Noname manuscript No.
(will be inserted by the editor)

Mining Location based Social Networks to Understand the Citizen's Check-in Patterns

Sonia Khetarpaul

Received: date / Accepted: date

Abstract Analyzing and understanding the movement patterns of the citizen's with in a city, plays an important role in urban and transportation planning. Though many recent research papers focused on mining LBSN services data and performed in-depth analysis of users' mobility patterns and their impact on their social inter-connections and friends. This paper focuses on understanding the Citizen's movement patterns of socially interconnected users in friendship networks, by analyzing their spatial-temporal footprints/check-ins. The aim of this paper is to find the impact of structural patterns hidden in the nodes of a friendship network and external environment changes on the check-in patterns of the users. First, we classify each spatial check-in event based on its cause into either self reinforcing behavior or social influence or external stimulus. Then we mine the collective behavior of the all the users during some special events.

Keywords spatial and temporal data · spatial influence · check-in patterns · Location based social networks

1 Introduction

The emergence of Location Based Social Network (LBSN) services such as - Facebook, Twitter, Flickr, Foursquare, etc, have created massive data sets with better spatial and temporal resolution than ever. These LBSN services have enabled researchers to perform in-depth analysis of users' mobility patterns and their impact on their social inter-connections and friends.

In Location Based Social Network services, users share their locations with their friends that they have visited. Most LBSNs give a unique identity number to each and every distinct location. Typically, a user checks-in to a particular location by using a smart mobile phone or a tablet. This information goes to the LBSN server and location based network services share the information with his friends.

Sonia Khetarpaul
Shiv Nadar University
E-mail: sonia.khetarpaul@snu.edu.in

Usually, the LBSN data consist of two sub-datasets. The primary dataset consists of check-in history of the users, where each check-in information consists of a user-id, location-id, latitude, longitude and the time of check-in. In addition, most LBSN services also provide secondary dataset that describes the underlying social network of the users.

Such collections of spatio-temporal location data comprise a rich source of information and have enabled researchers to study a variety of social behaviors. One particular type of behavior is when an individual visits a location (e.g., a shopping mall) due to the influence of another individual who visited that same location recently. We define this behavior as social spatial influence.

Human mobility patterns and their social interconnections which comes from LBSN services are of great importance in commerce. It contributes significantly to link prediction, targeted advertising, and item recommendation, which are crucial for most companies.

In order to understand the users' social behavior and distinguish check-ins that occur due to social spatial influence from others, it is important to understand the cause of each check-in. So, we categorize each spatial check-in event based on its cause into either self reinforcing behavior or social influence or due to external environmental changes.

The most noticeable application of social spatial influence (also discussed in [38]) is for identifying highly influential users: for target advertising. For example, by giving influential people vouchers/promotions so they can further spread the information to many other people connected to them, for political campaigns by making influential people the seeds of the campaigns, etc.

Social spatial influence also has its own utility when the above-mentioned applications are bounded to a certain geographical area. In such cases, we need to choose the seeds(users) who are closely related to the area and are influential to people in that area. For example, the president of a university is clearly more influential to the students at that university than some general students.

Analyzing the effect of external stimulus can also helpful for the preparation of future events. By analyzing the number of users participated in an event this year, their interest, etc, future arrangements can be made.

For this work, there is a constraint that all the check-ins should be geocoded and timestamped. A geocoded check-in means that each check-in constituted by a pair (Longitude, Latitude) and represents the position of the user at a particular moment of time. The precision of each value in the pair is up to six decimal places or more. A value in decimal degrees to 6 decimal places is accurate to 0.111 meter at the equator.

However, the data collected using LBSNs have their own limitations. Some of the most common limitations associated with the use of LBSNs refer to the lack of consistency in the provision of an acceptable amount of valid geocoded data for each user [34]. Some users share large amount of data whereas others share scarce data. Therefore, the amount of data are largely depends on user interest, availability of appropriate smartphones and internet connections. LBSN data retrieved about specific locations reveal important details about the everyday urban life in those places. LBSN data provide a representative sample of citizen preferences and mobility activities, which may sometimes biased towards users interest [7, 34]. Since users do not share their personal details, therefore, the sample data cannot be rigorously characterised in terms of user profiles. Evidently, some users on Social Networks like Twitter, Facebook, etc, represent their organisations, institutions, businesses, etc, thus, data shared by them is mostly biased in nature.

1.1 Contributions

The paper explores how network and spatial properties and external factors affect the number of check-ins and spatial influences. Using check-ins data from location based social networks and its users' friendship graph, the work shows how network and spatial properties like centrality, network neighborhood overlap, spatial check-ins overlaps, strong ties, etc, effects the check-ins and influential behavior of individuals. The paper also shows that how external environmental changes further impact on this behavior. The novelty of this work is, using intuitions derived from the real-world situations to analyze the different type of user check-ins, and using the existing metrics like degree centrality, closeness, centrality, etc, and new metrics like spatial overlap, spatial movement count and spatial-network centrality to determine various correlations among users' footprints and social networks structural properties.

With the application of different centrality measures and metrics, the paper shows the different type of associations among the social interconnections of users and their spatio-temporal footprints/ check-ins with the help of following three social processes:

- Homophily, or the formation of social ties due to matching individual traits;
- Social Influence or behavior correlation between adjacent users on the network; and
- External Stimulus, or correlation forged due to external influences from the environment.

We analyzed these social processes and correlations to understand the citizens movement patterns.

1.2 Related Work

There is a growing body of literature on LBSN analysis. Current research (related to our work) on LBSNs can be broadly classified into the following three areas: (i) analyzing social network to infer user location, next check-in prediction, and friend prediction, (ii) interrelationships between individual mobility patterns and social inter-connections, and (iii) mining urban mobility patterns. We briefly summarize results in these categories and put our results in context.

Some researchers have analyzed social networks information to infer user location, in [22,40]. Some other work [11,52] focus on prediction of user location exclusively based on the information available from the underlying social network. Spatio-temporal mining algorithms and analysis of spatial, temporal, social, and textual aspects of check-ins, are used to study unnoticed context between people and locations, in [17]. Cho et al. [18] explore human spatio-temporal movement in relation to social ties to analyze future check-ins of a user and effects of distance between users on future check-ins in a typical social network. Chang et al. [14] present a model for predicting future check-ins based on past check-ins, time of check-in, and user demographics. Cao et al. [13] introduced the linkage homophily principle and proposed an iterative framework for heterogeneous information system for prediction of multiple types of links. Chauhan et al. [16] predicted the next place of visit of a user, by calculating the probabilities of visiting different types of places using bank of binary classifiers and Markov models. In their other work, Chauhan et. al. [15] proposed an approach to predict next check-in location, by extracting various features from the historical tweets—for example, personality traits estimated from the past tweets and the actual words mentioned in the tweets.

Noulas et al. [36] analyze user check-in behavior to understand the spatio-temporal mobility patterns of users. In particular they use temporal check-in information to determine user movement patterns for a recommender system. Sudhir et al. [29] study the correlation among shared check-in locations and the structure and type of social ties. For the prediction of future check-ins, Cho et al. [19] focus on modeling the full temporal information of check-ins from a venue-centric perspective. Chang et al. [14] show that an increasing number of shared check-in locations results in increasing friendship probabilities. However, the correlation among social ties and number of shared check-ins is not studied. Pelechrinis et al. [37], using affiliation networks, also draw similar kind of conclusions. Their primary contribution is the interplay between the type of a check-in location and social ties. They show that check-in locations have higher clustering coefficients among friends than non-friends. Aris et al. [9] presented the methodology to measure social correlation and test whether influence is a source of such correlation or not.

The authors in [42] have proposed a memory-efficient on-line algorithm and a data structure called 'Influence Oracle' to determine a top-k set of influential locations. Liao et. al in [30] have proposed a Point of interest recommendation strategy using tensor factorization. They have used LDA topic modelling followed by user-topic-time tensor to determine the users' POIs preferences. The researchers in [43] have combined user activities and spatial features to recommend Point of Interests. The POIs recommendations varies with change in user activity patterns. Qiao et. al in [39] have proposed a framework called UP2VEC that contains heterogeneous LBSN graph. It is a joint representation learning model for users and POIs in Location Based Social Networks. The authors in [47] have determined the spatial distribution of citizens' demand for products and services creates patterns of emerging urban areas of activity.

The authors in [48], have given a general prospective of friendship prediction task in the LBSN domain with balanced depth. In [32], the authors have proposed a location recommendation method that incorporates geographical, categorical, and social preferences with location popularity. They experimentally showed that the Geographical preference generally more important than both categorical and social preferences. Minatel et al. in [35], presented a approach that uses the coarsening stage of a multilevel optimization scheme to build LBSNs by using stay points. The authors in [25], have proposed a new measure of centrality that both considers network and spatial properties, extends the influence maximization problem to the location-based social networks. In [20], the authors have proposed, a next-place prediction framework, which exploits LBSNs data to forecast the next location of an individual based on the observations of her mobility behavior over some period of time and the recent locations visited. The approach integrated frequent pattern mining and feature-based supervised classification to exploit a set of spatio-temporal features characterizing locations and movements among them.

Hasan et al. in [23], presented a survey of a wide variety of event detection methods applied to streaming Twitter data, classifying them according to shared common traits, and discussed different aspects of the subtasks and challenges involved in event detection. In [21], the authors have proposed an online algorithm that incrementally groups tweet streams into clusters. The approach summarizes the examined tweets into the cluster centroid by maintaining a number of textual and temporal features, and discovered groups of interest on particular themes.

In the direction of mining urban mobility patterns, different researchers have mined the traces of taxis and buses [44,8], users' footprints[49,33,10], mobile network traffic [41] and various kind of trajectories data to understand the different aspects of urban cities and citizens. Villatoro et al. in [49], considered citizens as sensors to obtain information about the state of the public transportation network and detected clusters of activity within the urban environment. In [33], authors explored how to use social media data to infer knowledge about urban dynamics and mobility patterns in a urban area. The authors [10] leveraged pervasive mobile sensing to uncover users' mobility patterns and constructed users' persuadability profiles. In [31], Liu et al. used cluster technique to qualitatively analyze the trip relationship between different locations, to correlate relationship between daily travel and land use, and calculated the pendulum value of daily travel. Sun et al. [45], used a multi-way probabilistic factorization model based on the concept of tensor decomposition and probabilistic latent semantic analysis to understand the human mobility patterns. On the other hand, in our research, the main focus is to understand and study the impact of human footprints on their social-network interconnections.

This research is an extension of previous papers presented in [27] and [28]. In previous work [28], we applied STS data model and its algebraic operations to analyze the correlation between the social interconnections of users and their spatio-temporal check-ins. In this work, we show that its not only social ties between a set of users that play a pivotal role, but also the centrality of users and strong network ties in the social network is also important and relate to check-ins and influence patterns of users. The main focus of this work differs from prior approaches, in the sense that we analyze the correlation among nodes check-ins information on the structure network with the help of social processes and relate this with network and spatial properties. We analyzed these social processes and correlations to understand the citizens movement patterns.

We start with defining the notations used:

- **Users:** Users $U = \{u_1, u_2, \ldots, u_n\}$, is a set of n users using social network services. Each user is identified by a unique user-id.

- **Friendship Network:** A friendship network is a graph whose nodes represent real world people/users and whose edges represent friendship among them. Formally, a friendship network is undirected graph (U, F), where U is a set of nodes(users) and $F \subseteq U \times U$ is a set of edges. An edge $(u_1, u_2) \in F$ is called a friendship edge between u_1 and u_2. In our analysis, we are considering friendship network as undirected graph so all the edges are symmetric. This means if u_1 is friend of u_2 then u_2 is also friend of u_1.

- **Check-in:** A Check-in, ch_i, is an action of registering u_i's presence at a location. It gives information about the date-time on which a location is visited by user. And, Check-ins Ch_i, is a set representing all the check-in done by user u_i. And, $Ch = \{Ch_1, Ch_2, \ldots, Ch_n\}$, represents a set of all the check-ins made by all the users.
  Formally, a check-in tuple consists of $<$ *user_id*, *latitude, longitude, timeStamp (date and time)*, *location_id* $>$.
  A check-in is again a kind of GPS point or geometric point of a user, where each spatial location is uniquely identified. Unlike trajectory data, a sequence of check-ins made by a user is not periodic (not within a fixed time interval).

Our datasets consist of friendship network of users and their spatio-temporal check-ins, over the period of time. We extract some important properties from the friendship graph and spatio-temporal footprints generated by users, that can be helpful for deeper understanding of network dynamics, its evolution and check-in patterns. So, first we define the three primary mechanisms to understand the check-ins dynamics. Then we show the correlation among spatial and network properties with the help of social processes like influence, homophily and external stimulus.

2 Analyzing Check-in Dynamics

Three primary factors shown in Figure 1 that can be reason for any check-in to occur are considered in our studies. We describe each in turn below. We are able to distinguish the check-ins based on these factors because of the rich information in the data that includes the user and location visited, along with social network interconnections among users.

Figure 2, unfolds three different kinds of check-in, i.e., self reinforcing behavior, social influence and external stimulus . Unfolding users' visits onto the time axis, we get more informative picture of social influence and type of check-ins. In Figure 2, each horizontal time axis corresponds to one place, and the time of check-in at any location by a user is represented by t_i, $\forall i$. A user checked-in to any location due to one of the following reasons:

2.1 Self-reinforcing Behavior

Analyzing the behavior of individual users reveals strongly predictable patterns. Many users return frequently and repeatedly to the same locations. A user who has recently visited a location is much more likely to visit it again soon. As shown in Figure 2 user u visits $Location1$ (e.g. restaurant) at times t_1, t_2 and t_3. This shows the self-reinforcing behavior of user u. In real life, a user visits his/her favorite restaurant, a shopping mall, or an amusement park repeatedly because of his/her interest or it may be less distant place from his/her home. These types of check-ins come under self-reinforcing behavior. Identifying check-in which occurs due to self reinforcing behavior is very helpful for target advertising.

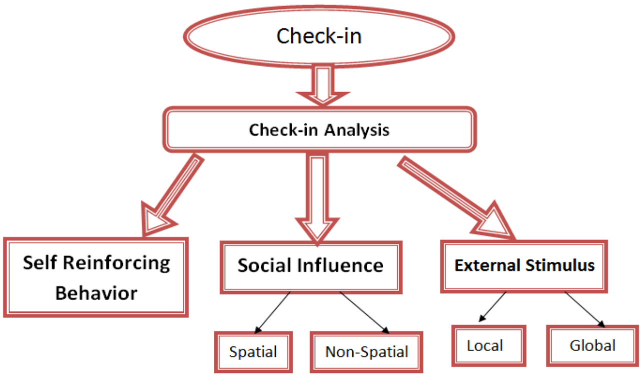


Fig. 1 Types of Check-ins

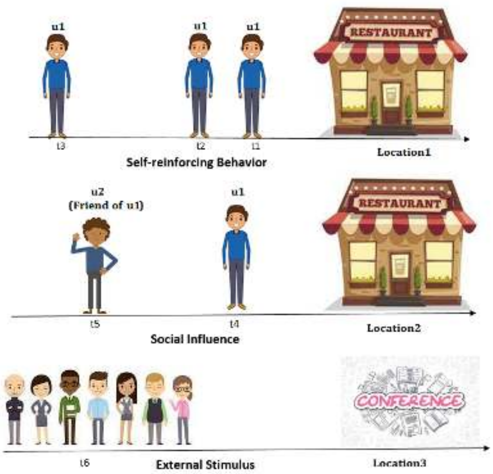


Fig. 2 Visits of Users Along Time Axis

Note that the aim of this study is to mine the *citizen's* check-in patterns within the given spatial region only. Therefore, the self-reinforcing behavior is defined considering the citizens of a given spatial region. This work is not taking care of tourists check-ins patterns.

2.2 Social Influence

Another factor affecting the check-in pattern is social influence. It refers to the phenomenon that the action of individuals can induce their friends to act in a similar way [9]. Influence is unidirectional relationship, when u influences v, it may not mean that v influences u. When u influences v, we say u is the *influencer* and v is the *influenced*. Social influence is important because of its many potential applications and it is also a critical part of the influence maximization problem

[26]. The social influence can be of two types: Social non-spatial influence and social spatial influence.

Social Non-Spatial Influence: When social influence is non-spatial, then spatial constraints do not exist. The social non-spatial influence can occur even though users are located in different cities, or in different countries or continents. If a user performs an action such as clicking the “like” button of a Facebook fan page, then if his/her friend will also perform the same action at a later time then this type of influence is referred as non-spatial influence. We are not analyzing non-spatial influence because of non-availability of data.

Social Spatial Influence: When social influence is spatial then spatial locations of users heavily matters. The chances of influence are low when users are living at distant places and high when users are close. As shown in Figure 1, user $u2$ visits $Location2$ (e.g. restaurant) at time t_5 after the visit of user $u1$ at time t_4. Then, this visit is considered as social spatial influence, if user $u2$ visits this place first time and $u1$ and $u2$ are friends. For example, if a person visits a restaurant or a monument or an amusement park then there are chances that his/her friends living nearby get influenced and also visit that place. These types of check-ins come under social spatial influence. This can be very useful measure of correlation for viral marketing.

2.3 External Stimulus

If a visit is not explained by either of the factors discussed above, we consider it to be caused by some external stimulus. It is the effect of external changes in the environment on the behavior changes in users and friends. External stimulus can be local or global. Local is one, which occurs independently for a particular user under consideration. If a user checked-in to some place because of some external factors, then this type of individual check-ins done because of any sudden changes in a short period of time, is called local-stimulus. In global external stimulus, changes occur due to changes in the external environment and it occurs for many number of users. For example, during festival time like Christmas or during any large scale event like wedding, etc, the changes occur for many users. As shown in Figure 2 at $Location3$ (e.g. conference site) many users check-ins at the same place nearly at the same time t_6, that can be the effect of global stimulus. So, these types check-ins come under external stimulus category.

Algorithm 1 describes the method to categorize each check-in into either self-reinforcing behavior or social influence or external stimulus and label them accordingly. Inputs to the algorithm are all check-ins (Ch), Δtt is threshold time, ch_i represents check-in to be labeled. And algorithm returns labeled check-in ch_i.

After identifying and labeling the check-ins, we analyze and explore the different structural properties of friendship network. We show how these properties relate to social-spatial influence and other types of check-ins. We show the impact social inter-connections of a friendship network on the check-in patterns of users.

Algorithm 1: Check-in-Analyzer(Check-in)

```
Data: check-ins Ch, Δtt, ch_i (check-in to be labeled), friendship network
Result: Labeled ch_i
begin
    ch_i.label = null
    while (ch_j ∈ Ch) AND (ch_j.location−id = ch_i.location-id) do
        if ch_j.timeStamp > ch_i.timeStamp then
            continue
        else
            if (ch_j.timeStamp < ch_i.timeStamp) then
                if (ch_j.user = ch_i.user) then
                    ch_i.label =' self'
                    exit
            else
                if (ch_j.user ≠ ch_i.user) AND (friend(ch_j.user, ch_i.user)) AND
                (|ch_i.timeStamp − ch_j.timeStamp| < Δtt) then
                    ch_i.label =' influence'
                    exit
    if (ch_i.label = null) then
        ch_i.label =' externalstimulus'
    return ch_i.label
```

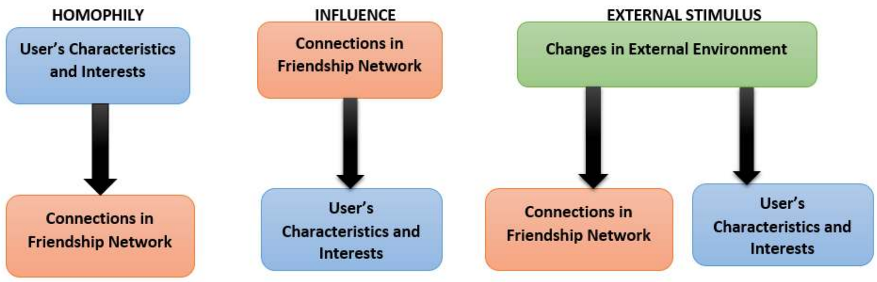


Fig. 3 Social Processes

3 Correlating Network and Spatial Properties with users' Check-ins by Analyzing Social Processes

Here, we show the correlation between the social interconnections of users and their spatio-temporal footprints/check-ins with the help of three social processes homophily, social spatial influence and external stimulus as described in Figure 3. We show how these social processes help to relate social network properties with footprints of users.

3.1 Correlating Centrality measures with Social-Spatial Influences

The notion of centrality is used to rank the users/nodes in the friendship graph in terms of how central or important they are. Roughly speaking, central nodes in a graph are important as they reach the network more quickly than non-central nodes and are very useful for fast circulation of information. We investigate, whether centrality of a node relates to social-spatial influence or not. Do more central nodes have more influence? Here, first we define degree centrality, closeness centrality [46, 50].

Degree Centrality: The simplest notion of centrality is the degree of a User u_i, the higher the degree, the more important or central the user in the social network. High degree of a user implies that it is more important or central in the relationships with other users in the social network.

$$\text{Degree Centrality:} \qquad C_d(u_i) = d_i = \sum_j A_{ij} \tag{1}$$

where C_d is Degree centrality of user u_i and A_{ij} is adjacency matrix of the friendship network. Normalized Degree Centrality:

$$\text{Normalized Degree Centrality:} \qquad C_d^{'}(u_i) = d_i/(n-1) \tag{2}$$

$C_d^{'}$ is normalized degree centrality of user u_i and n is number of users in the friendship network.

Closeness Centrality: closeness centrality is an important measure and determines how close a node is to other nodes in the network. It uses the sum of all the distances to rank centrality of a node in the friendship graph. Closeness centrality tries to minimize the total distance over all the other nodes, and thus a median node, has the highest closeness centrality.

$$\text{Average Distance:} \qquad D_{avg}(u_i) = \frac{1}{(n-1)} \sum_{j \neq i}^{n} d(u_i, u_j)$$

where n is number of users in the graph and $d(u_i, u_j)$ is shortest distance between two users u_i and u_j.

$$\text{Closeness Centrality:} \qquad C_c^{'}(u_i) = \frac{(n-1)}{\sum_{j \neq i}^{n} d(u_i, u_j)} \tag{3}$$

Betweenness Centrality: For every pair of vertices in a unweighted connected G graph, there exists at least one shortest path between the vertices such that either the number of edges that the path passes through is minimized. The betweenness centrality for each vertex is the number of these shortest paths that pass through the vertex.

$$\text{Betweenness Centrality:} \qquad C_b^{'}(u_i) = \sum_{u_j \neq u_i \neq u_k} \frac{\sigma_{u_j u_k}(u_i)}{\sigma_{u_j u_k}} \tag{4}$$

Where $\sigma_{u_j u_k}$ is the total number of shortest paths from node u_j to node u_k and $\sigma_{u_j u_k}$ is the number of those paths that pass through u_i.

Each of the centrality metrics has its own importance and usage, described in Table 1.

Social Influence: Social influence as discussed earlier refers to the phenomenon that the action of individual can stimulate his/her friends to act in a similar way [9]. For example, if a person visits a restaurant or a monument or an amusement park, then there are chances that his/her friends get influenced and also visit that

Table 1 Centrality measures

Centrality measures	Usage
Degree centrality	High Degree centrality indicates highly connected and popular users, who can quickly connect with the wider network
Closeness centrality	High Closeness Centrality for a user means that the user is well placed to influence the entire network quickly
Betweenness centrality	High value of Betwenness centrality for user means that the user can influence the flow around a system

place. Figure 3 (Influence) shows that connections in friendship network leads to the change in user's characteristics and interests.

All of the centrality three metrics effect the social influence in one or another way. Now, we try to correlate degree centrality, closeness centrality and betweenness centrality with Social-Spatial Influence.

Social-spatial influence of a user u_i on his/her friends is evaluated by counting the number of friends of those visited the same location visited by u_i with in threshold time period. It is important to understand that the social-spatial influence is high when users are spatially co-located. If a user and its friends live in the same city, then chances of the influence increases. Further, since, influence varies with time, therefore, it can be called as Social-Spatial-Temporal influence. So, we evaluate the Social-Spatial-Temporal influence:

$inf\text{-}count(u_i, l_k, \delta tt)$= number of friends of u_i who checked-in at l_k for the first time after u_i within a threshold time δtt. (had not checked in at l_k before u_i and having label='influence')

$$influence(u_i) = \sum_k inf\text{-}count(u_i, l_k), \qquad \forall k \tag{5}$$

where $influence(u_i)$ represents number of influences of user u_i and l_k is number of distinct check-ins of u_i. Note that in this work, the outliers are not considered in evaluating influence, because the LBSN data contain only check-ins and friendship connections. The data containing users' profiles or other information is not available, therefore, it is very difficult to distinguish among valid and invalid outliers.

We find the impact of degree and closeness centrality on the number of social-spatial influences. Later we show that high degree or closeness centrality of a node is observed to have high influence.

If a user check-ins to a limited number of locations repeatedly and frequently then this type of self reinforcing behavior does not affect the number of social-spatial influences. But if a user explores a number of different locations, then chances of increase in the number of influenced users are high. While taking this phenomena into the consideration we introduce another metric known as spatial Movement count.

Spatial Movement Count: Spatial movement count is the measure of how much a user visits new locations. It is evaluated by measuring the number of unique places visited or unique check-ins made by a user in a given time period.

Spatial Movement Count:

$$C_s(u_i) = S_i = \sum_k C(k) \tag{6}$$

$$C(k) = \begin{cases} 1 & \text{if } l_k \text{ is new location in } Ch_i(l_k) \\ 0 & \text{otherwise} \end{cases}$$

where $Ch_i(l_k)$ is a check-in of user u_i at location l_k.

Normalized Spatial Movement Count:

$$C_s^{'}(u_i) = \frac{S_i}{count(L)} \tag{7}$$

where L is the set of unique locations visited by all users.

When we consider both spatial and network centralities collectively, we get new notion of centrality called Spatial-Network centrality. It is defined as:

Spatial-Network Centrality: Spatial-Network centrality is the combination of Network measures like degree centrality and closeness centrality, and spatial measure like spatial movement count.

$$C_{sn}^{'}(u_i) = SN_i = \alpha(C_d^{'} + C_c^{'}) + (1 - \alpha)C_s^{'} \tag{8}$$

where $C_{sn}^{'}, C_d^{'}, C_c^{'}, C_s^{'}$ are normalized social-network, degree, closeness centralities and spatial movement count. And, α is a proportionality constant such that $0 < \alpha \leq 1$. The value $\alpha = 0.5$ indicates that both network and spatial measures are contributing equally in Spatial-Network centrality.

When, we consider both spatial and network centralities together, we get the better measure of social spatial influence. Later we show that high values of spatial-network centrality observed to have high social spatial influence and give a better measure of influence than degree centrality.

3.2 Correlating Strong ties with users' Check-ins by Analyzing Homophily

All the connections/links among the nodes in the friendship network are not of the same strength. Social Networks allow users to connect to many other users but not all are equally important. Friendships in social networks composed of strong ties means close friends and weak friends [9]. We estimate the strength of a tie from network topology and by analyzing user activities like check-ins. To measure strength of connection from network topology, we evaluate network neighborhood Overlap.

Network Neighborhood Overlap: Connection/tie strength in the friendship network can be measured based on neighborhood overlap. The neighborhood overlap is the number of shared friends of two users with respect to all the friends they have. The larger the overlap of number of friends the stronger the tie is.

$$nbOverlap(u_i, u_j) = \frac{\|N_i \cap N_j\|}{\|N_i \cup N_j\| - 2} \tag{9}$$

where N_i and N_j are number of friends of user u_i and u_j in the friendship network.

To measure strength of a connection from user activities, we evaluate spatial check-ins overlap.

Spatial Check-ins Overlap: Tie strength can also be measured by spatial Overlap. The spatial check-ins overlap is number of common locations visited by two users (within a given time interval) with respect to all the unique locations visited by both of them.

$$spatialOverlap(u_i, u_j) = \frac{\|Ch_i \cap Ch_j\|}{\|Ch_i \cup Ch_j\|} \tag{10}$$

where Ch_i and Ch_j are check-ins made by users u_i and u_j in the time interval (t_1, t_n).

To show the correlation between strong ties and users' footprints, we define homophily.

Homophily: Homophily is the tendency of individuals to associate and bond with similar individuals [9]. Figure 3(homophily) shows that similarity in user's characteristics and interests leads to new connections in friendship network.

We observe that users are similar in the sense when they have many common friends (large network neighborhood overlap) then they are more likely to check-in at same locations. This means larger network neighborhood overlap among users is indicative of their tendency to visit same locations. We show that user having high percentage of network neighborhood have more spatial check-ins overlap.

On the other hand, we later show that more the number of common places visited by users, more is their tendency to bond or friend with one another. We observe that more the spatial overlap exist among users, higher is the tendency that they are socially connected on the friendship network. So, a high degree of correlation exist between strong ties and users' footprints.

3.3 Correlating Users Footprints with Time by Analyzing External Stimulus

The number of footprints and social spatial influences change with the change in the external environment. It does not remain static over the period of time. We examine the affect of external stimulus on check-ins and influences of users.

External Stimulus: As discussed, external stimulus is the effect of external changes in environment on the behavior of users and friends [9]. An example of external stimulus is, during festival seasons people shop more and also influence more, which shows that changes in external environment cause changes in individual and social behavior. Even the changes in seasons bring changes in the check-in patterns of the users. During summer people like to check-in at ice-cream parlors, near water bodies, etc., on the other hand, during winters people often visit malls and closed areas. We later show the effect of external changes in environment on the behavior of users and friends. This shows that external environment conditions largely affect the check-in and influence of the users. Figure 3 (External Stimulus) shows that change in the external environment leads to the change in the connections in friendship network, and user's characteristics and interests.

Table 2 Datasets and Related Parameters

Dataset	♯Nodes	♯Edges	Average Degree	Density	♯Check-ins
Gowalla	57,073	635,388	22.26	0.00039	5,543,615
Brightkite	17,848	237,944	26.66	0.00149	4,032,866

Table 3 Datasets Statistics

Dataset	♯Users	♯Check-ins
Gowalla-TX	8,260	879,414
Gowalla-CA	6,530	669,214
Brightkite-NY	1,386	223,998
Brightkite-KS	339	62,520

4 Datasets Description

We evaluated the correlation among the users in social network in two large datasets. The first dataset, collected from Gowalla [4], which was a location-based social networking website where users shared their locations in the form of check-ins. The friendship network is an undirected graph collected using their public API, and consists of 196,591 nodes and 950,327 edges. Dataset consists of a total of 6,442,890 check-ins of these users over the period of Feb. 2009 - Oct. 2010.

The second dataset was collected from Brightkite [2], which was once a location-based social networking service provider where users shared their locations. The friendship network was collected using their public API, and consisted of 58,228 nodes and 214,078 undirected edges. Dataset also consisted of a total of 4,491,143 check-ins of these users over the period of Apr. 2008 - Oct. 2010.

In the experiments, we have chosen active users who have more than two friends and have at least 10 check-ins to ensure sufficient statistics for parameter estimation. These active users represented around 80% of the total number of check-ins. Table 2 describes two different datasets and related parameters.

We extracted all the check-ins of active users from Gowalla dataset for the states of Texas and California and two sets of check-ins of active users from Brighkite dataset for New York and Kansas. We collected all activities within a rectangular box of latitude-longitude coordinates around each of the selected state or city. Table 3 describes four sub-datasets. These datasets are analyzed and examined to find the correlations between friendship network properties and spatio-temporal check-ins of users. Figure 4 shows the check-ins distribution in four regions stated above.

5 Data Analysis and Results

We now show how the correlation of friendship network and spatial properties using social processes contain within itself insight of human behavior. First, we categorize the check-ins into different types. Then, results show the impact of network properties like degree centrality, closeness centrality and spatial-network

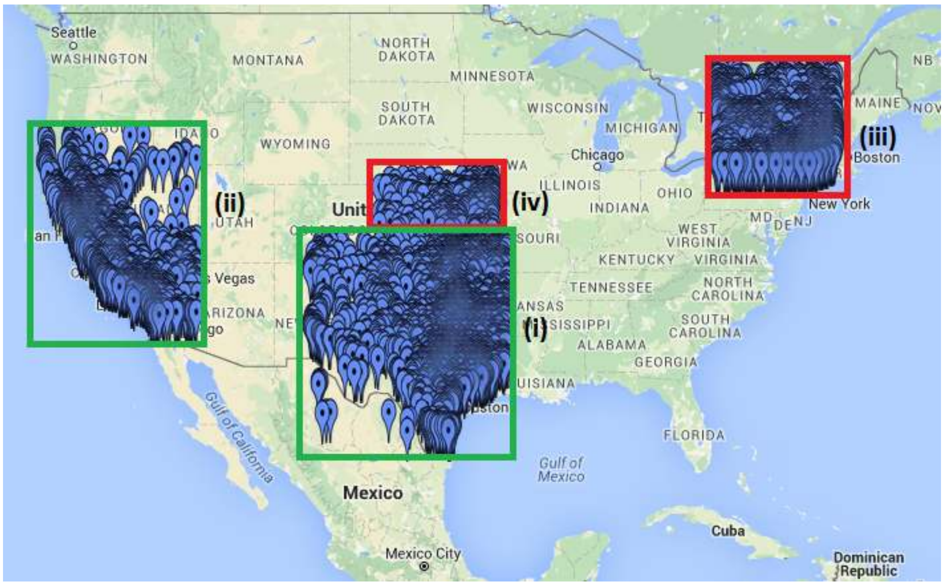


Fig. 4 Check-ins Distribution in Texas(i), California(ii), New York(iii) and Kansas(iv)

with social influence. The data is also analyzed to determine the correlation of neighborhood overlap with spatial overlap. Results show the change in the mobility pattern with the change in the external environment. These evaluation and detailed results are described below.

5.1 Check-in Analysis

As stated earlier, each check-in can be categorized into one of the following self reinforcing behaviors of the user or social influence of the friend or some external factors i.e. external stimulus. So we apply the Algorithm 1 (Check-in-Analyzer) for labeling each check-in as 'self' or 'influence' or 'external stimulus'. We have considered threshold time (Δtt) of 15 days, as input to the algorithm. It is important to note that the social spatial influences only occur at geographic local scale, when users are spatially co-located. For example, if two friends live in the same city then only there are chances that one gets influenced by other, and checked-in at the same location. Table 4 shows the result of applying Algorithm 1 (Check-in-Analyzer) on all four subsets of Gowalla and Brightkite given in Table 3. It is observed that percentage of social spatial influences are less than 10% in each. It is very less as compared to percentages due to external stimulus and self reinforcing behavior. The results show that social spatial influences are very less because it needs users to be spatially co-located in the specific region. At geographic global scale influences are very less, nearly zero percent.

In paper [51], Zhang et al. indicate that the similarity of friends' spatial trails at a geographically global scale cannot be attributed to spatial influence, it is up to 40% of the geographically localized similarity between friends. In our analysis

Table 4 Datasets and Check-ins Analysis

Check-in Analysis			
Dataset	Self Reinforcing Behavior(%)	external stimulus (%)	Social Spatial Influence(%)
GW-TX	51	48	1
GW-CA	46	52	2
BK-NY	72	21	7
BK-KS	79	17	4

we show that actual social spatial influence is less than 10% for the given datasets after localizing the data at city and state level.

5.2 Correlating Centrality with Influence

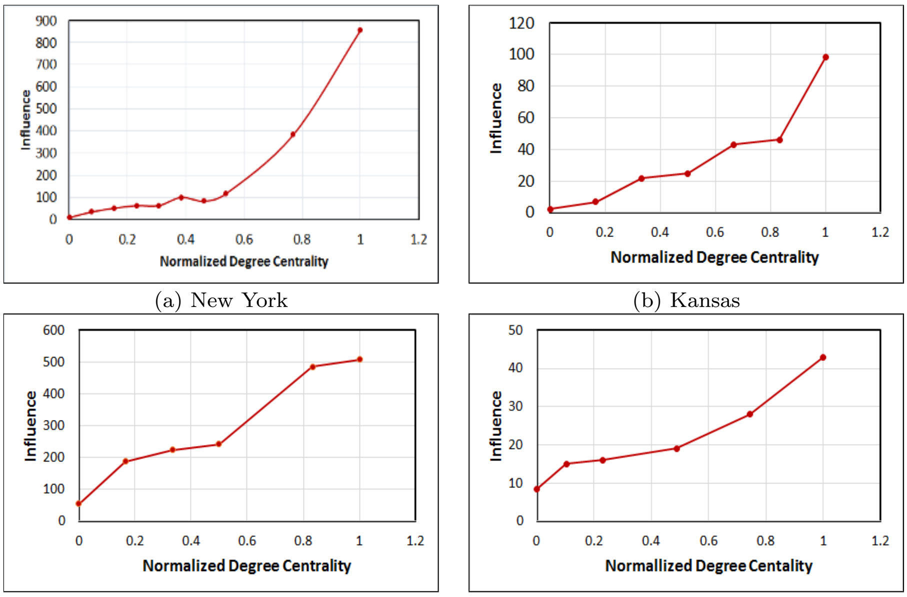
Here, we show how central nodes stimulate influences. We show the relationship of degree and closeness centrality with influences.

The procedure to compute social spatial influences for a specific centrality value is as follows:

- Evaluate the centrality of each node.
- Compute the Social-Spatial influence for each user, for each possible friends pair and for each location
- For a specific range of centrality values, calculate the sum of influences and number of users and
- Divide the sum of influence by the total number of users with in the given centrality range
- Similarly find the average influences for all the possible ranges of centrality.

5.2.1 Correlating Degree Centrality with Influence

We evaluated the correlation among the users based on their social network property,i.e., degree centrality and their spatio-temporal influences. The results are shown in Figures 5(a), 5(b), 5(c) and 5(d). The x-axis represents the normalized degree centrality in the scale of 0 to 1, and y-axis indicates the number of influences. TTable 3 shows the sub-datasets from four different cities/states, the number of users, and their check-ins in the respective city/state. It is observed from New York city data that most of the users have many social network interconnections which results in high degree centrality hence more number of check-ins and influences. On the other hand, from Kansas state data, it is observed that maximum users have low degree centrality values and which results in less number of check-ins and influences. We removed the outliers and plotted the graphs for the centrality values belong to majority of the users. When we plotted the graphs for majority of the spatial co-located users shown in Figures 5(a), 5(b), 5(c) and 5(d), it is observed that influence exhibits a positive correlation with degree centrality and hence it is proportional to degree centrality. This proportionality is observed in all sub-datasets. This means if a user has more number of friends, then his/her influence is more as compared to the user with less number of friends on social networks. But, degree centrality is not the only and independent factor affecting influence.



(a) New York (b) Kansas

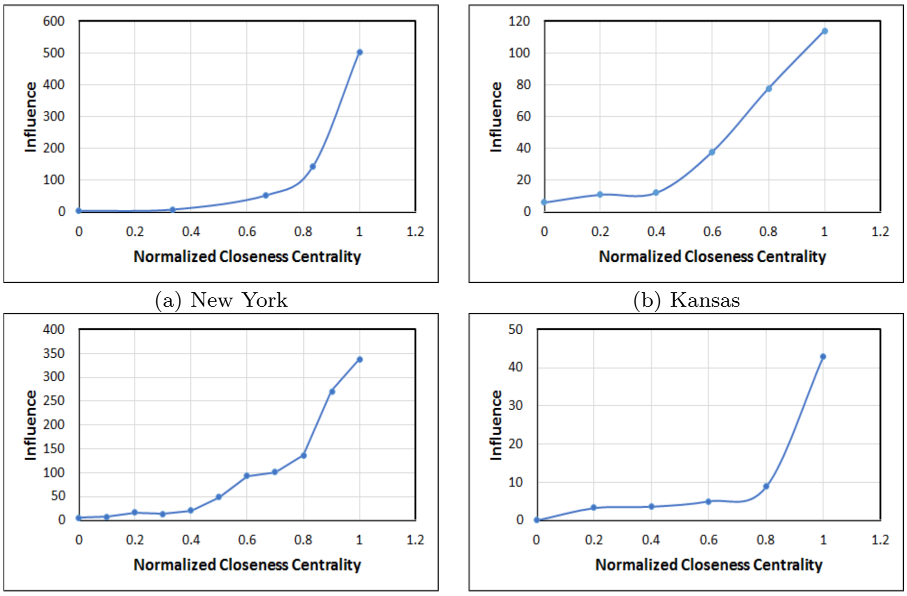
(c) Texas (d) California

Fig. 5 Degree Centrality Vs Influence

5.2.2 Correlating Closeness Centrality with Influence

We also evaluated the effect of closeness centrality on social influence. Dijkstra algorithm [3] is used to determine the shortest distance between two points. Data is pruned, for considering reachability in the following way to avoid considering infinite distance. Let r_i denote the number of nodes reachable from node u_i (i.e. not considering nodes which are at distance infinity from u_i). Then for evaluating results, we considered all the u_i's those have same r_i's values. It is examined that most of the nodes have same values for r_i and the remaining nodes have very less value for r_i.

The result is shown in Figure 6. The x-axis represents the normalized closeness centrality and y-axis indicates the number of influences. It is observed that influence exhibits a positive correlation with closeness centrality and hence it is proportional to closeness centrality. This proportionality is observed in all sub-datasets and shown in Figures 6(a), 6(b), 6(c) and 6(d). This means if a user is more central in the friendship network, then he or she may be more influential.



(a) New York (b) Kansas

(c) Texas (d) California

Fig. 6 Closeness Centrality vs Influence

5.2.3 Correlating Betweenness Centrality with Influence

We also evaluated the effect of betweenness centrality on social influence. The betweenness centrality was evaluated for the Gowalla dataset. A faster algorithm for large and sparse networks, proposed by Ulrik Brandes in [12], is used to calculate the Betweenness centrality.

After removing the outliers, result is shown in Figure 7. The x-axis represents the normalized Betweenness centrality and y-axis indicates the number of influences. It is evident from the data that a large number of nodes have very low values for betweeness centrality. This can be perceived from the graph that there are more data points near the zero value, on the other hand, there are less data points for values more than 0.2. It is observed that influence exhibits a positive correlation with betweenness centrality and hence it is proportional to betweenness centrality.

5.2.4 Correlating Spatial-Network Centrality with Influence

As discussed above, when we consider the social network properties like degree and closeness centralities, it relates to social influences. It is observed that when we consider both spatial properties as well as network properties results get improved. We apply equation 8 for different values of α. For $\alpha = 0.5$ and $C_c = 0$, result is shown in Figure 8. The x-axis represents the degree centrality and social-network centrality, and y-axis indicates the influences. It is observed that influence is more proportional to the spatial-network centrality than degree centrality.

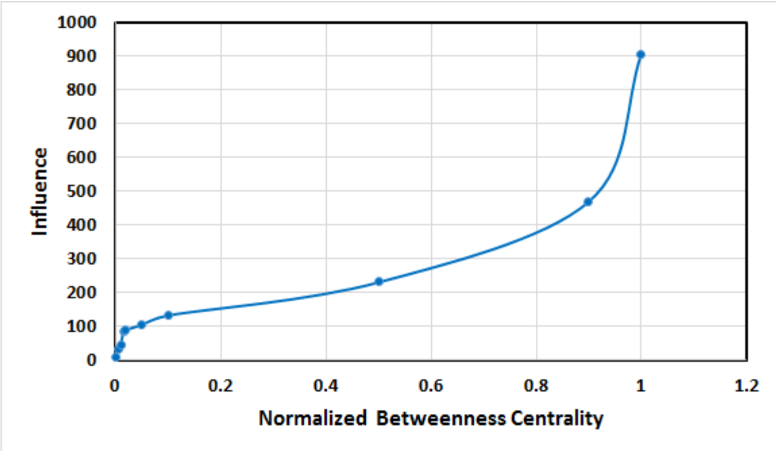


Fig. 7 Betweenness Centrality Vs Influence

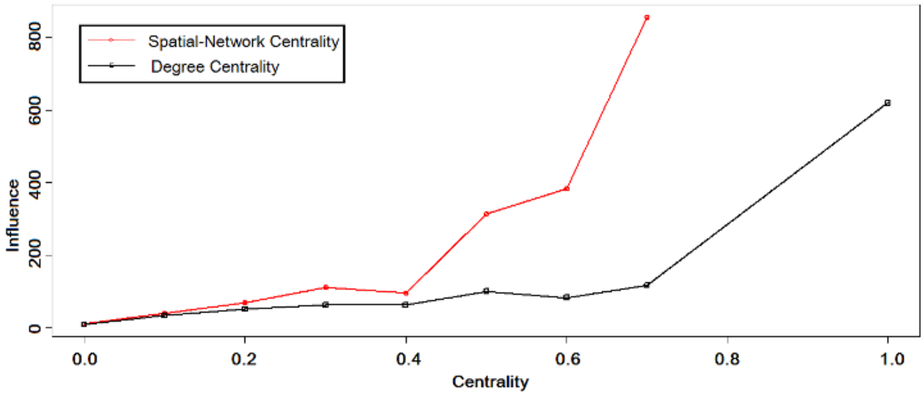


Fig. 8 Spatial Network Centrality and Degree Centrality Vs Influence

5.3 Correlating Strong Ties with users' Footprint by Analyzing Homophily

As stated earlier, homophily is the tendency of individuals to bond with other users having similar interests. We observe that when users have many common friends then they are more likely to check-in at similar locations. Table 5 shows that user having neighborhood overlap above 50% have more spatial check-ins overlap than the user having lesser. This pattern is observed for both Gowalla and Brightkite datasets.

On the other hand, we can also say that more the number of common places visited by users, more is their tendency to be friend with one another. Figure 9 shows that all the users who have more than 300 shared check-ins are friends and this friendship percentage keeps on decreasing as the number of shared check-ins decreases.

Table 5 Homophily

Neighborhood Overlap	>50%	30-50%	30-10%
Average number of shared check-ins for GW	43	27	22.4
Percentage of shared check-ins for GW	10.8%	8.4%	3.1 %
Average number of shared check-ins for BK	123.5	64.5	35.8
Percentage of shared check-ins for BK	6.6%	4.5%	1.9%

The procedure to compute friendship percentage for a given spatial overlap range is as follows:

- Compute the spatial check-in overlap between all the possible pairs of users
- For a specific range of overlap values, count how many of the users are actually connected in the social network and
- Divide the above number by the total number of users with in the overlap range, and find the friendship percentage.

Hasan et al. [24] show that the check-in similarity between two users increases with the increase of the friendship probability. They used the Cosine Similarity Metric and Jensen–Shannon Divergence Metric to show the relationship between check-in similarity and friendship probability. We are using Jaccard similarity coefficient for evaluating spatial overlap and show the relationship between friendship percentage and spatial check-in overlap.

As stated earlier, Sudhir et al. [29] prove that large number of shared check-in means there is high probability that social ties exist but reverse is not true. We show that although a large number of spatial overlaps are indicative of social ties as shown in Figure 9 but also a large number of social ties or connection also gives rise to many important parameters/ network properties, those are indicative of common check-ins and influences made by users. We claim that if two users have large number of network neighborhood overlap then it increases the probability that they have more number of shared check-ins. Figure 9 shows that more the number spatial overlap two users have, more is the probability that the social tie/friendship exist between them. Table 5 shows that more the percentage of network neighborhood overlap between two users more the chances that they have more number of shared check-ins.

5.4 Correlating Users Check-ins with Time and Analyzing external stimulus

As stated earlier, external stimulus is the effect of external changes in environment on the behavior changes in users and friends. Figure 11 shows a graph that represents regular check-ins pattern for a month of Austin (Texas) users. It is observed that there are more number of check-ins in the week-ends as compared to number of check-ins in week-days as shown in Figure 10. The Gowalla-TX has very less number of distinct users' check-ins before January, 2010. So, we analyzed check-ins and spatial influences patterns of Austin users for the months January, 2010

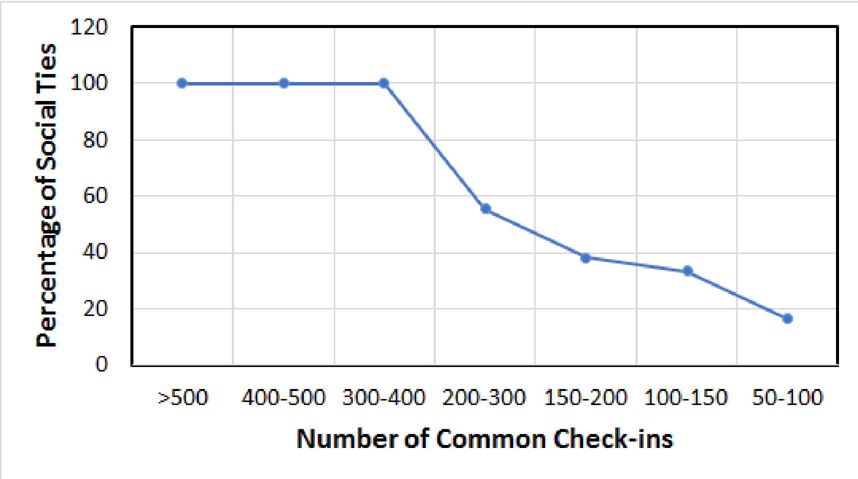


Fig. 9 Correlating Spatial Overlap with Social Connections

to October 2010 (the graph is shown in Figure 13). It is observed that month of March doesn't follow regular monthly check-ins pattern, there is maximum number of check-ins and influences in this month. As March is the spring season, with pleasant weather conditions and a season of festivals like music, film, kite flying, hot air balloons festivals, etc. Figure 12 shows the graph of March. We analyzed that there was heavy check-ins in the month between 12^{th} to 21^{st} March. And, during 13^{th} to 15^{th} March, it check-ins are nearly 10,000 and above. This is because there was SXSW [6] music, film and interactive festival during that period. It is mention in the wikipedia link [1] that in the 2010 music festival, which took place March 12–21, had an estimated 12–13,000 paying attendees. Our graph shown in Figure 12, also proves that maximum number of check-ins from Gowalla LBSN users, during this period, are above 10,000. This shows that external environment conditions largely effects the check-ins and spatial influences behavior of the users. It is observed that many of these spatial influences are actually due to external stimulus.

Table 6 shows the some of events/festivals occurred in Austin, Texas, during period January, 2010 to October, 2010. It shows that the number of check-ins during festival period is higher than the average number of the check-ins during the corresponding month (excluding event days). Based on the list of identified festivals and number of check-ins from the data, we can obtain several insights about the behavior and characteristics of the users. Out of the listed events three of them are music events, we can categorically say that most of participants are fond of music. Two of the events are athletic events named 'Austin Marathon and Half Marathon' and 'Statesman Capitol 10,000' (races), where the number of check-ins are only 704 and 1200 respectively, but, actual number of participants is above 8000 [1,5]. So, we can also state that many of the Gowalla-Austin users are not very fond of athletic events. This shows that by analyzing external stimulus we can study the behavior patterns of users.

[1] https://en.wikipedia.org/wiki/South_by_Southwest # 2010

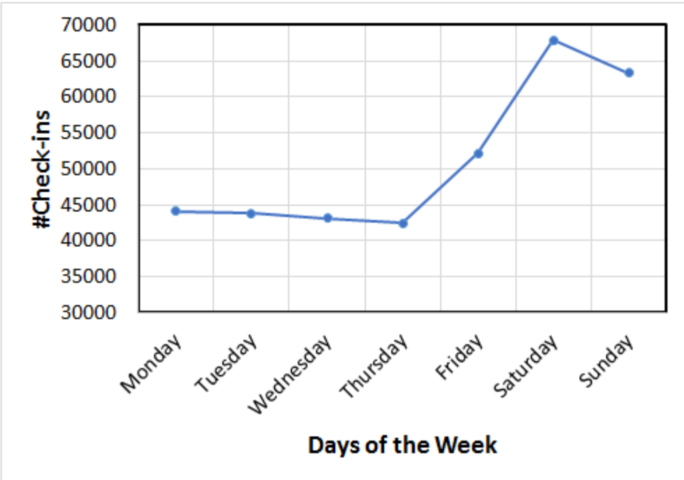


Fig. 10 Weekly Distribution of Check-ins

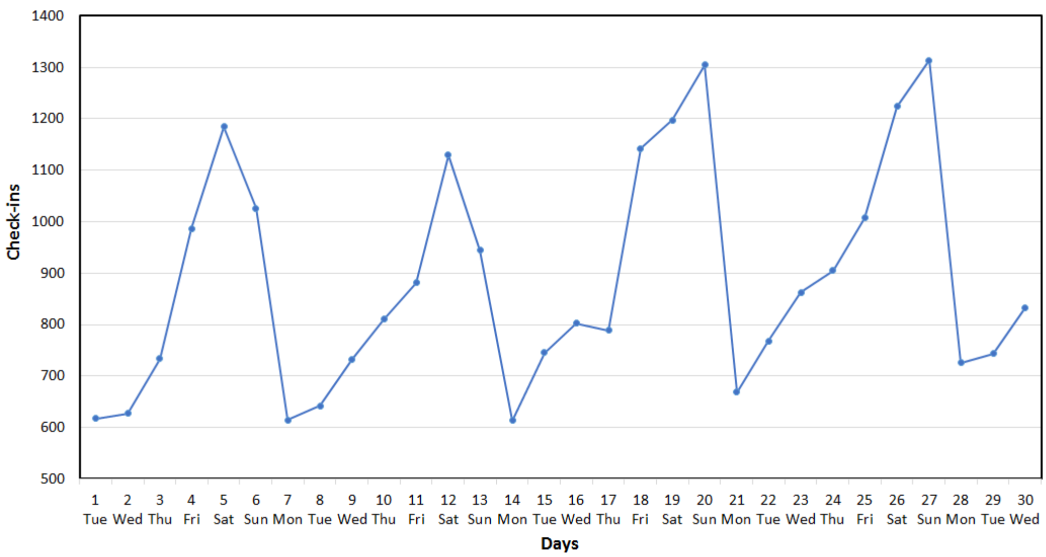


Fig. 11 Monthly Distribution of Check-ins

Figure 14 shows the check-ins and influences in the year 2010, in California state. It is observed that months June to September have maximum check-ins and spatial influences as it is best time to visit California. At this time people celebrate various festivals and events across various cities of California.

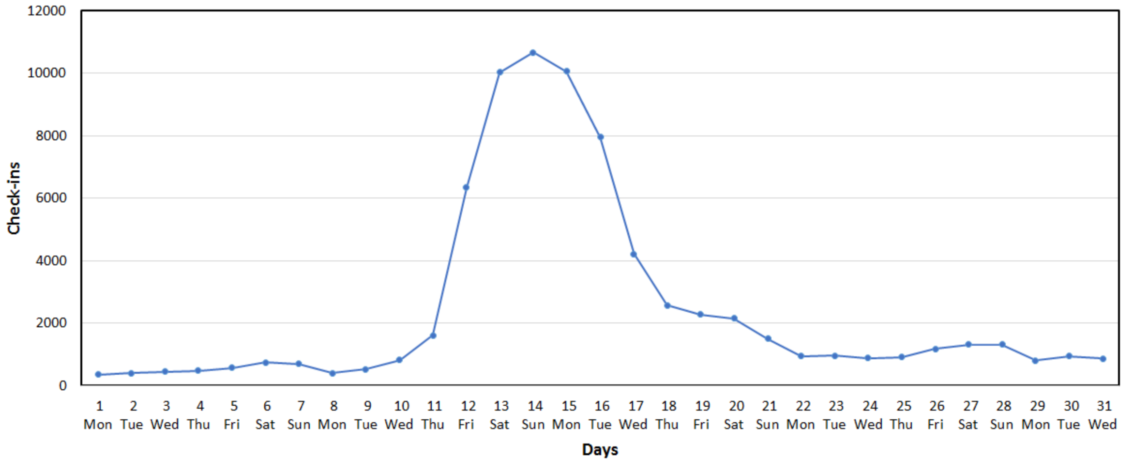


Fig. 12 Distribution of Check-ins in the Month of March

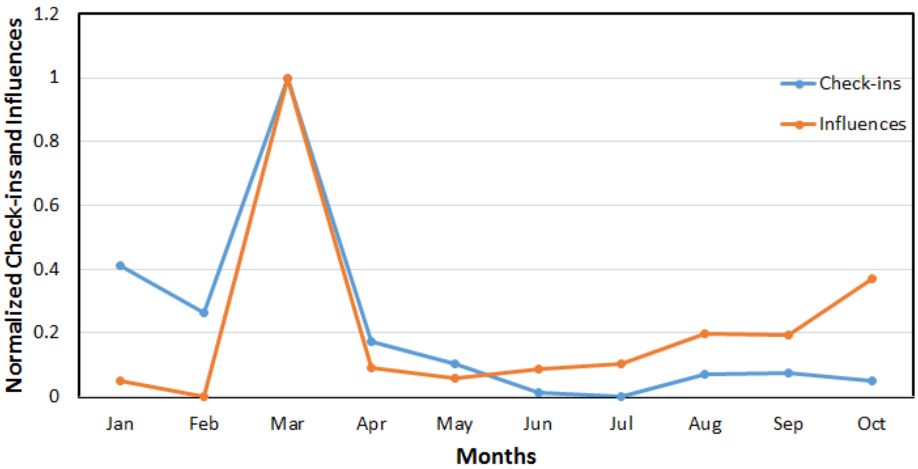


Fig. 13 Variations in number of Check-ins for Austin (January-October)

6 Conclusions and Future Work

In this paper, we have analyzed the data pertaining to Location Based Social Networks to study the effect of friendship in a social network on users check-in behavior. We evaluated fine-grained view of check-ins based on self reinforcing behavior, social network friend influence and external factors. We analyzed the correlation between social friendship network properties and users' check-in behavior with the help of three social processes- homophily, influence and external stimulus. It is observed that various topological properties like degree centrality, closeness centrality, betwenness centrality and neighborhood overlap, and spatial property like spatial movement count exhibit a positive relationship with social spatial influences. Spatial overlap plays an important role to determine spatial

Table 6 Variations in number of Check-ins due to External Stimulus

Event Date	Event Name	Average Check-ins during Month	Average Check-ins during Event
14^{th} February, 2010	Austin Marathon and Half Marathon	398	704
$12^{th} - 21^{st}$ March, 2010	South by Southwest Conferences and Festivals	805	57956
10^{th} April, 2010	Austin Family Music Festival	820	1554
11^{th} April, 2010	Statesman Capitol 10,000	820	1200
17^{th} July, 2010	Bastille Day in Austin	796	1577
25^{th}-26^{th} September, 2010	Pecan Street Festivals	886	2135
8^{th}-10^{th} October, 2010	ACL Music Festival Day	842	2458
15^{th}-17^{th} October, 2010	Austin Record Convention	842	2353

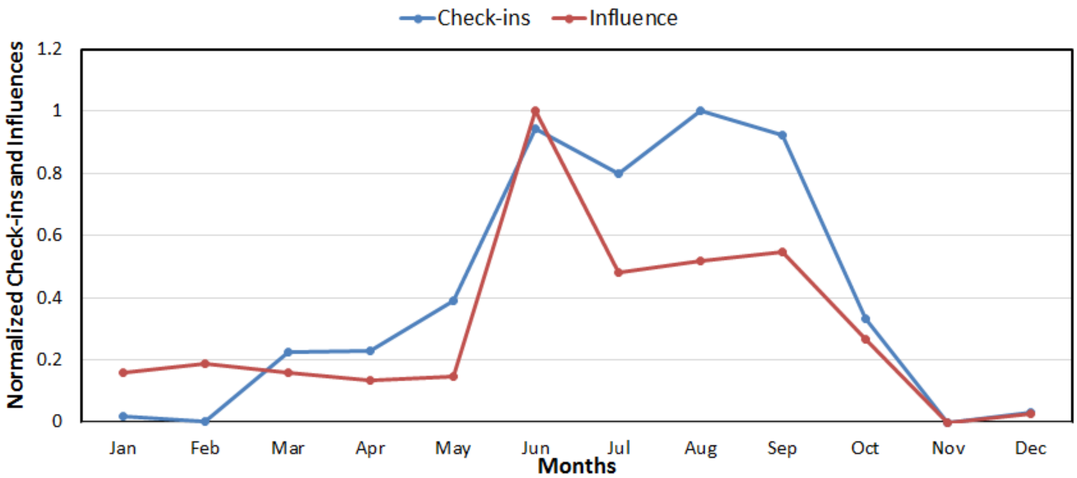


Fig. 14 Variations in number of Check-ins for California (January-December)

influences. It is also observed that the external changes in the environment impact the check-ins and influences of users.

The observed correlations between spatio-temporal check-ins of the users and their friendship network properties can further be used to solve influence maximization problem. It can also be used for community detection at different levels.

References

1. Austin - half marathon. http://results.active.com/events/2010-austin-marathon-half-marathon/half-marathon. Accessed: 2019-10-14.
2. Brightkite Dataset download link. https://snap.stanford.edu/data/loc-brightkite.html. Accessed: 2019-01-14.
3. Dijkstra algorithm-shortest path. https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm. Accessed: 2019-01-14.

4. Gowalla Dataset download link. https://snap.stanford.edu/data/loc-gowalla.html. Accessed: 2019-01-14.
5. Statesman capitol 10,000. http://www.cap10k.com/10k-history/. Accessed: 2019-10-14.
6. Sxsw - austin festival. https://en.wikipedia.org/wiki/South_by_Southwest. Accessed: 2019-01-14.
7. Taras Agryzkov, Almudena Nolasco-Cirugeda, Jose L Oliver, Leticia Serrano-Estrada, Leandro Tortosa, and Jose F Vicent. Using data from foursquare web service to represent the commercial activity of a city. *International Journal of Computer, Control, Quantum and Information Engineering. World Academy of Science, Engineering and Technology*, 9(1):69–76, 2015.
8. Álvaro Aguilera-García, Juan Gomez, Natalia Sobrino, and Juan José Vinagre Díaz. Moped scooter sharing: Citizens' perceptions, users' behavior, and implications for urban mobility. *Sustainability*, 13(12):6886, 2021.
9. Aris Anagnostopoulos, Ravi Kumar, and Mohammad Mahdian. Influence and correlation in social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 7–15. ACM, 2008.
10. Evangelia Anagnostopoulou, Jasna Urbančič, Efthimios Bothos, Babis Magoutas, Luka Bradesko, Johann Schrammel, and Gregoris Mentzas. From mobility patterns to behavioural change: leveraging travel behaviour and personality profiles to nudge for sustainable transportation. *Journal of Intelligent Information Systems*, 54(1):157–178, 2020.
11. Lars Backstrom, Eric Sun, and Cameron Marlow. Find me if you can: improving geographical prediction with social and spatial proximity. In *Proceedings of the 19th international conference on World Wide Web*, pages 61–70. ACM, 2010.
12. Ulrik Brandes. A faster algorithm for betweenness centrality. *Journal of mathematical sociology*, 25(2):163–177, 2001.
13. Guofeng Cao, Shaowen Wang, Myunghwa Hwang, Anand Padmanabhan, Zhenhua Zhang, and Kiumars Soltani. A scalable framework for spatiotemporal analysis of location-based social media data. *Computers, Environment and Urban Systems*, 51:70–82, 2015.
14. Jonathan Chang and Eric Sun. Location 3: How users share and respond to location-based data on social networking sites. In *Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media*, pages 74–80, 2011.
15. Arun Chauhan, Krishna Kummamuru, and Durga Toshniwal. Prediction of places of visit using tweets. *Knowledge and Information Systems*, 50(1):145–166, 2017.
16. Arun Chauhan, Ravi Tejwani, and Durga Toshniwal. Twitter can predict your next place of visit. *International Journal of Distributed Sensor Networks*, 12(2):9274715: 1–10, 2016.
17. Zhiyuan Cheng, James Caverlee, Kyumin Lee, and Daniel Z Sui. Exploring millions of footprints in location sharing services. *ICWSM*, 2011:81–88, 2011.
18. Eunjoon Cho, Seth A Myers, and Jure Leskovec. 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*, pages 1082–1090. ACM, 2011.
19. Yoon-Sik Cho, Greg Ver Steeg, and Aram Galstyan. Where and why users" check in". In *AAAI*, pages 269–275. Citeseer, 2014.
20. Carmela Comito. Next: a framework for next-place prediction on location based social networks. *Knowledge-Based Systems*, 204:106205, 2020.
21. Carmela Comito, Agostino Forestiero, and Clara Pizzuti. Bursty event detection in twitter streams. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 13(4):1–28, 2019.
22. Huiji Gao, Jiliang Tang, and Huan Liu. 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*, pages 1582–1586. ACM, 2012.
23. Mahmud Hasan, Mehmet A Orgun, and Rolf Schwitter. A survey on real-time event detection from the twitter data stream. *Journal of Information Science*, 44(4):443–463, 2018.
24. Samiul Hasan, Satish V Ukkusuri, and Xianyuan Zhan. Understanding social influence in activity-location choice and life-style patterns using geo-location data from social media. *Frontiers in ICT*, 3:10, 2016.
25. Mohammad Hosseinpour, Mohammad Reza Malek, and Christophe Claramunt. Socio-spatial influence maximization in location-based social networks. *Future Generation Computer Systems*, 101:304–314, 2019.
26. David Kempe, Jon Kleinberg, and Éva Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM, 2003.

27. Sonia Khetarpaul, SK Gupta, and L Venkata Subramaniam. Correlating social interconnections of users with spatio-temporal check-ins behavior. In *Proceedings of the 11th International Workshop on Spatial and Spatiotemporal Data Mining (SSTDM-16) in Co-operation with IEEE ICDM 2016*, pages 1133–1138. IEEE, 2016.
28. Sonia Khetarpaul, SK Gupta, and L Venkata Subramaniam. Spatiotemporal social (sts) data model: correlating social networks and spatiotemporal data. *Social Network Analysis and Mining*, 6(1):81, 2016.
29. Sudhir B Kylasa, Giorgos Kollias, and Ananth Grama. Social ties and checkin sites: Connections and latent structures in location based social networks. In *Proceedings of the 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining 2015*, pages 194–201. ACM, 2015.
30. Guoqiong Liao, Shan Jiang, Zhiheng Zhou, Changxuan Wan, and Xiping Liu. Poi recommendation of location-based social networks using tensor factorization. In *2018 19th IEEE international conference on mobile data management (MDM)*, pages 116–124. IEEE, 2018.
31. Liang Liu, Assaf Biderman, and Carlo Ratti. Urban mobility landscape: Real time monitoring of urban mobility patterns. In *Proceedings of the 11th International Conference on Computers in Urban Planning and Urban Management*, pages 1–16. Citeseer, 2009.
32. Yaxue Ma, Jin Mao, Zhichao Ba, and Gang Li. Location recommendation by combining geographical, categorical, and social preferences with location popularity. *Information Processing & Management*, 57(4):102251, 2020.
33. Matteo Manca, Ludovico Boratto, Victor Morell Roman, Oriol Martori i Gallissà, and Andreas Kaltenbrunner. Using social media to characterize urban mobility patterns: State-of-the-art survey and case-study. *Online Social Networks and Media*, 1:56–69, 2017.
34. Pablo Martí, Leticia Serrano-Estrada, and Almudena Nolasco-Cirugeda. Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74:161–174, 2019.
35. Diego Minatel, Vinícius Ferreira, and Alneu de Andrade Lopes. Local-entity resolution for building location-based social networks by using stay points. *Theoretical Computer Science*, 851:62–76, 2021.
36. Anastasios Noulas, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare. *ICwSM*, 11:570–573, 2011.
37. Konstantinos Pelechrinis and Prashant Krishnamurthy. Location affiliation networks: Bonding social and spatial information. In *Machine Learning and Knowledge Discovery in Databases*, pages 531–547. Springer, 2012.
38. H. Pham and C. Shahabi. Spatial influence - measuring followship in the real world. In *2016 IEEE 32nd International Conference on Data Engineering (ICDE)*, pages 529–540, May 2016.
39. Yaqiong Qiao, Xiangyang Luo, Chenliang Li, Hechan Tian, and Jiangtao Ma. Heterogeneous graph-based joint representation learning for users and pois in location-based social network. *Information Processing & Management*, 57(2):102151, 2020.
40. Adam Sadilek, Henry Kautz, and Jeffrey P Bigham. Finding your friends and following them to where you are. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 723–732. ACM, 2012.
41. Günther Sagl, Martin Loidl, and Euro Beinat. A visual analytics approach for extracting spatio-temporal urban mobility information from mobile network traffic. *ISPRS International Journal of Geo-Information*, 1(3):256–271, 2012.
42. Muhammad Aamir Saleem, Rohit Kumar, Toon Calders, Xike Xie, and Torben Bach Pedersen. Location influence in location-based social networks. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 621–630, 2017.
43. Yali Si, Fuzhi Zhang, and Wenyuan Liu. An adaptive point-of-interest recommendation method for location-based social networks based on user activity and spatial features. *Knowledge-Based Systems*, 163:267–282, 2019.
44. Thiago Sobral, Teresa Galvão, and José Borges. Visualization of urban mobility data from intelligent transportation systems. *Sensors*, 19(2):332, 2019.
45. Lijun Sun and Kay W Axhausen. Understanding urban mobility patterns with a probabilistic tensor factorization framework. *Transportation Research Part B: Methodological*, 91:511–524, 2016.
46. Lei Tang and Huan Liu. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1):1–137, 2010.

47. Taner Üsküplü, Fatih Terzi, and Hüma Kartal. Discovering activity patterns in the city by social media network data: a case study of istanbul. *Applied Spatial Analysis and Policy*, pages 1–14, 2020.
48. Jorge C Valverde-Rebaza, Mathieu Roche, Pascal Poncelet, and Alneu de Andrade Lopes. The role of location and social strength for friendship prediction in location-based social networks. *Information Processing & Management*, 54(4):475–489, 2018.
49. Daniel Villatoro, M Pus, and Marc Torrent-Moreno. Citizen as a sensor: The barcelona urban mobility use-case. In *Proceedings of the Smart City Expo World Congress (Barcelona, 2012)*, 2012.
50. Mohammed J Zaki and Wagner Meira Jr. *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press, 2014.
51. Ke Zhang and Konstantinos Pelechrinis. Understanding spatial homophily: the case of peer influence and social selection. In *Proceedings of the 23rd international conference on World wide web*, pages 271–282. ACM, 2014.
52. Yuan Zhong, Nicholas Jing Yuan, Wen Zhong, Fuzheng Zhang, and Xing Xie. You are where you go: Inferring demographic attributes from location check-ins. In *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining*, pages 295–304. ACM, 2015.