Social Network Analysis and Mining

Spatio-Temporal-Social(STS) Data Model - Correlating Social Networks and Spatio-Temporal Data

--Manuscript Draft--

Manuscript Number:	SNAM-D-16-00070
Full Title:	Spatio-Temporal-Social(STS) Data Model - Correlating Social Networks and Spatio-Temporal Data
Article Type:	Original Article
Corresponding Author:	Sonia Khetarpaul Indian Institute of Technology Delhi Delhi, INDIA
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	Indian Institute of Technology Delhi
Corresponding Author's Secondary Institution:	
First Author:	Sonia Khetarpaul
First Author Secondary Information:	
Order of Authors:	Sonia Khetarpaul
	Shyam Kumar Gupta
	L Venkata Subramaniam
Order of Authors Secondary Information:	
Funding Information:	
Abstract:	A location based social network (LBSN) is a network representation of social relations among actors, which not only allows them to connect to other users/friends but also users can share and access their physical locations. Here, the physical location consists of instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. This paper aim to capture this Spatio-Temporal Social Network (STS) data of Location based social networks and model it. In this paper, we propose a STS data model which captures both non spatial and spatial properties of moving users, connected on social network. In our model, we define data types and operations that make querying spatio-temporal-social network data easy and efficient. We extend spatio-temporal data model for moving objects proposed in [9] for social networks. The data model infers individual's location history, and helps in querying social network users for their spatio-temporal locations, social links, influences, their common interests, behavior, activities, etc. We show the some results of applying our data model on a spatio-temporal dataset (GeoLife) and two large real life spatio-temporal social network data sets (Gowalla, brightkite) collected over a period of two years. We Apply propose model to determine interesting locations in the city and correlate the impact of social network relationships on the spatiotemporal behavior of the users.
Suggested Reviewers:	Karine Reis Ferreira karine@dpi.inpe.br Works in Spatio-temporal data mining.
	Kotagiri Rao kotagiri@unimelb.edu.au Works is data mining and big data analytics.
	ARNAB BHATTACHARYA arnabb@iitk.ac.in Works in spatio-temporal data mining.

Social Network Analysis and Mining manuscript No.
(will be inserted by the editor)

Spatio-Temporal-Social(STS) Data Model - Correlating Social Networks and Spatio-Temporal Data

Sonia Khetarpaul
S K Gupta · L Venkata Subramaniam

the date of receipt and acceptance should be inserted later

Abstract A location based social network (LBSN) is a network representation of social relations among actors, which not only allows them to connect to other users/friends but also users can share and access their physical locations. Here, the physical location consists of instant location of an individual at a given timestamp and the location history that an individual has accumulated in a certain period. This paper aim to capture this Spatio-Temporal Social Network (STS) data of Location based social networks and model it.

In this paper, we propose a STS data model which captures both non spatial and spatial properties of moving users, connected on social network. In our model, we define data types and operations that make querying spatio-temporal-social network data easy and efficient. We extend spatio-temporal data model for moving objects proposed in [9] for social networks. The data model infers individual's location history, and helps in querying social network users for their spatio-temporal locations, social links, influences, their common interests, behavior, activities, etc. We show the some results of applying our data model on a spatio-temporal dataset (GeoLife) and two large real life spatio-temporal social network data sets (Gowalla, brightkite) collected over a period of two years. We Apply propose model to determine interesting locations in the city and correlate the impact of social network relationships on the spatio-temporal behavior of the users.

Department of Computer Science and Engineering
IIT Delhi, India
E-mail: sonia@cse.iitd.ac.in
Department of Computer Science and Engineering
IIT Delhi, India
E-mail: skg@cse.iitd.ac.in
IBM Research-India
E-mail: lvsubram@in.ibm.com

Keywords spatio-temporal data · social network · check-ins

1 Introduction

Social network sites like Foursquare, Facebook, Twitter, etc, give facilities to their users to share their location-based data and these sites advance in geo-spatial data gathering. They have created massive data sets with better spatial and temporal resolution than ever. These large data sets generated by their socially connected users, have motivated a challenge to develop a model which can capture spatio-temporal-social behavior of the users easily and precisely. For various purposes, we need to know when two persons co-occur in the same physical location or share similar location histories, knowledge about their common interests, behavior, and activities, interesting locations they visited, the influences they made on each other, importance of each other on social network, etc. We need a model that represents spatio-temporal-social datasets from different social network services so that managing, querying and analyzing of the data becomes easy and efficient.

Spatio-Temporal-Social data models help to mine human mobility patterns and their social interconnection attributes. Such type of mining is of great importance in two aspects (as discussed in [31]) : 1) commerce: it contributes significantly to link prediction, targeted advertising, and item recommendation, which are crucial for most companies; 2) uses: it is directive for information sharing, membership attachment, and trust establishment, which are all-encompassing for various personalized services.

In this paper, we extend the data model and algebra for spatio-temporal moving objects proposed in [9] for

social networks. There are many different types of moving objects that change their spatial and/or non spatial properties over period of time. These moving objects can be people, animals, buses, aircraft, cars, etc. This paper considers the moving objects as moving users of social networks.

We propose a data model to handle the spatio-temporal and social network information of objects using algebraic data types and functions. The main contribution of this work is an extensible algebra to represent relationship among the objects, where objects are changing their location in space and time. We extend the model proposed by Ferreira et al. [9], in two ways. First, we extend it by adding new operations in existing data types for spatio-temporal model for moving objects. Secondly, we define relationships among objects and various operations on it. The proposed data model with little modification can capture changes in many areas, including location-based social network services, environmental changes, behavior analysis of moving objects, etc, where objects are related with other objects and having spatial-temporal properties. Algebra describes data types and their operations in a formal way, independent of programming languages.

The rest of the paper is organized as follows. Section 2 presents the proposed spatio-temporal social network data model. Section 3 describes three datasets used for analysis and related parameters. Section 4 describes the results of applying STS data model and algebra to determine interesting locations. Section 5 shows results on spatio-temporal social network data and determine the impact of social interconnection on spatio-temporal mobility pattern of users. Finally section 6 concludes the paper.

1.1 Literature Review

There is a growing body of literature on querying and modeling Spatio-temporal-social network data. Current research related to our work can be broadly classified into the following three areas (i) spatio-temporal data modeling, (ii) querying and mining location based social network data and (iii) correlation between individual mobility patterns and social inter-connections. We briefly summarize the work in these categories and put our model and results in context.

Spatio-Temporal Data Modeling: Growth of mobile computing has motivated much work on moving objects. Gting et al. [12] propose a model that represents continuous changes in the spatial location or extent of objects. They define an algebra, data types and operations over them, for moving objects. Spaccapietra et al. [24] propose a conceptual model for trajectories of moving objects. They define trajectories as countable journeys that are semantically divided by defining a temporal sequence of time intervals when the object changes its position and stays fixed. Erwig et al. [8] proposed a new methodology where moving points and moving regions are viewed as 3-D entities and captured their structure and behavior by modeling them as abstract data types. Our work focuses on defining a spatio-temporal data model for moving users, those are socially inter-connected and affects each others movements.

Ferreira et al. in [9] focus on defining an algebra that covers the whole process to obtain events generated by objects from raw observations. The Geospatial Event Model (GEM), proposed by Worboys et al. [26], introduces concept of an event and relationships between events and geographical-objects in their model. It defines two types of relationships, object-event and event-event. Their model is based on the idea that an event can affect or be associated to one or more objects or events of different types. Galton et al. [10] improve such relationships for events, states, and processes in dynamic networks. The proposed STS Model define the relationship among objects (not the events) because in real life objects are inter-connected or related with one other.

In the spatio-temporal data mining research area, many algorithms, techniques and languages[8, 10, 12, 16, 17, 26, 28–30] have been proposed to detect patterns of trajectories and mine meaningful information.

Querying and Mining Location Based Social Network Data: Doytsher et al. [7] presented the concept of a socio-spatial graph and a set of operators that form a query language. Kang et al. [15] proposed a model called spatio-temporal uncertain networks (STUN) and the concept of STUN subgraph matching (or SM) queries. Many algorithms and techniques have been proposed to predict future check-in location of a user using their location based social network. Some Researchers have analyzed social network information to infer user location, in [11, 22]. Authors in [2, 31] 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 [4]. For the prediction of future check-ins, Cho et al. [6] focus on modeling the full temporal information of check-ins from a venue-centric perspective. Chang et al. [3] present a model for predicting future check-ins based on past check-ins, time of check-in, and user demographics. In this paper, we define a high level, programming language independent STS data model

with data types and operations that will make querying and mining location based social network data easy and efficient.

Correlating Individuals mobility patterns with social inter-connections:

Noulas et al. [20] analyze user check-in behavior to study the spatio-temporal mobility patterns of users. In particular they use temporal check-in information to determine user mobility patterns for a recommender system. Kylasa et al. [19] study the relationships between shared check-in locations and the structure and nature of social ties in the network. Cho et al. [5] 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. [3] show that an increasing number of shared check-in locations results in increasing friendship probabilities. However, the relationship between social ties and number of shared check-in locations is not investigated. Pelechrinis et al. [21], using affiliation networks, also draw similar conclusions. Their primary contribution is the interplay between the nature of a check-in location and social ties. They show that check-in locations have higher clustering coefficients among friends, when compared to non-friends. Aris et al. [1] presented the methodology to measure social correlation and test whether influence is a source of such correlation or not. Our, Spatio-Temporal-Social data model is high level data model and defines data types and operations that help to determine correlation among individual's mobility patterns and social interconnections.

2 The Proposed Spatio-Temporal-Social(STS) Data Model and Algebra

In this section, we define the proposed STS model for socially inter-connected moving objects, its architecture, data types and operations. First, we define the three primitive data types used in the proposed model.

2.1 Primitive Data-types

There are three primitive types: *Value*, *Time* and *Geometry* [9].

- *Value* is a generic type to express attribute values that can be an *Integer, Float, String* or *Boolean.*
- *Time* is a generic type that can be an *Instant* or a *Period* defined in the ISO temporal model [14].
- *Geometry* is a generic type compliant with the Geometry type defined in the OGC Geometry Model [13]. It can be a *Point, Line, Polygon, MultiPoint, MultiLineString,* or *MultiPolygon* type.

These types and their operations are well-known and have already been defined in the adopted standard.

2.2 Architecture of STS Data Model

The architecture of STS data model is shown in Figure 1. The model has three levels of data abstraction. First level, as shown in the Figure 1, represents the raw data or observations. The second level, contains the four data types which represents raw observations in more formalized way. And, the third level, is formalized and structured level, represents user's view of data. It shows the data as connected objects with spatial and non-spatial properties associated with it. Detail description of each of three levels is explained below.

2.2.1 Observations:

First, we describe the first level which is the raw spatio-temporal-social network observations. Spatio-temporal-social network observations further are of two types - spatio-temporal observations and social network interconnections. Social network interconnections contain information about related/connected objects in the social network. In the following section, we define the data type named *Relation* to handle these observations.

On the other hand, spatio-temporal data or observations contain information about physical movements of objects in the real world. These observations are the basis for spatio-temporal modeling [18]. Following Sinton's approach [23], a spatio-temporal observation should have three attributes: space, time and theme (the term "theme" refers to object being observed). We can create generalizations of spatio-temporal information by fixing one attribute, controlling another and measuring the third. This means to: (1) keep one attribute constant; (2) vary the second attribute in a controlled way; and (3) measure the third attribute, taking into consideration the constraints of the second attribute. In the second level, the spatio-temporal observations are mapped to three data types. Previous work [9] proposes three data types, time series, coverage and trajectory, to represent the combinations (1), (2) and (3):

1. Fixing location, controlling time, and measuring theme results in a time series.
2. Fixing theme, controlling time, and measuring location results in a trajectory.
3. Fixing time, controlling location, and measuring theme results in a coverage.

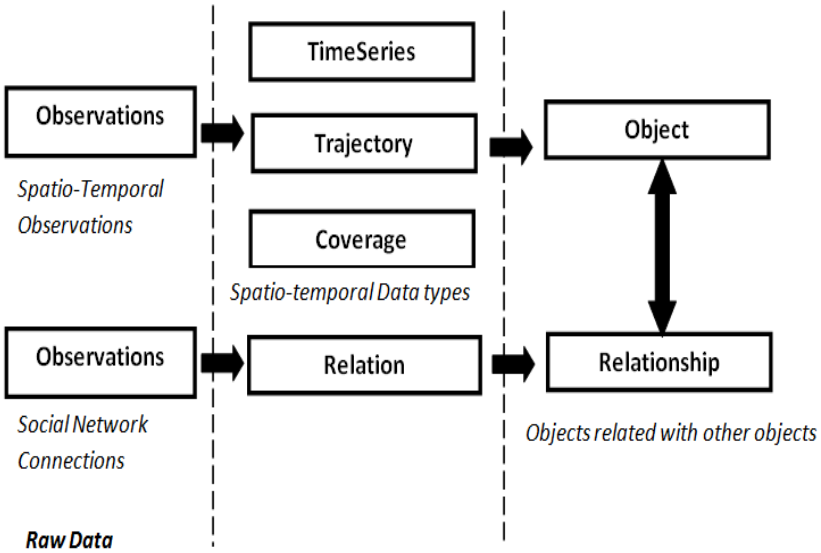

Fig. 1: The Proposed STS Model (Three Levels of Data Abstraction)

2.2.2 Data Types

This is the second level and defines three data types [9] as abstractions built on spatio-temporal observations: *Timeseries*, *Trajectory*, and *Coverage* and one data type: *Relation* on social network inter-connections. Using these types, we can do different types of analysis on the same observation set and address different application needs. These data types are defined as:

1. **TimeSeries:** A time series represents the variation of a property over time. It is obtained from observations that measure values at controlled times in a fixed location. For example, number of times a location visited by a user over the period of time, etc. TimeSeries is parametrized by Geometry (G), Time (T) and Value (V) types. These observations have a fixed geometry (G) and measured values (V) at controlled times (T).

$$type : \textbf{TimeSeries}[G : Geometry, T : Time, V : Value] \quad (1)$$

2. **Trajectory:** A trajectory represents how locations or boundaries of an object change over time. For example, change in the location of car over a period of time, number of unique places visited by user over a period of time, etc. Trajectory is parametrized by Value (V), Time (T) and Geometry (G) types. The observations have a fixed identity (V) and measured geometries (G) at controlled times (T).

$$type : \textbf{Trajectory}[V : Value, T : Time, G : Geometry] \quad (2)$$

New operation on trajectory data type is:

$$\textbf{Mode} : Trajectory \times time\_period \rightarrow G \quad (3)$$

where G is a maximum times visited Geometry point.

3. **Coverage:** A coverage represents the variation of a property in a spatial extent at a time. For every location within such extent, it is possible to compute a value of this property. For example, all locations visited by a person in the city during one day or variation of air pollution in the city districts during one week are represented by a coverage. Coverage is parametrized by Time (T), Geometry (G) and Value (V).

$$type : \textbf{Coverage}[T : Time, G : Geometry, V : Value] \quad (4)$$

Operations and axioms on spatio-temporal data types are defined in [9].

4. **Relation:** A Relation represents link or connection between two users in the social network. For example, a user is a friend of another user in the social network like Facebook, a user is following a celebrity user on social network Twitter, etc. Relation is parametrized by Value (ID_i) and Value

(ID_j). The observation represents unidirectional connection or relation (from ID_i to ID_j) between two users with fixed identities, ID_i and ID_j.

$$type : \textbf{Relation}[ID_i : Value, ID_j : Value] \quad (5)$$

Operations:

new : $ID_i \times ID_j \rightarrow Relation$
id : $Relation \rightarrow ID_i$
link : $Relation \rightarrow ID_j$

2.3 Objects and Relationships

The third level is the main contribution of this paper. We start with defining the object as defined in [9], then we define the new unary and binary operations involving one object or two objects respectively. We consider moving objects which are connected with some other objects and having some movement histories. For our analysis, we take objects as users. Each object has interconnections with some other objects on social network. Every object has some spatial and non-spatial parameters associated with it, which change over the period of time. We define a new data type named *Relationship*, which represents inter-connections of each object with other objects of similar type. Then, we define the various unary and binary operations on it.

2.3.1 Object:

An object is an identifiable entity whose spatial and non-spatial properties can change. The movement of a car shows change in the spatial parameters but changing the color of car shows change in non-spatial parameters. An object can be fixed like a restaurant, a mall, a building, etc, or it can be moving object like a person, a car, an animal, etc. This paper considers object as user of social network. Changes in the object can be instantaneous or continuous. Cases of instantaneous changes arise mostly due to change in legal rules that demand an immediate change in an object. When a government creates new laws or alter existing ones then changes take effect instantaneously and further effect objects involved. Continuous changes refer to a variation over time and space. Examples include the movement of cars on highways or movement of migratory birds. In these cases, the spatial locations of cars and birds change over time.

The Object type is parametrized by its identity type (ID), a TimeSeries (TS) that represents the variation of its non-spatial parameters and a Trajectory (TJ) that describes the change of its spatial parameters. An object can have one or more non-spatial properties, but we consider only one in the type definition for simplicity.

$$type : \textbf{Object}[ID : Value, TS : TimeSeries, TJ : Trajectory] \quad (6)$$

Defined Operations:

new : $ID \times TS \times TJ \rightarrow Object$
id : $Object \rightarrow ID$
timeseries : $Object \rightarrow TS$
trajectory : $Object \rightarrow TJ$
state : $Object \times Time \rightarrow (Value, Geometry)$

Unary operations Involving one Object:

geoPoints : $Object \times time\_period \rightarrow G$
where G is a set of unique Geometry points visited by $Object$ within $time\_period$ $[t_1, t_n]$ where $t_n > t_1$. (7)

Stay Point: A Stay point as defined in [16], represents a geographical location where an object spent a significant amount of time within a predefined time interval. If we consider object as a person, then stay point can be a work place, home, restaurant, historical place, congestion area, shopping mall, a stadium, worship places, etc. Stay points are determined for objects based on their spatio-temporal observations. Given the spatio-temporal observations of an object, we use this information to determine it's stay points at different time intervals. Since the log contains both Trajectories and Timeseries information, it is possible to determine how much time a person spent within spatial extent in a given time period.

For calculating stay points, two important parameters, $threshold\_distance$ (Δtd) and $threshold\_time$ (Δtt) are used. If a user spends more than Δtt within a Δtd then the mean of the GPS points within that region is considered as his/her stay point and the time when user entered in that region and time when leave that region are recorded. Equation 8 defines unary operation stayPoint. Algorithm 1, describes the stay-point determination method [16] from given trajectory of an object, where G represents geoPoints of object O for the given time interval.

$$\textbf{stayPoint} : Object \times Value(threshold\_time) \times Value(threshold\_distance) \rightarrow G \quad (8)$$
where G is set of Geometry points.

Algorithm 1: StayPointCalculation(geoPoints, Δtd, Δtt)

Data: geoPoints $G = g_1, g_2, \ldots, g_n$ in given time interval T_i of Object, Δtt, Δtd
Result: A set of stay points $S(T_i)$
begin
  $geoPointsCount \leftarrow |G|$ //number of given geoPoints
  $i \leftarrow 0$
  while $i < |G|$ **do**
    $j \leftarrow i + 1$
    $count \leftarrow 0$ //to count number of stay points
    while $j < geoPointsCount$ **do**
      $dist \leftarrow CalculateDistance(lat_i, lng_i, lat_j, lng_j)$
    if $dist < \Delta td$ **then**
      $timediff \leftarrow time_j - time_i$
    if $timediff > \Delta tt$ **then**
      $sumlng \leftarrow 0$
      $sumlat \leftarrow 0$
      $pt \leftarrow 0$ //count number of geoPoints satisfying condition
    $arvtime \leftarrow time_i$ //arrival time
    $deptime \leftarrow time_j$ //departure time
    $k \leftarrow i$
    while $k < j$ **do**
      $sumlat \leftarrow sumlat + lat_k$
      $sumlng \leftarrow sumlng + lng_k$
      $pt \leftarrow pt + 1$
      $k \leftarrow k + 1$
    $StayPtslat \leftarrow sumlat/pt$ //finding mean
    $StayPtslng \leftarrow sumlng/pt$
    $S \leftarrow (StayPtslat, StayPtslng, arvtime, depTime, \Delta tt, \Delta td)$
    $stayPoints.insert(S)$
    $Count \leftarrow Count + 1$
    $i \leftarrow j$ break
    $j \leftarrow j + 1$
    $i \leftarrow i + 1$
  $Return(stayPoints)$

Binary Operations Involving two Objects: In order to uniquely identify each *object*, in binary operations, we add a unique subscript to each.

$$\textbf{spatialIntersection} : Object_i \times Object_j \times time\_period \rightarrow G \quad (9)$$

where $G = \{G_1, G_2, \ldots, G_k\}$ are Geometry points and represent common places visited by object O_i and O_j.

$$\textbf{spatialUnion} : Object_i \times Object_j \times time\_period \rightarrow G \quad (10)$$

where $G = \{G_1, G_2, \ldots, G_n\}$ are Geometry points and represent places visited by O_i or O_j or both.

Spatial Overlap: The spatial overlap is number of common locations visited by two different objects with respect to all the unique locations they both visited.

$$spatialOverlap(O_i, O_j) = \frac{\text{number of common places visited by both } O_i \text{ and } O_j}{\text{number of unique places visited by either by } O_i \text{ or } O_j}$$

$$= \frac{count(spatialIntersection(O_i, O_j, Time))}{count(geoPoints(O_i)) + count(geoPoints(O_j))} \quad (11)$$

where operation *count* returns the number of Geometry points.

2.4 Relationship

Objects interact with other objects. An object is connected or related with some other objects through a relationship. So, a relationship is any association, linkage, or connection between the objects of interest. For example, a person is friend with other person through

social network, or a person is an employee of an organization, etc.

These connections in the social network may be symmetric or asymmetric. There are two ways to model human relationships in social networks. An asymmetric relationship is how Twitter currently works. You can follow someone else without him/her following you back. It is a unidirectional relationship that may not be bidirectional. Facebook, on the other hand, has always used a symmetric relationship, where each time you add someone as a friend they have to add you as a friend as well.

$$Type : \textbf{Relationship}[O_1 : Object, \{Rl_2, Rl_3, \ldots, Rl_n\} : Relation] \tag{12}$$

where $n > 1$ and all $\{Rl_2, Rl_3, \ldots, Rl_n\}$ are different unidirectional Relations. First value in each of these Relations, contain ID of Object O_1. Second value contains ID of object with which object O_1 is connected.

If Relationship is symmetric then $Relation[ID_1, ID_j]$ and $Relation[ID_j, ID_1]$ both exists, otherwise Relationship is asymmetric. Where ID_1 and ID_j are identity types of Objects O_1 and O_j respectively.

2.4.1 Operations on Relationships:

$$\begin{aligned} &\textbf{new} : Object \times \{Rl_2, Rl_3, \ldots, Rl_n\} \\ &\rightarrow Relationship(n > 1) \\ &\textbf{friend} : Relationship \rightarrow ID_2, ID_3, \ldots, ID_n \\ &\textbf{degree} : Relationship_1 \rightarrow n - 1 \\ &\text{(number of relations of } O_1\text{)} \\ &\textbf{distance} : Relationship(O_i, ID_j) \rightarrow Value \end{aligned} \tag{13}$$

where Value can take only be either 1 or 0, depending upon whether O_i is friend of O_j or not.

2.4.2 Binary Operations:

In order to uniquely identify each *object* and *Relationship*, in binary operations, we add a unique subscript to each.

$$\textbf{neighborhoodIntersection} : Relationship_i \times Relationship_j \rightarrow \{id_1, id_2, \ldots, id_k\} \tag{14}$$

id's of those objects that have relations with both the objects O_i and O_j

$$\textbf{neighborhoodUnion} : Relationship_i \times Relationship_j \rightarrow \{id_1, id_2, \ldots, id_k\} \tag{15}$$

id's of all objects that have relations with either objects O_i or O_j or both.

Neighborhood Overlap: The neighborhood overlap is number of shared friends of two users with respect to all the friends they have. The larger the overlap the stronger the connection is.

$$nbOverlap(R_i, R_j) = \frac{\text{number of unique friends of both } R_i \text{ and } R_j}{\text{number of friends who are adjacent to at least } R_i \text{ or } R_j}$$

$$= \frac{count(neighborhoodIntersection(R_i, R_j))}{count(neighborhoodUnion(R_i, R_j) - 2)} \tag{16}$$

where R_i and R_j are of type *Relationship* and operation *count* returns the number of Geometry points.
Influence: Influence is a process that causes behavioral correlations between adjacent users/objects in the social network. This refers to the phenomenon that the action of individuals can induce their friends to act in a similar way [1]. For example, if a person visits a restaurant or a monument or an amusement park, then there are the chances that his/her friends get influenced and also visit that place. This can be very useful for viral marketing.

$$\textbf{influenceSpatial} : Relationship_i \times Object_j \times time\_period \times Geometry\_point(G_k) \rightarrow 1 \text{ or } 0 \tag{17}$$

where G_k belongs to TJ of O_j, (function returns 1 if Geometry points G_k visited by given Object O_j first time after Object O_i visited within time-period, 0 otherwise)

For example, if a user visits a restaurant and his friend gets influence by him/her and within few days visits same restaurant.

$$\textbf{influenceNonspatial} : Relationship_i \times Object_j \times time\_period \times Value(non\_spatial\_Property) \rightarrow 1 \text{ or } 0 \tag{18}$$

where G_k belongs to TJ of $Object$, (function returns 1 if same non spatial property observed in given Object O_j first time after Object O_i have that property within time-period, 0 otherwise)

For example, if a user likes a particular political party webpage and his/her friend gets influence by him/her and within few days like same political party webpage.

Spatio-temporal social network data model and algebra for connected objects has been defined. In the following section, detailed descriptions of three different datasets used for analysis are given.

3 Datasets Description

For the validity of STS model, analysis was done on three datasets. First dataset contains only spatio-temporal trajectory information of users. The other two datasets consists of spatio-temporal data of visited locations by the users as well as their social network connectivities.

GPS trajectory dataset (GeoLife) [28–30] was collected by Microsoft Research. The data was collected for 165 users over a period of over two years (from April 2007 to August 2009). A GPS trajectory of this dataset is represented by a sequence of time-stamped points, each of which contains the information of latitude, longitude, height, speed and direction, etc. These trajectories were recorded by different GPS loggers or GPS-phones, and have a variety of sampling rates. 95 percent of the trajectories are logged in a dense representation, e.g., every $2 \sim 5$ seconds or every $5 \sim 10$ meters per point, while a few of them do not have such a high density being constrained by the devices. This dataset recorded a broad range of users' outdoor movements at different time intervals over a day, including not only life routines like going home and going to work but also some entertainment and sports activities, such as shopping, sightseeing, dining, hiking, and cycling etc. In the data collection program, a portion of users have carried a GPS logger for more than two years, while some of them may only have a trajectory dataset of a few weeks. We took a sample of on 126 active users from this dataset and worked on their GPS traces. Table 1 describes related parameters. The majority of the data was created in Beijing, China. A large part also came from Hangzhou.

Table 1: Dataset and Related Parameters

♯Users	♯Days	♯GPS Points	Area Covered
126	68,612	5,832,020	3,880,951 Sq. kms

We analyzed two LBSN datasets to find the correlation between users' social network inter-connections and their mobility patterns. First dataset, collected from Gowalla [34], which was a location-based social networking website where users shared their locations in the form of check-ins. The social/friendship network is an undirected graph and collected using their public API, and consists of 196,591 nodes and 950,327 edges. Dataset consisted of a total of 6,442,890 check-ins of these users over the period of Feb. 2009 - Oct. 2010.

Second dataset was collected from Brightkite [32], which was once a location-based social networking service provider where users shared their locations. The social/friendship network was collected using their public API, and consists of 58,228 nodes and 214,078 undirected edges. Dataset also consists of a total of 4,491,143 check-ins of these users over the period of Apr. 2008 - Oct. 2010.

For the data analysis, 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 social network properties and spatio-temporal check-ins of users. Figure 2 shows the check-ins distribution in four regions stated above.

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

We have introduced the three different dataset used for analysis. In the following section, new operations defined on *object* are applied to determine spatio-temporal interesting locations in the city from trajectory dataset (GeoLife) of users.

4 Data Analysis and Results on Spatio-Temporal Dataset

In this section, we apply proposed STS data model and algebra to determine interesting locations in the city from the GPS trajectory dataset (explained in previous section). In the following subsections, interesting locations are defined and a method is given for determination. Then, we show the results of interesting locations and temporal interesting locations.

Table 2: Datasets and Related Parameters(Gowalla and Brightkite)

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

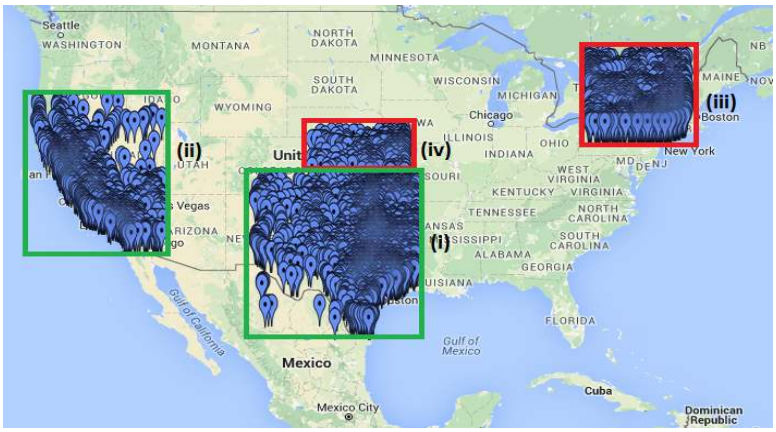

Fig. 2: Check-ins Distribution in Texas(i), California(ii), New York(iii) and Kansas(iv)

4.1 Interesting Location Determination by Applying STS Model

An interesting location *IntLoc(T_{ij})* is one that is visited by many individuals during time interval t_i to t_j. *IntLoc(T_{ij})* is a geographical region where the number of stay points of distinct users within a time interval T_i exceeds a given *ThresCount*. *ThresCount* is a user defined parameter. An interesting location can be a historical place, a good restaurant, a shopping complex, a stadium, a garden or any place that is visited by many users during the same time interval.

First, we evaluate the stay points of each user for given time interval using equation 8. Then, to determine interesting locations, we aggregate these bags of stay points for all the users.

4.1.1 Spatial Union of bags for predefined Time Interval:

We denote the spatial Union of all the stay points of all the users for a given time interval by $SPUnion(T_i)$. It is the set of all distinct elements from two bags.

$$SPSetUnion(T_i, T_j) = spatialUnion \ldots (spatialUnion (spatialUnion(S_1, S_2), S_3) \ldots, S_n) \tag{19}$$

where S_k is the set of staypoints of object O_k for the given time period $[T_i, T_j]$.

4.1.2 Temporal Interesting Location Determination:

The Temporal mode of a set of geometry points is the value in the set that occurs most often during a specified time interval. The Mode of Stay Points $SPMode(T_i)$ is the set of values that occur more than *ThresCount* times within the time period T_i. So, we evaluate temporal mode using equation 3. For applying the mode operation, Longitude and Latitude values of each stay point are truncated so that it points to the defined geographical region. For example truncating to three decimal places corresponds to a physical distance of about 100 meters. We apply the mode operation repeatedly, each time removing the points that contribute to it. In this way, we iteratively remove the most frequent stay points from the set.

$M_1 = mode(SPUnion(T_i))$
$M_2 = mode(SPUnion(T_i) - \{M_1\})$
$M_3 = mode(SPUnion(T_i) - \{M_1, M_2\})$
$M_n = mode(SPUnion(T_i) - \{M_1, M_2, \ldots, M_{n-1}\})$
$SPMode(T_i) = \{M_1, M_2, M_3, \ldots, M_n\}$

In the above M_1 is the most frequently visited stay point. An interesting location is the collection of all the stay points contributing to a particular mode calculation during specified interval of time and is represented by the centroid of all these points.

To ensure that an interesting location does not show up just because one person visited it repeatedly, we require that this location is the stay point of more than *ThresUserCount* number of people. We later show the

Table 4: Summarizes different statistics derived for 126 total users

Users	# Days	Avg days per user	# GPS Points	# Stay Points
126	68,612	545	5,832,020	23,989

results of applying STS data model to determine interesting locations on GPS trajectory dataset of moving users.

4.2 Results of Interesting Locations

First, we determine stay points of users from their GPS trajectories. Then, we apply the method explained above to determine interesting locations and show the results. We then apply K-means clustering algorithm and compare our results.

4.2.1 Stay point detection

The first step is to detect the stay points of every user with appropriate arrival time *(arrTime)* and departure time *(depTime)*. For this, we set ThresTime to 20 minutes and ThresDistance to 0.2 K.M. for stay point detection. In other words, if an individual stays over 20 minutes within a distance of 200 meters, this location is marked as a stay point. These two parameters enable us to find out some significant places, such as offices, restaurants and shopping malls, etc., at different interval of time over a day while ignoring the geo-regions without semantic meaning, like the places where people wait for traffic lights or encounter traffic congestion. As a result, we extracted total 23989 stay points from the dataset of 126 users. In Table 4, we shows the different statistics of our data. As one can see, the stay points give a compressed representation of the GPS traces comprising of millions of points. We also give the number of stay points and the total number of interesting locations detected for each of these users.

4.2.2 Interesting Location Determination

Some of the top interesting locations found using our method are shown in Table 5. This table lists the position of the interesting location (lat, long). We list the name of the place such as the building name or location name. Based on the list of interesting locations identified from the data, we can obtain several insights about the behavior and characteristics of the users. Specifically, for the points identified, after mapping them to real-world labels (as seen in Table 5), we can categorically say that most of participants are highly educated and perhaps students or faculty members as they seem to be visiting the main building of the university most often. Further, we can also state that many are possibly technologists as they are visiting the two buildings often where Computer Scientists can be found. Interesting location Tea House, which lies between the University and the Technology center, suggests that it is favorite spot for people from both locations. One could use such information to identify a good spot to recruit new employees, spread awareness of products, etc. Such information when collected for a large number of people is also useful for planning traffic and various crowd management activities.

Table 5: Interesting Locations Determined Using Mode of Stay Points

Latitude	Longitude	Location
40	116.327	Main Building Tsinghua University
39.976	116.331	China Sigma Center Microsoft, China R & D Group
40.01	116.315	Da Yi Tea Culture Center, Tea House
39.975	116.331	Cuigong Hotel
39.985	116.32	Loongson Technology Service Center

Table 6: Temporal Staypoints detected

Time Interval	# Staypoints in Bag Union
12 am to 6 am	2638
6 am to 12 pm	6257
12 pm to 6 pm	9003
6 pm to 12 am	6091

4.2.3 Temporal Data Analysis

It is also possible to analyze the data for specific time periods. In Table 6, we show the number of stay points calculated for four different time intervals: 12 am to 6 am, 6 am to 12 pm, 12 pm to 6 pm and 6 pm to 12 am .

As expected people would be at their homes from 12 am to 6 am and this time period has the least number of stay points as most people are likely to be at their home at this time.

Applying the mode operation on the result obtained from time interval based bag union gives us high frequency locations visited by many users during that time interval. A sample visualization of mode operation is shown in Figure 3. A location having high frequency at any interval of time can be an interesting location in that time period. For calculating high frequency lo-

Fig. 3: Result of Mode Operation

Table 7: Temporal Division of Interesting Locations

Time Interval	# Interesting Locations
12 am to 6 am	4
6 am to 12 pm	13
12 pm to 6 pm	25
6 pm to 12 am	17

cations, the parameter *ThresCount* is set to 10. For a given location if the mode is more than *ThresCount* then that can be an interesting location. Table 7 gives the number of interesting locations in our dataset for the different time periods. All low frequency regions with count less than ThresCount are eliminated because they can be particular a single user's workplace or home. A total of 59 locations are detected as interesting locations.

Some of the top temporally analyzed interesting locations found using our method are shown in Table 8. This table lists the position of the interesting location (lat, long). The frequency here refers to the number of stay points that comprise this interesting location. After mapping them to real-world labels, it is observed that Tsinghua University is most visited place from morning 6am to night 12pm. This means many among the users are students of the University. Out of top 5 visited locations three are common in the morning and midday time. During time period 6pm-12am, Wudaokou is the most popular place. It may be noted that according to [37], Wudaokou is the popular student hangout place in northwestern Beijing at night. During time period 12am-6am, Yuanmingyuan Park Area and some fast food restaurants are most visited places.

4.2.4 Comparison With Clustering

We compared the results obtained from the mode operation with those obtained using simple k-means clustering of the stay points obtained from the bag union. We set the k value at 59 so that we can make a direct comparison of the two results. The top two interesting locations identified by mode operation and k-means clustering match. Further, among the top ten interesting locations found by the two methods, four locations overlap. While we have not gone into the merits of one approach over the other, the comparison does point to important similarity and even more interestingly sufficient dissimilarity between the two approaches. Clustering as the first solution that comes to mind for this problem. However, clustering has its own challenges, the first obvious one being to decide the number of clusters. The mode operation is linear as it finds the most frequently appearing points. In contrast k-means is NP hard [35].

We applied STS model on spatio-temporal trajectory dataset of moving users. In the following section, we apply STS model and algebra on spatio-temporal social network datasets (Gowalla, Brightkite), to find the correlation among users' social-interconnections and mobility patterns.

5 Data Analysis and Results on Spatio-Temporal Social Network Dataset

In this section, we apply STS data model and algebra to determine the correlation between social network properties and spatio-temporal behavior of users. To determine this correlation, LBSN datasets (Gowalla, Brightkite) are analyzed and three different types of correlations are evaluated. 1) Datasets are analyzed to determine correlation of network properties like degree centrality, closeness centrality with influences. 2) The data is analyzed to determine the correlation of neighborhood overlap with spatial overlap. 3) Analysis is also done to determine the effects of external changes in environment on the mobility patterns and hence on influences of users. We start by defining the centrality of a user in the social network.

5.1 Centrality Analysis in Social Networks using STS Algebra

Before defining the centrality of a user in the social network, we formally define the term *check_in*.

Check_in: A *Check_in*, ch_l, is an action of registering u_i's presence at a location l. It gives information about the date-time on which a location is visited by a user. Check-ins Ch_i, is a set representing all the check-in done by user u_i.

Table 8: Interesting Locations Determined Using Temporal Mode of Stay Points

Time Period	Latitude	Longitude	Frequency	Location
6am-12pm	39.982	116.321	304	Keyu Community Residential area
	39.973	116.334	210	Zhonghang Reconnaissance Design and Research Institute, Miaoweilai Chinese Type Fast-food, Mahua Stretched Noodles
	40	116.327	207	Tsinghua University
	39.976	116.331	148	China Academy of Space Technology
	39.978	116.301	119	Yiduofu Home Cooking, Beijing Xiangyu Xingye Technology and Trade Center
12pm-6pm	40	116.327	344	Tsinghua University
	39.973	116.334	279	Zhonghang Reconnaissance Design and Research Institute, Miaoweilai Chinese Type Fast-food, Mahua Stretched Noodles
	39.978	116.301	157	Yiduofu Home Cooking, Beijing Xiangyu Xingye Technology and Trade Center
	39.974	116.334	91	Motel 168, Home Inn Mei Lin GE(Chinese Restaurant)
	39.982	116.326	64	Chinese Academy of Sciences, Jisuan Institute of Technology
6pm-12am	39.99	116.33	163	Wudaokou Haidian District,Beijing
	39.991	116.33	154	Tsinghua Park
	40.003	116.344	122	Beijing Forestry University Multiple-use Bldg
	40	116.327	107	Tsinghua University
	39.981	116.329	82	Ami'er Restaurant
	40.058	116.404	80	Xinchuan Noodle Restaurant,Qinkang Restaurant
12pm-6am	40.01	116.315	382	Yuanmingyuan Park Area
	40.058	116.404	86	Shanxi Noodle Restaurant
	39.951	116.362	85	Henan Liuji Stewed Noodles, Tianye Yunnan Cai
	39.981	116.329	67	Chengdu Snack Food

5.1.1 Centrality Analysis

The notion of centrality is used to rank the objects in a social network in terms of how central or important they are.

Degree Centrality: The simplest notion of centrality is the degree of an Object O_i. The higher the degree, the more important or central the Object in the social network [25,27]. High degree of an object implies that it is more important or central in the relationships with other objects of a social network.

Degree Centrality:

$$C_D(O_i) = d_i = degree(Relationship_i) \tag{20}$$

where C_D is Degree centrality of Object O_i.

Normalized Degree Centrality:

$$C_D^{'}(O_i) = d_i/(n-1) \tag{21}$$

$C_D^{'}$ is normalized degree centrality of object O_i and n is number of objects in the social network.

Closeness Centrality: Closeness centrality uses the sum of all the distances to rank how central an object in the social network is [25,27]. Central objects in the graph are important as they can be reached in the whole network more quickly than non-central nodes.

Average Distance

$$D_{avg}(O_i) = \frac{1}{(n-1)} \sum_{j \neq i}^{n} d(O_i, O_j)$$

where n is the number of objects in the graph and $d(O_i, O_j)$ is shortest distance between objects O_i and O_j. Shortest distance between two objects is 1 (equation 13), if direct relation exists between them.

Closeness Centrality

$$C_c^{'}(O_i) = \frac{(n-1)}{\sum_{j \neq i}^{n} d(O_i, O_j)} \tag{22}$$

After defining the centrality measures in the social networks, we correlate these measures with influences users' have on each other.

5.2 Correlating Centrality with Influence

We show the relationship of degree and closeness centrality with influences.

Influence: Influence of a user u_i on his/her friends is evaluated using equation 17. Since check-ins are spatial locations, so we use $influenceSpatial$ operation to determine influences. In our results, we have considered threshold time period of 15 days.

It is important to note that maximum number of spatial(check-ins) influences occur only when users are spatially co-located. On the other hand, non-spatial influences are independent of spatial distances. For example, visiting a restaurant depends on spatial location while purchasing a book is non spatial.

5.2.1 Correlating Degree Centrality with Influence

We evaluated the correlation among the users based on their social network property, degree centrality and their spatio-temporal influences. The results are shown in Figures 4,5,6 and 7. The x-axis represents the degree centrality and y-axis indicates influence. It is observed that influence is proportional to degree centrality. This proportionality is observed in all four datasets.

5.2.2 Correlating Closeness Centrality with Influence

We also evaluated the effect of closeness centrality on social influence. The results are shown in the Figures 8,9,10 and 11 for the four datasets. The x-axis represents the closeness centrality and y-axis indicates influence. Dijkstra algorithm [33] is used to determine the shortest distance between two points.

It is observed that influence is also proportional to closeness centrality. This proportionality is observed in all the datasets.

The strength of a social interconnection is measured by evaluating neighborhood overlap. So, in the following subsection, we discuss how strong connections affect check-ins pattern of users.

5.3 Correlating Strong Connections with users' Check-ins by Analyzing Homophily

Homophily is the tendency of individuals to bond/connect with other users having similar interests. To determine the strength or bonding of social connection between users, we determine neighborhood overlap using equation 16. Also, if two users have similar interests, then there are chances that they check-ins at same locations. So, to measure similar interest, we evaluate spatial overlap using equation 11. We observe that when users have many common friends (more neighborhood overlap), then they are more likely to check-in at similar locations. Table 9 shows that users having neighborhood overlap above 50% have more number of shared check-ins than the users 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 means more spatial overlap, more is their tendency to friend with one another. Figure 12 shows that all the users who have more than 300 shared check-ins are friends and this friendship percentage keeps on decreasing as number of shared check-ins decrease.

Sudhir et al. [19] show that large number of shared check-ins means there is high probability that social inter-connections exist but reverse is not true. We show that although a large number of spatial overlaps are indicative of social connections as shown in Figure 12 but also a large number of social connections also give many important parameters/ network properties, which 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 12 shows that more the number spatial overlap two users have, more is the probability that the social connection/friendship exist between them. Table 9 shows that more the percentage of network neighborhood overlap between two users more the chances that they have more number of shared check-ins. We also show that not only the number of shared check-ins are indicative of influences but degree centrality, closeness centrality also indicate influences.

Table 9: Correlating Neighborhood Overlap with Shared Check-ins

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%

In the following subsection, we show how external environment changes affect the check-ins and influence behavior of users.

5.4 Correlating Users Check-ins and Influences with Time and Analyzing Confounding

Confounding is the effect of external changes in environment on the behavior changes in users and friends.

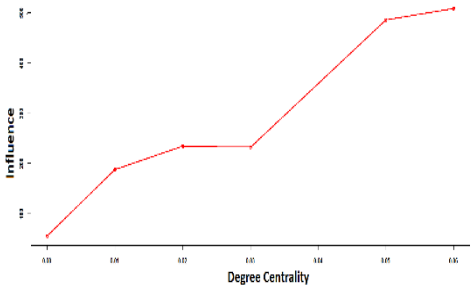

Fig. 4: Degree Centrality Vs Influence(GW-TX Dataset)

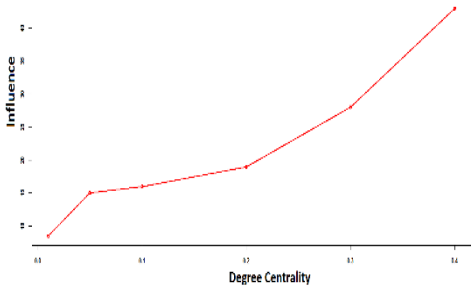

Fig. 5: Degree Centrality Vs Influence (GW-CA Dataset)

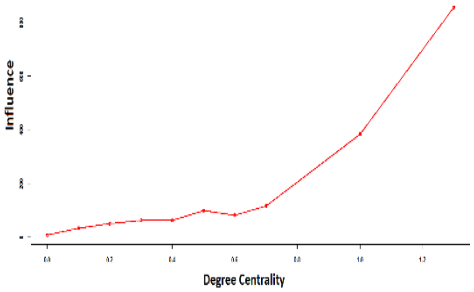

Fig. 6: Degree Centrality Vs Influence (BK-NY Dataset)

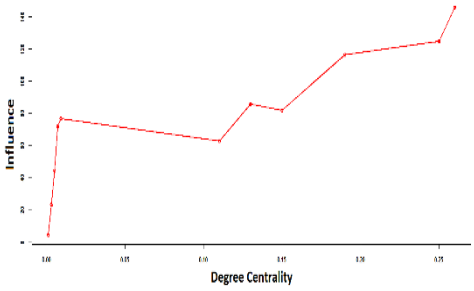

Fig. 7: Degree Centrality Vs Influence (BK-KS Dataset)

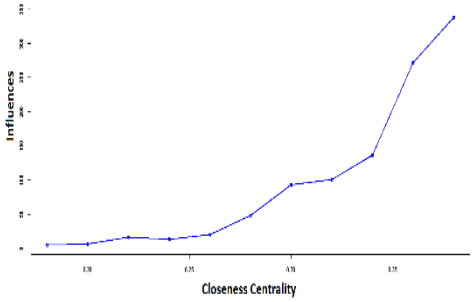

Fig. 8: Closeness Centrality Vs Influence (GW-TX Dataset)

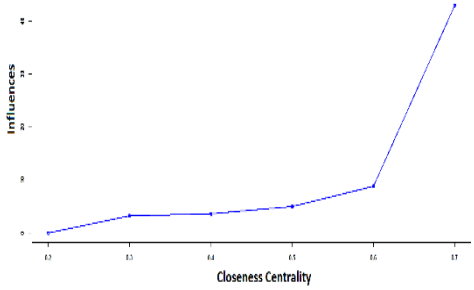

Fig. 9: Closeness Centrality Vs Influence (GW-CA Dataset)

Figure 13 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. The Gowalla-TX data has very less number of distinct users' check-ins before January, 2010. We therefore analyzed check-ins and influences patterns of Austin users for the months January, 2010 to October 2010. The graph is shown in figure 15. 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. It is noticed that 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 14 shows the graph of March. We analyzed that there was heavy check-ins in the month between 12^{th} to 21^{st} March. This is because there was SXSW [36] music, film and interactive festival during that period. This shows that external environment conditions largely effects the check-ins and influences behavior of the users.

Figure 16 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 influ-

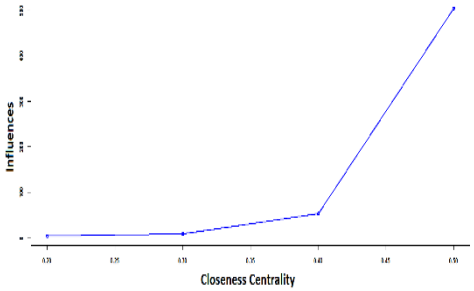

Fig. 10: Closeness Centrality Vs Influence (BK-NY Dataset)

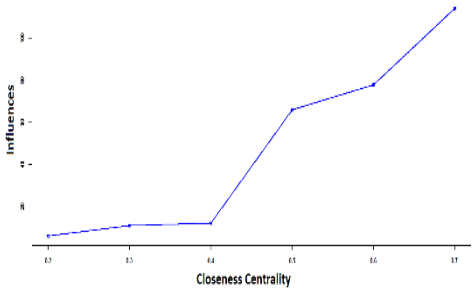

Fig. 11: Closeness Centrality Vs Influence (BK-KS Dataset)

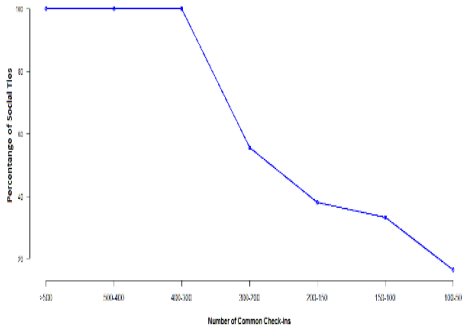

Fig. 12: Correlating Spatial Overlap with Social Connections

ences as it is best time to visit California. At this time people celebrate various festivals and events across various cities of California.

6 Conclusions

We have presented a spatio-temporal data model for social networks. It is a high level model with language independent data types and operations. This model not only captures spatio-temporal information of moving users but also their inter-connections in social network. It is useful for querying and mining spatio-temporal social network data. By applying STS data model and algebra, we analyzed and mined three large datasets (Geolife, Gowalla, Brightkite). Analysis is done and results are evaluated to determine the temporal interesting locations in a city on Geolife trajectory dataset. We also studied the impact of social inter-connections on spatio-temporal behavior of moving users on Gowalla and Brightkite datasets.

References

1. Anagnostopoulos, Aris, 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.
2. Backstrom, Lars, 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.
3. Chang, Jonathan 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.
4. Cheng, Zhiyuan, James Caverlee, Kyumin Lee, and Daniel Z Sui. Exploring millions of footprints in location sharing services. *ICWSM*, 2011:81–88, 2011.
5. Cho, Eunjoon, 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.
6. Cho, Yoon-Sik, Greg Ver Steeg, and Aram Galstyan. Where and why users" check in". In *AAAI*, pages 269–275. Citeseer, 2014.
7. Doytsher, Yerach, Ben Galon, and Yaron Kanza. Querying geo-social data by bridging spatial networks and social networks. In *Proceedings of the 2nd ACM SIGSPATIAL International Workshop on Location Based Social Networks*, LBSN '10, pages 39–46, New York, NY, USA, 2010. ACM.
8. Erwig, Martin, Ralf Hartmut Gu, Markus Schneider, and Michalis Vazirgiannis. Spatio-temporal data types: An approach to modeling and querying moving objects in databases. *GeoInformatica*, 3(3):269–296, 1999.
9. Ferreira, Karine Reis, Gilberto Camara, and Antnio Miguel Vieira Monteiro. An algebra for spatiotemporal data: From observations to events. *Transactions in GIS*, 18(2):253–269, 2014.
10. Galton, Antony and Michael Worboys. Processes and events in dynamic geo-networks. In *Geospatial semantics*, pages 45–59. Springer, 2005.
11. Gao, Huiji, 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*

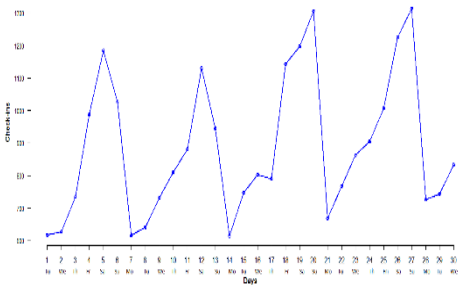

Fig. 13: Monthly Distribution of Check-ins Pattern

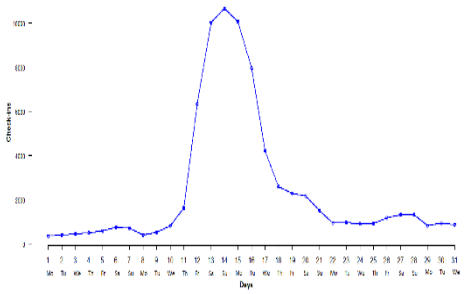

Fig. 14: Check-ins in the Month of March

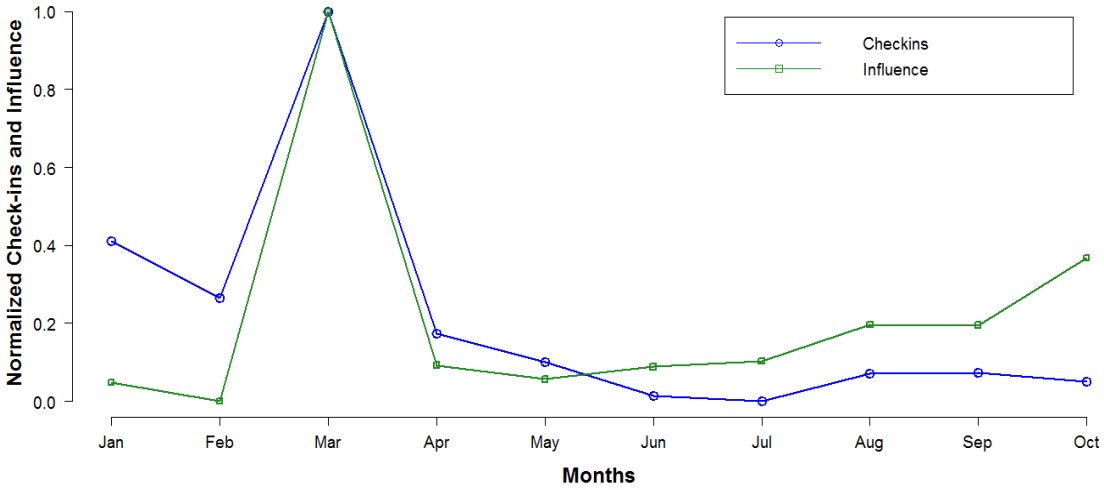

Fig. 15: Variations in number of Check-ins and Influences (January-October)

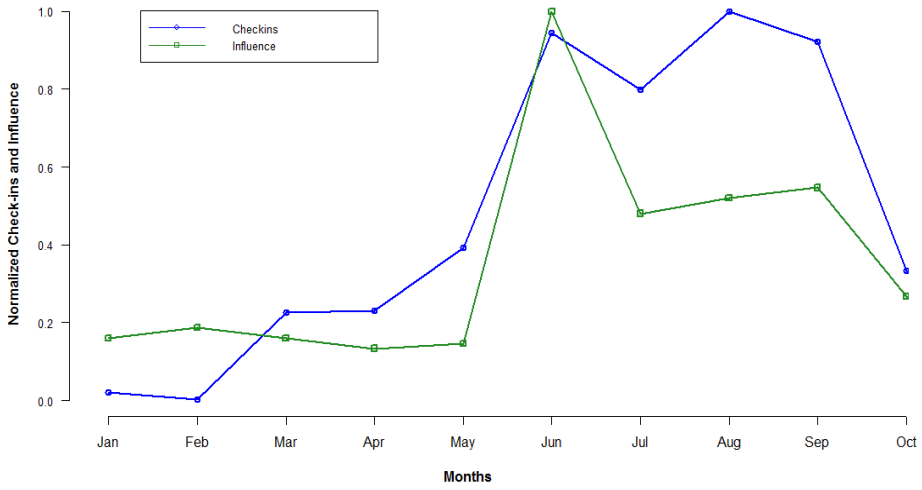

Fig. 16: Variations in number of Check-ins and Influences for California (January-October)

international conference on Information and knowledge management,pages 723–732. ACM, 2012.
12. Gting, Ralf Hartmut, and Markus Schneider. *Moving objects databases*. Elsevier, 2005.
13. Herring, John R. Opengis implementation specification for geographic information-simple feature access-part 2: Sql option. *Open Geospatial Consortium Inc*, 2006.
14. TC ISO. 211 sc, 2002: Iso 19108. *Temporal schema*.
15. Kang, Chanhyun, Andrea Pugliese, John Grant, and VS Subrahmanian. Stun: querying spatio-temporal uncertain (social) networks. *Social Network Analysis and Mining*, 4(1):1–19, 2014.
16. Khetarpaul, Sonia, Rashmi Chauhan, SK Gupta, L Venkata Subramaniam, and Ullas Nambiar. Mining gps data to determine interesting locations. In *Proceedings of the 8th International Workshop on Information Integration on the Web: in conjunction with WWW 2011*, page 8. ACM, 2011.
17. Khetarpaul,Sonia, SK Gupta, and L Venkata Subramaniam. Analyzing travel patterns for scheduling in a dynamic environment. In *Availability, Reliability, and Security in Information Systems and HCI*, pages 304–318. Springer, 2013.
18. Kuhn, Werner. A functional ontology of observation and measurement. In *GeoSpatial Semantics*, pages 26–43. Springer, 2009.
19. Kylasa, Sudhir B, 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.
20. Noulas, Anastasios, Salvatore Scellato, Cecilia Mascolo, and Massimiliano Pontil. An empirical study of geographic user activity patterns in foursquare. *ICwSM*, 11:70–573, 2011.
21. Pelechrinis, Konstantinos and Prashant Krishnamurthy. Location affiliation networks: Bonding social and spatial information. In *Machine Learning and Knowledge Discovery in Databases*, pages 531–547. Springer, 2012.
22. Sadilek, Adam, 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.
23. Sinton, David. The inherent structure of information as a constraint to analysis: Mapped thematic data as a case study. *Harvard papers on geographic information systems*, 6:1–17, 1978.
24. Spaccapietra, Stefano, Christine Parent, Maria Luisa Damiani, Jose Antonio de Macedo, Fabio Porto, and Christelle Vangenot. A conceptual view on trajectories. *Data & knowledge engineering*, 65(1):126–146, 2008.
25. Tang, Lei, and Huan Liu. Community detection and mining in social media. *Synthesis Lectures on Data Mining and Knowledge Discovery*, 2(1): 1-137, 2010.
26. Worboys, Michael, and Kathleen Hornsby. From objects to events: Gem, the geospatial event model. *Geographic Information Science*, pages 327–343. Springer, 2004.
27. Zaki, Mohammed J., and Wagner Meira Jr. Data mining and analysis: fundamental concepts and algorithms. *Cambridge University Press*, 2014.
28. Zheng, Yu, Quannan Li, Yukun Chen, Xing Xie, and Wei-Ying Ma. Understanding mobility based on gps data. In *Proceedings of the 10th International Conference on Ubiquitous Computing*, UbiComp '08, pages 312–321, New York, NY, USA, 2008. ACM.
29. Zheng, Yu, Lizhu Zhang, Xing Xie, and Wei-Ying Ma. Mining interesting locations and travel sequences from gps trajectories. In *Proceedings of the 18th International Conference on World Wide Web*, WWW '09, pages 791–800, New York, NY, USA, 2009. ACM.
30. Zheng, Yu. Trajectory data mining: an overview. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 6(3):29, 2015.
31. Zhong, Yuan, 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.
32. Brightkite Dataset download link. https://snap.stanford.edu/data/loc-brightkite.html. Accessed: 2016-01-14.
33. Dijkstra algorithm-shortest path. https://en.wikipedia.org/wiki/Dijkstra%27s_algorithm. Accessed: 2016-01-14.
34. Gowalla Dataset download link. https://snap.stanford.edu/data/loc-gowalla.html. Accessed: 2016-01-14.
35. K-means Clustering-NP hard https://en.wikipedia.org/wiki/K-means_clustering. Accessed: 2016-06-12.
36. Sxsw - austin festival. https://en.wikipedia.org/wiki/South_by_Southwest. Accessed: 2016-01-14.
37. Wudaokou. https://en.wikipedia.org/wiki/Wudaokou. Accessed: 2016-06-02.