Expert Systems With Applications 86 (2017) 113–124

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

Discovering socially important locations of social media users

Ahmet Sakir Dokuz[a,*], Mete Celik[b]

[a] Department of Computer Engineering, Omer Halisdemir University, 51245 Nigde, Turkey
[b] Department of Computer Engineering, Erciyes University, 38039 Kayseri, Turkey

ARTICLE INFO

Article history:
Received 6 January 2017
Revised 8 May 2017
Accepted 27 May 2017
Available online 29 May 2017

Keywords:
Socially important locations mining
Spatial social media mining
Historical social media data analysis
Social media networking sites
Twitter

ABSTRACT

Socially important locations are places that are frequently visited by social media users in their social media life. Discovering socially interesting, popular or important locations from a location based social network has recently become important for recommender systems, targeted advertisement applications, and urban planning, etc. However, discovering socially important locations from a social network is challenging due to the data size and variety, spatial and temporal dimensions of the datasets, the need for developing computationally efficient approaches, and the difficulty of modeling human behavior. In the literature, several studies are conducted for discovering socially important locations. However, majority of these studies focused on discovering locations without considering historical data of social media users. They focused on analysis of data of social groups without considering each user's preferences in these groups. In this study, we proposed a method and interest measures to discover socially important locations that consider historical user data and each user's (individual's) preferences. The proposed algorithm was compared with a naïve alternative using real-life Twitter dataset. The results showed that the proposed algorithm outperforms the naïve alternative.

1. Introduction

With the development of location-aware mobile devices, social media networking sites started collecting location information of their users when they are online. By using location data and social interactions (e.g., location-location interactions, user-location interactions, and user-user interactions), valuable information can be obtained, such as, which social friends are also real friends, which user likes what types of locations, which people spatially follow each other, which person's recommendation about a location gets more responses, which locations are important for a group of people, and so on. The aim of this study is to discover socially important locations that can be defined as places that are frequently visited by social media users in their social media life.

Discovering socially important locations is important for several application domains, including recommender systems, marketing (e.g., targeted advertisement applications), and urban planning. By discovering socially important locations, spatial preferences of a social group could be revealed.

However, discovery of socially important locations is challenging for several reasons. First, current interest measures are not sufficient to mine such locations. Second, the size and dimension of the data are growing over time. Third, human behaviors are unpredictable and can be affected by several factors.

Important locations mining problem is studied by many researchers and several solutions are proposed. However, these studies have several limitations. Some of these studies are based on GPS or Call Detail Record (CDR) data, thus proposed solutions are dependent to these factors. Another part of studies used social media datasets. Although several algorithms and frameworks are present for user-level recommendation and discovery tasks (Kefalas, Symeonidis, & Manolopoulos, 2016; Zheng, 2015), relatively small number of studies are performed for discovering group-level important/interesting locations of a group of social media users.

Example socially important locations for a social media user (twitter user) and a social media user group (1000 twitter users) can be seen in Fig. 1(a) and (b), respectively, in Istanbul, Turkey. Fig. 1 (a) shows three socially important locations of the social media user. The dataset used to discover the socially important locations of the social media user includes 1765 social media activities (i.e., location and time information of tweets) that belong to the period of December 2012 and December 2015. As can be seen in Fig. 1 (a), these locations are spatially unrelated locations and

* Corresponding author.
E-mail addresses: adokuz@ohu.edu.tr (A.S. Dokuz), mcelik@erciyes.edu.tr (M. Celik).

http://dx.doi.org/10.1016/j.eswa.2017.05.068
0957-4174/

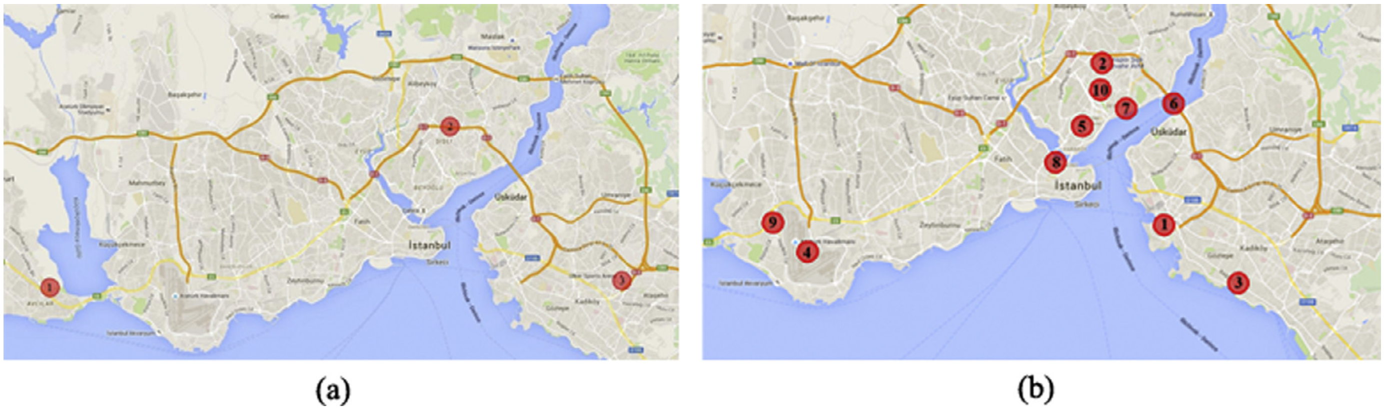

Fig. 1. Socially important locations of (a) a selected user and (b) a social media user group.

may have different meanings for the user. In location 1, there is a university campus, in location 2, there is a big shopping mall nearby, and in location 3, there is a big company's head office. We can make several predictions about importance of these locations for the user. For example, location 1 might be the user's university, location 2 might be a working place or shopping place of the user, and location 3 might be a working place. Fig. 1 (b) shows ten socially important locations of 1000 social media users in Istanbul, Turkey. The details of social media user group is presented in Section 5.2.2. Socially important locations of the group are clustered around the places where city social life is active. Comparison of the socially important locations of the social media user given in Fig. 1 (a) and that of social media user group given in Fig. 1 (b) shows that, the location 2, (i.e., the shopping mall) is a common socially important location for the social media user and the social media user group. In contrast, locations 1 and 3 are socially important locations only for the social media user not for the social media user group.

The three socially important locations of the user in Fig. 1 (a) are discovered based on his/her daily social media activity records. When social media users are brought together, the analysis of these users' datasets will lead to the discovery of important location(s) of that user group. For instance, if social media data of music lovers are analyzed, live music cafes, concert areas, and music shops are more likely to appear as socially important locations. If social media data of football fans are analyzed, then stadiums and sports-related locations are more likely to appear as socially important locations.

This paper deals with socially important locations mining problem over social media datasets. In this paper, we formulized interest measures of location density, visit lifetime and user prevalence to quantify socially important locations and proposed an algorithm, which is called SocioSpatially Important Locations Mining algorithm (SS-ILM), for discovering socially important locations efficiently. The proposed approach has two stages. In the first stage, user-level socially important locations are discovered using location density and visit lifetime interest measures. In the second stage, socially important locations of a social user group are discovered using user prevalence interest measure. The proposed algorithm, SS-ILM, was compared with a naïve algorithm. Both algorithms are experimentally evaluated using real dataset of Istanbul, Turkey Twitter users.

The rest of this paper is organized as follows. Section 2 discusses related work. Section 3 presents formal model of socially important locations mining problem. Section 4 introduces socially important locations mining algorithms, a naïve algoritm and proposed SS-ILM algorithm. Section 5 presents the experimental evaluation and Section 6 gives conclusions and future directions.

2. Related work

Spatial and spatio-temporal data mining have found a new research area after the introduction of Location Based Social Network (LBSN) datasets (Kefalas et al., 2016).

Spatial and spatio-temporal data mining is an important domain of data mining which aims to handle spatial and temporal data. Spatial co-location mining (Celik, 2015; Celik, Shekhar, Rogers, & Shine, 2008; Dadaser-Celik, Celik, & Dokuz, 2012; Yu, 2016) and spatial clustering (Hu & Sung, 2005; Tung, Hou, & Han, 2001) are some of the main topics (Shekhar, Lu, & Zhang, 2001). A co-location in social media datasets can be defined as a group of users or locations that are located neighboring spatial areas (Kefalas & Symeonidis, 2015; Vasuki, Natarajan, Lu, Savas, & Dhillon, 2011; Weiler, Schmid, Mamoulis, & Renz, 2015). The studies on mining classical spatial co-location mining assume that objects are boolean, that is, objects are present or not in a space or time. However, this assumption is not valid for social media datasets since a social media user may have more than one social media activity (e.g., sending tweets or posting messages on Facebook) at a location. Because of this reason, the algorithms and interest measures proposed in these studies may not be suitable to mine socially important locations out of social media datasets.

The important locations mining studies can be classified into two categories: 1) mining user-level socially important locations and 2) mining group-level socially important locations. Table 1 presents a summary of the literature related to this study.

The studies on mining user-level socially important locations include location/route recommendation, friend recommendation, activity/tag recommendation, interesting/important locations discovery, route mining, and discovering urban and mobility patterns (Table 1).

User-level location/route recommendation studies include the ones conducted by Bao, Zheng, and Mokbel (2012), Doytsher, Galon, and Kanza (2011), Kefalas and Symeonidis (2015), Levandoski, Sarwat, Eldawy, and Mokbel (2012), Liu and Seah (2015), Liu and Xiong (2013) and Ying, Lu, Kuo, and Tseng (2012). Liu and Seah (2015) proposed a method for important location (i.e., Point of Interest - POI) recommendation based on GPS trajectory data of mobile users and proposed density-based clustering approach to discover important locations. To determine the recommendation score of an important location (i.e., POI) three different factors (popularity, temporal and geographical features) have been used. Doytsher et al. (2011) presented a graph model that supports the representation of frequently-traveled routes in a socio-spatial network, proposed a query language to traverse the graph for querying social-based route recommendations, and used GPS trajectory dataset. In the studies conducted by Levandoski et al. (2012) and

Table 1
Literature work.

		User	Group
Recommendation	Location/route	Bao et al. (2012), Doytsher et al. (2011), Kefalas and Symeonidis (2015), Levandoski et al. (2012), Liu and Xiong (2013), Ying et al. (2012)	Zheng et al. (2008), Zheng et al. (2010), Yin et al. (2011)
	Friend	Hannon et al. (2010), Kefalas and Symeonidis (2015), Ying et al. (2010), Zheng et al. (2011)	Gaete-Villegas et al. (2012), Symeonidis et al. (2011)
	Activity/tag	Budak et al. (2013), Feng et al. (2015), Sattari et al. (2012)	Guy et al. (2010)
Discovery	Location	Cao et al. (2010), Isaacman et al. (2011), Pavan et al. (2015), Ying et al. (2011), Zhou et al. (2007)	Khetarpaul et al. (2011), Zheng et al. (2009)
	Route	Cai et al. (2014), Cenamor et al. (2017), Chen et al. (2009), Chen et al. (2011), Doytsher et al. (2011)	Chen et al. (2011), Wei et al. (2012)
	Urban pattern	Chen et al. (2016), Cho et al. (2011), Liu et al. (2012)	Cranshaw et al. (2012), Ferrari et al. (2011)

Bao et al. (2012), location-aware recommendation systems were proposed by considering location-based ratings of users. Ying et al. (2012) proposed UPOI-Mine algorithm for POI recommendation by mining check-in behaviors of users and used regression tree for estimating relevance score of POIs of each user. Liu and Xiong (2013) proposed a topic and location aware POI recommender system by exploiting associated textual and context information of location based social networks. Kefalas and Symeonidis (2015) proposed a friend and location recommendation method by incorporating the time dimension into recommender systems. The proposed method can capture user preferences that change over time. These studies mainly deal with GPS data and needs user ratings for location/route recommendations.

User-level friend recommendation studies include the ones conducted by Hannon, Bennett, and Smyth (2010), Kefalas and Symeonidis (2015), Zheng, Zhang, Ma, Xie, and Ma (2011) and Ying, Lu, Lee, Weng, and Tseng (2010). Zheng et al. (2011) discovered important locations of a user and then recommends new locations based on the other similar users' important locations. In the study GPS data were used and hierarchical-graph-based similarity measurement was used as the user similarity measure. Ying et al. (2010) used semantic trajectories of mobile users for friend recommendation. Hannon et al. (2010) studied a range of different profiling and recommendation strategies for followers/followees recommendation and used twitter dataset. This study does not take into account location and time information of social media users. These studies mainly deal with similarity search among the users for friend recommentations.

User-level activity/tag recommendation studies include the ones conducted by Budak, Georgiou, Agrawal, and El Abbadi (2013), Feng et al. (2015) and Sattari et al. (2012). Feng et al. (2015) proposed a data structure to support hierarchical spatio-temporal hashtag clustering. This data structure supports to query the twitter data with different time and space granularity. Budak et al. (2013) proposed GeoScope for detection of correlations between topics and locations. In this study twitter data were used to discover location-topic pairs. Sattari et al. (2012) proposed a method which recommends activities for a given location. They used Singular Value Decomposition (SVD) to make recommendations from a sparse location-activity matrix.

User-level interesting/important locations discovery studies include the ones conducted by Cao, Cong, and Jensen (2010), Ying, Lee, Weng, and Tseng (2011), Zhou, Bhatnagar, Shekhar, and Terveen (2007), Isaacman et al. (2011) and Pavan, Mizzaro, Scagnetto, and Beggiato (2015). Cao et al. (2010) proposed a framework for the mining of semantically meaningful and significant locations. In this study, GPS data collected from cars were obtained and clustering based approaches were used to discover significant semantic locations. Ying et al. (2011) proposed a clustering-based strategy for predicting the next location of a user's movement by taking into account the geographic and semantic features of users' trajectories. Zhou et al. (2007) proposed methods to mine personally important places. In this study, GPS data were used and clustering based approaches were developed. Isaacman et al. (2011) proposed clustering and regression based approaches to mine important personal locations. In this study, Call Detail Records (CDRs) data were used. Pavan et al. (2015) proposed a feature based approach to find personal places of interest. The studies on interesting/important locations discovery mainly use GPS (trajectory) and CDR datasets.

User-level route mining studies include the ones conducted by Cai, Hio, Bermingham, Lee, and Lee (2014), Cenamor, de la Rosa, Nez, and Borrajo (2017) and Chen, Jiang, Zheng, Li, and Yu (2009); Chen, Shen, and Zhou (2011). Chen et al. (2009) proposed a method which simplifies trajectories from large number of points to limited ones without loosing its skeleton and semantic meaning. Cai et al. (2014) proposed methods for the detection of region of interests and they combined the proposed algorithm with sequential pattern mining to discover sequential trajectory patterns. Chen et al. (2011) discovered most popular route between two locations by considering previous users' traveling trajectories using Absorbing Markov Chain model. Cenamor et al. (2017) proposed a system that builds personalized plans for tourists using the dataset collected from minitube travelling social network. These studies mainly deal with trajectories to discover routes.

User-level urban and mobility pattern discovery studies include the ones conducted by Chen, Chiang, and Peng (2016), Cho, Myers, and Leskovec (2011) and Liu, Huang, Chen, Shen, and Yan (2012). Chen et al. (2016) proposed approaches to mine and cluster daily mobility evolution patterns out of check-in data. The aim in this study is to capture the daily movement behavior of users in a city by taking into account both spatial and temporal information. Cho et al. (2011) investigated patterns of human mobilities and proposed human mobility models to predict the locations and future movement of an individual. Liu et al. (2012) proposed a method to discover areas of interest by analyzing both geo-tagged images and check-in data. These studies mainly focused on predicting locations and future movements of users.

The studies on mining group-level socially important locations include location/route recommendation, friend recommendation, activity/tag recommendation, interesting/important locations discovery, route mining, and discovering urban and mobility patterns (Table 1).

Group-level location/route recommendation studies include the ones of Zheng, Wang, Zhang, Xie, and Ma (2008), Zheng, Xie, and Ma (2010) and Yin, Cao, Han, Luo, and Huang (2011). Zheng et al. (2008,2010) proposed GeoLife platform to index, visualize, query and mine GPS data. In these studies, travel sequences among the locations were detected based on user experience and location interest. Yin et al. (2011) focused on mining frequent trajectory patterns from geo-tagged social media data. In this study, mined trajectory patterns were ranked in order to find the representative trajectories.

Group-level friend recommendation studies include the ones of Gaete-Villegas, Cha, Lee, and Ko (2012) and Symeonidis, Papadimitriou, Manolopoulos, Senkul, and Toroslu (2011). Gaete-Villegas et al. (2012) proposed an approach to match individuals as travel partner by leveraging location data in online social networks. Symeonidis et al. (2011) proposed Geo-social online recommender system that provides friends, locations and activities recommendations based on user check-in data.

Group-level activity/tag recommendation study conducted by Guy, Zwerdling, Ronen, Carmel, and Uziel (2010) proposed recommender system based on people and tags. In this study tag-based and people-based recommenders were compared and proposed a people-tag-based recommender system. This study deals with tags and people for recommendations and does not take into account location and time information of the dataset.

Group-level interesting/important locations discovery studies include the ones of Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) and Zheng, Zhang, Xie, and Ma (2009). Zheng et al. (2009) proposed approaches and models to discover top n interesting locations and top m travel sequences by considering users' travel experiences and the correlation between locations. In the study, GPS data were used. Khetarpaul et al. (2011) proposed an approach to discover interesting locations by aggregating GPS data and analyzing multiple users' trajectories.

Group-level route mining studies include the studies of Chen et al. (2011) and Wei, Zheng, and Peng (2012). Chen et al. (2011) proposed a route mining approach to discover most popular path from a trajectory dataset by taking into account users' traveling trajectories. Wei et al. (2012) proposed an approach to discover the popular (i.e., top-k) routes from uncertain trajectories.

Group-level urban and mobility patterns discovery studies include the ones of Ferrari, Rosi, Mamei, and Zambonelli (2011) and Cranshaw, Schwartz, Hong, and Sadeh (2012). Ferrari et al. (2011) proposed an approach for the discovery of urban patterns from location-based social networks. In this study, twitter dataset were used and a probabilistic topic model of Latent Dirichlet Allocation (LDA) were adopted to identify mobility patterns of people. Cranshaw et al. (2012) proposed clustering approaches for discovering the distinct geographic areas of the city and movement patterns of the people in the city. In this study, a check-in dataset collected from users of a location-based online social network were used.

The studies in the literature have several limitations. First, some of these studies use GPS or CDR data. Social media datasets have more information than GPS or CDR data and contain more interaction among users. Second, some of the studies need user voting for determining important locations. Third, some of the studies used text mining methods and tags for location identification and did not consider spatial information. And finally, majority of the studies do not identify important locations of social media groups by considering individual user's preferences.

To fill the gap in the literature, this study aims to discover important locations from user's social media activity histories without requiring descriptive information about locations such as voting or location type. Our study falls into group-level interesting/important locations discovery category. The results of this study are extracted with the user group's data, not with the cultural or public's opinions. In this study, new interest measures (location density, visit lifetime, and user prevalence) are proposed to quantify socially important locations and a new algorithm, named as SocioSpatially-Important Locations Mining (SS-ILM) algorithm, is proposed to efficiently mine such locations from social media datasets.

3. Basic concepts and problem definition

In this section, first, the general idea of socially important locations mining from social media datasets is presented and then, problem definition is given.

3.1. Basic concepts

The problem of socially important locations mining for a user group has two stages. The first stage is to determine user-level socially important locations and the second stage is to determine group-level socially important locations of social media users. The definitions related to socially important location mining problem are given below. Definitions 1–5 formulizes the discovery of user-level socially important locations. Definitions 6–7 formulizes the discovery of group-level socially important locations.

Definition 1. Given the social media activity history of a social media user u and a location l, the **location density** of l for u is the fraction of number of occurrences of u at l to the total number of occurrences of u.

$$location\ density_l^u = \frac{\#\ of\ occurrences\ of\ u\ at\ l}{\#\ of\ all\ occurrences\ of\ u} \tag{1}$$

Location density is used to determine how frequently the location l occurs in the social media activity history of the social media user u.

Definition 2. Given location density value of social media user u at location l and minimum location density threshold value $min\_density$, the location l is called **frequent location**, if location density value satisfies the $min\_density$ threshold, such as, $location\ density_l^u \geq min\_density$.

Table 2
Notations used in the paper.

Notation	Description
u	Social media user
l	Location
min_density	Location density threshold
min_visit	Visit lifetime threshold
SILU	Socially Important Location for User
UP	User Prevalence
min_UP	Minimum user prevalence threshold
SIL	Socially Important Location

Definition 3. Given the social media activity history of a social media user u and a location l, the **visit lifetime** of l for u is the fraction of user's time difference between last visit and first visit of location l to the time difference between first and last occurrence in social media data, i.e., social media lifetime, of social media user u.

$$visit\ lifetime_l^u = \frac{time\ difference\ between\ last\ and\ first\ visit\ of\ l}{social\ media\ lifetime\ of\ u} = \frac{Last\_Visit_{l\ in\ u} - First\_Visit_{l\ in\ u}}{Last\_Time_u - First\_Time_u} \quad (2)$$

Visit lifetime of a location l for user u shows the time pervasiveness of l over u's social media lifetime.

Definition 4. Given visit lifetime value of a social media user u at location l and minimum visit lifetime threshold value *min_visit*, the location l is called **time pervasive location**, if visit lifetime value satisfies the *min_visit* threshold, such as, $visit\ lifetime_l^u \geq min\_visit$.

min_visit threshold controls the propagation of location l in user us social media lifetime.

Definition 5. Given a location l for a social media user u, the location l is called **Socially Important Location for User (SILU)**, if l is frequent and time pervasive location for u as defined in Definitions 2 and 4, respectively.

Definition 6. Given a location l, **User Prevalence (UP)** of l is the fraction of the number of social media users who have l as SILU (as defined in Definition 5) to the total number of social media users.

$$UP_l = \frac{\#\ of\ users\ that\ have\ l\ as\ SILU}{Total\ number\ of\ users} \quad (3)$$

User prevalence value presents location l's popularity among the social media user group.

Definition 7. Given UP value and minimum UP threshold *min_UP*, a location l is **Socially Important Location (SIL)**, if UP value of l satisfies *min_UP* threshold, such as, $UP_l \geq min\_UP$.

min_UP threshold controls percentage of users who have a location as SILU.

Table 2 summarizes the notations used throughout this paper.

3.2. Problem definition

We formulize socially important locations mining problem as follows: Given a set of location and time information (which can be extracted from social media activities) of social media user collection D, a minimum location density threshold *min_density*, a minimum visit lifetime threshold *min_visit*, and a minimum user prevalence threshold,*min_UP*, the problem is to efficiently discover correct socially important locations that satisfy given threshold values.

4. Discovery of socially important locations

In this section, first, a naïve algorithm is discussed and then novel SS-ILM algorithm is proposed to mine socially important locations. We also present an execution trace of the SS-ILM algorithm.

For the discovery of socially important locations, first, user-level socially important location mining and then, group-level socially important location mining stages are applied. The time consuming part of this process is the discovery of user-level socially important locations which include two main steps such as, calculation of location densities and calculation of visit lifetimes. The algorithms which is proposed in this section use both calculations of location density and visit lifetime. However, the design of the algorithms differ based on location pruning strategy.

4.1. Naïve algorithm

A naïve algorithm calculates location densities and visit lifetimes separately and then discovers frequent and time pervasive locations that are satisfying *min_density* and *min_visit* thresholds to discover user-level socially important locations. After that, group-level socially important locations that satisfy *min_UP* threshold can be discovered. In this approach, frequent locations and time pervasive locations are discovered separately and then socially important locations are discovered. The pseudo-code of the naïve algorithm is presented in Algorithm 1.

Algorithm 1 Naïve algorithm.

Inputs:
D: set of location and time information of social media activities of social media users
min_density: a location density threshold
min_visit: a visit lifetime threshold
min_UP: a user prevalence threshold
L: set of extracted and labeled locations
Output: A set of Socially Important Locations, *SIL*

```
1  initialization: allLocations = null
2  for each user social media activity u in D
3      SILU = null
4      for each location l in L[u]
5          location_density = calculate-location-density(l, u)
6          visit_lifetime = calculate-visit-lifetime(l, u)
7          if location_density ≥ min_density and visit_lifetime ≥ min_visit
8              SILU ← l
9          end if
10     end for
11     allLocations ← SILU
12 end for
13 locations = calculate-UP(allLocations)
14 SIL = extract-socially-important-locations(locations, min_UP)
15 return SIL
```

In the algorithm, step 1 initializes the parameters, steps 2 through 12 give an iterative process to discover user-level socially important locations, and step 13 and 14 are used for discovering group-level socially important locations. The steps between 2 and 12 run for each user social media activity and steps between 4 and 10 run for each location of users. In step 5, location density of location l for social media user u is calculated in *calculate-location-density* function and in step 6, visit lifetime of location l for social media user u is calculated in *calculate-visit-lifetime*. In steps between 7 and 9 frequent and time pervalent locations are discovered that satisfy *min_density* and *min_visit* thresholds. In step

11, frequent and time pervasive locations of social media user *u* are added into location set. In step 13, in *calculate-UP* function, user prevalence values of locations are calculated. In step 14, in *extract-socially-important-locations* function, socially important locations that satisfy *min_UP* threshold are discovered. Finally, algorithm outputs socially important locations *SIL* in step 15.

4.2. SS-ILM algorithm

Naïve algorithm is a straight forward way to discover socially important locations of social media users. It discovers frequent locations and time pervasive locations separately which leads to unnecessary computational cost. To eliminate the limitation of naïve algorithm, we propose SocioSpatially Important Locations Mining (SS-ILM) algorithm to discover socially important locations. The algorithm, first, discovers frequent locations that satisfy the *min_density* threshold and then uses these frequent locations to discover time pervasive locations that satisfy the *min_visit* threshold. In SS-ILM algorithm, the pruning procedure is applied as early as possible to discover frequent and time pervasive locations and so unnecessary calculations are avoided. The pseudo code of the SS-ILM algorithm is given in Algorithm 2.

Algorithm 2 SS-ILM algorithm.

Inputs:
D: set of location and time information of social media activities of social media users
min_density: a location density threshold
min_visit: a visit lifetime threshold
min_UP: a user prevalence threshold
L: set of extracted and labeled locations
Output: A set of Socially Important Locations, SIL

```
1  initialization: allLocations = null
2  for each user social media activity u in D
3      SILU = null
4      for each location l in L[u]
5          location_density = calculate-location-density(l, u)
6          if location_density ≥ min_density
7              visit_lifetime = calculate-visit-lifetime(l, u)
8              if visit_lifetime ≥ min_visit
9                  SILU ← l
10             end if
11         end if
12     end for
13     allLocations ← SILU
14 end for
15 locations = calculate-UP(allLocations)
16 SIL = extract-socially-important-locations(locations, min_UP)
17 return SIL
```

In Algorithm 2, location density of location *l* for social media user *u* is calculated in step 5. In step 6, location *l* checked if it satisfies *min_density* threshold and so it is frequent location. If it is a frequent location, steps between 7 and 10 run to check if it is a time pervasive location. In step 7, the visit lifetime of location *l* for the social media user *u* is calculated and, in step 8, it is checked if it is a time pervasive location. If so, it is recorded into the set of socially important locations for user, *SILU*. Avoiding unnecessary calculation of visit lifetime value provides speedup to SS-ILM algorithm.

4.3. Execution trace

This section presents execution trace of SS-ILM algorithm. User-level socially important locations are discovered based on a sample dataset of a single user and group-level socially important locations are discovered based on a sample dataset of 5 users. Fig. 2 shows the number of social media activity of a user (e.g., twitter or facebook activities) for 20 days in different locations (e.g., L0, L1, ... , L8). Days are shown at the bottom of the figure and the locations are shown at the left hand side of the figure. The social media activity of the user occurs in 9 different locations (locations L0... L8) which are shown in Fig. 3. As can be seen in Fig. 2, in Day 1, one social media activity (e.g., sending a tweet at a location) was recorded at location L0 and in Day 8, three social media activities (e.g., sending three tweets) were recorded at location L4 by the user.

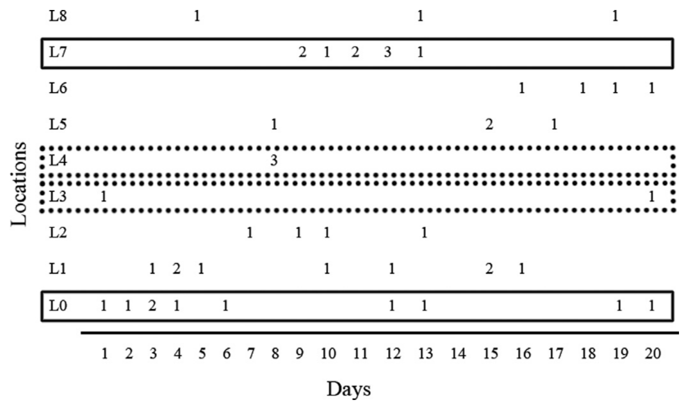

Fig. 2. Social media activity of a user for 20 days in different locations. In the figure each number shows the number of social media activity of the user in a specific day and at a specific location.

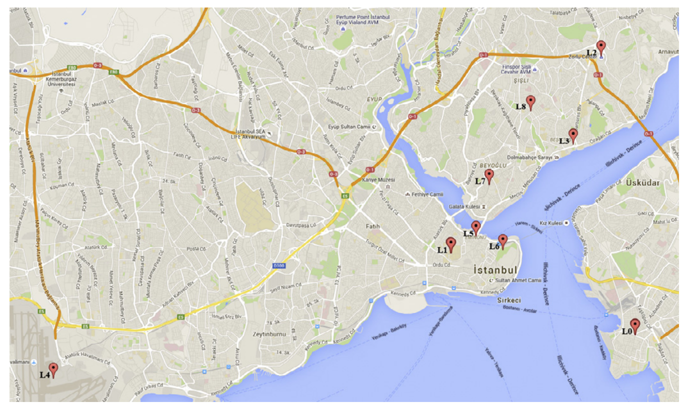

Fig. 3. Location distribution for one user on a map.

As can be seen in Fig. 2, most of the social media activity in 20 days occurred at locations L0 and L7 which are indicated with straight lines in the figure. L0 is an example of time pervasive and important location. The user has social media activities in location L0 for almost half of total days (social media lifetime) and, in contrast to this, the number of social media activity in each day is low (i.e., one or two). L7 is an example of frequent and important location. In location L7, the user has social media activity for 5 days and the number of social media activity in each day is high (i.e., two and three). As can be seen in Fig. 2, location L0 was visited many times with a low social media activity rate in each day and location L7 was visited in a short period of time with a high social media activity rate in each day.

There are two critical locations in Fig. 2 which are indicated with dotted lines, such as, L3 and L4. In location L3, the user has two social media activities for a period of 20 days. In location L4, three social media activities occurred only in one day. These critical locations are pruned using *min_density* and *min_visit* threshold

Table 3
Location density and visit lifetime values and discovered/pruned locations for User 1 (Disc. = Discovered, Prun. = Pruned).

	L0	L1	L2	L3	L4	L5	L6	L7	L8
Loc density	10/48	9/48	4/48	2/48	3/48	4/48	4/48	9/48	3/48
Is frequent	Yes	Yes	Yes	No	No	Yes	Yes	Yes	No
Visit lifetime	20/20	14/20	7/20	–	–	10/20	5/20	5/20	–
Is time pervasive	Yes	Yes	Yes	–	–	Yes	Yes	Yes	–
Discover/Prune	Disc.	Disc.	Disc.	Prun.	Prun.	Disc.	Disc.	Disc.	Prun.

Table 4
Daily visited locations for 5 users for 20 days (Daily distinct locations are separated with comma).

Day	User 1	User 2	User 3	User 4	User 5
1	L0, L3	L0	L10	L0	L7
2	L0	L1	L0, L1, L3	L2, L10	L5
3	L0, L0, L1	L2, L3	L2, L6	L2, L2	L1, L11
4	L1, L0, L1	L6	L0, L7	L4	L0
5	L1, L8	L0, L1	L1, L9	L6	L2
6	L0	L6, L9	L6	L4, L4	L7, L11
7	L2	L10	L0, L3	L10	L0, L5
8	L4, L4, L5, L4	L2, L6	L4, L12	L4, L6	L9
9	L2, L7, L7	L0, L3	L6	–	L2
10	L1, L7, L2	L0, L1	L1, L3	L0, L4	L9
11	L7, L7	L0, L0, L2	L2, L12	L7, L10	L1, L5
12	L7, L0, L7, L7, L1	L7	L0, L4	L0, L12	L7
13	L0, L8, L2, L7	L2, L3	L4, L12	L10	L1, L1
14	–	L0	L1, L2, L3	L6	L1, L1
15	L5, L1, L1, L5	L1, L6	L6	L10	L1, L11
16	L1, L6	L6, L12	L1, L1, L4	L6	L5
17	L5	L9	L7	L6, L6, L8	L11
18	L6	L1, L3	L9	L8	L6, L7
19	L6, L8, L0	–	L10	L8	–
20	L0, L3, L6	L7, L10	L10, L10	L8	L11

Table 5
Discovered and pruned locations for 5 users, and UP values and SIL for sample dataset (Disc. = Discovered, Prun. = Pruned).

Location	User 1	User 2	User 3	User 4	User 5	UP	SIL
L0	Disc.	Disc.	Disc.	Disc.	Prun.	4/5 = 0.8	Yes
L1	Disc.	Disc.	Disc.	Prun.	Disc.	4/5 = 0.8	Yes
L2	Disc.	Disc.	Disc.	Prun.	Prun.	3/5 = 0.6	Yes
L3	Prun.	Disc.	Disc.	Prun.	Prun.	2/5 = 0.4	No
L4	Prun.	Prun.	Disc.	Disc.	Prun.	2/5 = 0.4	No
L5	Disc.	Prun.	Prun.	Prun.	Disc.	2/5 = 0.4	No
L6	Disc.	Disc.	Disc.	Disc.	Prun.	4/5 = 0.8	Yes
L7	Disc.	Prun.	Prun.	Prun.	Disc.	2/5 = 0.4	No
L8	Prun.	Prun.	Prun.	Disc.	Prun.	1/5 = 0.2	No
L9	Prun.	Prun.	Prun.	Prun.	Prun.	0/5 = 0.0	No
L10	Prun.	Prun.	Disc.	Disc.	Prun.	2/5 = 0.4	No
L11	Prun.	Prun.	Prun.	Prun.	Disc.	1/5 = 0.2	No
L12	Prun.	Prun.	Disc.	Prun.	Prun.	1/5 = 0.2	No

values. For this execution trace, *min_density* and *min_visit* are set to 4/48 and 4/20, respectively. Table 3 shows discovered and pruned locations for the user shown in Fig. 2 after these thresholds are applied. As can be seen in the table, L3 and L4 are pruned because they do not satisfy the thresholds of *min_density* and *min_visit*, respectively.

Table 4 presents the locations of social media activities of 5 users for 20 days. A social media user may conduct social media activity at more than one location each day, thus these locations are separated with comma. Table 5 presents discovered locations for each user and User Prevalence (UP) values of each locations. In this execution trace, minimum user prevalence (*min_UP*) threshold was set to 3/5. The locations that satisfy *min_UP* threshold are discovered as socially important locations. As can be seen in Table 5, locations of L0, L1, L2, and L6 satisfy the given thresholds and so the algorithm outputs these locations as socially important locations.

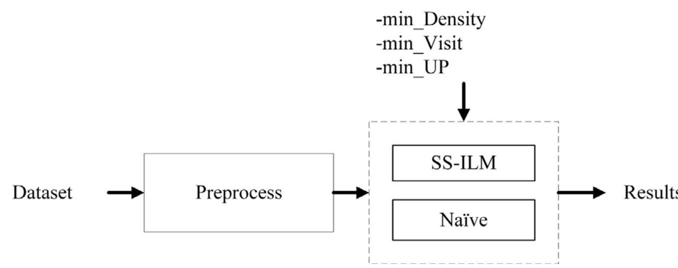

Fig. 4. Experimental setup.

5. Experimental evaluation

In this section, we present our experimental evaluations. In the experiments, a real-world Twitter dataset was used as social media activity dataset. The dataset preparation is explained in Sections 5.1 and 5.2. Fig. 4 shows the steps of experimental setup for this study. First, dataset is given to the system and then a preprocessing step is performed (Section 5.2). At the preprocessing step, data cleaning (Section 5.2.1), user selection (Section 5.2.2), temporal overweighting prevention (Section 5.2.3), and location labeling (Section 5.2.4) processes are performed. After preprocessing step, SS-ILM and naïve algorithms are run onto the dataset. The algorithms require the thresholds of *min_density, min_visit*, and *min_UP* to discover socially important locations. Location density threshold *min_density* is used for discovering user-level frequent locations, visit lifetime threshold *min_visit* is used for discovering user-level time pervasive locations, and user prevalence threshold *min_UP* is used for discovering group level socially important locations.

We evaluated the behaviors of the SS-ILM and the naïve algorithm to answer the following questions;

- What is the effect of number of social media users (dataset size) on the performances of algorithms?
- What is the effect of *min_density* threshold?
- What is the effect of *min_visit* threshold?
- Which locations are socially important for city of Istanbul, Turkey residents?
- How successful is our proposed algorithms based on related studies?

The experiments were conducted on an Intel Core i7 CPU with 3.40 GHz, and 8 GB of RAM.

5.1. Dataset

Social media networking sites, such as Twitter, provide data collecting Application Programming Interfaces (APIs) for developers. In this study, we used Twitter dataset that have geographical information as social media activity dataset. To collect data from Twitter, REST API and Streaming API were used (Twitter, 2017). In

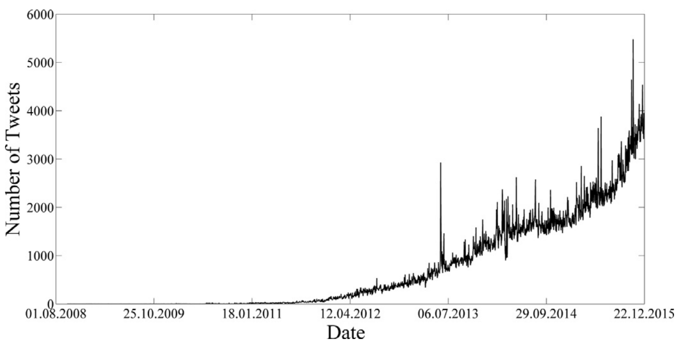

Fig. 5. Temporal distribution of dataset.

addition to these, Twitter4j open source Java library (Yamamoto, 2017) was used for performing queries and getting results from Twitter APIs. Streaming API provides geographical boundary search on streaming tweets. To create the dataset, first, Istanbul, Turkey based geographical search is performed and then users were collected. Approximately 2500 users were collected in this step. Then, REST API was used to gather all tweets of these users. Three parameters were collected for each tweet, such as, date/time, latitude, and longitude. The dataset belongs to the period of September 2008 and December 2015.

5.2. Preprocessing

In this section, data cleaning, user selection, temporal overweighting prevention, and location labeling procedures that are used in this study are explained.

5.2.1. Data cleaning

In the experiments, we used the data from active Twitter users. In this study, active Twitter users are defined as users that send tweets no less than 50 between 2008 and 2015. If the number of tweets of a user is low, then the user is either passive or a new user. However, active user definition may change based on user preferences or applications. In the dataset, a proportion of Twitter users are spam users. To avoid spam users, we used two criteria; followers count and follower/friends ratio. If a users' followers count is less than 10 and follower/friends ratio is below 0.1, then this user is labeled as spam user and so that user was not included in the dataset. The values for these parameters are assigned according to many spam user detection literature and detailed information can be found in Benevenuto, Magno, Rodrigues, and Almeida (2010) and Zheng, Zeng, Chen, Yu, and Rong (2015).

5.2.2. User selection

In this study, the main focus is to discover socially important locations of Twitter users that send tweets in Istanbul, Turkey (Istanbul residents). Because users are collected by Streaming API and this API do not return historical data, we could not guarantee that users have sufficient tweets inside Istanbul at the data collection step. To assign a user as Istanbul resident, latitude and longitudes of his/her tweets are differentiated from Istanbul's central latitude and longitude. If more than half of the tweets of the user are within Istanbul borders (i.e. 40 km from Istanbul center), then this user is selected as Istanbul resident.

By applying this process 1000 users were determined as real and active Istanbul Twitter users for the experiments. 1.886.065 tweets were present in the dataset and the dataset belongs to the period of September 2008 and December 2015. Temporal distribution of the dataset is shown in Fig. 5. As can be seen in the figure, most of the tweets are belong to the period of 2013–2015. The main reason for this is, Twitter has gained importance among Turkey social media users after 2012.

5.2.3. Preventing temporal overweighting

When we analyzed the dataset, we realized that users may tweet (i.e., conduct social media activity) more than once at a location at the same time. If this behavior becomes common, then a location might have more presence than its correct presence because the user was at that place once but tweeted several times. We defined this problem as temporal overweighting of a location. This problem is sometimes unintentional, such as a user has a conversation via tweeting to his/her friends and tweets several times within a short time span. To prevent temporal overweighting of a location, we defined a threshold, which is 60 minutes. If a user tweets more than once at the same location within 60 minutes, then we assume that this location information is not new and these tweets should be counted as once. With this approach, temporal overweighting of a location is prevented. However different approaches/criteria can also be applied to prevent temporal overweighting.

5.2.4. Location labeling method

The Twitter APIs provide accurate latitude-longitude pairs of user tweets. This approach is beneficial for getting fine-grained results, but also a problem for location labeling. For example, a shopping mall or a stadium might be located in 1 km^2 area but we could define many distinct locations for this shopping mall or stadium because the accurate latitude and longitude pairs do not match. To overcome this problem, we defined a threshold for being same location for different latitude-longitude pairs. As used before in Cho et al. (2011) and Pavan et al. (2015), we defined this threshold as 100 m. If two locations are closer than 100 m, same labels are assigned to these two locations.

5.3. Experimental results

In this section, first, we present the experiments to evaluate the performances of naïve and SS-ILM algorithms. After, we present discovered socially important locations of social media users in Istanbul. Finally, we compared our algorithms with a related study.

5.3.1. Effect of number of users

In this experiment, we evaluated the effect of number of users on the runtimes of both algorithms. *min_density, min_visit* and *min_UP* were set at 0.01, 0.05 and 0.01, respectively. Number of users was increased by 200 from 200 to 1000. Fig. 6 shows the effect of the number of users for both algorithms. As can be seen in Fig. 6, both algorithms are sensitive to the number of users and so their execution times increase as the number of users increase. The result shows that SS-ILM algorithm is more computationally efficient than naïve algorithm.

5.3.2. Effect of minimum location density threshold

In this experiment, we evaluated the effect of minimum location density (*min_density*) threshold on runtimes of both algorithms. The number of users, *min_visit* and *min_UP* were set at 1000, 0.05, and 0.01, respectively. *min_density* value was increased by 0.005 from 0.005 to 0.05. Fig. 7 shows the runtimes of both algorithms as *min_density* threshold increases. Naïve algorithm discovers frequent locations and time pervasive locations separately. In contrast, SS-ILM algorithm discovers frequent locations and then out of these frequent locations, it discovers time pervasive locations. Because of this reason, as can be seen in Fig. 7, the execution time of SS-ILM algorithm outperforms naïve algorithm due to the early pruning strategy of frequent locations. The SS-ILM algorithm consumes 25% less time than naïve algorithm.

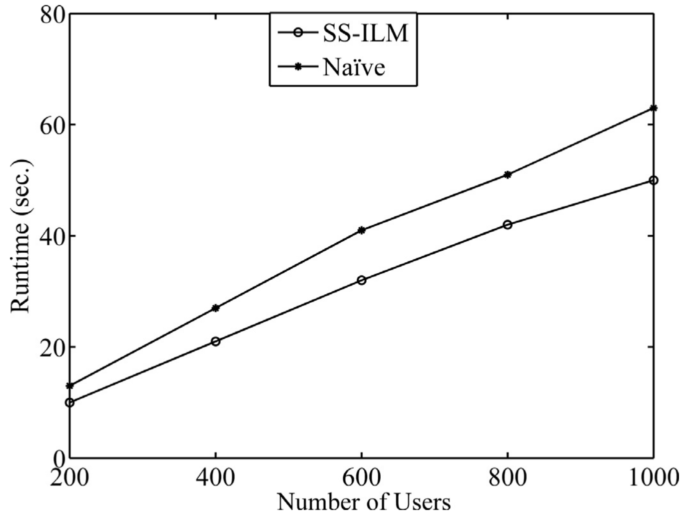

Fig. 6. Effect of number of users.

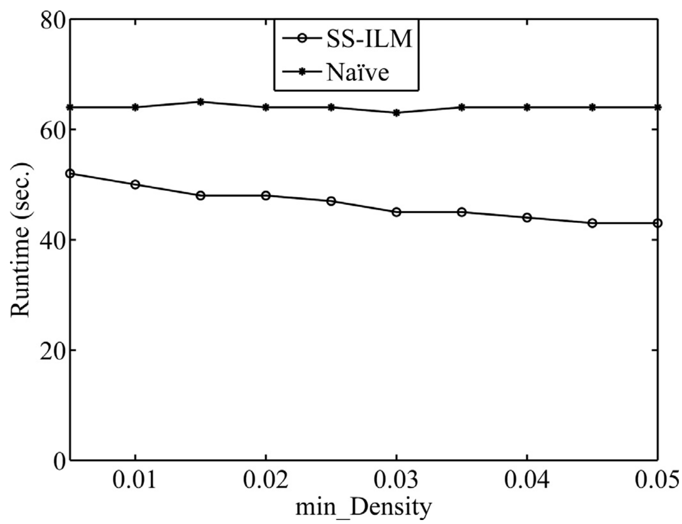

Fig. 7. Effect of minimum location density (*min_density*) threshold.

5.3.3. Effect of minimum visit lifetime threshold

In this experiment, we evaluated the effect of minimum visit lifetime (*min_visit*) threshold on runtimes of both algorithms. The number of users, *min_density* and *min_UP* were set at 1000, 0.01, and 0.01, respectively. *min_visit* value was increased by 0.01 from 0.01 to 0.1. Fig. 8 shows runtimes of both algorithms as the value of *min_visit* threshold increases. The runtime trends of both algorithms are the same since *min_visit* based pruning is applied at the final step of user-level important locations discovery. The SS-ILM algorithm outperforms naïve algorithm due to the early pruning strategy of frequent locations.

5.3.4. Discovery of socially important locations in istanbul

In this section, we discuss the discovery of socially important locations in Istanbul. The number of users is 1000, and *min_density, min_visit* and *min_UP* were set at 0.01, 0.05 and 0.01, respectively. We presented top 10 socially important locations for the sake of simplicity and presentation. The results are shown in a map in Fig. 9 and also Table 6 shows location coordinates and descriptions. The location descriptions are manually extracted, none of the named entity recognition systems are used since they are out of the scope of this study.

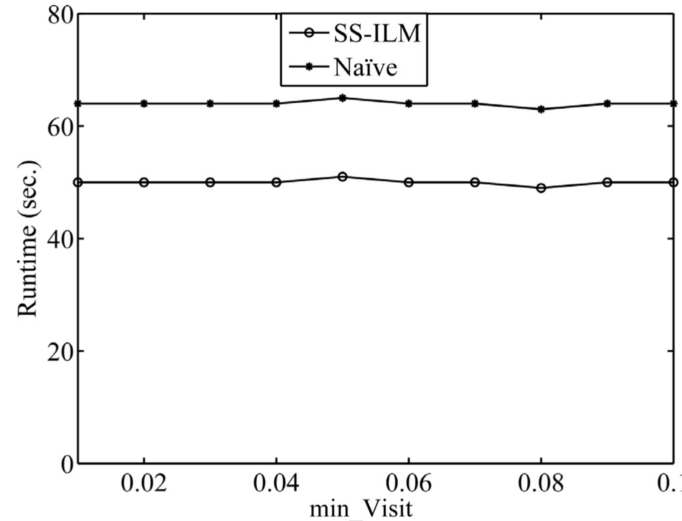

Fig. 8. Effect of minimum visit lifetime (*min_visit*) threshold.

Fig. 9. Socially Important locations of 1000 users in Istanbul.

Table 6
Socially important locations discovered by SS-ILM algorithm.

Order	Discovered socially important locations
1	Kadikoy Bull Statue
2	Sisli Cevahir Shopping Center
3	Bagdat St., Goztepe
4	Ataturk Airport
5	Istiklal Avenue
6	15 Temmuz Sehitler Bridge
7	Besiktas Culture Center
8	Ragip Gumuspala St., Eminonu
9	Istanbul Aydin Uni.
10	Vali Konagi St., Tesvikiye St., Abdi Ipekci St.

In Table 6, discovered 10 socially important locations are listed. The first location is the central of Kadikoy district, and it is a socially important location for Istanbul residents, in general. The second location is a popular shopping center in Sisli district. The third location is Bagdat Street in Goztepe district. In Bagdat street, there are many restaurants and shopping areas. The fourth location is Ataturk Airport, which hosts huge amounts of passengers daily. The fifth location is a popular avenue in Beyoglu district. The sixth location is one of the bridges that connects Anadolu and Europe banks of Istanbul. The seventh location is a culture center which is visited by many audiences. The eighth location is a street which goes along the Istanbul Bosphorus. The ninth location is a univer-

Table 7
Socially important locations discovered by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011).

Order	Discovered socially important locations
1	Kadikoy Bull Statue
2	Ataturk Airport
3	Bagdat St., Goztepe
4	Sisli Cevahir Shopping Center
5	15 Temmuz Sehitler Bridge
6	Besiktas Culture Center
7	Buyukdere St., Sisli
8	Vali Konagi St., Tesvikiye St., Abdi Ipekci St.
9	Istiklal Avenue
10	Atasehir Avenue, Atasehir

sity campus. Finally, the tenth location is a popular location which is in the intersection of districts of Beyoglu, Sisli, and Besiktas.

According to the results, we can conclude that our proposed algorithm can successfully discover socially important locations of Istanbul (resident) Twitter users. However, we see that some popular and touristic locations such as Sultanahmet Square, or Topkapi Palace were not discovered by the algorithms. The reason is related to the characteristics of the dataset used. The users selected for this study did not send significant amount of tweets in these locations. We should also mention that we included the residents of Istanbul, not the tourists in our analysis.

5.3.5. Comparison with related studies

As discussed in the related work, our study falls into group-level interesting/important location discovery. To the best of our knowledge, group-level interesting/important location discovery studies were conducted by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) and Zheng et al. (2009).

The study conducted by Zheng et al. (2009) employs user ratings to determine important locations and since there is no user ratings about locations in our dataset, we excluded this study to compare it with the proposed algorithm of SS-ILM.

We compared our study with the study conducted by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) based on three criteria; discovered top 10 important locations, number of discovered results, and execution time. The first and the second criteria reveals which algorithm discovers more reasonable results. The third criterion reveals which algorithm consumes less time to extract important locations.

When we compare the discovered top 10 important locations of both algorithms, we realized that the results are different. The top 10 important locations discovered by SS-ILM algorithm is presented in Table 6 and top 10 important locations discovered by the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) is presented in Table 7. When the results were compared, it can be seen that the order of two locations, such as Kadikoy Bull Statue and Bagdat St., Goztepe, are same in both algorithms. The locations of Ragip Gumuspala St., Eminonu and Istanbul Aydin University were listed in top 10 list of SS-ILM algorithm and were not listed in the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011). Both algorithms also discovered same locations, such as, Ataturk Airport, Istiklal Avenue, and 15 Temmuz Sehitler Bridge, however their ordering are different in both top 10 lists. The reason of the ordering difference between algorithms caused Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) orders locations based on their visit frequencies, while SS-ILM algorithm orders locations based on their user prevalence. The problem of ordering locations with frequency value raises when a proportion of the users visit a location with a very high frequency, while other users never visit. Thus true ordering of locations for a user group cannot be performed based on frequency. These results showed that SS-ILM algorithm is more successful at both discovering important locations and ordering them.

In terms of number of discovered results, the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) discovered 179 locations as important for 1000 users. The results of this algorithm depend on the minimum number of user threshold (ThresCount). In this experiment minimum number of users (ThresCount) is assigned as 80, which is the same proportion used in study conducted by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011). SS-ILM algorithm discovered 27 locations as socially important locations. The results show that SS-ILM algorithm is more efficient on pruning spurious locations.

In terms of execution time, the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) consumes 26 seconds to extract results for 1000 users, while our proposed SS-ILM algorithm consumes 52-44 seconds. In terms of execution time, our experiments show that the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) consumes less time than the SS-ILM algorithm.

As a result, SS-ILM algorithm discovers socially important locations by eliminating most of the spurious locations. So it is more successful at extracting reasonable and appropriate number of socially important locations than the study conducted by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011). However, SS-ILM algorithm takes more time than the algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011). This is due to the differences in both approaches. The algorithm proposed by Khetarpaul, Chauhan, Gupta, Subramaniam, and Nambiar (2011) takes into account the visit frequencies of the locations, in contrast, SS-ILM algorithm discovers locations based on their user prevalence.

6. Contributions and limitations

In this section, the contributions of this study and limitations are discussed.

6.1. Contributions

The contributions of this paper can be listed as follows:

- The interest measures of location density, visit lifetime and user prevalence are proposed to quantify socially important locations.
- A new algorithm, which is called SocioSpatially Important Locations Mining algorithm, is proposed for discovering socially important locations efficiently.
- Proposed algorithm is experimentally evaluated using real dataset of Istanbul, Turkey Twitter users.
- Proposed algorithm is compared with a related study.

6.2. Limitations

This paper focused on developing solutions to socially important locations mining problem over social media datasets. The following subjects are beyond the scope of this paper:

- Temporal analysis of socio-spatially important locations,
- Automatic labeling of discovered locations, such as named entity recognition system,
- New approaches to location labeling problem.

7. Conclusions and future work

In this study, we developed techniques to discover socially important locations of social media users by analyzing their historical social media activities. Discovering socially important locations are essential for analyzing a user group's spatial preferences. We proposed interest measures of location density, visit lifetime, and user prevalence to quantify socially important locations and a novel SocioSpatially Important Locations Mining (SS-ILM) algorithm to discover socially important locations. Proposed algorithms can be easily adopted to any social media activity dataset which has geographical and temporal information. The performance of SS-ILM algorithm is compared with that of a naïve algorithm using real dataset of Istanbul Twitter users. The results showed that the proposed algorithm outperforms the naïve algorithm. Also proposed algorithm is compared with a related study in context of runtime and quality and quantity of results. The comparison showed that although SS-ILM algorithm is less efficient at runtime, it is more successful at discovering and ordering qualified and appropriate number of important locations.

In the future, we are planning to extent this study with temporal aspects and to generate results for spatio-temporally social important locations from social media activity datasets. We are also planning to develop new algorithms to discover socially important locations more efficiently. A potential application area of this study could be to discover socially important locations of a dedicated social media user group, such as a brand's followers or a friend group.

References

Bao, J., Zheng, Y., & Mokbel, M. F. (2012). Location-based and preference-aware recommendation using sparse geo-social networking data. In *Proceedings of the 20th international conference on advances in geographic information systems. SIGSPATIAL '12* (pp. 199–208). New York, NY, USA: ACM. doi:10.1145/2424321.2424348.

Benevenuto, F., Magno, G., Rodrigues, T., & Almeida, V. (2010). Detecting spammers on twitter. *Collaboration, electronic messaging, anti-abuse and spam conference (ceas)*.

Budak, C., Georgiou, T., Agrawal, D., & El Abbadi, A. (2013). Geoscope: Online detection of geo-correlated information trends in social networks. *Proceedings of the VLDB Endowment, 7*(4), 229–240. doi:10.14778/2732240.2732242.

Cai, G., Hio, C., Bermingham, L., Lee, K., & Lee, I. (2014). Sequential pattern mining of geo-tagged photos with an arbitrary regions-of-interest detection method. *Expert Systems with Applications, 41*(7), 3514–3526. http://dx.doi.org/10.1016/j.eswa.2013.10.057.

Cao, X., Cong, G., & Jensen, C. S. (2010). Mining significant semantic locations from gps data. *Proceedings of the VLDB Endowment, 3*(1–2), 1009–1020. doi:10.14778/1920841.1920968.

Celik, M. (2015). Partial spatio-temporal co-occurrence pattern mining. *Knowledge and Information Systems, 44*(1), 27–49. doi:10.1007/s10115-014-0750-2.

Celik, M., Shekhar, S., Rogers, J. P., & Shine, J. A. (2008). Mixed-drove spatiotemporal co-occurrence pattern mining. *IEEE Transactions on Knowledge and Data Engineering, 20*(10), 1322–1335. doi:10.1109/TKDE.2008.97.

Cenamor, I., de la Rosa, T., Nez, S., & Borrajo, D. (2017). Planning for tourism routes using social networks. *Expert Systems with Applications, 69*, 1–9. http://dx.doi.org/10.1016/j.eswa.2016.10.030.

Chen, C.-C., Chiang, M.-F., & Peng, W.-C. (2016). Mining and clustering mobility evolution patterns from social media for urban informatics. *Knowledge and Information Systems, 47*(2), 381–403. doi:10.1007/s10115-015-0853-4.

Chen, Y., Jiang, K., Zheng, Y., Li, C., & Yu, N. (2009). Trajectory simplification method for location-based social networking services. In *Proceedings of the 2009 international workshop on location based social networks. LBSN '09* (pp. 33–40). New York, NY, USA: ACM. doi:10.1145/1629890.1629898.

Chen, Z., Shen, H. T., & Zhou, X. (2011). Discovering popular routes from trajectories. In *Proceedings of the 2011 IEEE 27th international conference on data engineering. ICDE '11* (pp. 900–911). Washington, DC, USA: IEEE Computer Society. doi:10.1109/ICDE.2011.5767890.

Cho, E., Myers, S. A., & 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. KDD '11* (pp. 1082–1090). New York, NY, USA: ACM. doi:10.1145/2020408.2020579.

Cranshaw, J., Schwartz, R., Hong, J., & Sadeh, N. (2012). The livehoods project: Utilizing social media to understand the dynamics of a city. In *The international AAAI conference on weblogs and social media (ICWSM-12)* (pp. 58–65).

Dadaser-Celik, F., Celik, M., & Dokuz, A. S. (2012). Associations between stream flow and climatic variables at Kzlrmak river basin in Turkey. *Global NEST Journal, 14*(3), 354–361.

Doytsher, Y., Galon, B., & Kanza, Y. (2011). Storing routes in socio-spatial networks and supporting social-based route recommendation. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. LBSN '11* (pp. 49–56). New York, NY, USA: ACM. doi:10.1145/2063212.2063219.

Feng, W., Zhang, C., Zhang, W., Han, J., Wang, J., Aggarwal, C., et al. (2015). *Streamcube: Hierarchical spatio-temporal hashtag clustering for event exploration over the twitter stream. 2015 IEEE 31st international conference on data engineering* (pp. 1561–1572). doi:10.1109/ICDE.2015.7113425.

Ferrari, L., Rosi, A., Mamei, M., & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. LBSN '11* (pp. 9–16). New York, NY, USA: ACM. doi:10.1145/2063212.2063226.

Gaete-Villegas, J., Cha, M., Lee, D., & Ko, I.-Y. (2012). Tramsnet: A mobile social network application for tourism. In *Proceedings of the 2012 acm conference on ubiquitous computing. UbiComp '12* (pp. 1004–1011). ACM. doi:10.1145/2370216.2370433.

Guy, I., Zwerdling, N., Ronen, I., Carmel, D., & Uziel, E. (2010). Social media recommendation based on people and tags. In *Proceedings of the 33rd international ACM SIGIR conference on research and development in information retrieval. SIGIR '10* (pp. 194–201). ACM. doi:10.1145/1835449.1835484.

Hannon, J., Bennett, M., & Smyth, B. (2010). Recommending twitter users to follow using content and collaborative filtering approaches. In *Proceedings of the fourth ACM conference on recommender systems. RecSys '10* (pp. 199–206). ACM. doi:10.1145/1864708.1864746.

Hu, T., & Sung, S. Y. (2005). Clustering spatial data with a hybrid em approach. *Pattern Analysis and Applications, 8*(1), 139–148. doi:10.1007/s10044-005-0251-8.

Isaacman, S., Becker, R., Cáceres, R., Kobourov, S., Martonosi, M., Rowland, J., et al. (2011). Identifying important places in people's lives from cellular network data. In K. Lyons, J. Hightower, & E. M. Huang (Eds.), *Pervasive computing: 9th International conference, Pervasive 2011, San Francisco, USA, June 12–15, 2011. Proceedings* (pp. 133–151). Berlin, Heidelberg: Springer. doi:10.1007/978-3-642-21726-5_9.

Kefalas, P., & Symeonidis, P. (2015). Recommending friends and locations over a heterogeneous spatio-temporal graph. In L. Bellatreche, & Y. Manolopoulos (Eds.), *Model and data engineering: 5th International conference, MEDI 2015, Rhodes, Greece, September 26–28, 2015, proceedings* (pp. 271–284)). Springer International Publishing. doi:10.1007/978-3-319-23781-7_22.

Kefalas, P., Symeonidis, P., & Manolopoulos, Y. (2016). A graph-based taxonomy of recommendation algorithms and systems in lbsns. *IEEE Transactions on Knowledge and Data Engineering, 28*(3), 604–622. doi:10.1109/TKDE.2015.2496344.

Khetarpaul, S., Chauhan, R., Gupta, S. K., Subramaniam, L. V., & Nambiar, U. (2011). Mining gps data to determine interesting locations. In *Proceedings of the 8th international workshop on information integration on the web: conjunction with www 2011*. In *IIWeb '11* (pp. 1–6). New York, NY, USA: ACM. doi:10.1145/1982624.1982632.

Levandoski, J. J., Sarwat, M., Eldawy, A., & Mokbel, M. F. (2012). Lars: A location-aware recommender system. In *Proceedings of the 2012 IEEE 28th international conference on data engineering. ICDE '12* (pp. 450–461). Washington, DC, USA: IEEE Computer Society. doi:10.1109/ICDE.2012.54.

Liu, B., & Xiong, H. (2013). Point-of-interest recommendation in location based social networks with topic and location awareness. In *Proceedings of the 2013 SIAM international conference on data mining* (pp. 396–404)). doi:10.1137/1.9781611972832.44.

Liu, J., Huang, Z., Chen, L., Shen, H. T., & Yan, Z. (2012). Discovering areas of interest with geo-tagged images and check-ins. In *Proceedings of the 20th ACM international conference on multimedia. MM '12* (pp. 589–598). New York, NY, USA: ACM. doi:10.1145/2393347.2393429.

Liu, Y., & Seah, H. S. (2015). Points of interest recommendation from gps trajectories. *International Journal of Geographical Information Science, 29*(6), 953–979. doi:10.1080/13658816.2015.1005094.

Pavan, M., Mizzaro, S., Scagnetto, I., & Beggiato, A. (2015). Finding important locations: A feature-based approach. In *2015 16th IEEE international conference on mobile data management: 1* (pp. 110–115). doi:10.1109/MDM.2015.11.

Sattari, M., Manguoglu, M., Toroslu, I. H., Symeonidis, P., Senkul, P., & Manolopoulos, Y. (2012). Geo-activity recommendations by using improved feature combination. In *Proceedings of the 2012 ACM conference on ubiquitous computing. UbiComp '12* (pp. 996–1003). ACM. doi:10.1145/2370216.2370432.

Shekhar, S., Lu, C.-T., & Zhang, P. (2001). Detecting graph-based spatial outliers: Algorithms and applications (a summary of results). In *Proceedings of the seventh ACM SIGKDD international conference on knowledge discovery and data mining. KDD '01* (pp. 371–376). New York, NY, USA: ACM. doi:10.1145/502512.502567.

Symeonidis, P., Papadimitriou, A., Manolopoulos, Y., Senkul, P., & Toroslu, I. (2011). Geo-social recommendations based on incremental tensor reduction and local path traversal. In *Proceedings of the 3rd ACM SIGSPATIAL international workshop on location-based social networks. LBSN '11* (pp. 89–96). ACM. doi:10.1145/2063212.2063228.

Tung, A. K. H., Hou, J., & Han, J. (2001). Spatial clustering in the presence of obstacles. In *Proceedings 17th international conference on data engineering* (pp. 359–367). doi:10.1109/ICDE.2001.914848.

Twitter (2017). Twitter developers web site. https://dev.twitter.com/.

Vasuki, V., Natarajan, N., Lu, Z., Savas, B., & Dhillon, I. (2011). Scalable affiliation recommendation using auxiliary networks. *ACM Transactions on Intelligent Systems and Technology, 3*(1), 1–20. doi:10.1145/2036264.2036267.

Wei, L.-Y., Zheng, Y., & Peng, W.-C. (2012). Constructing popular routes from uncertain trajectories. In *Proceedings of the 18th ACM SIGKDD international conference on knowledge discovery and data mining. KDD '12* (pp. 195–203). ACM. doi:10.1145/2339530.2339562.

Weiler, M., Schmid, K. A., Mamoulis, N., & Renz, M. (2015). Geo-social co-location mining. In *Second international ACM workshop on managing and mining enriched geo-spatial data. GeoRich'15* (pp. 19–24). New York, NY, USA: ACM. doi:10.1145/2786006.2786010.

Yamamoto, Y. (2017). Twitter4j java library. http://twitter4j.org/en/index.html.
Yin, Z., Cao, L., Han, J., Luo, J., & Huang, T. (2011). Diversified trajectory pattern ranking in geo-tagged social media. In *Proceedings of the 2011 SIAM International Conference on Data Mining* (pp. 980–991)). doi:10.1137/1.9781611972818.84.
Ying, J. J.-C., Lee, W.-C., Weng, T.-C., & Tseng, V. S. (2011). Semantic trajectory mining for location prediction. In *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems. GIS '11* (pp. 34–43). New York, NY, USA: ACM. doi:10.1145/2093973.2093980.
Ying, J. J.-C., Lu, E. H.-C., Kuo, W.-N., & Tseng, V. S. (2012). Urban point-of-interest recommendation by mining user check-in behaviors. In *Proceedings of the ACM SIGKDD international workshop on urban computing. UrbComp '12* (pp. 63–70). New York, NY, USA: ACM. doi:10.1145/2346496.2346507.
Ying, J. J.-C., Lu, E. H.-C., Lee, W.-C., Weng, T.-C., & Tseng, V. S. (2010). Mining user similarity from semantic trajectories. In *Proceedings of the 2nd ACM SIGSPATIAL international workshop on location based social networks. LBSN '10* (pp. 19–26). New York, NY, USA: ACM. doi:10.1145/1867699.1867703.
Yu, W. (2016). Spatial co-location pattern mining for location-based services in road networks. *Expert Systems with Applications, 46*, 324–335. http://dx.doi.org/10.1016/j.eswa.2015.10.010.
Zheng, X., Zeng, Z., Chen, Z., Yu, Y., & Rong, C. (2015). Detecting spammers on social networks. *Neurocomputing, 159*, 27–34. http://dx.doi.org/10.1016/j.neucom.2015.02.047.
Zheng, Y. (2015). Trajectory data mining: An overview. *ACM Transactions on Intelligent Systems and Technology, 6*(3), 1–41. doi:10.1145/2743025.
Zheng, Y., Wang, L., Zhang, R., Xie, X., & Ma, W.-Y. (2008). Geolife: Managing and understanding your past life over maps. In *Proceedings of the ninth international conference on mobile data management. MDM '08* (pp. 211–212). Washington, DC, USA: IEEE Computer Society. doi:10.1109/MDM.2008.20.
Zheng, Y., Xie, X., & Ma, W.-Y. (2010). Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Engineering Bulletin, 33*, 32–39.
Zheng, Y., Zhang, L., Ma, Z., Xie, X., & Ma, W.-Y. (2011). Recommending friends and locations based on individual location history. *ACM Transactions on the Web, 5*(1), 1–44. doi:10.1145/1921591.1921596.
Zheng, Y., Zhang, L., Xie, X., & Ma, W.-Y. (2009). Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th international conference on world wide web*. In *WWW '09* (pp. 791–800). New York, NY, USA: ACM. doi:10.1145/1526709.1526816.
Zhou, C., Bhatnagar, N., Shekhar, S., & Terveen, L. (2007). Mining personally important places from gps tracks. In *2007 IEEE 23rd international conference on data engineering workshop* (pp. 517–526). doi:10.1109/ICDEW.2007.4401037.