An Exploration of Trajectory Data using Complex Networks
Summary

- Introduction
- Methodology
  - Experiments
- Open issues
- Next steps
The growth of GPS-equipped devices.

Spatio-temporal data have been used by several applications:

- traffic management;
- displacement of animals (bird migration);
- vehicles monitoring (trucks).

These applications need to analyze a huge amount of positioning data.
On the other hand, complex network has been receiving much attention to analyze data.

Besides, complex network works on large data.

Large networks:

- Tweets: 17,069,982 nodes and 476,553,560 edges
- WWW Altavista: 203,549,046 nodes and 2,130,000,000 edges

Then, we want to analyze spatio-temporal data from a complex network point of view.
Trajectory is a spatio-temporal evolution of a moving object [Spaccapietra et al, 2008]

- Tuple (id, x, y, t)

Complex network is represented as a graph with nodes (vertices, points) and edges (links, connections) [Newman, 2003].
Important properties of complex networks:

- Node degree: number of nodes linked to a node;
- Average shortest path length: average of all shortest path;
- Diameter: maximum shortest path;
- Component: subgraph in which any two vertices are connected by a path;
- Power law: distribution (degree) follows a function: \( p(x) = a \times x^{-\alpha} \) ("The rich get richer")
Introduction

- Clustering coefficient: if A links B and B links C, then there is a high probability of A links C ("Friend of my friend is my friend");

- Community detection: groups of nodes in a network that are more densely connected internally than with the rest of the network
Methodology

- Combine complex networks and trajectories.
- We have to map trajectory data to a complex network (graph).
- Some possibilities:
  - Each point in the dataset is a node;
  - each trajectory is a node;
  - each location(event) is a node;
  - and so on.
Methodology

We have thought about two approaches:

1. Each trajectory as a node:
   - Edge is defined by a similarity function.

2. Each event as a node:
   - Edge is defined by shared trajectories.
First Approach

- It aims at creating a network of trajectories with a similarity function.

- A set $S$ of trajectories is represented as network $(N,E)$ with a similarity function $f$ between trajectories and a threshold constant $c$.
  - Each node in $N$ represents a trajectory;
  - Two nodes $n,m$ form an edge iff $f(n,m) \geq c$;
First approach

- We define a function $f(s,t,k)$:
  - $s,t,k$ are spatial, temporal and frequency parameters of $f$ respectively.

- They induce a buffer $B[s,t](T)$.

- Let $T$ and $U$ be two trajectories;
  - meet (or collide): $T$ and $U$ meet iff $B[s,t](T)$ and $B[s,t](U)$ overlap;
  - encounter: there is an encounter between $T$ and $U$ iff $T$ and $U$ meet more than $k$ times.
First approach

- Other similarity function of trajectory can be used:
  - speed;
  - shape;
  - and so on.

- Therefore, the network will represent the relationships between trajectories through a given similarity function.
Experiment 1

- Experiments on displacement of vehicles in Milan, Italy.
- The data are “raw” - [id, latitude, longitude, time] - and it has already been preprocessed (elimination of outliers, noises).
- Trajectories of each day of the week are presented in 7 distinct files.
- For each day, we have used 4 parameters.

<table>
<thead>
<tr>
<th>Day</th>
<th>Number of trajectories</th>
<th>Average number of point per trajectory</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sunday</td>
<td>23535</td>
<td>8.290461</td>
</tr>
<tr>
<td>Monday</td>
<td>34812</td>
<td>8.927956</td>
</tr>
<tr>
<td>Tuesday</td>
<td>36824</td>
<td>9.206279</td>
</tr>
<tr>
<td>Wednesday</td>
<td>36023</td>
<td>9.467285</td>
</tr>
<tr>
<td>Thursday</td>
<td>35340</td>
<td>9.871647</td>
</tr>
<tr>
<td>Friday</td>
<td>33822</td>
<td>8.697179</td>
</tr>
<tr>
<td>Saturday</td>
<td>25576</td>
<td>7.74695</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Experiment Minimum Frequency Spatial Distance Temporal Distance</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 3 0.3 30</td>
</tr>
<tr>
<td>2 3 0.3 15</td>
</tr>
<tr>
<td>3 AVG 0.3 30</td>
</tr>
<tr>
<td>4 AVG 0.3 15</td>
</tr>
</tbody>
</table>
In fact, we have found many features.

However, the interpretation is not easy.

### Experiment 1

| Experiment Minimum Frequency Spatial Distance Temporal Distance |
|------------------|------------------|------------------|------------------|
| 1                | 3                | 0.3              | 30               |
| 2                | 3                | 0.3              | 15               |
| 3                | AVG              | 0.3              | 30               |
| 4                | AVG              | 0.3              | 15               |

![Trajectory Network - Tuesday (3 | 0.3 | 30)](image)
Experiment 1

Clustering coefficient  Average shortest path length  Diameter
Experiment I
Open Issues

- It is not cheap to compute the encounters among trajectories;
- Interpretation is not easy.
- What does the network offer that existing trajectory methods do not offer?
- Different kinds of network can be built.
Next steps

- Community detection on the network:
  - Maybe it does not make sense for customers;
  - However, we can use other dataset to detect communities of locations: recommendations and “life style”.

- Build a network from T-patterns:
  - What is the main advantage of having global view of T-patterns

- Investigate recommendation coefficient.
Thanks!