utilized such data. These studies only focus on the application of LBSNs in a specific field, such as recommendation, urban planning, etc. However, the purpose of this article is to retrieve a large number of papers and apply an effective text analysis method to clarify the research trends and research hotspots related to LBSNs while also discussing research challenges and potential research directions. The contributions of this study are detailed as follows.

(1) Prior surveys focus on one area of LBSNs, such as recommendation or link prediction. In contrast, the current survey is more comprehensive, focuses on the applications of LBSN data, and analyzes main research topics and methods of LBSN datasets mining.
(2) To conduct an efficient review of the existing work, this study applies a topic analysis method namely latent Dirichlet allocation (LDA) to process the collected papers.
(3) Based on the results of LDA, we detect the main research directions (topics) hidden in literature and sum up the challenging research problems and methods of each direction in detail.

The rest of this paper is organized as follows. In Sect. 2, we introduce the main process of LDA and explain the result in detail. In Sect. 3, we categorize and summarize the existing research from different aspects based on the topics derived from LDA. Section 4 describes the limitations and challenges of existing works and potential research directions. Section 5 presents the conclusions of this work.

2 Materials and methods

2.1 Selecting the articles

We synthesize the existing research by using the integrative literature review method. Articles are chosen from multiple disciplines, including information systems, computer science, sociology, and marketing. The databases include IEEE Xplore, ACM Digital Library, ABI/INFORM Complete (ProQuest), Business Source Complete (EBSCO), Emerald, Springer Link, Elsevier, ScienceDirect, Wiley Online Library, Taylor & Francis ST, and Web of Science (SCIE). In addition to electronic database search, the references listed on the retrieved articles are also examined to check whether relevant studies can be retrieved. Studies are included if they use data from LBSNs. The search includes articles written in English from January 2009 to 2020, as LBSNs appeared around 2009. Boolean search is used to limit the search to only those articles containing the keywords "location-based social network," "geo-social network," "check-ins," "spatiotemporal social network," "mobile social network," "event-based social network," and "location-based social media".

Attention is articles containing the keywords "location-based social network," "geo-social network," "check-ins," "spatiotemporal social network," "mobile social network," and "location-based social media". Attention is given to commonly used synonyms, such as "LBSNs," "LBSN," and "geographic social network." We also sift the title, abstract, and introduction of each retrieved article to ensure compliance with the inclusion criteria. Studies that do not fit the criteria are excluded. Eventually, 1,381 papers are collected. The distribution of republication time is shown in Fig. 1. As can be seen, the yearly number of publications rises with the popularity of LBSNs.
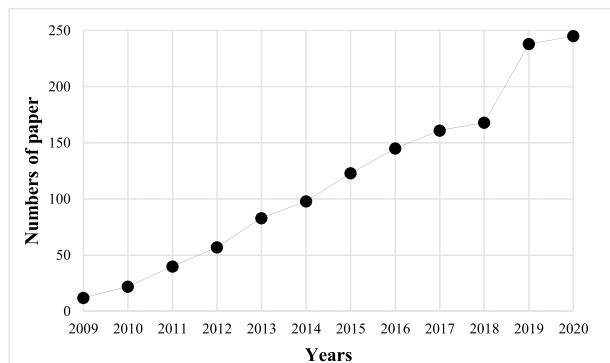
2.2 Text mining

To identify current research trends and interesting future directions, we apply LDA, which has been proven to be highly suitable for dealing with documents with multiple topics for text mining [1]. We extract the title, abstract, and keywords of each paper and transform them according to the best practices in the text-mining analysis [7]. The transformation process is as follows. (1) All the stop words, such as "a," "the," "at," and "from," which are common in English, are deleted. (2) Corpus is stripped of whitespaces, punctuations, and numbers. (3) All words are lemmatized and converted to lower case. After that, the processed results are used as inputs of LDA.

One necessary parameter of LDA is the number of topics: fewer topics correspond to broader topics, whereas a large number of topics may reduce their interpretability. Following the prior studies[8, 9], we select the perplexity score as the criterion for topic number selection. Perplexity score is a standard measure for assessing how well one topic model fits the data. To determine the number of topics objectively, we compute perplexity scores using threefold cross-validation for 5, 10, 15, and 20 topics, respectively. From the results, we finally choose 15 topics. However, empirical applications have shown that LDA prefers very large topic numbers [10, 11]. In addition, LDA may infer overlaps topics. To solve this problem, following the literature [12], we use the symmetrized Kullback-Liebler (KL) distance to measure the similarity between topics and finally obtain eight summative topics. We hope that future research will propose more objective and effective methods to optimize the number of topics. In Fig. 2, we calculate the weight of each topic by integrating the information of the document-topic distribution.

Through further analysis, we observe that many meaningless, high-frequency words, such as "show," "method," and "model," always have high rankings in each topic, which

Fig. 1 Distribution of papers from January 2009 to 2020

Fig. 2 Probability of papers per topic

may impair discovered topics. Meanwhile, some low-frequency words are also useless, because they lack representativeness. Therefore, similar to the literature [13], we calculate the TF-IDF value for each word in every topic and select the most representative words to describe the topic. The corresponding word cloud for each topic is displayed in the following sections.

3 Results and analysis

3.1 T1 location recommendation

The words displayed in Fig. 3, such as "location," "POIs," and "recommendation," imply that this topic is highly associated with location recommendation. Location recommendation aims to mine users' movement patterns or rules from their trajectories and explores their location preferences to make recommendations. According to [14], whether the recommended locations meet users' interests depends on three factors, namely, user preference, social influence, and geographical influence. Users' activities are time regular, and the temporal factors also affect their behaviors. Therefore, researchers have proposed various methods to capture the impacts of these factors on users' check-in behaviors and make effective location recommendations.

Fig. 3 Word cloud for topic 1

The most common approach to identify user preferences is to analyze their historical trajectories, which reflect the users' mobility patterns and location preferences. In prior literature, trajectory data is represented as different structures, and different models and methods have been proposed for them. One of the most common structures is the matrix, including user–location matrix and user–time–location three-dimension (3D) matrix. Matrix-based methods, such as hidden Markov model (HMM) [15, 16], matrix factorization [17–19], and collaborative filtering [20–22], can be used to analyze user mobility. Based on the check-in distribution, some researchers use probability models [13, 23, 24] or clustering methods [25, 26] for location recommendation. However, those methods only focus on the users' check-in locations without considering the sequence correlation between check-ins. The existing study has pointed out that trajectory sequences reflect personal interests and preferences [27]. Markov chains technique [28] can capture the sequence relation between locations, but they require higher computation. Embedding models [29–33] and Deep Neural Networks (DNN) can directly mine user features from their trajectory, thus, achieve relatively better performance. They can map users or locations to low-dimensional vectors, which contain implicit features. Especially, DNN can not only incorporate various context information but also have high-resolution structures, thus achieving the state-of-the-art result in location recommendation. Considering the sequential properties of user check-ins, Recurrent Neural Network (RNN) has become the most common approach for trajectory information mining. To capture the spatial–temporal contextual information, Spatial–Temporal Recurrent Neural Networks (ST-RNN) [34] and Contextual Attention Recurrent Architecture (CARA) [35] were proposed. Both of them learn user dynamic preference from the time interval and geographical distance among user check-in sequences. However, both of them focus on short-term preferences behind user movement patterns and ignore users' long-term preferences. To better capture the variability of user preferences, the Hierarchical Variational Attention Model (HVAM)[36] and ARNPP-GAT[37] were constructed to learn users' long-term and short-term preferences simultaneously. They exploited different attention mechanisms to capture the effect of distant check-ins on users' current location preferences. Besides, considering the dynamics of user preferences, recommending locations for a given user at a specified time [38–40] has also be discussed.

Users' preferences are easily influenced by surrounding people. Trust and interest similarity carried along in social relationships amongst users can enhance personalized search and recommendations [41] and the addition of friend information can also alleviate the cold-start problem. Therefore, friend-based collaborative filtering is commonly used to integrate social relationships for location recommendation, in which the target user's check-in behavior can be predicted from his/her friends' check-in records or ratings [42, 43]. However, some users may have a limited number of friends, and non-friends who share common locations also contribute to the recommendation. In [44], authors defined three types of friends, i.e., social friends, location friends, and neighboring friends based on the online social connection, check-ins records, and geographic distance between users. Then they learned a set of potential locations from those friends for the target user. Similarly, [45] defined four types of geo-social correlations, i.e., local friends, local non-friends, distant friends, and distant non-friends, then, investigated the correlations between a user's check-ins and his geo-social circles. [46, 47] used multidimensional tensors to model various contextual information in users' check-ins data, in [46], user relationships were used as regularization terms of the factorization, while in [47], a Gauss radial basis function (RBF) based model named SVR is used to extract the influence weights of different users. Embedding models [29, 31] and DNN [48, 49] can automatically learn user social features from social networks. Both [29] and [31] used random walk-based methods to sample

user nodes, and then learn node embeddings from them by preserving node proximity in the embedding space. DNN can integrate user social information with other features to improve recommendation results. For example, [48] constructed a Context Graph Attention Model (CGA), which can learn the influence weights of different friends and neighboring locations by attention mechanism. [49] proposed a random walk masked self-attention option to balance the influence of higher-order neighbors and first-order neighbors on target user's check-in behavior.

Geographical characteristics are the most essential and effective information for location recommendation. Through the analysis of the literature, we find that existing works mainly focus on the following location characteristics: location semantic, location distance, and location popularity. Different locations, such as hotels, cinemas, companies, etc., have unique semantic tags, which reflect the category or function of different locations. To analyze characteristics of user trajectories and predict the locations that users may be interested in, some researchers extract the category information from users' check-in sequences and convert users' trajectories to semantic sequences [50–53]. The addition of semantic information can alleviate the sparsity problem of users' check-ins and improve the recommendation effect. Location distance is also a major feature in predicting locations that users may visit. It mainly includes two kinds of distances: the distance from the user's current location and the distance from the user's frequent activity areas or home address. Some works have proven that people tend to visit POIs (Point of interest) close to their homes or offices and that people might also be interested in exploring POIs surrounding those they are in favor of, even if they are far from home [2, 54, 55]. Therefore, many methods based on the probability distribution of user check-ins and spatial clustering methods have been proposed [23, 56, 57]. Meanwhile, location popularity can also affect user check-in behaviors to a great extent. Usually, people prefer to visit places with a better reputation [23, 58]. In addition to the location explicit features mentioned above, DNN can learn the implicit features of locations. Different locations have different geographical influences on user preferences. It is necessary to model the relationships between locations in user trajectories[59]. POI-POI graphs are commonly used to represent the relationships between locations, based on that, [48, 49] applied graph attention networks to learn location features, and [60] exploited graph auto-encoder to capture highly non-linear geographical influences of different locations. But, in [48] edges of the POI-POI graph only represent the geographical proximity between locations, while the edges in [49] containe the temporal, spatial, and preference information behind in users' trajectory. In [59], the authors used three factors, i.e., the geo-influence of POIs, the geo-susceptibility of POIs, and the distance between POIs, to model the geographical co-influence between POIs and constructed a Geographical-Temporal Awareness Hierarchical Attention Network (GT-HAN) to capture the geographical influence of them.

Note that, it is less effective to make recommendations relying on a single feature. Therefore, several methods jointly use rating, geotagged photos, and tweets to build a multi-factor model to improve the recommendation performance [33, 61–63]. In Table 2, we summarize the existing literature according to the methods and information they used.

3.2 T2 route planning

The words in Fig. 4 including "trip," "tourist," and "itinerary," point to route planning or trip recommendation. Route planning aims to recommend consecutive location sequences under user preferences and other constraints. It is more complicated than stand-alone

Table 2 Summary and comparison of different literature for location recommendation

Property Technology	User Trajectory	Social Relationship	Geographical Features	Time-Aware
Tensor Factorization	[16–19, 21, 13, 30]	[44, 46, 47, 54, 61]	[17, 46, 47, 54, 57, 58, 61]	[16, 18, 19, 21, 47]
Collaborative Filtering	[20-22, 62]	[38, 42, 43, 55, 62]	[38, 43, 55, 62]	[38, 62]
Probability Model	[23, 13, 24, 26, 39]	[26, 43, 45]	[23, 24, 26, 43, 45, 64]	[13, 40, 64]
Clustering Method	[25, 26, 39]	[52]	[52, 57]	[39]
Markov Model	[15, 16, 28]	[65]	[15]	[15, 16, 28]
Embedding Model	[29–32, 60]	[29, 31–33, 63]	[33, 63]	[31-33, 63]
Deep Neural Networks	[20, 34, 35, 59]	[20, 37, 48, 49, 53]	[48, 49, 53, 60, 59]	[34-37, 39; 49, 53, 59]

Fig. 4 Word cloud for topic 2
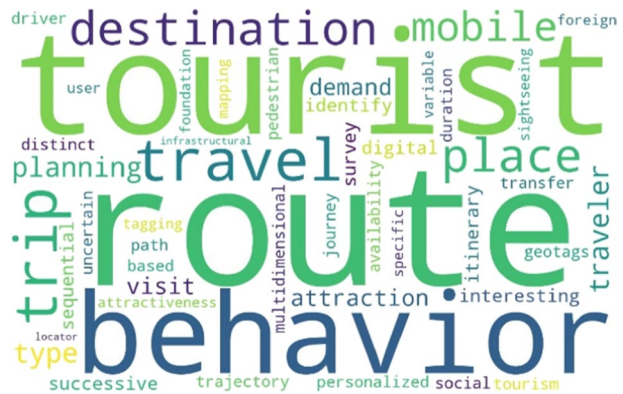
location recommendation since it needs to incorporate various real-world constraints, such as travel time and traffic conditions.

According to different scenarios, route planning can be divided into daily commuting planning and trip recommendation. The daily commuting route is more restricted by traffic conditions and transit time, while trip recommendations pay more attention to user interest and the influence of location features, such as POI popularity. For different constraints, different studies have been conducted.

Similar to stand-alone location recommendation in Sect. 3.1, the recognition of user preferences mainly depends on the user's historical trajectory. The matrix factorization technique [66, 67] and probabilistic model [68] have been used to capture the transition of user check-in in time and location. Represent methods [69] and deep neural network architecture have also shown superior results [32, 70].

Time budget constraint mainly refers to three types of time, i.e., total travel time, transition time between locations, and visiting time for each location. Concerning these, time-aware route recommendation methods have been proposed [71–75]. In [71], the route-planning problem was regarded as an optimization problem, and the time factor was viewed as a constraint condition for the objective function. Then, heuristic algorithms were applied to find the optimal route. The authors in [72] proposed a tensor-decomposition-based

time-sensitive route recommendation system, which considered users' preference, visiting time, and transition time. The sparsity of user time data was considered in [15], it proposed a time-aware route recommendation method based on the Markov model and used the data of similar friends to solve the problem of data sparsity. Traffic conditions are also an important factor for route planning, as it essentially determines the transit time. In [74], the author proposed a personalized, interactive, and traffic-aware trip planning service for users by leveraging the Foursquare check-in and taxi GPS data, where taxi GPS data is used to estimate the transit time between any two locations at any time period. To capture the complex spatio-temporal characteristics of route planning, the article [75] proposed heuristic workflow-aware trip planning algorithms whose generated location set not only meets the user activity intent but also has reasonable spatial distribution and satisfies the temporal constraints.

POI characteristics, such as POI popularity, availability, and diversity, affected the effectiveness of the recommendation. For example, users might want to visit various locations during a trip, and POIs might be available only at certain times. Therefore, the author of [73] set a POI diversity threshold to meet controls the minimum number of POI categories in candidate routes. In [74], the total number of visitors and the total number of check-ins were used to compute the popularity of a given venue. In [76], the authors proposed a travel-route-generating algorithm considering the location popularity and distance between each location pair. A travel route was constructed as a path tree, and the starting location was viewed as the root. The locations in the POI list of different times were presented as tree nodes of different levels. Travel route planning could be treated as the problems of strolling amongst tree nodes and generating feasible paths. In [77], the author measured the diversity by using the accumulation of the maximal POI similarity between a POI and another one in the trip, POI similarity is calculated by using the hierarchy of a POI category, after that, trip diversity was used as an integrated objective function during trip planning. In recent years, DNNs have been increasingly applied to route recommendations [78, 79], and they focus more on the learning of location features and long-range dependence exhibited in user trajectories than on single location recommendations.

In Table 3, we have given a summary of existing route planning belonging to different categories.

3.3 T3 friendship prediction

Words in topic 3, such as "link," "friendship," "relationship," and "inference," pertain mostly to friendship prediction or link reference (Fig. 5). Past studies [80, 81] have proven that geographical proximity has an apparent correlation with user relationships. Friendship prediction in LBSNs aims to discover the similarities amongst users by considering spatial proximity and social closeness. The mainstream of methods in this area is always based on the following information sources: check-in location, trajectory, and contextual information.

3.3.1 Methods based on check-in location information

The methods based on check-in location information focus on the check-in points of each user rather than on sequences or trajectories. They identified users' preferences or habits in a physical world from their check-in locations or regions and then combined such data with users' online social property to make high-quality predictions. The most

Fig. 5 Word cloud for topic 3

Table 3 Summary and comparison of different works on route planning

		Traffic condition	Time constraint	Location feature	Trajectory history
Daily Route	[66, 68, 67]				√
	[69, 72]		√		√
	[70, 78, 79]		√	√	√
Travel Route	[71, 75]		√	√	
	[32]				√
	[73, 76]		√	√	√
	[74]	√	√	√	√
	[77]		√		√

common and direct method is to define and extract features to characterize users' homophily and social proximity. On this basis, different classifiers can be applied to discover potential friends [82–84]. The co-location events, which refer to a pair of users visiting the same place, are an important signal for friendship. In [82], the authors found that approximately 30% of new links appeared amongst users who visited the same places. Thus, several studies judged user closeness by extracting user co-location features, such as the frequency of co-location, closeness of significant locations, and the probability of co-location. In [83], the author proposed a variety of check-in location features, including GeoDist and check-in observation, and then compared their predictive capabilities. To integrate multiple information, various works combined online and offline features and then compared user similarities by defining multiple features. In [84], the authors constructed three types of networks, namely, social network, co-location network, and co-located friend networks. For each network, they also defined five categories containing approximately 67 features to describe the characteristics. Numerous user-generated contents, such as a post for a POI, were also generated in LBSNs. Such semantic information can also be used to assist link prediction [85].

In addition to explicit features, the implicit features of user check-in information were also analyzed and used for friendship prediction. Classical methods include matrix decomposition [86] and representation learning [29, 31].

3.3.2 Methods based on trajectory information

Although the addition of check-in location information effectively improves the accuracy of friendship prediction, there are some limitations. First, a pair of friends who seldom co-occurrence or a pair of strangers who often visit the same locations will receive an unreliable estimation of the real social strength between them. Second, these studies do not consider how the overall trajectory patterns of users change with the varying living styles [87]. On the other hand, factors that the positioning errors of GPS devices, the crowded distribution of locations in a city, and the time difference of users' check-ins, make it difficult to define user co-occurrence [88]. Therefore, trajectory similarity has recently been used for user mobility homophily analysis. Related methods mine the trajectory pattern behind user trajectory rather than geo-positions, and analyze the mobility similarity between users. The underlying intuition is that the trajectory sequences reflect the "lifestyle" of users, and people have similar "lifestyle" are more likely to be friends [88]. Meanwhile, trajectory-based methods address the problem caused by using explicit co-occurrence patterns [89] or predicting new friendships in different cities. Different locations in urban areas have different themes, and geographically close users may not perform the same activities. For example, when two users appear in a business district at the same time, one user may go to the movies, whereas the other may go shopping. On the contrary, users with similar preferences may live far away, and a geographic overlap may be difficult. Therefore, many scholars have attempted to mine semantic information in user locations to compare user similarities. In [90], the authors used the tags of the landmarks that the users passed by as the semantic labels of their trajectories. Such a trajectory, namely, < school, park, restaurant >, could then be expressed as a semantic sequence. Then, they extracted the maximal semantic trajectory pattern of each user's trajectory and used it to measure the similarity between them. Authors of [91] converted each stay region into a feature vector and clustered the stay points into different categories. Each category had a semantic meaning. User trajectories were transformed into location histories amongst those categories named semantic location history (SLH). After that, user similarity could be calculated from their SLH.

In addition to trajectory matching methods, topic models are also common methods for mining location semantics and user preferences. The authors in [87] proposed a probabilistic generative model to learn lifestyle-related patterns from users' trajectories and measure similarities amongst different users. This model considered user preference, functionality dependence between different locations, service duration, and selected lifestyle during the trajectory generation process.

3.3.3 Methods based on context

Users' mobile behaviors are affected by multiple contextual information, including user-, trajectory-, and location-levels as well as temporal contexts. Thus, contextual factors are crucial for studying the similarity of user movement patterns. The most common and useful methods to incorporate contextual information are embedding models and deep neural networks. The main idea behind these methods is to represent each entity in a network as a feature vector (embedding vector) containing considerable contextual information in a low-dimensional vector space. The relationships amongst different entities can then be derived from those vectors.

Many researchers have constructed different neural network architectures to learn the representation of different entities in LBSNs (e.g., user, location, trajectory) from the context of user check-in data. In [92], authors proposed a multi-context trajectory-embedding model, which can flexibly characterize various contexts, including user-, trajectory-, and location-levels, as well as temporal contexts in a unified manner. Each contextual feature could be represented as a distributed vector, and the final user representation can be obtained by concatenating those vectors. The authors in [93] proposed a generic embedding model called geographical convolutional neural tensor network (GeoCNTN) to extract geographical features of entities. In this model, the input of the embedding layer was obtained from the original geographical data by a curvature-sensitive fuzzy clustering algorithm (Geo-CMeans). A neural network-based learning model was then used to extract local patterns and global geographical features. Finally, a neural tensor network was applied to discover higher-order correlations between global and local features. In [3], authors viewed LBSNs as a hypergraph, including user–user edges (friendships) and user–time–POI–semantic hyperedges (check-ins), and proposed an embedding method named LBSN2vec. This method first applied a random-walk-with-stay scheme to jointly sample user check-ins and social relationships, after which it learned nodes embeddings from those edges by preserving n-wise node proximity [94] also built a graph prototype to represent LBSN and learned node embeddings by integrating the geographical influence, user social relationship, and temporal information.

To categorize the existing methods, we show the different information sources and features in Fig. 6. Based on that, various methods have been proposed to resolve the prediction task.
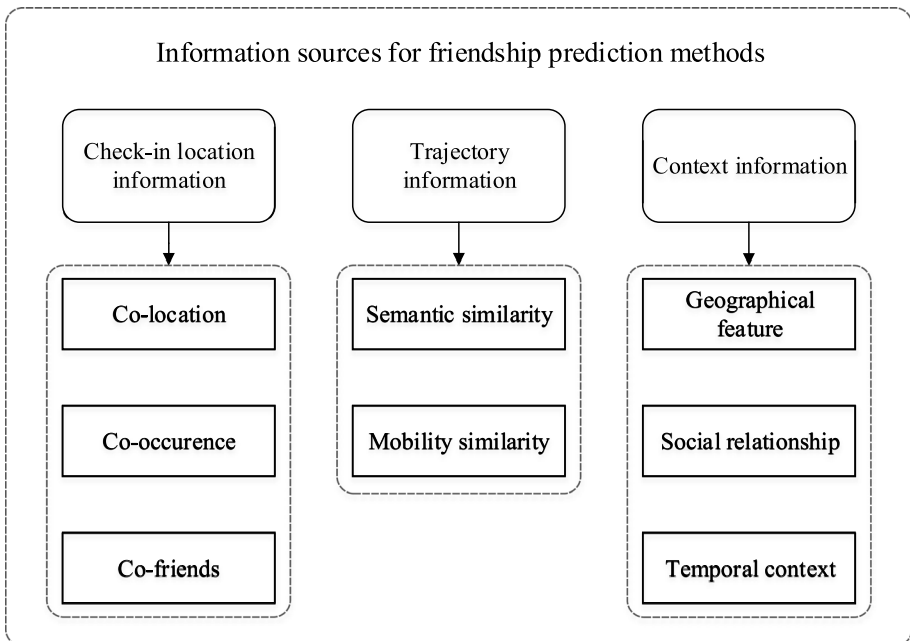

Fig. 6 Different information and features used in existing methods for friendship prediction

3.4 T4 influence maximization

In topic 4, words such as "influence," "propagation," "spread," and "maximization" suggest that this topic is related to the influence maximization issues. Influential maximization refers to the selection of a group of nodes that can maximize the scope of information spread under a certain propagation model (Fig. 7). In online social networks, many metrics and methods can be used to calculate the influence of nodes. However, in LBSNs, information can be spread not only through online social networks but also between neighbors who are close to one another in the physical world. Thus, certain methods may be inappropriate for LBSNs. Given that these LBSNs are heterogeneous networks, they have two kinds of nodes, namely, users and locations. Researchers have proposed various methods to mine the influential users or locations of LBSNs.

3.4.1 Influential users

In LBSNs, social influence is also called spatiotemporal social influence, because a user's behaviors in the physical world can also affect other people. Many researchers have proposed different methods to measure the offline influence or constructed models to integrate user influence in the online and offline worlds and then obtain the influential nodes.

Several studies have attempted to find influential nodes using pre-defined features. For example, in [95], influential users were defined as people with high social and location centralities, and the location centrality meant having the highest number of neighbors inside a query region. Similarly, in [96], users were viewed as influential users if the total number of their friends was close to the average number of nodes in the network.

Geo-social multilayer models, which considered the influence of online and offline interactions, were proposed in [97, 98] for influential node detection. Reference [97] proposed a method to unify multiple interaction layers into an integrated source of the degree of nodes within a geo-social network and applied a Susceptible Infected Recovered Model (SIR) model to evaluate the influence of nodes. In [98], researchers proposed a double-layer scheme to integrate the online social network layer and the offline mobile network layer and then compressed the double-layer networks into a single-layer network by considering the weight of both layers. After that, to select the seeds, a practical greedy heuristic was used, which considered nodes' influences in the double layers.

Fig. 7 Word cloud for topic 4

Also, the budget restricts the effect of influential node mining. Therefore, numerous scholars have explored the problem of influential nodes mining with budget constraints. For example, authors in [99] proposed a multi-objective optimization-based influence spread framework to balance the trade-off between the objective of influence spread maximization and the objective of promotion cost minimization.

3.4.2 Influential locations

Influential location identification is also an important issue, as it is useful for many applications, such as outdoor marketing. In LBSNs, the influence of locations can be judged by users' social relationships or their movement patterns across locations. For example, in [100], authors defined the influence of a location by its capacity to spread its visitors to other locations, they thought that these visitors could spread the information obtained from that area to other areas. They constructed an interaction graph in which nodes and edges represented locations and users traveling between locations, respectively. If a user visited both locations within a limited time, then the user was a "bridging visitor". An influence was exhibited when many bridging visitors from one location to another exist. To alleviate the sparsity of check-in data, they also defined "friendship-based influence between locations" and used the friends of "bridging visitor" as "potential visitors" to calculate location influence. The work of [101] directly used users' influence in social networks and check-ins numbers to calculate location influence, the behind intuition was that check-ins made by high-influential users were more valuable than that by low-influential users because the former could be seen by more users. Some studies directly convert the relationship between locations into a location network, and then directly identify influential locations using traditional node influence calculation methods. In [102], authors constructed a location–location graph to capture correlations between locations, similar to [100], edges in this graph represented users' transitions among different locations. Then, they used betweenness centrality to measure the influence of locations.

3.4.3 Location promotion

Location promotion is a unique and valuable application of influential nodes in LBSNs. Therefore, the promotion effect is an important factor in evaluating the influence of nodes. This problem can be described as follows: given an LBSN G and an integer K, in which K seed nodes (user nodes or location nodes) should be selected to maximize the number of users eventually moving to the promoted location under influence diffusion models. Several researchers have formulated this problem as an influence maximization problem in LBSNs and constructed various information propagation models to solve this problem [103–106]. [103] proposed two user mobility models, namely, Gaussian- and distance-based mobility models, to capture check-in behaviors of users. Location-aware propagation probabilities could be derived based on those models. The authors also extended the independent cascade model to a location-aware independent cascade model, which could describe the process of information propagation and identify nodes with marketing value. However, authors in [103] just considered one location at a time, which may be ineffective. Therefore, [104] focused on a bundle of location promotion. Using information propagation in an LBSN, authors selected k users who checked in at a location in a given bundle of locations to maximize the number of users. A multi-location-aware independent cascade model and

a greedy algorithm were proposed to describe information propagation and select users with the greatest marginal gain.

The influence of communities was also considered in [105]. The authors believed that users belonging to the same geo-community might have a few well-visited locations, and conducting initial propagation at such locations could influence many more seed users. Thus, they tended to select a set of groups as influence seeds rather than a set of users. They did so by grouping users based on check-in records and extracting the common mobility model of different groups. By considering the social relationship and group-level moving probability of each group, they proposed a greedy algorithm to effectively select *K* seed groups over the graph. In [106], researchers focused on the importance of distance between users and promoted locations. They proposed that the variety of distances between users and promoted locations resulted in variations among users. Thus, when promoted location changed, the vital nodes set also changed. They then developed an index-based heuristic approach called MIA-DA to approximate the influence calculation.

Several methods and algorithms can be used to study information maximization. The authors of [107] presented a unified system called IMaxer that provided a complete pipeline of novel models and algorithms for influence maximization, after which they evaluated and compared IM techniques for a particular scenario (Fig. 8). In Table 4, we present a summary of existing influence maximization methods belonging to different categories.

3.5 T5 community detection

Words for topic 7 contain such terms as "community," "overlap," and "detection," which point to spatial community detection. Community detection is a popular topic in modern network science. A community is a group of people with tighter interactions with its members than outside the community. In LBSNs, users interact through the online and offline worlds. Thus, intimacy and similarities in both lines must be considered to identify communities. In [108], authors emphasized that structural properties of location-based communities were very different from those of common social networks, demonstrating that two community structures did not yield the same user groupings. Edges were more likely to form within local communities than within social communities, meanwhile, friends in the same local community were more likely to visit the same places. In terms of the dynamics of social and local communities, they also showed that local communities were more dynamic and volatile. Thus, given that the approaches of community detection on

Fig. 8 Word cloud for topic 5

Table 4 Summary and comparison of influence maximization in LBSNs

		Target node		Receive node		Method/feature
		User	Location	User	Location	
Influence User	[95]	√		√		Socio-spatial centrality
	[96]	√		√		Divide-and-conquer
	[97]	√		√		SIR model
	[98]	√		√		Greedy heuristics
	[99]	√		√		Multi-objective optimization
Influence Location	[100]		√		√	Greedy algorithm
	[101]		√	√		DBSCAN
	[102]		√	√		Betweenness centrality
Location Promotion	[103]		√	√		Independent cascade model
	[104]	√		√		MLICM
	[105]		√	√		GLP
	[106]		√	√		MIA-DA

traditional social networks might fail to capture the information found in LBSNs, the location information must be considered to infer communities.

Methods of community detection can be divided into the following categories: methods based on optimization, community searching, clustering methods, and generative models. We now introduce the related works on these methods.

3.5.1 Methods based on optimization

In optimization methods, researchers first defined a community segmentation objective function according to designed metrics or rules. Then, the best community segmentation result could be achieved by maximizing the objective function.

Modularity is the main metric to evaluate the quality of network partition. Many researchers have added spatial constraints to traditional modularity and then achieved community detection by optimizing it. For example, Liu et al. [109] proposed new modularity based on network and connection localities to find communities. Nodes in the community had tight connections and were distributed in a small geographical scope. Authors of [110] introduced "spatially near modularity," which combined the optimization function of the *k*-means clustering algorithm and Newman–Girvan (NG) modularity. In this method, k-means clustering was used to control the influence of the distance between nodes. Chen [111] modified modularity to geo-modularity by introducing an edge weight, which was the inverse of the geographic distance to the power of *n*. The power value *n* of the distance was determined using a proposed spatial clustering coefficient.

Several works have also constructed their own defined optimization functions. For instance, in [112], researchers viewed LBSN as a multi-modality non-uniform hypergraph, vertices of the hypergraph were all entities, and heterogeneous interactions (e.g., posting comments and uploading photos whilst visiting places) were regarded as hyperedges. Overlapping community detection task was formulated as an optimization problem of density subgraph detection over heterogeneous hypergraph.

3.5.2 Methods based on community search

The community search problem, which aims to find dense subgraphs containing a query vertex, has attracted considerable attention. Authors in [113] proposed a hybrid framework to integrate users' attributes and geolocations. First, they constructed users' interest vectors by extracting keywords associated with places they visited. Then, they constructed an attributed social graph by computing interest-based similarities. Finally, they attempted to find a *k*-core subgraph in which all vertices were intensively linked and distributed within a given range by core decomposition.

Given a query vertex, the authors in [114, 115] attempted to find a dense subgraph that contained it whilst satisfying social and spatial constraints on large geo-social networks. Authors in [114] solved this problem by conducting the best-co-located community searching. Here, they constructed an index structure called AC-Tree based on the spatial information, local structure information, and attribute information of the social network. After that, they pruned it to determine the community that satisfies structure cohesiveness, spatial cohesiveness, and quality optimality requirements. In [115], researchers defined this as an RB-k-core search problem and used k-core to insure social cohesiveness. They also utilized a radius-bounded circle to restrict the locations of users in an RB-k-core. In the end, the rotating-circle-based algorithm was proposed to solve this problem.

3.5.3 Methods based on clustering

Clustering methods can cluster nodes into different groups according to different distance measurement indicators and clustering rules. Unlike traditional social networks, the clustering algorithm in LBSNs must consider the online and offline distance between users. Therefore, different community clustering methods are designed according to the differences in distance measurement methods.

Authors in [116, 117] used edge clustering to detect overlapping communities. They formulated LBSNs as attributed bipartite graphs and proposed an edge-centric co-clustering framework to cluster communities, in which users were not only structurally close to one another but also shared similar attributes [116]. However, authors in [117] highlighted that the variety of clustering centers might result in different clustering results. After that, they proposed a fuzzy subtractive clustering method that determined cluster centers with a probability based on the potential of each data point. This method has higher accuracy than the method in [116]. In [118], a density-based geo-community detection model was proposed, which considered users' social connections and Euclidean distance to detect communities. authors used the common geo-social neighborhood to measure the similarity of two vertices and replaced the Euclidean distance in the DBSCAN algorithm with geo-social similarity. Except for explicit features for the measurement of user similarity, the work of [119] devised a deep autoencoder algorithm to obtain latent representations of users and then utilized the *k*-means algorithm for community detection.

The authors in [120] emphasized that a user's check-in location might be inaccurately mapped to a specific place but might also correspond to the distribution of a set of places within the neighborhood of the check-in position. Therefore, they transformed venue-less check-ins into a distribution of a set of places within the neighborhood of the check-in position and proceeded to model user behavior patterns from their check-in sequences. They then applied a fuzzy clustering technique to group users with similar behavior patterns.

Similar to [120], the work in [121] regarded users in the same community as having similar movement behaviors that were represented as trajectory patterns. They used a probabilistic suffix tree as a user's trajectory profile and then formulated a similarity measurement to calculate the distance between users. They also identified communities by using a clustering algorithm based on distance. [122] used a trajectory clustering algorithm to classify trajectories into different kinds of clusters according to semantic and geographical distances.

3.5.4 Methods based on the probabilistic generative model

Multiple factors should be considered when addressing the community discovery problem in LBSNs. However, identifying communities according to artificially defined metrics may lead to incomplete or inaccurate solutions. Many scholars used generative models to model the generation process of the community, explain the formation process of the community, and then efficiently identify them.

A unified probabilistic generative model, also known as User-Community-Geo-Topic, was proposed in [123] to simulate the generative process of communities as a result of network proximities, spatiotemporal co-occurrences, and semantic similarity. In this model, a community has three components: a multinomial distribution over topics, which represent the probability of each community on different topics, (2) a multinomial distribution over spatial items (locations), it means the probability of each community visits different spatial items, (3) and a probability vector, each component of it is the mean of a Bernoulli distribution representing the interaction probability between different communities. Each user is related with a multinomial distribution over communities, thus suggesting his/her membership with the community. User check-in and interaction behaviors were considered in this model, and the Gibbs Sampling was used for model inference. Zhang [124] emphasized that data from LBSNs contained rich knowledge on users' social interactions in communities, geographical mobility patterns between regions, categorical preferences on activities, interests in topics, and opinion expressions for sentiments. The latent communities, regions, activities, topics, and sentiments of users were deemed interdependent. Then, an LDA-based model called CRATS was used to jointly mine those latent variables by mimicking the decision-making process of users who visit check-in at venues. Xu [125] also described the generative process of check-ins by using a social relation and temporal-spatial topic model to identify communities. They defined two types of check-ins in LBSNs based on dynamic user interaction. The first type of check-ins meant visiting a place with friends, and the second type of check-ins meant visiting a place alone.

Figure 9 shows a summary of the main categories of community detection methods and representative algorithms.

3.6 T6 event detection

As shown in Fig. 10, words such as "event," "detect," "geotagged," and "discover" suggest that topic 5 is related to event detection from the geo-tagged text. LBSNs allow users to post what they are involved in with location information in a real-time manner. Those geo-located data contain large amounts of information related to local events, which can help obtain patterns of the dynamics of crowds throughout urban areas. Event detection can support (1) the detection of unexpected behaviors in the city and (2) the analysis of the posts to infer what is happening [126].
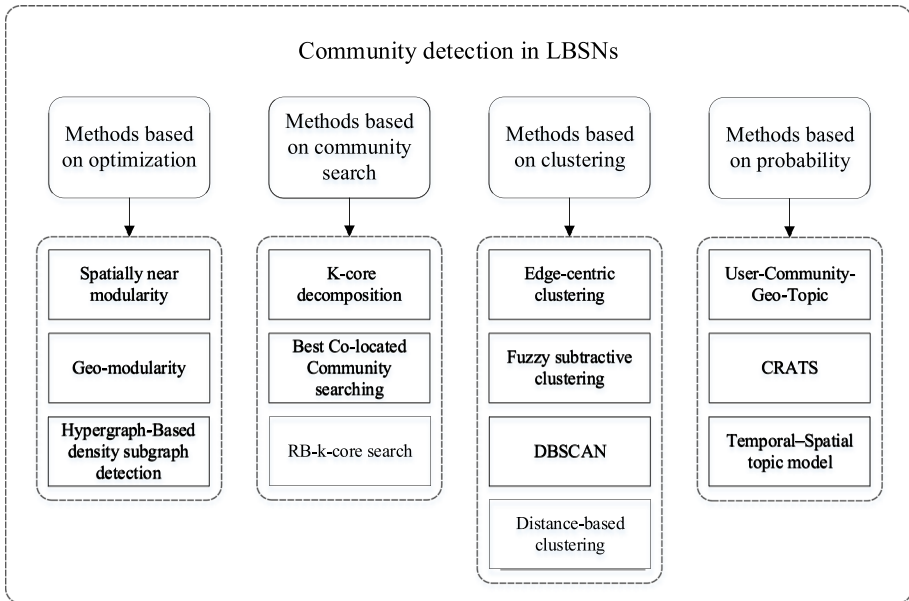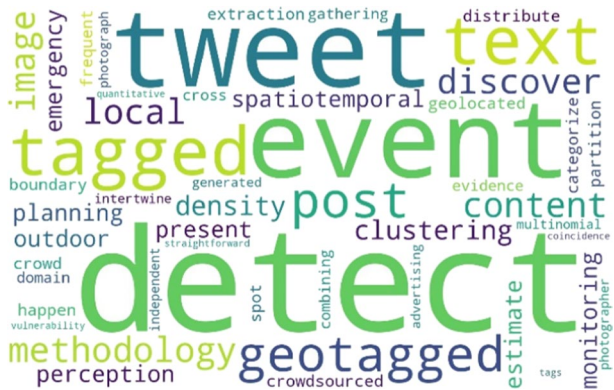
Fig. 9 Different methods for community detection

Fig. 10 Word cloud for topic 6

To identify unusual events, the most general idea is to deduce geographical regularities from the usual behavioral patterns of crowds with geo-tagged information, then, compare these regularities with estimated ones. [127] built a socio-geographical event-monitoring system, in which authors extracted three feature indicators from geo-tagged tweets, namely, numbers of tweets, crowds, and moving users, to establish the regularity of user crowd patterns. They identified usual events when users' crowd patterns changed compared with regularity. Authors in [128] constructed mobility and social activity tensors, which can be decomposed to extract the basic patterns. Then, they identified significant and unusual irregularities from those basic patterns and used them to detect the related urban event. Thereafter, they described the event from the aspects of influenced regions, gathering time, and popular activities. The clustering method is an effective method to group a large amount of user data to derive regular patterns and identify abnormal activity. In [129],

clusters referred to dense groups of tweets posted by different users that talk about an event happening near their location. Then, the DBSCAN based technique called Tweet-SCAN was applied to group event-related tweets by considering content, time, location, and user of each tweet.

The above research focuses on discovering abnormal mobility patterns within the city. However, they pay less attention to the specific content of events, i.e., "what is happening", which can be obtained through text-based content aggregation technology. In [130], authors utilized K-means clustering to group geo-tagged tweets and extracted Spatio-temporal outliers, then they analyzed corresponding content of the abnormal event from the outlier tweets by topic model.

Some studies also analyze "who is involved or likely to be involved in the event" [131–135]. In [133], to identify the on-site users of an event, the author used a fused feature Gaussian process regression model to infer the relevance between a user and a social event by integrating the factors of mobility influence, content similarity, and social relationship. In [134], based on the data from Meetup, the author analyzed the dynamic mutual influence among users on their offline event participation, the authors pointed out that users' preferences for social events are influenced by their friends, and this influence is mutual and dynamic. Authors of [135] proposed a multimodal to collect user information from multi-source platforms for event detection, which could enrich event content for LBSNs and EBSNs (Event-based Social Networks).

Figure 11 summarizes the main research architecture of event detection in the literature.

3.7 T7 urban mobility

For topic 7, words such as "city," "urban," "dynamic," and "mobility" suggest that this topic is related to the urban dynamic and human mobility. The existing Spatio-temporal data in LBSNs provides a resource for large-scale urban mobility research. It can help to solve various problems faced by cities as well as in understanding the characteristics and functions of different regions of the city.

The data used for assessing urban mobility typically come from three sources: physical sensors, the infrastructure of cities, and statistical data. However, accessing those data requires authorization from the city governments. Besides, data from those sources may
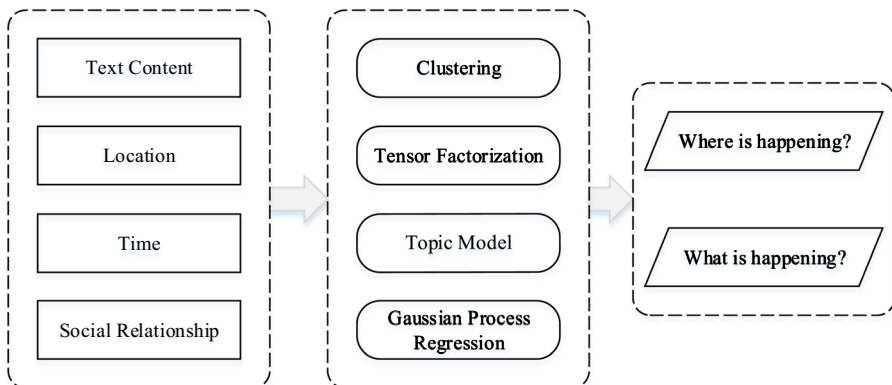

Fig. 11 The research architecture of event detection in LBSNs

not be complete. The worldwide LBSNs data make it possible to analyze the activities of residents in real-time and observe city dynamics. Past studies [136, 137] revealed that the LBSNs system can be used as a source of large-scale participatory sensing to acquire knowledge about city dynamics and urban social behaviors. We summarize the existing research based on the main application scenarios of urban mobility analysis, including urban planning, urban traffic, and public safety and security (Fig. 12).

3.7.1 Urban function identification

The exploration of geographical mobility patterns of citizens can help understand the urban spatial structure, identify functions and connections of different areas of a city, and enhance a knowledge-based urban development process.

Clustering algorithms are effective to identify the functions of different urban areas by clustering the locations or visitors with the same characters [138–140]. A past study [138] proposed a spectral clustering-based model by considering the spatial proximity and social proximity, wherein spatial proximity was measured by the distance between venues, and social proximity was derived from the distribution of people who check-in to certain venues. The authors of [140] took the temporal variations into account to identify the dynamics of functional regions. To do this, three clustering algorithms were used to cluster the keywords of different regions at varying time intervals. The work of [139] also applied a spectral clustering method to recognize urban land use. However, the author just considered the number of geo-tagged twitter about each location in a week. Thus, although the method can reflect the dynamics of different areas at different times, there was less information on the semantics of location types.

There is a large amount of user-generated information in LBSNs. Based on that, the LDA model is commonly applied to analyze semantic information [141, 142]. In [141], researchers applied the LDA and self-organizing maps (SOM) to explore the patterns of human activities and characterize underlying urban morphological structures by considering the geospatial, temporal, and semantic characteristics of georeferenced tweets. Authors of [142] regarded the function of location as a topic and applied an improved probabilistic topic model to infer functions of regions.

Several studies are based on predefined categories and view the location function identification as a multi-label classification task [143, 144]. Authors of [143] extracted 19 variables from the characteristics of check-ins and users, temporal features of user check-in

Fig. 12 Word cloud for topic 7

behaviors, and then built environment features of POI. Then, they trained three classification models, namely, random forest, decision tree, and support vector machine, for the classification task.

3.7.2 Urban traffic

As the population grows, the transportation problem has become an urgent issue. Data from the LBSNs system contain real-time trajectory information of citizens, which can help in analyzing traffic conditions. The authors in [145] used the real-time information of massive users in LBSNs to study human traffic in cities. They extracted the temporal patterns of human traffic at different types of venues and analyzed characteristics of those patterns using time-series decomposition. Thereafter, an ARIMA model was constructed to predict the temporal patterns by historical data. The model can be used by governments to schedule transportation traffic.

Meanwhile, in [146], researchers used digital footprints of users in LBSNs to model a microsimulation, which could reproduce not only trips of citizens but also traffic volumes at main roads. More directly, the author in [147] designed a system that could help drivers find an alternative route to avoid an incident based on solutions provided by other drivers in LBSNs. In [148], the author analyzed the impact of online social interaction ("social learning") on driver behavior, which could help to predict the driving routes or patterns of taxi drivers. The authors in [149] focused on the evacuation problem in cities, that is, finding a good path to a safe place without resulting in congestion. They proposed a method that used time- and geo-stamped microblog posts to measure demographic snapshots of a city and visualized high-risk evacuation roads based on geographical characteristics and demographics. Considerable research [150, 151] works on discovering the dense regions in cities, which is useful for governments to predict and prevent traffic jams and chaos.

3.7.3 Public security

LBSNs data reveal the nature of human activities in cities, and considerable studies have proven that those data can be used for public security research. For example, the work in [152] proved the usefulness of LBSNs data in urban crime analysis. The authors first extracted features about an ambient population from LBSNs data; then, based on those features, linear and spatial econometric models of crime were built for crime description. In [153], the researchers utilized LBSNs data to assist crime event prediction. The research assumed that, if many visitors in a region often visited unsafe regions at an unsafe time, their visits implied the region with a high safety risk. Then, they extracted prediction-specific dynamic features based on human mobility and used the random forest to predict whether a crime event would happen at a target region. A common method to identify the spatio-temporal dynamics of crime is to apply agent-based modeling to simulate crime patterns; however, incomplete crime data can lead to inaccurate results. To solve this problem, [154] used open data, such as those on road networks, to simulate the urban structure and used LBSNs data to represent activity nodes as proxies for human activities. Then, it analyzed the offender mobility under different mobility strategies. The authors of [155] proposed a directed weighted graph to present the user flow among regions, in which the graph nodes referred to regions, and directed edges were the movement directions of humans. After that, they defined the region risk factor under the hypothesis that human

mobility from a high crime risk area implies a high crime risk in the arrival area, after which they proposed a set of graph flow-based features for crime rate prediction.

Except for the crime analysis, disaster emergency response is also the main research direction, including natural disasters [156, 157] and epidemics [158]. One study [156] pointed out that LBSNs can be used as a tool for disaster information sharing and mutual assistance among nearby people and then proposed a Hybrid Disaster Response System (HyDRS), which allowed users to effectively cope with a disaster by using their mobile devices. To assist decision-makers to respond quickly in the mitigation of damage, one study [157] analyzed the temporal evolution characteristics of fine-grained topics related to natural disasters by combining LDA topics and the Markov transition probability matrix. In terms of the spread of the epidemic, another study [158] found that flu-related traffic on social media was closely related to actual flu outbreaks. Therefore, they pointed out that the integration of social media and real-world clinical could enhance the surveillance systems of contagious diseases.

In Table 5, we category the existing research about urban mobility according to the information source and method they used.

3.8 T8 privacy protection

Words in Fig. 13, such as "privacy", "preserving", and "protection", suggest that topic 8 is related to the privacy protection problem. Unlike traditional networks, the main method for users in LBSNs to socialize and enjoy various services is location sharing. Users' privacy refers to the information beyond their public data and spans the additional personal details

Table 5 Summary and comparison of urban mobility research based on LBSNs

		Information sources				Method
		Spatial	Social	Time	Text	
Urban function	[138]	√	√			Spectral clustering
	[139]				√	Spectral clustering
	[140]	√		√		Hierarchical clustering
	[141]	√		√	√	LDA/Neural network
	[142]	√				LDA
	[143]	√	√	√		Multi-Label classifier
Urban traffic	[145]	√		√		ARIMA model
	[146]	√		√		Radiation model
	[147]	√		√		Client–server system
	[149]	√		√	√	Path searching
	[150]	√		√	√	EBSCAN
	[151]	√		√	√	Spatial clustering
Public security	[152]	√	√	√	√	Linear model
	[153]	√		√	√	Random Forest
	[154]	√			√	Agent-based model
	[155]	√				Graph flow-based features
	[156]	√				HyDRS
	[157]	√				LDA/Markov
	[158]	√				multi-method data analytics

Fig. 13 Word cloud for topic 8

that may be inferred from their location data. Such data may be used improperly by attackers and untruth third-party service providers. Therefore, the protection of personal privacy information has also attracted researchers' attention.

When users provide their location information to an LBSNs server as they check-in in various locations, their personal information may be exposed. The authors in [5] indicated that location data could indeed reveal privacy information, including not only location coordinates but also personal information, such as daily pursuit and religious/political/sexual inclinations. Three types of user information exist in LBSNs, namely, location information, user profile, and social information. The existing privacy protection mechanisms are mainly aimed at such information.

3.8.1 Location information

In LBSNs, the main method for users' interaction is to share location or geo-related content in public, which provides resources to adversaries to design attacks and bring threats. The common threats can be categorized into the following types: (1) Direct sharing attacks [159]: extraction and exploitation of location information from a user's check-in history; (2) Tracking [160]: monitoring a user's movement trajectories; and (3) Inference attacks [161]: obtaining the knowledge of users unlawfully by analyzing their locations. In response to these threats, various approaches are proposed: sharing-decision-, location-modification-, and anonymity-based methods.

For the first type, the decision of whether to publish the current location can be determined by users or some techniques. Many researchers, such as those in [162], have proposed mechanisms that could make decisions automatically by analyzing contextual features and users' privacy preferences. In [163], the authors also considered the dynamics and spatiotemporal correlations of users' privacy requirements and proposed a context-aware location privacy protection. Similarly, the work of [164] considered the variability of users' privacy requirements and proposed a user-defined location-sharing scheme (ULSS) to achieve privacy preservation under different contexts.

The second type of method is obfuscating or encrypting location information. Location obfuscation means making the location information inaccurate by spatial cloaking or perturbation. The work in [165] focused on sensitive locations, such as hospitals, and then cloaked those locations following user preference. However, attackers might prune this region to find the accurate location. Meanwhile, when users share continuous location

information, an attacker may infer the real location from the overlapping regions. Thus, the authors in [160] hid user activity by generating semantically robust cloaking regions containing several semantically heterogeneous locations, such as bars and shopping malls. For continuous check-ins, the work of [166] constructed a circle region to cloak the target region which can avoid overlapping circle attacks effectively. However, the obfuscation methods may damage the quality of user data and affect the effect of personal services like recommendations. To solve this problem, one study [167] proposed PrivRank, a customizable and continuous privacy-preserving data publishing framework, by considering both the specific requirements of user privacy and data utility for enabling high-quality personalized recommendation. Location perturbation is an effective method against location-history-aware attackers. In [168], the authors proposed data perturbation techniques that users can employ to provide noisy locations of shared geo-tags. For location encryption, many cryptographic protocols have been defined by different mechanisms to ensure that an LBSNs platform can encrypt users' location information following the protocols. A representative mechanism has been presented in [169], in which two protocols provide complete privacy for the service provider and controllable privacy for the buddies. This mechanism represented the space of user locations as a finite number of non-overlapping cells. Users could define the size of the cells to control the uncertainty with which their friends can locate them. The location information shared by users was then expressed as an encrypted index of the cell containing their actual location.

The third method is anonymity. In general, the ultimate goal of an anonymization approach is to preserve social media user privacy while ensuring the utility of published data. The location privacy of users can be protected by replacing users' identities with other pseudonyms or by enlarging real locations of users to include additional location information corresponding to geographical locations of k other users. The most representative method is K-anonymity. This produces the K-anonymity region containing k different location points as an obfuscated region to hide the user's precise location. The traditional k-anonymity approach regards each user's query as an isolated event. Authors in [170] considered the spatiotemporal relationship of users' location history and location attribution, then proposed a varied k-anonymity and l-diversity location privacy-preserving model. A trusted anonymizer server, which could process users' location information before it reached the service provider, is necessary to implement k-anonymity. However, some research has proven that a third-party server cannot guarantee user privacy and that it forms a single point of failure. The authors in [171, 172] proposed K-anonymity-based algorithms without involving a trusted anonymizer server.

3.8.2 User profile

Users have to provide personal profiles and preference information to LBSN providers to receive high-quality services. Such information typically includes user identity and demographic information. Unfortunately, detailed personal information may be used maliciously. Meanwhile, user identity information can also be inferred from their trajectories [173]. A dangerous situation occurs if the users' personal information and trajectory information are identified and matched by malicious attackers or untrusted service providers. Common threats to user profile information include identify theft, impersonation, and hacking. The most common method to avoid those hazards is to implement the distributed storage of user information. In [174], researchers proposed a decentralized matching service, in which a matching service was used to match the encrypted location information and users'

interest. The matching service entity was unaware of the actual value of users' information but was able to answer user queries for nearby POIs without knowing the actual user identities or their locations. The work in [175] also proposed a two-part privacy-preserving matching service. One of the components (the "matcher") was used to store identity information and attribute values of entities (users and POIs) without knowing their relationship. The other component (the "combiner"), which came from a different provider, was used to store the relationship between identities and attributes of users and POIs. The two components implemented a matching service by running a privacy-preserving protocol on pseudo-ids of entities and their attributes to enable various services.

The *K*-anonymity [176] and Dynamic pseudonymity [177] are also effective methods for user profile protection. In [177], the author introduced a Dynamic Pseudonymity Mechanism (DPP) to protect users' query privacy, in which, location-related query information of the user was divided into chunks and assigned different pseudo identities while interacting with different service providers. This ensured that the adversary cannot link user information to a query result or obtain a true user ID from any one of the service providers.

3.8.3 Social information

Users can build their friend lists on LBSNs and interact with each other by sharing locations or tips on specific venues. Some people do not want to expose their friend lists. That can be achieved by setting access permissions manually. However, social information is still in danger of leaking. The service provider in LBSNs will recommend friends or communities based on user preferences to attract new users and increase user engagement. Users' friend lists and interaction information with the other users will be extracted and used frequently, which may result in the leakage of personal social information. Attackers may use users' social information to spread spam and malicious links.

Several mechanisms concerning data-distributed storage have been proposed by numerous researchers to solve this problem. For instance, in [175], user and relationship data were encrypted and stored on different servers. In [178], the author proposed a cryptographic primitive, functional pseudonym, for location sharing in LBSN. This method protects both location and friendship privacy and does not need a fully trusted server. However, social ties may still be inferred from check-in data, because friends may have similar location preferences and mobility patterns. To overcome this problem, authors in [179] guided users to check-in safe locations such that the adversary could not correctly identify the friends of users. They developed a two-step framework named check-in shielding scheme. It first estimated social strength amongst users and then generated a list of secure places. Then, users could select places from the list based on their preferences.

In Table 6, we present an overview of threats and solutions for different privacy information in LBSNs, along with the representative methods mentioned above.

4 Future research

As we have discussed above, LBSNs data provide opportunities for research in multiple fields, and many scholars have conducted in-depth and extensive research on LBSNs. However, numerous open questions and challenges remain and need to be solved due to data and technology limitations. In this section, we discuss possible limitations and potential research directions in the future.

Table 6 An overview of privacy threats and corresponding solutions in LBSNs

	Threats	Solutions	Representative method
Location Information	Direct sharing attacks Tracking Inference attacks	Sharing-decision methods Location modification Anonymity	ULSS[164] Obfuscating[165, 160, 167, 166] Perturbation[168] Encrypting[169] K- anonymity [171, 172, 170]
User Profile	Identify theft Impersonation Hacking	Data-distributed storage Anonymity Dynamic pseudonymity	Privacy-preserving matching service[174, 175] K-anonymity[176] DPP [177]
Social Information	Spam Malicious links	Data-distributed storage Cryptographic primitive	Encrypted storage[175] Functional pseudonym[178] Check-in shielding scheme [179]

4.1 Data simulation and fusion

Publicly available data and crawled data are the main data for LBSNs research. However, those data are increasingly unavailable and suffer several limitations. Most of the available data are too old (2008–2013) to be mapped to the current existing users or locations. Meanwhile, due to technical limitations, these data are sparse and noisy. In contrast, data crawling methods can obtain novel datasets, but many platforms do not provide open APIs or if they do, they have high restrictions on the use of APIs. As different users have varying degrees of willingness to share their locations, the crawling data do not reflect the behaviors of all the population. Incomplete data may influence the effectiveness of experimental results. Therefore, how to use limited data resources for efficient research is worthy of further attention. To achieve this goal, we point out the following research directions:

(1) Data simulation: Based on the limited LBSN data, various scientific methods can be designed to mine users' movement patterns, simulate their check-in location distributions, and extract their check-in characteristics, thereby generating synthetic data similar to the real data. Recently, several academics have conducted pioneering research in this area. For example, in [180], the author proposed a scalable and efficient agent-based framework for geo-social simulation, which not only considered user mobility patterns but also integrated psychological and social theories.

(2) Fusion of multiple data: With the development of the Internet, users can publish not only text data but also photos, videos, and sounds. Extracting knowledge from multiple or cross-platform data can enrich user and location profiles as well as alleviate data sparsity problems. In this regard, the work of [181] has made a good attempt by proposing a novel seed-driven approach that uses the points of one source as the seeds to extract and combine data from the others. In [182], the author proposed a cross-domine method that used transfer learning to fuse knowledge from different domains, effectively alleviating the data scarcity.
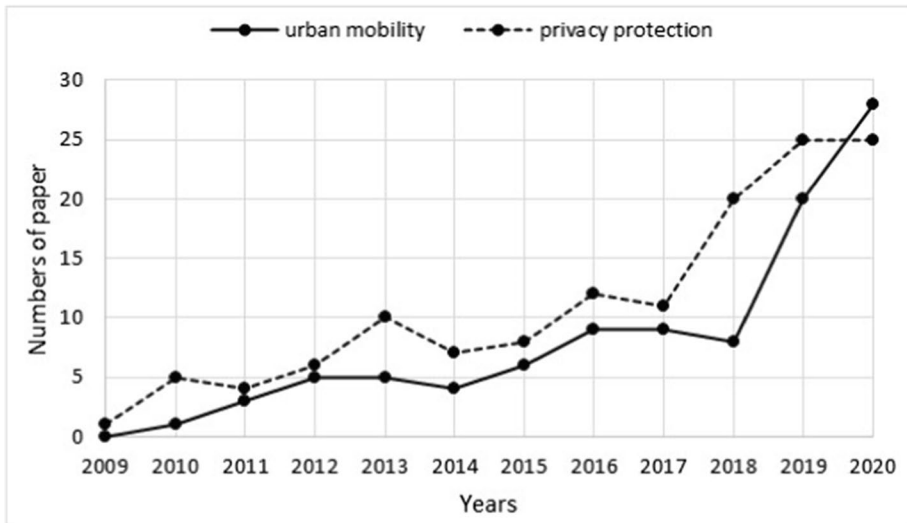
Fig. 14 The trends of published works on urban mobility and privacy protection

4.2 Privacy-aware method

Research on user sharing behavior has shown that privacy concerns are the main reason affecting user check-ins [183]. As shown in Fig. 14, privacy-related issues have attracted increasing research attention in recent years. However, most of the research focused on how to protect users' privacy by establishing effective protocols and data encryption or obfuscation methods. In terms of specific application scenarios, such as event detection and community detection, researchers have ignored the privacy needs of users when improving the effect of the model.

Different applications involve different user information. Therefore, privacy-aware methods for various application tasks are urgently needed. Utilizing less user-sensitive information and considering user privacy demand as one of the constraints of models are both effective ways to preserve privacy under specific application scenarios. In fact, in the field of location recommendation, a few scholars have studied the privacy-preserving methods for recommendations. For instance, in [167], authors considered both the specific requirements of user privacy and the data utility for enabling high-quality personalized recommendations. Then, they proposed a customizable and continuous privacy-preserving framework, which could protect users against inference attacks while enabling personalized ranking-based recommendations. Several works identify users' privacy locations from their trajectory histories and protect such information by adding noise information [184], but this method will reduce the quality of data. Therefore, privacy-aware methods that can effectively consider both the task requirements and user privacy requirements are of great research value.

4.3 New applications and services

Through the analysis of the paper publication tendency about different research topics, we found some research hotspots and novel research topics in recent years.

In general, location recommendation and friendship prediction are the most popular research fields, and the articles about those topics account for the largest proportion in the total number of articles published annually. In contrast, urban mobility is an emerging research hotspot. As shown in Fig. 14, in the past two years, the number of articles on urban mobility has soared.

Also, we find some interesting research points, which have not been widely studied, but have great application value:

(1) Temporal dynamic analysis: LBSNs are active platforms in which data and structure are changing over time. However, most of the existing research is based on static networks, and few studies have considered network dynamics, such as the dynamic changes in community structure and the roles of users. Limited research also exists on incremental analysis of user data.

(2) Group recommendation in LBSNs: Although a few studies have analyzed the problem of location recommendation for a group of users [185, 186], relatively limited studies on group recommendation are available. An increasing number of users prefer traveling in groups; therefore, group recommendations may have great research value and broad application prospects.

(3) AR-based LBSNs: The development of technology enriches the functions of existing LBSNs and produces more novel LBSNs, such as Sekai Camera and Wallame, which integrate augmented reality (AR) technology into LBSNs and enrich the users' presence feeling and interactive content. As far as we know, a few works have studied the impacts of AR-based LBSNs on GIScience research [187], and the relationship between local visual information and user-generated content[188]. Nevertheless, this field needs to be further explored in the future.

4.4 Technological innovations

The vast plethora of information in LBSNs led to the development of technologies and models that can analyze and learn users' behaviors for multiple applications. In recent years, deep learning technology has shown superior ability in many tasks. For example, they can automatically extract features from a large number of user data and have shown high computational efficiency and strong scalability. However, although has achieved state-of-the-art results in various tasks, deep learning technology still has some drawbacks that can be improved:

(1) Most neural network architectures are based on structured data, whereas LBSNs data is heterogeneous unstructured data. There are multiple entities in LBSNs, such as locations, users, etc., and the interactions among those entities are complex. Therefore, traditional deep learning methods may not fully capture network features, and novel deep learning architectures based on heterogeneous graph data are needed. The graph neural network that is used to process heterogeneous graphs and hypergraphs has developed rapidly in recent years, but in the field of LBSNs, relatively little research has been done.

(2) Interpretable neural network: One of the main purposes of analyzing LBSNs data is to guide various applications. For example, the research about user mobility can be used for city planning, traffic guidance, etc. Therefore, the method must have high interpretation capabilities and be able to effectively mine the law behind the data. In

this sense, deep learning technologies are limited. The training and learning process of the neural networks is based on the optimization of an objective function aimed at minimizing the error in the reconstructed output, and the parameters in each layer are auto-tuned [189]. The model obtained through training with a large amount of data is more like an empirical result. Interpretability is a basic problem in deep learning. In recent years, many studies have tried to solve this problem, such as using mathematical tools to diagnose and evaluate the feature expression ability of networks [190] or using a small sample of weakly supervised learning methods for model training.Conclusion

In this article, we provide a systematic survey of the existing works about LBSNs. We collect 1,381 documents from January 2009 to 2020 and applied a text-mining method, the LDA, to extract the main topics from a large number of papers. According to the results of LDA, we categorized the existing research about LBSNs into the following topics: (1) location recommendation, (2) route planning, (3) friendship prediction, (4) urban mobility, (5) community detection, (6) event detection, (7) influence maximization, and (8) privacy protection. Then, we discuss in detail the key research problems, methodologies, and applications for each topic. Based on the survey, we also find some limitations and promising research direction in four areas: data simulation and fusion, privacy-aware method, new applications and services, and technological innovation. We hope that this survey motivates the development of new initiatives concerning LBSNs research.

References

1. Moro S, Pires G, Rita P, Cortez P (2019) A text mining and topic modelling perspective of ethnic marketing research. J Bus Res 103:275–285. https://doi.org/10.1016/j.jbusres.2019.01.053
2. Cho E, Myers SA, Leskovec J (2011) Friendship and mobility: user movement in location-based social networks. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. https://doi.org/10.1145/2020408.2020579
3. Yang D, Qu B, Yang J, Cudre-Mauroux P (2019) Revisiting user mobility and social relationships in lbsns: a hypergraph embedding approach. In: The World Wide Web Conference. ACM, pp 2147–2157. https://doi.org/10.1145/3308558.3313635
4. Bao J, Zheng Y, Wilkie D, Mokbel M (2015) Recommendations in location-based social networks: a survey. GeoInformatica 19(3):525–565. https://doi.org/10.1007/s10707-014-0220-8
5. Riaz Z, Dürr F, Rothermel K (2018) Location privacy and utility in geo-social networks: survey and research challenges. In: 16th annual conference on privacy, security and trust. IEEE, pp 1–10. https://doi.org/10.1109/pst.2018.8514193
6. Silva TH, Viana AC, Benevenuto F, Villas L, Salles J, Loureiro A, Quercia D (2019) Urban computing leveraging location-based social network data: a survey. ACM Comput Surv 52(1):17. https://doi.org/10.1145/3301284
7. Loureiro SMC, Guerreiro J, Eloy S, Langaro D, Panchapakesan P (2019) Understanding the use of virtual reality in marketing: a text mining-based review. J Bus Res 100:514–530. https://doi.org/10.1016/j.jbusres.2018.10.055
8. Teh YW, Jordan MI, Beal MJ, Blei DM (2004) Sharing clusters among related groups: hierarchical Dirichlet processes. Paper presented at the proceedings of the 17th international conference on neural information processing systems,

9. Qian S, Zhang T, Xu C, Shao J (2016) Multi-modal event topic model for social event analysis. IEEE Trans Multimed 18:233–246. https://doi.org/10.1109/tmm.2015.2510329

10. Yuchi Zhang WWM, Schweidel DA (2017) Modeling the role of message content and influencers in social media rebroadcasting. Int J Res Mark 34(1):110–119. https://doi.org/10.1016/j.ijresmar.2016.07.003

11. Jia Liu OT (2018) A semantic approach for estimating consumer content preferences from online search queries. Mark Sci 37(6):930–952. https://doi.org/10.1287/mksc.2018.1112

12. Steyvers MaG T (2007) Latent semantic analysis: a road to meaning. Probabilistic topic model. Laurence Erlbaum

13. Yezheng Liu JW, Jiang Y (2016) PT-LDA: a latent variable model to predict personality traits of social network users. Neurocomputing 210:155–163. https://doi.org/10.1016/j.neucom.2015.10.144

14. Lim KH, Chan J, Karunasekera S, Leckie C (2018) Tour recommendation and trip planning using location-based social media: a survey. Knowl Inf Syst. https://doi.org/10.1007/s10115-018-1297-4

15. Shen T, Chen H, Ku W-S (2018) Time-aware location sequence recommendation for cold-start mobile users. In: Proceedings of the 26th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. ACM, pp 484–487. https://doi.org/10.1145/3274895.3274958

16. He J, Li X, Liao L, Song D, Cheung WK (2016) Inferring a personalized next point-of-interest recommendation model with latent behavior patterns. In: Proceedings of the thirtieth AAAI conference on artificial intelligence. pp 137–143

17. Lian D, Zheng K, Ge Y, Cao L, Chen E, Xie X (2018) GeoMF++: scalable location recommendation via joint geographical modeling and matrix factorization. ACM Trans Inf Syst 36(3):33. https://doi.org/10.1145/3182166

18. Zhao X, Li X, Liao L, Song D, Cheung WK (2015) Crafting a time-aware point-of-interest recommendation via pairwise interaction tensor factorization. In: International conference on knowledge science, engineering and management. Springer, Berlin. pp 458–470. https://doi.org/10.1007/978-3-319-25159-2_41

19. Ying Y, Chen L, Chen G (2017) A temporal-aware POI recommendation system using context-aware tensor decomposition and weighted HITS. Neurocomputing 242:195–205. https://doi.org/10.1016/j.neucom.2017.02.067

20. Yang C, Bai L, Zhang C, Yuan Q, Han J (2017) Bridging collaborative filtering and semi-supervised learning: a neural approach for poi recommendation. In: Proceedings of the 23rd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1245–1254. https://doi.org/10.1145/3097983.3098094

21. Jiao X, Xiao Y, Zheng W, Wang H, Hsu C-H (2019) A novel next new point-of-interest recommendation system based on simulated user travel decision-making process. Futur Gener Comput Syst. https://doi.org/10.1016/j.future.2019.05.065

22. Lian D, Ge Y, Zhang F, Yuan NJ, Xie X, Zhou T, Rui Y (2018) Scalable content-aware collaborative filtering for location recommendation. IEEE Trans Knowl Data Eng 30(6):1122–1135. https://doi.org/10.1109/tkde.2018.2789445

23. Liu B, Fu Y, Yao Z, Xiong H (2013) Learning geographical preferences for point-of-interest recommendation. In: Proceedings of the 19th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 1043–1051. https://doi.org/10.1145/2487575.2487673

24. Liu B, Xiong H, Papadimitriou S, Fu Y, Yao Z (2014) A general geographical probabilistic factor model for point of interest recommendation. IEEE Trans Knowl Data Eng 27(5):1167–1179. https://doi.org/10.1109/tkde.2014.2362525

25. Su C, Li N, Xie X-Z (2018) Point-of-Interest recommendation based on spatial clustering in LBSN. In: 4th annual international conference on network and information systems for computers. IEEE, pp 7–12. https://doi.org/10.1109/icnisc.2018.00011

26. Xiong X, Qiao S, Han N, Xiong F, Bu Z, Li R-H, Yue K, Yuan G (2020) Where to go: an effective point-of-interest recommendation framework for heterogeneous social networks. Neurocomputing 373:56–69. https://doi.org/10.1016/j.neucom.2019.09.060

27. Zheng Y, Zhang L, Ma Z, Xie X, Ma W-Y (2011) Recommending friends and locations based on individual location history. ACM Trans Web 5(1):5. https://doi.org/10.1145/1921591.1921596

28. Cheng C, Yang H, Lyu MR, King I (2013) Where you like to go next: successive point-of-interest recommendation. Paper presented at the proceedings of the 23rd international joint conference on artificial intelligence

29. Zhao WX, Fan F, Wen J-R, Chang EY (2018) Joint representation learning for location-based social networks with multi-grained sequential contexts. ACM Trans Knowl Discov Data 12(2):22. https://doi.org/10.1145/3127875

30. Pourali A, Zarrinkalam F, Bagheri E (2019) Neural embedding features for point-of-interest recommendation. In: IEEE/ACM international conference on advances in social networks analysis and mining. IEEE, pp 657–662. https://doi.org/10.1145/3341161.3343672

31. Yang D, Qu B, Yang J, Cudre-Mauroux P (2020) LBSN2Vec++: heterogeneous hypergraph embedding for location-based social networks. IEEE Trans Knowl Data Eng 99:1–1. https://doi.org/10.1109/tkde.2020.2997869

32. Gao Q, Trajcevski G, Zhou F, Zhang K, Zhong T, Zhang F (2019) DeepTrip: adversarially understanding human mobility for trip recommendation. In: Proceedings of the 27th ACM SIGSPATIAL international conference on advances in geographic information systems. pp 444–447. https://doi.org/10.1145/3347146.3359088

33. Xiong X, Xiong F, Zhao J, Qiao S, Li Y, Zhao Y (2020) Dynamic discovery of favorite locations in spatio-temporal social networks. Inf Process Manag 57(6):102337. https://doi.org/10.1016/j.ipm.2020.102337

34. Liu Q, Wu S, Wang L, Tan T (2016) Predicting the next location: a recurrent model with spatial and temporal contexts. In: 30th AAAI conference on artificial intelligence

35. Manotumruksa J, Macdonald C, Ounis I (2018) A contextual attention recurrent architecture for context-aware venue recommendation. In: 41st international ACM SIGIR conference on research and development in information retrieval. ACM, pp 555–564. https://doi.org/10.1145/3209978.3210042

36. Zhao J, Zhao P, Liu Y, S. Sheng V, Li Z, Zhao L (2020) Hierarchical variational attention for sequential recommendation. In: Cham. Database systems for advanced applications. Springer, Berlin, pp 523–539. https://doi.org/10.1007/978-3-030-59419-0_32

37. Liang W, Zhang W (2020) Learning social relations and spatiotemporal trajectories for next check-in inference. IEEE Trans Neural Netw Learn Syst 99:1–11. https://doi.org/10.1109/tnnls.2020.3016737

38. Yuan Q, Cong G, Ma Z, Sun A, Thalmann NM (2013) Time-aware point-of-interest recommendation. In: Proceedings of the 36th international ACM SIGIR conference on research and development in information retrieval. ACM, pp 363–372. https://doi.org/10.1145/2484028.2484030

39. Liu Y, Liu C, Liu B, Qu M, Xiong H (2016) Unified point-of-interest recommendation with temporal interval assessment. In: Proceedings of the 22nd ACM SIGKDD iInternational conference on knowledge discovery and data mining. ACM, pp 1015–1024. https://doi.org/10.1145/2939672.2939773

40. Liu Y, Ester M, Qian Y, Hu B, Cheung DW (2017) Microscopic and macroscopic spatio-temporal topic models for check-in data. IEEE Trans Knowl Data Eng 29(9):1957–1970. https://doi.org/10.1109/tkde.2017.2703825

41. Song Y, Hu Z, Leng X, Tian H, Yang K, Ke X (2015) Friendship influence on mobile behavior of location based social network users. J Commun Netw 17(2):126–132. https://doi.org/10.1109/jcn.2015.000026

42. Ye M, Yin P, Lee W-C (2010) Location recommendation for location-based social networks. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, pp 458–461. https://doi.org/10.1145/1869790.1869861

43. Zhang J-D, Chow C-Y (2015) CoRe: Exploiting the personalized influence of two-dimensional geographic coordinates for location recommendations. Inf Sci 293:163–181. https://doi.org/10.1016/j.ins.2014.09.014

44. Li H, Ge Y, Hong R, Zhu H (2016) Point-of-interest recommendations: learning potential check-ins from friends. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 975–984. https://doi.org/10.1145/2939672.2939767

45. Gao H, Tang J, Liu H (2012) gSCorr: modeling geo-social correlations for new check-ins on location-based social networks. In: Proceedings of the 21st ACM international conference on Information and knowledge management. ACM, pp 1582–1586. https://doi.org/10.1145/2396761.2398477

46. Yao L, Sheng QZ, Qin Y, Wang X, Shemshadi A, He Q (2015) Context-aware point-of-interest recommendation using tensor factorization with social regularization. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. ACM, pp 1007–1010. https://doi.org/10.1145/2766462.2767794

47. Gao R, Li J, Li X, Song C, Chang J, Liu D, Wang C (2018) STSCR: Exploring spatial-temporal sequential influence and social information for location recommendation. Neurocomputing 319:118–133. https://doi.org/10.1016/j.neucom.2018.07.041

48. Zhang S, Hong C (2018) Exploiting context graph attention for POI recommendation in location-based social networks. Paper presented at the international conference on database systems for advanced applications

49. Lim N, Hooi B, Ng S-K, Wang X, Goh YL, Weng R, Varadarajan J (2020) STP-UDGAT: spatial-temporal-preference user dimensional graph attention network for next POI recommendation. Paper presented at the proceedings of the 29th ACM international conference on information and knowledge management

50. Li J, Wang Y, McAuley J (2020) Time interval aware self-attention for sequential recommendation. Paper presented at the proceedings of the 13th international conference on web search and data mining, Houston, TX, USA

51. Liu X, Liu Y, Aberer K, Miao C (2013) Personalized point-of-interest recommendation by mining users' preference transition. In: Proceedings of the 22nd ACM international conference on information & knowledge management. pp 733–738. https://doi.org/10.1145/2505515.2505639

52. Zhu L, Xu C, Guan J, Zhang H (2017) SEM-PPA: a semantical pattern and preference-aware service mining method for personalized point of interest recommendation. J Netw Comput Appl 82:35–46. https://doi.org/10.1016/j.jnca.2016.12.033

53. Zhan G, Xu J, Huang Z, Zhang Q, Xu M, Zheng N (2019) A semantic sequential correlation based lstm model for next poi recommendation. In: 20th IEEE international conference on mobile data management IEEE, pp 128–137. https://doi.org/10.1109/mdm.2019.00-65

54. Zhang Z, Liu Y, Zhang Z, Shen B (2019) Fused matrix factorization with multi-tag, social and geographical influences for POI recommendation. World Wide Web 22(3):1135–1150. https://doi.org/10.1007/s11280-018-0579-9

55. Liu W, Lai H, Wang J, Ke G, Yang W, Yin J (2019) Mix geographical information into local collaborative ranking for POI recommendation. World Wide Web. https://doi.org/10.1007/s11280-019-00681-1

56. Zhang J-D, Chow C-Y (2015) Geosoca: Exploiting geographical, social and categorical correlations for point-of-interest recommendations. In: Proceedings of the 38th international ACM SIGIR conference on research and development in information retrieval. ACM, pp 443–452. https://doi.org/10.1145/2766462.2767711

57. Lian D, Zhao C, Xie X, Sun G, Chen E, Rui Y (2014) GeoMF: joint geographical modeling and matrix factorization for point-of-interest recommendation. In: Proceedings of the 20th ACM SIGKDD international conference on knowledge discovery and data mining. ACM, pp 831–840. https://doi.org/10.1145/2623330.2623638

58. Wang W, Chen J, Wang J, Chen J, Gong Z (2019) Geography-aware inductive matrix completion for personalized Point-of-Interest recommendation in smart cities. IEEE Internet Things J 7(5):4361–4370. https://doi.org/10.1109/jiot.2019.2950418

59. Liu T, Liao J, Wu Z, Wang Y, Wang J (2020) Exploiting geographical-temporal awareness attention for next point-of-interest recommendation. Neurocomputing. https://doi.org/10.1016/j.neucom.2019.12.122

60. Chang B, Jang G, Kim S, Kang J (2020) Learning graph-based geographical latent representation for point-of-interest recommendation. In: Proceedings of the 29th ACM international conference on information and knowledge management. pp 135–144. https://doi.org/10.1145/3340531.3411905

61. Xing S, Liu F, Zhao X, Li T (2018) Points-of-interest recommendation based on convolution matrix factorization. Appl Intell 48(8):2458–2469. https://doi.org/10.1007/s10489-017-1103-0

62. Kefalas P, Manolopoulos Y (2017) A time-aware spatio-textual recommender system. Expert Syst Appl 78:396–406. https://doi.org/10.1016/j.eswa.2017.01.060

63. Hu X, Xu J, Wang W, Li Z, Liu A (2020) A graph embedding based model for fine-grained POI recommendation. Neurocomputing. https://doi.org/10.1016/j.neucom.2020.01.118

64. Li W, Liu X, Yan C, Ding G, Sun Y, Zhang J (2020) STS: spatial–temporal–semantic personalized location recommendation. ISPRS Int J Geo Inf 9(9):538. https://doi.org/10.3390/ijgi9090538

65. Zhang C, Zhang K, Yuan Q, Zhang L, Hanratty T, Han J (2016) Gmove: group-level mobility modeling using geo-tagged social media. In: Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. pp 1305–1314. https://doi.org/10.1145/2939672.2939793

66. Lin I-C, Lu Y-S, Shih W-Y, Huang J-L (2018) Successive POI recommendation with category transition and temporal influence. In: IEEE 42nd annual computer software and applications conference IEEE, pp 57–62. https://doi.org/10.1109/compsac.2018.10203

67. Chen J, Li X, Cheung WK, Li K (2016) Effective successive POI recommendation inferred with individual behavior and group preference. Neurocomputing 210:174–184. https://doi.org/10.1016/j.neucom.2015.10.146

68. Sang J, Mei T, Sun J-T, Xu C, Li S (2012) Probabilistic sequential POIs recommendation via check-in data. In: Proceedings of the 20th international conference on advances in geographic information systems. ACM, pp 402–405. https://doi.org/10.1145/2424321.2424375

69. Ying H, Wu J, Xu G, Liu Y, Liang T, Zhang X, Xiong H (2019) Time-aware metric embedding with asymmetric projection for successive POI recommendation. World Wide Web 22(5):2209–2224. https://doi.org/10.1007/s11280-018-0596-8

70. Fang Y, Wang Z, Cheng R, Li X, Luo S, Hu J, Chen X (2019) On spatial-aware community search. IEEE Trans Knowl Data Eng 31(4):783–798. https://doi.org/10.1109/tkde.2018.2845414

71. Hsieh H-P, Li C-T, Lin S-D (2014) Measuring and recommending time-sensitive routes from location-based data. ACM Trans Intell Syst Technol 5(3):45. https://doi.org/10.1145/2542668

72. Zhu X, Hao R, Chi H, Du X (2017) Fineroute: personalized and time-aware route recommendation based on check-ins. IEEE Trans Veh Technol 66(11):10461–10469. https://doi.org/10.1109/tvt.2017.2764999

73. Zhang C, Liang H, Wang K (2016) Trip recommendation meets real-world constraints: POI availability, diversity, and traveling time uncertainty. ACM Trans Inform Syst 35(1):5. https://doi.org/10.1145/2948065

74. Chen C, Zhang D, Guo B, Ma X, Pan G, Wu Z (2014) TripPlanner: Personalized trip planning leveraging heterogeneous crowdsourced digital footprints. IEEE Trans Intell Transp Syst 16(3):1259–1273. https://doi.org/10.1109/tits.2014.2357835

75. Xu J, Chen J, Zhou R, Fang J, Liu C (2019) On workflow aware location-based service composition for personal trip planning. Futur Gener Comput Syst 98:274–285. https://doi.org/10.1016/j.future.2019.03.010

76. Yu Z, Feng Y, Xu H, Zhou X (2014) Recommending travel packages based on mobile crowdsourced data. IEEE Commun Mag 52(8):56–62. https://doi.org/10.1109/mcom.2014.6871670

77. Luan W, Liu G, Jiang C, Zhou M (2018) MPTR: a maximal-marginal-relevance-based personalized trip recommendation method. IEEE Trans Intell Transp Syst 19(11):3461–3474. https://doi.org/10.1109/tits.2017.2781138

78. Lian D, Wu Y, Ge Y, Xie X, Chen E (2020) Geography-aware sequential location recommendation. Paper presented at the proceedings of the 26th ACM SIGKDD international conference on knowledge discovery and data mining

79. Yu F, Cui L, Guo W, Lu X, Li Q, Lu H (2020) A category-aware deep model for successive POI recommendation on sparse check-in data. Paper presented at the proceedings of the web conference 2020

80. Goldenberg J, Levy M (2009) Distance is not dead: social interaction and geographical distance in the internet era. arXiv preprint https://arxiv.org/09063202

81. Scellato S, Noulas A, Lambiotte R, Mascolo C (2011) Socio-spatial properties of online location-based social networks. In: 5th international AAAI conference on weblogs and social media

82. Hsieh H-P, Yan R, Li C-T (2015) Where you go reveals who you know: Analyzing social ties from millions of footprints. In: Proceedings of the 24th ACM international on conference on information and knowledge management. ACM, pp 1839–1842. https://doi.org/10.1145/2806416.2806653

83. Valverde-Rebaza JC, Roche M, Poncelet P, de Andrade LA (2018) The role of location and social strength for friendship prediction in location-based social networks. Inf Process Manag 54(4):475–489. https://doi.org/10.1016/j.ipm.2018.02.004

84. Cranshaw J, Toch E, Hong JI, Kittur A, Sadeh NM (2010) Bridging the gap between physical location and online social networks. In: UbiComp 2010: ubiquitous computing, 12th international conference, UbiComp 2010,. https://doi.org/10.1145/1864349.1864380

85. Zhang Z, Zhao X, Wang G (2017) FE-ELM: a new friend recommendation model with extreme learning machine. Cogn Comput. https://doi.org/10.1007/s12559-017-9484-2

86. Li Z, Fang X, Sheng ORL (2017) A survey of link recommendation for social networks: Methods, theoretical foundations, and future research directions. ACM Trans Manag Inf Syst 9(1):1–26. https://doi.org/10.3724/sp.j.1016.2011.01820

87. Zhao WX, Zhou N, Zhang W, Wen J-R, Wang S, Chang EY (2016) A probabilistic lifestyle-based trajectory model for social strength inference from human trajectory data. ACM Trans Inf Syst 35(1):8. https://doi.org/10.1145/2948064

88. Xiao X, Yu Z, Luo Q, Xing X (2014) Inferring social ties between users with human location history. J Ambient Intell Humaniz Comput 5(1):3–19. https://doi.org/10.1007/s12652-012-0117-z

89. Alharbi B, Zhang X (2016) Learning from your network of friends: a trajectory representation learning model based on online social ties. In: IEEE 16th international conference on data mining. https://doi.org/10.1109/icdm.2016.0090

90. Ying JJ-C, Lu EH-C, Lee W-C, Weng T-C, Tseng VS (2010) Mining user similarity from semantic trajectories. In: Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks. ACM, pp 19–26. https://doi.org/10.1109/cloudcom-asia.2013.74

91. Xiao X, Zheng Y, Luo Q, Xie X (2010) Finding similar users using category-based location history. In: Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems. ACM, pp 442–445. https://doi.org/10.1145/1869790.1869857

92. Zhou N, Zhao WX, Zhang X, Wen J-R, Wang S (2016) A general multi-context embedding model for mining human trajectory data. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/tkde.2016.2550436

93. Daizong Ding MZ, Xudong Pan, Duocai Wu, Pearl Pu (2018) Geographical feature extraction for entities in location-based social networks. Paper presented at the international World Wide Web conference

94. Qiao Y, Luo X, Li C, Tian H, Ma J (2020) Heterogeneous graph-based joint representation learning for users and POIs in location-based social network. Inf Process Manag 57(2):102151. https://doi.org/10.1016/j.ipm.2019.102151

95. Hosseinpour M, Malek MR, Claramunt C (2019) Socio-spatial influence maximization in location-based social networks. Futur Gener Comput Syst 101:304–314. https://doi.org/10.1016/j.future.2019.06.024

96. Agah A, Dasari PK (2017) Measuring influence in online social networks: with a focus on Gowalla and Brightkite. In: IEEE international conference on communications. IEEE, pp 1–5. https://doi.org/10.1109/icc.2017.7997037

97. Zhou Y, Li Y, Wang Z, Luo Y, Yang X (2017) Identification of influential spreaders in geo-social network. In: 25th international conference on geoinformatics. IEEE, pp 1–4. https://doi.org/10.1109/geoinformatics.2017.8090941

98. Yang Y, Xu Y, Wang E, Lou K, Luan D (2018) Exploring influence maximization in online and offline double-layer propagation scheme. Inf Sci 450:182–199. https://doi.org/10.1016/j.ins.2018.03.048

99. Wang L, Yu Z, Xiong F, Yang D, Pan S, Yan Z (2019) Influence spread in geo-social networks: a multiobjective optimization perspective. IEEE Trans Cybern. https://doi.org/10.1109/tcyb.2019.2906078

100. Saleem MA, Kumar R, Calders T, Xie X, Pedersen TB (2017) Location influence in location-based social networks. In: Proceedings of the tenth ACM international conference on web search and data mining. ACM, pp 621–630. https://doi.org/10.1145/3018661.3018705

101. Liu J, Hao F, Wang Y (2018) Discovering influential areas according to check-in records and user influence in social networks. In: IEEE 20th international conference on high performance computing and communications; IEEE 16th international conference on smart city; IEEE 4th international conference on data science and systems IEEE. https://doi.org/10.1109/hpcc/smartcity/dss.2018.00193

102. Yu F, Jiang S (2018) Mining location influence for location promotion in location-based social networks. IEEE Access 6:73444–73456. https://doi.org/10.1109/access.2018.2882057

103. Zhu WY, Peng W-C, Chen L-J, Zheng K, Zhou X (2016) Exploiting viral marketing for location promotion in location-based social networks. ACM Trans Knowl Discov Data 11(2):1–28. https://doi.org/10.1145/3001938

104. Li G, Wen Z-Y, Zhu W-Y (2016) Promoting a bundle of locations via viral marketing in location-based social networks. In: Conference on technologies and applications of artificial intelligence. IEEE, pp 32–39. https://doi.org/10.1109/taai.2016.7880165

105. Wu X, Fu L, Yao Y, Fu X, Wang X, Chen G (2018) GLP: a novel framework for group-level location promotion in geo-social networks. IEEE/ACM Trans Netw 26(6):2870–2883. https://doi.org/10.1109/TNET.2018.2879437

106. Wang X, Zhang Y, Zhang W, Lin X (2017) Efficient distance-aware influence maximization in geo-social networks. IEEE Trans Knowl Data Eng 29(3):599–612. https://doi.org/10.1109/tkde.2016.2633472

107. Saleem MA, Kumar R, Calders T, Xie X, Pedersen TB (2017) Imaxer: a unified system for evaluating influence maximization in location-based social networks. In: Proceedings of the 2017 ACM on conference on information and knowledge management. ACM, pp 2523–2526. https://doi.org/10.1145/3132847.3133184

108. Brown C, Nicosia V, Scellato S, Noulas A, Mascolo C (2013) Social and place-focused communities in location-based online social networks. Eur Phys J B. https://doi.org/10.1140/epjb/e2013-40253-6

109. Liu Z, Huang Y (2014) Community detection from location-tagged networks. In: Proceedings of the 22nd ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, pp 525–528. https://doi.org/10.1145/2666310.2666496

110. Joseph Hannigan HG, Medina R, Roos P, Shakarian P (2013) Mining for spatially-near communities in geo-located social networks. AAAI Ffall symposium - Technical Report. https://doi.org/10.21236/ada590263

111. Chen Y, Xu J, Xu M (2015) Finding community structure in spatially constrained complex networks. Int J Geogr Inf Sci 29(6):889–911. https://doi.org/10.1080/13658816.2014.999244

112. Zhao Y-L, Chen Q, Yan S, Chua T-S, Zhang D (2013) Detecting profilable and overlapping communities with user-generated multimedia contents in LBSNs. ACM Trans Multimed Comput Commun Appl 10(1):1–22. https://doi.org/10.1145/2502415

113. Alaqta I, Wang J, Awrangjeb M (2019) Effective community search over location-based social networks: conceptual framework with preliminary result. In: Australasian database conference. Springer, pp 119–131. https://doi.org/10.1007/978-3-030-12079-5_9

114. Luo J, Cao X, Xie X, Qu Q (2019) Best co-located community search in attributed networks. In: Proceedings of the 28th ACM international conference on information and knowledge management. pp 2453–2456. https://doi.org/10.1145/3357384.3358107

115. Wang K, Cao X, Lin X, Zhang W, Qin L (2018) Efficient computing of radius-bounded k-cores. In: IEEE 34th international conference on data engineering. IEEE, pp 233–244. https://doi.org/10.1109/icde.2018.00030

116. Wang Z, Zhou X, Zhang D, Yang D, Yu Z (2013) Cross-domain community detection in heterogeneous social networks. Personal UbiquitComput 18(2):369–383. https://doi.org/10.1007/s00779-013-0656-0

117. Ghane'i-Ostad M, Vahdat-Nejad H, Abdolrazzagh-Nezhad M (2018) Detecting overlapping communities in LBSNs by fuzzy subtractive clustering. Social Network Analysis and Mining 8 (1). https://doi.org/10.1007/s13278-018-0502-5

118. Yao K, Papadias D, Bakiras S (2019) Density-based community detection in geo-social networks. In: Proceedings of the 16th international symposium on spatial and temporal databases. pp 110–119. https://doi.org/10.14711/thesis-991012710864903412

119. Li X, Sun C, Zia MA (2020) Social influence based community detection in event-based social networks. Inf Process Manag 57(6):102353. https://doi.org/10.1016/j.ipm.2020.102353

120. Li H, Deng K, Cui J, Dong Z, Ma J, Huang J (2018) Hidden community identification in location-based social network via probabilistic venue sequences. Inf Sci 422:188–203. https://doi.org/10.1016/j.ins.2017.09.019

121. Hung C-C, Chang C-W, Peng W-C (2009) Mining trajectory profiles for discovering user communities. In: Proceedings of the 2009 International Workshop on Location Based Social Networks. ACM, pp 1–8. https://doi.org/10.1145/1629890.1629892

122. Xu C, Zhu L, Liu Y, Guan J, Yu S (2019) DP-LTOD: Differential privacy latent trajectory community discovering services over location-based social networks. IEEE Transactions on Services Computing:1–1. https://doi.org/10.1109/tsc.2018.2855740

123. Yin H, Hu Z, Zhou X, Wang H, Zheng K, Nguyen QVH, Sadiq S (2016) Discovering interpretable geo-social communities for user behavior prediction. In: IEEE 32nd international conference on data engineering. IEEE, pp 942–953. https://doi.org/10.1109/icde.2016.7498303

124. Zhang JD, Chow C-Y (2016) CRATS: An LDA-based model for jointly mining latent communities, regions, activities, topics, and sentiments from geosocial network data. IEEE Trans Knowl Data Eng. https://doi.org/10.1109/tkde.2016.2594772

125. Xu S, Cao J, Zhu X, Dong Y, Liu B (2018) Community discovery based on social relations and temporal-spatial topics in LBSNs. In: Pacific-Asia conference on knowledge discovery and data mining. Springer, pp 206–217. https://doi.org/10.1007/978-3-319-93040-4_17

126. Cerezo-Costas H, Fernández-Vilas A, Martín-Vicente M, Díaz-Redondo RP (2018) Discovering geo-dependent stories by combining density-based clustering and thread-based aggregation techniques. Expert Syst Appl 95:32–42. https://doi.org/10.1016/j.eswa.2017.11.019

127. Lee R, Sumiya K (2010) Measuring geographical regularities of crowd behaviors for Twitter-based geo-social event detection. In: Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks. ACM, pp 1–10. https://doi.org/10.1145/1867699.1867701

128. Chen L, Jakubowicz J, Yang D, Zhang D, Gang P (2017) Fine-Grained urban event detection and characterization based on tensor cofactorization. IEEE Trans Hum Mach Syst 47(3):380–391. https://doi.org/10.1109/thms.2016.2596103

129. Capdevila J, Cerquides J, Nin J, Torres J (2016) Tweet-SCAN: an event discovery technique for geo-located tweets. Pattern Recogn Lett. https://doi.org/10.1016/j.patrec.2016.08.010

130. Xu S, Li S, Huang W (2020) A spatial-temporal-semantic approach for detecting local events using geo-social media data. Trans GIS 24(1):142–173. https://doi.org/10.1111/tgis.12589

131. Georgiev P, Noulas A, Mascolo C (2014) The call of the crowd: Event participation in location-based social services. In: Proceedings of the International AAAI Conference on Web and Social Media. vol 1

132. Du R, Yu Z, Mei T, Wang Z, Wang Z, Guo B (2014) Predicting activity attendance in event-based social networks: content, context and social influence. Paper presented at the Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing,

133. Yu Z, Yi F, Lv Q, Guo B (2018) Identifying on-site users for social events: mobility, content, and social relationship. IEEE Trans Mob Comput. https://doi.org/10.1109/tmc.2018.2794981

134. Xu T, Zhu H, Zhong H, Liu G, Xiong H (2018) Exploiting the dynamic mutual influence for predicting social event participation. IEEE Trans Knowl Data Eng 6:1122–1135. https://doi.org/10.1109/tkde.2018.2851222

135. Yeo J, Cho H, Park J-W, Hwang S-w (2017) Multimodal KB harvesting for emerging spatial entities. IEEE Trans Knowl Data Eng 29(5):1073–1086. https://doi.org/10.1109/tkde.2017.2651805

136. Silva T et al (2014) Large-scale study of city dynamics and urban social behavior using participatory sensing. IEEE Wirel Commun 21(1):42–51. https://doi.org/10.1109/mwc.2014.6757896

137. Smarzaro R, Lima TFMD, Jr CAD (2017) Could data from location-based social networks be used to support urban planning? In: International World Wide Web conference. https://doi.org/10.1145/3041021.3051700

138. Cranshaw J, Schwartz R, Hong J, Sadeh N (2012) The livehoods project: Utilizing social media to understand the dynamics of a city. In: 6th international AAAI conference on weblogs and social media

139. Frias-Martinez V, Frias-Martinez E (2014) Spectral clustering for sensing urban land use using Twitter activity. Eng Appl Artif Intell 35:237–245. https://doi.org/10.1016/j.engappai.2014.06.019

140. Assem H, Lei X, Buda TS, O'Sullivan D (2017) Spatio-temporal clustering approach for detecting functional regions in cities. In: IEEE 28th international conference on tools with artificial intelligence. IEEE. https://doi.org/10.1109/ictai.2016.0063

141. Steiger E, Resch B, Zipf A (2016) Exploration of spatiotemporal and semantic clusters of Twitter data using unsupervised neural networks. Int J Geogr Inf Sci 30(9):1694–1716. https://doi.org/10.1080/13658816.2015.1099658

142. Guo Z, Zheng Z, Liu J, Wang S, Zhong P, Zhu M, He Y, Jiang L, Zhou G, Zhang H (2018) Urban functional regions using social media check-Ins. In: IEEE international geoscience and remote sensing symposium. IEEE, pp 5061–5064. https://doi.org/10.1109/igarss.2018.8517974

143. Zhang X, Sun Y, Zheng A, Wang Y (2020) A new approach to refining land use types: predicting point-of-interest categories using weibo check-in data. ISPRS Int J Geo-Inform 9(2):124. https://doi.org/10.3390/ijgi9020124

144. Terroso-Saenz F, Munoz A (2020) Land use discovery based on Volunteer Geographic Information classification. Expert Syst Appl 140:112892. https://doi.org/10.1016/j.eswa.2019.112892

145. Zhao H, Gong Q, Chen Y, Chen J, Li Y, Fu X (2018) This place is swarming: using a mobile social app to study human traffic in cities. In: IEEE international conference on pervasive computing and communications workshops. IEEE, pp 259–264. https://doi.org/10.1109/percomw.2018.8480234

146. Mcardle G, Furey E, Lawlor A, Pozdnoukhov A (2014) Using digital footprints for a city-scale traffic simulation. ACM Trans Intell Syst Technol 5(3):1–16. https://doi.org/10.1145/2517028

147. AlDwyish A, Tanin E, Karunasekera S (2015) Location-based social networking for obtaining personalised driving advice. In: Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems. pp 1–4. https://doi.org/10.1145/2820783.2820789

148. Xu T, Zhu H, Xiong H, Zhong H, Chen E (2019) Exploring the social learning of taxi drivers in latent vehicle-to-vehicle networks. IEEE Trans Mob Comput 99:1–1. https://doi.org/10.1109/tmc.2019.2915228

149. Masaki Kanno YE, Masaharu Hirota, Shohei Yokoyama, Hiroshi Ishikawa (2016) Visualizing high-risk paths using geo-tagged social data for disaster mitigation. In: LBSN16 proceedings of the 9th ACM SIGSPATIAL workshop on location-based social networks. ACM. https://doi.org/10.1145/3021304.3021308

150. Yokoyama S, Bogárdi-Mészöly Á, Ishikawa H (2015) Ebscan: an entanglement-based algorithm for discovering dense regions in large geo-social data streams with noise. In: Proceedings of the 8th ACM SIGSPATIAL international workshop on location-based social networks. ACM, p 7. https://doi.org/10.1145/2830657.2830661

151. Hu Q, Bai G, Wang S, Ai M (2019) Extraction and monitoring approach of dynamic urban commercial area using check-in data from Weibo. Sustain Cities Soc 45:508–521. https://doi.org/10.1016/j.scs.2018.11.039

152. Kadar C, Brüngger R, Pletikosa I (2017) Measuring ambient population from location-based social networks to describe urban crime. In: International conference on social informatics. https://doi.org/10.1007/978-3-319-67217-5_31

153. Rumi SK, Deng K, Salim FD (2018) Theft prediction with individual risk factor of visitors. In: Proceedings of the 26th ACM SIGSPATIAL international conference on advances in geographic information systems. ACM, pp 552–555. https://doi.org/10.1145/3274895.3274994

154. Rosés R, Kadar C, Gerritsen C, Rouly C (2018) Agent-based simulation of offender mobility: integrating activity nodes from location-based social networks. In: Proceedings of the 17th international conference on autonomous agents and multiAgent systems. International Foundation for Autonomous Agents and Multiagent Systems, pp 804–812

155. Rumi SK, Salim FD (2020) Modelling regional crime risk using directed graph of check-ins. In: Proceedings of the 29th ACM international conference on information and knowledge management. pp 2201–2204. https://doi.org/10.1145/3340531.3412065

156. Jung Y (2019) Community-based localized disaster response through temporary social overlay networks. Mob Netw Appl 24(5):1641–1653. https://doi.org/10.1007/s11036-017-0892-z

157. Wu W, Li J, He Z, Ye X, Zhang J, Cao X, Qu H (2020) Tracking spatio-temporal variation of geo-tagged topics with social media in China: a case study of 2016 hefei rainstorm. Int J Disast Risk Reduction 50:101737. https://doi.org/10.1016/j.ijdrr.2020.101737

158. Zadeh AH, Zolbanin HM, Sharda R, Delen D (2019) Social media for nowcasting flu activity: spatio-temporal big data analysis. Inform Syst Front 21(4):743–760. https://doi.org/10.1007/s10796-018-9893-0

159. Li H, Zhu H, Du S, Liang X, Shen XS (2018) Privacy leakage of location sharing in mobile social networks: attacks and defense. IEEE Trans Dependable Secure Comput 15(4):646–660. https://doi.org/10.1109/tdsc.2016.2604383

160. Lee B, Oh J, Yu H, Kim J (2011) Protecting location privacy using location semantics. In: Proceedings of the 17th ACM SIGKDD international conference on knowledge discovery and data mining. https://doi.org/10.1145/2020408.2020602

161. Sai AMVV, Li Y (2020) A survey on privacy issues in mobile social networks. IEEE Access 8:130906–130921. https://doi.org/10.1109/access.2020.3009691

162. Dong C, Jin H, Knijnenburg BP (2015) Predicting privacy behavior on online social networks. In: 9th international AAAI conference on web and social media. pp 91–100

163. Zhang H, Xu Z, Zhou Z, Shi J, Du X (2015) CLPP: context-aware location privacy protection for location-based social network. In: IEEE international conference on communications. https://doi.org/10.1109/icc.2015.7248480

164. Peng T, Liu J, Wang G, Liu Q, Chen J, Zhu J (2020) A user-defined location-sharing scheme with efficiency and privacy in mobile social networks. Sci Program. https://doi.org/10.1155/2020/7832875

165. Damiani ML, Bertino E, Silvestri C (2010) The PROBE framework for the personalized cloaking of private locations. Trans Data Privacy 3(2):123–148

166. Ma T, Jia J, Xue Yu, Tian Y, Al-Dhelaan A (2018) Protection of location privacy for moving kNN queries in social networks. Appl Soft Comput 66(3):525–532. https://doi.org/10.1016/j.asoc.2017.08.027

167. Yang D, Qu B, Cudré-Mauroux P (2018) Privacy-preserving social media data publishing for personalized ranking-based recommendation. IEEE Trans Knowl Data Eng 31(3):507–520. https://doi.org/10.1109/tkde.2018.2840974

168. Luceri L, Andreoletti D, Tornatore M, Braun T, Giordano S (2020) Measurement and control of geo-location privacy on Twitter. Online Soc Netw Med 17:100078. https://doi.org/10.1016/j.osnem.2020.100078

169. Mascetti S, Freni D, Bettini C, Wang XS, Jajodia S (2010) Privacy in geo-social networks: proximity notification with untrusted service providers and curious buddies. VLDB J 20(4):541–566. https://doi.org/10.1007/s00778-010-0213-7

170. Lee C, Yin L-H, Dong L (2014) A modified-k-anonymity towards spatial-temporal historical data in location-based social network service. In: Asia-Pacific web conference. Springer, Berlin, pp 301–311. https://doi.org/10.1007/978-3-319-11119-3_28

171. Sun G, Song L, Liao D, Yu H, Chang V (2019) Towards privacy preservation for "check-in" services in location-based social networks. Inform Sci 481:616–634. https://doi.org/10.1016/j.ins.2019.01.008

172. Lu R, Lin X, Shi Z, Shao J (2014) PLAM: a privacy-preserving framework for local-area mobile social networks. In: IEEE INFOCOM 2014-IEEE conference on computer communications. IEEE, pp 763–771. https://doi.org/10.1109/infocom.2014.6848003

173. Seglem E, Züfle A, Stutzki J, Borutta F, Faerman E, Schubert M (2017) On privacy in spatio-temporal data: user identification using microblog data. In: International symposium on spatial and temporal databases. Springer, Berlin. pp 43–61. https://doi.org/10.1007/978-3-319-64367-0_3

174. Jaiswal SAN (2010) Trust no one: A decentralized matching service for privacy in location based services. In: Proceedings of the 2nd ACM SIGCOMM workshop on networking, systems, and applications for mobile handhelds. https://doi.org/10.1145/1851322.1851336

175. Guha S, Jain M, Padmanabhan VN (2012) Koi: a location-privacy platform for smartphone apps. In: Proceedings of the 9th USENIX conference on networked systems design and implementation. USENIX Association, pp 14–14

176. Siddula M, Cai Z, Miao D (2018) Privacy preserving online social networks using enhanced equicardinal clustering. In: IEEE 37th international performance computing and communications conference. IEEE, pp 1–8. https://doi.org/10.1109/pccc.2018.8710844

177. Zhang S, Wang G, Bhuiyan MZA, Liu Q (2018) A dual privacy preserving scheme in continuous location-based services. IEEE Internet Things J 5(5):4191–4200. https://doi.org/10.1109/trustcom/bigdatase/icess.2017.264

178. Son J, Kim D, Bhuiyan MZA, Tashakkori R, Seo J, Lee DH (2018) Privacy enhanced location sharing for mobile online social networks. IEEE Trans Sustain Comput. https://doi.org/10.1109/tsusc.2018.2842788

179. Chen B-H, Li C-T, Chuang K-T (2018) A check-in shielding scheme against acquaintance inference in location-based social networks. World Wide Web. https://doi.org/10.1007/s11280-018-0653-3

180. Kim JS, Jin H, Kavak H, Rouly OC, Zufle A (2020) Location-based social network data generation based on patterns of life. Paper presented at the 21st IEEE international conference on mobile data management

181. Isaj S, Pedersen TB (2019) Seed-driven geo-social data extraction. In: Proceedings of the 16th international symposium on spatial and temporal databases. pp 11–20. https://doi.org/10.1145/3340964.3340973

182. Zhang H, Wei S, Hu X, Li Y, Xu J (2020) On accurate POI recommendation via transfer learning. Distrib Parallel Databases 38(3):585–599. https://doi.org/10.1007/s10619-020-07299-7

183. Li S, Shen H, Sang Y (2019) A survey of privacy-preserving techniques on trajectory data. International symposium on parallel architectures, algorithms and programming. Springer, Berlin, pp 461–476

184. Yin C, Ju X, Yin Z, Wang J (2019) Location recommendation privacy protection method based on location sensitivity division. EURASIP J Wirel Commun Netw 1:266. https://doi.org/10.1186/s13638-019-1606-y

185. Logesh R, Subramaniyaswamy V, Vijayakumar V, Li X (2019) Efficient user profiling based intelligent travel recommender system for individual and group of users. Mobile Netw Appl 24(3):1018–1033. https://doi.org/10.1007/s11036-018-1059-2

186. Purushotham S, Kuo C-CJ (2016) Personalized group recommender systems for location-and event-based social networks. ACM Trans Spatial Algorithms Syst 2(4):16. https://doi.org/10.1145/2987381

187. Liu C, Fuhrmann S (2018) Enriching the GIScience research agenda: fusing augmented reality and location-based social networks. Trans GIS 22(3):775–788. https://doi.org/10.1111/tgis.12345

188. Liu C, Fuhrmann S (2020) Analyzing relationship between user-generated content and local visual information with augmented reality-based location-based social networks. Trans GIS. https://doi.org/10.1111/tgis.12630

189. Christoforidis G, Kefalas P, Papadopoulos AN, Manolopoulos Y (2019) Recommending points of interest in LBSNs using deep learning techniques. In: IEEE international symposium on innovations in intelligent systems and applications. IEEE, pp 1–6

190. Guan C, Wang X, Zhang Q, Chen R, He D, Xie X (2019) Towards a deep and unified understanding of deep neural models in nlp. In: International conference on machine learning. pp 2454–2463

Xuemei Wei is working toward a Ph.D. degree in Management Science and Engineering from the Hefei University of Technology. She received her bachelor's degree in Management Science from Inner Mongolia University, Hohhot, China. Her research interests include machine learning and data mining, social network analysis, social networking services and recommendations.

Yang Qian is a Postdoctoral Fellow at the School of Management, Hefei University of Technology. He received his Ph.D. in management science and engineering from Hefei University of Technology and in 2020. His research interests include electronic commerce, online marketing, and text mining.

Jianshan Sun is an Associate Professor in the School of Management at the Hefei University of Technology. Her research focuses on online consumer behavior, social networks, and user generated content. Her works have been published in journals such as Scientific Reports, Discrete Dynamics in Nature and Society, Future Generation Computer Systems, and IEEE Transactions on Professional Communication..

Jianshan Sun is an Associate Professor of Electronic Commerce at the Hefei University of Technology. He received his Ph.D. in Information Systems from the City University of Hong Kong in 2014. His research interests include big data analytics, personalized recommendation, and social network analysis. His research has been presented at international conferences such as ICIS, HICSS, and PACIS and published in several journals including Tourism Management, Journal of the Association for Information Science and Technology, Decision Support Systems, International Journal of Production Economics, Information Processing and Management, amongst others.

Yezheng Liu is a Professor of Electronic Commerce at the Hefei University of Technology. He received his Ph.D. in Management Science and Engineering from the Hefei University of Technology in 2001. His main research interests include data mining, decision science, electronic commerce, and intelligent decision support systems. His current research focuses on big data analytics, online social network, personalized recommendation system, and outlier detection. He is the author and co-author of numerous papers in scholarly journals, including Marketing Science, Decision Support Systems, International Journal of Production Economics, Knowledge-Based Systems, Journal of Management Sciences in China. He is a national member of the New Century Talents Project.