

A Framework for Discovering Spatio-temporal Cohesive Networks

Jin Soung Yoo¹ and Joengmin Hwang²

¹ Department of Computer Science, Indiana University-Purdue University,
Fort Wayne, Indiana, USA
yooj@ipfw.edu

² Computer Science & Engineering Department, University of Minnesota,
Minneapolis, Minnesota, USA
jhwang@cs.umn.edu

Abstract. A *spatio-temporal cohesive network* represents a social network in which people often interact closely in both space and time. Spatially and temporally close people tend to share information and show homogeneous behavior. We discuss modeling social networks from spatio-temporal human activity data, and alternative interest measures for estimating the strength of subgroup cohesion in spatial and temporal space. We present an algorithm for mining spatio-temporal cohesive networks.

1 Introduction

The recent revolution in mobile aware technology, e.g., GPS, mobile phones, and in-car navigation systems, has allowed rich data to be collected about the activities of individuals. The daily activity data can form complex and dynamic networks of spatial and spatio-temporal interactions of people. Spatial social science [1] recognizes the key role that spatial concepts, such as location, distance, proximity, neighborhood, and region, play in human society. Spatially and temporally close social groups tend to share information and have homogeneous behavior in space and time. The identification of the interesting patterns can provide important insights into many application domains such as homeland defense, public health, ecology, business and education.

Social networks, in the most general sense, refer to relationships that shape a society's social interaction [11]. Sociologists have long studied spatially complete social networks based on local interaction [5,2,13]. However, there are very few theoretical works that study space and time simultaneously. In the data mining literature, there is increasing interest in mining social networks. Most works have concentrated on identifying social networks based on non spatial context. In other hand, [4] views the movements of people among specific locations as a spatial interaction problem. [7] worked on mining social networks using spatio-temporal events. However, the event is defined to a semantic event which is any social collectivity of actors, e.g., conferences and games, not having actual geographic location.

One of the key problems in mining spatio-temporal social networks is how to appropriately model human activity data to find social networks. We discuss

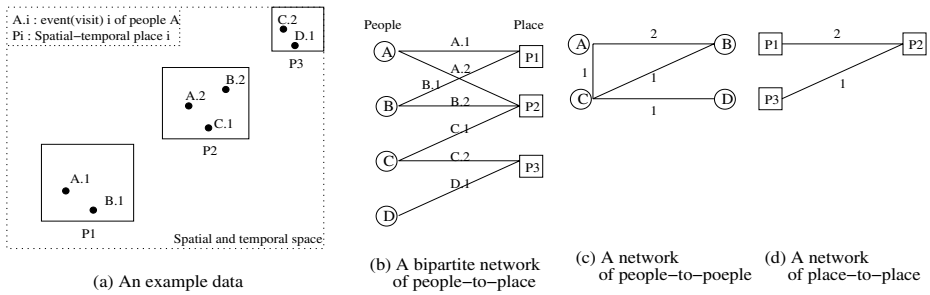


Fig. 1. Spatio-Temporal Semantic Model

the two approaches, *spatio-temporal semantic model* and *spatio-temporal location model*. In this paper, we focus on the problem to discover *spatio-temporal cohesive networks* based on the spatio-temporal location model. One important cohesive network is a *clique* [8]. A clique refers to a network in which there is a relationship between any two people. Our spatio-temporal cohesive network could be interpreted as one in which has a relationship with everyone in his/her spatial neighborhood within a time interval. Another important problem for discovering the patterns is to use proper interest measures for estimating the strength of subgroup cohesion in spatial and temporal space. We present alternative interest measures and compare them. Spatio-temporal cohesive network mining also presents computational challenges since the activity objects are embedded in continuous space and time. We extend our algorithm proposed for mining co-located itemsets in space [12] to discover spatio-temporal cohesive networks from spatio-temporal datasets.

The remainder of the paper is organized as follows. Section 2 discusses social network models of spatio-temporal human activity data. Section 3 defines the problem of mining spatio-temporal cohesive networks and presents the algorithm. Section 4 is concluded with future work.

2 Social Network Modeling

We first discuss two different approaches for modeling social networks from spatio-temporal human activity data. Then we define spatio-temporal cohesive networks and present the interest measures.

2.1 Spatio-temporal Semantic Model

The spatio-temporal semantic model considers the visit activity of people to specific places and time (intervals). Figure 1 (a) shows an example of visited places {P1, P2, P3} of four persons {A, B, C, D}. A place means a semantic location which is visited by people. For instance, a conference is an example of a semantic place in which researchers gather and exchange their work and thought. A name of a locality and a ZIP code are another example of semantic

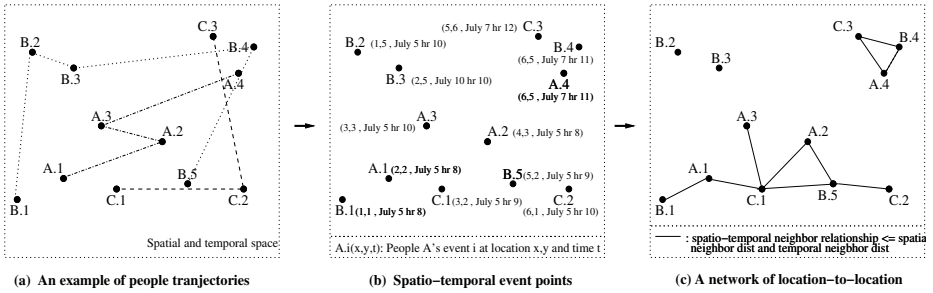


Fig. 2. Spatio-Temporal Location Model

location. The places can be different geographic locations or can be the same location with different time intervals of visit. In Figure 1 (a), each visit activity is represented by the visitor and an unique activity id per person, e.g., A.1.

People-to-Place. An individual and a place have a relationship if the person visits the place in a time interval. Bipartite graphs are often used to model people and place relations [3]. In a bipartite graph, vertices are divided into two disjoint sets, e.g., a set of people and a set of places, as Figure 1 (b). An edge may be labeled by an activity event to link its people and place. A bipartite network can be transformed into a one-mode social network which depends on the emphasis of a specific type of interaction, i.e. people to people or place to place.

People-to-People. In a people-to-people model, the people are linked in a social network based on their common visits to places. The number of common visits can be used as the weight of each link. For example, in Figure 1 (c), people A and B are linked with a weight of 2 due to their visits to places P1 and P2.

Place-to-Place. In a place-to-place network, two places are connected if they share at least one visitor. Figure 1 (d) shows the weight of connection between two places which is defined by the number of visitors.

2.2 Spatio-temporal Location Model

The spatio-temporal semantic models are limited to interactions based on specific semantic places rather than the contact of people in arbitrary geographic location. In contrast, the spatio-temporal location model captures the interaction of people in the geographic context with time. For example, suppose Figure 2 (a) shows a simplified example of people moving in space and time. An activity event can be defined with a geographic location where people stay for a while. For example, an activity event can be modeled as a tuple \langle person, event, location (x, y) , time $($ or time interval $)$, other attributes \rangle , where person is an individual who engages in an event and event is a distinguished event per person, x and y represent a geographic location where the event happens, and time is the start time (optionally including duration) of the event. The activity event can also include non spatial and temporal attributes, e.g., activity type. Figure 2 (b) shows the spatio-temporal event location points from Figure 2 (a).

Location-to-Location. The activity events in spatial and temporal space can be connected by a spatial relationship, e.g., metric relationship (e.g., Euclidean distance), topology relationship (e.g., within, nearest), and a temporal relationship, e.g., before, overlap, contain. Figure 2 (c) shows a network of location-to-location in spatial and temporal space. Two events (e.g., A.1 and B.1) are connected because they occur close to one another within a spatial distance and a temporal distance.

2.3 Spatio-temporal Cohesive Network

We model our spatio-temporal cohesive social network based on the spatio-temporal location model. Despite the variability in semantics, social networks share a common structure in which social entities are generally termed *actors* and the relationships between a pair of social entities are known as *ties* [9]. Let us suppose that A is a set of actor and E is a set of spatio-temporal events of the actors. We define a spatio-temporal cohesive network as follows.

Definition 1. A spatio-temporal cohesive network N is a subset of actors, $N \subseteq A$, whose events $I \subseteq E$ often form cliques using a spatial neighbor relationship and a temporal neighbor relationship.

We introduce two different interest measures for measuring cohesion of networks. First, cohesive index represents a probability that makes cliques from all possible events of the actors.

Definition 2. The cohesive index $CI(N)$ of a network $N = \{a_1, \dots, a_k\}$ is defined as $\frac{|cohesive\ instances\ of\ N|}{\prod_{i=1}^k |events\ of\ a_i|}$, where a_i is an actor of N and Π is the multiplication function.

In the definition, the cohesive instance of a network N is a subset of events, $I \subseteq E$, that includes the events of all actors in the network N and forms a clique using a spatial relationship and a temporal relationship.

Definition 3. The participation index $PI(N)$ of a network $N = \{a_1, \dots, a_k\}$ is defined as $\min_{a_i \in N} \left\{ \frac{|\pi_{a_i}(cohesive\ instances\ of\ N)|}{|events\ of\ a_i|} \right\}$, where π_{a_i} is selection of distinct events of actor a_i .

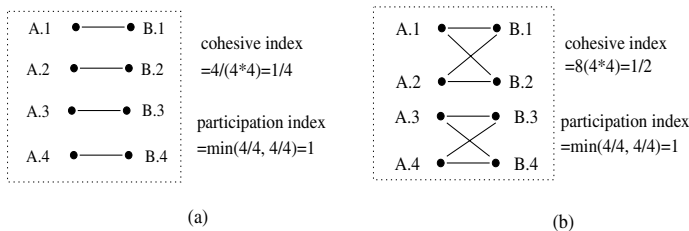


Fig. 3. Comparison of Interest Measures

The participation index was first introduced for spatial co-location mining in [10]. We adopt the measure for measuring the cohesion of objects in space and time. The participation index considers the ratio of actor events who participate in a cohesive network. A high cohesive index or participation index indicates that the actors in the social network likely have ties together through their activities. However, the two interest measures have different characteristics. Figure 3 shows a comparison of them. In the case of Figure 3 (a), each event of actor A (or B) has a neighbor relationship with the corresponding event of actor B(or A). The participation index seems to perfectly capture the co-occurrence relationship with the value 1. In contrast, the cohesive index shows a low strength value since only four pairs among all possible 16 pairs have neighbor relationships. Next, let us consider the case of Figure 3 (b). The cohesive index reflects the strength of neighbor relationships with the increase of neighbor pairs. In contrast, the participation index does not reflect the strength well. In both datasets, the participation index shows the same prevalence values. The choice of interest measure leaves to the application.

3 Spatio-temporal Cohesive Network Mining

We define the problem of mining spatio-temporal cohesive networks as follows. Given a set of actor $A = \{a_1, \dots, a_n\}$, a set of spatio-temporal events $E = E_1 \cup \dots \cup E_n$, where $E_i (1 \leq i \leq n)$ is a set of events of an actor a_i , a spatial distance neighbor relationship, a temporal distance neighbor relationship, and a minimum prevalent threshold, we want to find spatio-temporal cohesive networks whose prevalence values (CI or PI) are greater than the prevalence threshold.

Spatio-temporal cohesive network mining presents computational challenges since the event objects are embedded in continuous space and time. It is hard to transactionize a spatial-temporal dataset to apply traditional data mining techniques. Another way is to find all spatio-temporal neighboring object pairs from the input dataset, represent them to a neighborhood graph, and then find all cohesive instances from the graph for calculating the prevalence values. However, finding all cliques from the graph is NP-complete in the graph-theory. It is also non-trivial to reuse subgraph mining [6]. The subgraph mining was used to find frequent subgraphs in a large graph database (i.e., a set of graphs). We extend our algorithm proposed for mining co-located feature sets in space [12] to discover spatio-temporal cohesive networks from spatio-temporal datasets. The following shows the pseudo code of the algorithm.

4 Conclusion

This paper presents a framework for mining cohesive networks to study the interaction of people in space and time. We need detail optimization of the proposed algorithm according to spatial and temporal characteristics of data, and the evaluation of the scalability with large real datasets. In the future, we also plan to explore more relaxed structures as well as clique as spatio-temporal cohesive subgroups, and develop the mining algorithm.

Inputs

A : a set of actors, E : a spatio-temporal event dataset
 SR : a spatial neighbor relationship, TR : a temporal neighbor relationship, min_prev : minimum prevalence threshold

Output

Spatio-temporal cohesive networks whose prevalence $\geq min_prev$

Variables

NR : a neighborhood relation table, C_k : a set of size k candidate networks, CI_k : candidate cohesive instances of size k networks, TI_k : true cohesive instances of size k networks, P_k : a set of size k prevalent spatio-temporal cohesive networks

Method

```

1)  $NR = \text{gen\_star\_neighborhood\_relation\_table}(E, SR, TR)$ ;
2)  $P_1 = A$ ;  $k = 2$ ;
3) while (not empty  $P_{k-1}$ ) do
4)    $C_k = \text{gen\_candidate\_networks}(P_{k-1})$ ;
5)   for  $r \in NR$  do
6)      $CI_k = \text{filter\_star\_instances}(C_k, r)$ ;
7)   end do
8)   if  $k = 2$  then  $TI_k = CI_k$ 
9)   else do  $C_k = \text{filter\_coarse\_prev\_networks}(C_k, CI_k, min\_prev)$ ;
10)       $TI_k = \text{filter\_clique\_instances}(C_k, CI_k)$ ;
11)   end do
12)   $P_k = \text{find\_prev\_cohesive\_networks}(C_k, TI_k, min\_prev)$ ;
13)   $k = k + 1$ ;
14) end do
15) return  $\bigcup(P_2, \dots, P_k)$ ;

```

References

- Center for spatially integrated social science. Center for Spatially Integrated Social Science, <http://www.csiss.org/>
- Coleman, J.S.: Foundations of Social Theory. Harvard University Press (1990)
- Eubank, S., Guclu, H., Anil Kumar, A., Marathe, M., Srinivasan, A., Toroczkai, Z., Wang, N.: Modeling Disease Outbreaks in Realistic Urban Social Networks. *Nature* 429, 180–184 (2004)
- Guo, D.: Mining and Visualizing Spatial Interaction patterns for Pandemic Response. In: Workshop on Spatial Data Mining, SIAM Conf. on Data Mining (2006)
- Johnson, C., Gilles, R.P.: Spatial Social Networks. *Review of Economic Design* 5, 273–300 (2000)
- Kuramochi, M., Karypis, G.: Frequent Subgraph Discovery. In: Proc. of IEEE Intl. Conf. on Data Mining (2001)
- Lauw, H.W., Lim, E.P., Tan, T.T., Pang, H.H.: Mining Social Network from Spatio-Temporal Events. In: Workshop on Link Analysis, Counterterrorism and Security (2005)
- Luce, R., Perry, A.: A Method of Matrix Analysis of Group Structure. *Psychometrika* 14, 95–116 (1949)
- Scott, J.: *Social Network Analysis: A Handbook*, 2nd edn. Sage, Thousand Oaks (2000)
- Shekhar, S., Huang, Y.: Co-location Rules Mining: A Summary of Results. In: Proc. of Intl. Sym. on Spatio and Temporal Database (2001)
- Wasserman, S., Faust, K.: *Social Network Analysis*. Cambridge University Press, Cambridge (1994)
- Yoo, J.S., Shekhar, S., Celik, M.: A Join-less Approach for Spatial Co-location Mining: A Summary of Results. In: Proc. of Fifth IEEE Intl. Conf. on Data Mining (2005)
- Goodchid, M.F., Janelle, D.G.: *Spatially Integrated Social Science*. Oxford University Press, Oxford (2004)