Profiling users of the Vélo ’v bike sharing system
Albrecht Zimmermann, Mehdi Kaytoue, Marc Plantevit, Céline Robardet, Jean-François Boulicaut

To cite this version:
Albrecht Zimmermann, Mehdi Kaytoue, Marc Plantevit, Céline Robardet, Jean-François Boulicaut. Profiling users of the Vélo ’v bike sharing system. 2nd International Workshop on Mining Urban Data (MUD), Ioannis Katakis, François Schnitzler, Thomas Liebig, Dimitrios Gunopulos, Katharina Morik, Gennady L. Andrienko, Shie Mannor, Jul 2015, Lille, France. pp.63-64. hal-01193017

HAL Id: hal-01193017
https://hal.archives-ouvertes.fr/hal-01193017
Submitted on 4 Sep 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
profiling users of the véló’v bike sharing system

albrecht zimmermann, mehdi kaytoue, cédile robardet, jean-françois boulicaut

insa-lyon, cnrs, liris umr5205, f-69621, france

marc plantévit

université lyon 1, cnrs, liris umr5205, f-69622, france

abstract

detecting and characterizing geographical areas that are attractive places for specific people, in specific contexts, is an important but challenging new problem. mobility traces and their related circumstances can be modeled thanks to an augmented graph in which nodes denote geographic locations and edges are represented by a set of transactions that describe users’ demographic information (e.g. age, gender, etc.) as well as the conditions of the movement (e.g. day/night, holiday, transportation mode, etc.). we propose to extract connected subgraphs that are related to some user profiles, and use it to understand the usages of the véló’v bike sharing system.

1. introduction

the problem considered hereafter is how to detect and characterize geographical areas that are attractive places and routes for specific contexts. such areas are frequently accessed together in certain conditions by users of similar profiles compared to all contexts and users. starting from a relational database that gathers information on people movements – such as origin, destination, date and time of travel, means of transport, reasons for traveling, etc. – as well as demographic data, we adopt a graph-based representation that results from the aggregation of individual travels. in such a graph, the vertices are locations or points of interest (poi) and the edges stand for user’s co-visitations. travel information as well as user demographics are labels associated to the edges of the graph.

2. travel patterns in the véló’v system

vélo’v is the bicycle sharing and renting system run by the city of lyon (france) and the company jcdexaux. the véló’v dataset contains movement data collected between jan. 2011 and dec. 2012. each movement includes both bicycle stations and timestamps for departure and arrival, as well as some basic demographics about the user of the bike. we aggregated all movements a user performed between any two stations for the entire time period. hence, the véló’v stations are nodes in the graph (342 in total),

http://www.velo.grandlyon.com/
Identifying demographic specific areas from mobility profiles

<table>
<thead>
<tr>
<th>User</th>
<th>Gender</th>
<th>Age</th>
<th>Time</th>
<th>Travels</th>
</tr>
</thead>
<tbody>
<tr>
<td>u1_1</td>
<td>F</td>
<td>20</td>
<td>Day</td>
<td>(A,C),(B,A),(C,D)</td>
</tr>
<tr>
<td>u2_1</td>
<td>F</td>
<td>20</td>
<td>Night</td>
<td>(D,E),(E,D),(D,A)</td>
</tr>
<tr>
<td>u3_2</td>
<td>M</td>
<td>23</td>
<td>Day</td>
<td>(B,C),(C,D),(C,E)</td>
</tr>
<tr>
<td>u4_2</td>
<td>M</td>
<td>23</td>
<td>Night</td>
<td>(D,A),(D,E),(E,D)</td>
</tr>
<tr>
<td>u1_3</td>
<td>F</td>
<td>45</td>
<td>Day</td>
<td>(A,B),(B,C),(C,B)</td>
</tr>
<tr>
<td>u3_3</td>
<td>F</td>
<td>45</td>
<td>Night</td>
<td>(B,D),(D,E),(E,B)</td>
</tr>
<tr>
<td>u5_4</td>
<td>M</td>
<td>50</td>
<td>Day</td>
<td>(A,B),(B,C),(C,B)</td>
</tr>
<tr>
<td>u5_4</td>
<td>M</td>
<td>50</td>
<td>Night</td>
<td>(D,E),(E,D),(E,B)</td>
</tr>
</tbody>
</table>

Figure 1. Example of contextualized trajectories: (a) Transactional view; (b) Aggregate graph w.r.t. the most general context \( * = (\text{Age} \in [20, 50], \text{Gender} \in \{F, M\}, \text{Time} \in \{\text{Day}, \text{Night}\}) \); (c) Aggregate graph w.r.t. context \( (\text{Age} \in [45, 50], \text{Time} = \text{Day}) \); (d) Aggregate graph w.r.t. context \( (\text{Gender} = M) \); (e) Aggregate graph w.r.t. context \( (\text{Age} \in [45, 50], \text{Time} = \text{Day}, \text{Gender} = M) \);

and edges link two stations if a VÉLO’V customer checked out a bicycle at the first station and returned it at the second one. We treat the edges as undirected. Customers are described by nominal attributes such as gender, type of membership card, ZIP code and country of residence, as well as a numerical one: year of birth. There are a total 50,601 customers. The data set comprises around 2 million contextualized edges.

Given the characteristics of different users, we aim to identify populations that use the rental bicycles in a particular manner. Figure 2 shows 4 different DCSA from VÉLO’V. Pattern (a) identifies people born after 1968, living in a city (Saint Chamond) located approximately 50km from Lyon. It is therefore not surprising that the edges involve the two main train stations of Lyon: Perrache (south-west) and Part-Dieu (center), from which users take bicycles to areas that are not easily reached by metro or tram, such as the 1st and 4th arrondissements. The edges of pattern (b) radiate from all of Lyon’s train stations, not only the major ones. Its description refers to holders of a regional train subscription, and the pattern notably involves 200 nodes, almost 60% of the stations. It is very likely that this pattern identifies commuters. Pattern (c) involves users born after 1980 and can identify three main areas: the scientific campus in the north, the Presqu’île and its pubs, and the shopping area in the city center. It is notable that several of the long edges correspond to very comfortable cycling routes. Pattern (d) does not seem to be very exciting: young people that live in the 3rd arrondissement use VÉLO’V bicycles to move around in their area. At a second glance, however, this is the closest that we will come to a ground truth in real-world data: the ZIP code of users aligns with the area where the bicycles are used!

3. Conclusion

The problem of finding DCSA in edge-attributed graphs has many applications in location based social networks and recommendation systems. It allows to find connected components highly characteristic of a given category of users. The proposed inductive approach is solved thanks to an efficient data mining algorithm that avoids materializing all contexts/induced-graph pairs and benefits from pruning and upper bound computations techniques. Its use for the analysis of the bicycle sharing system VÉLO’V demonstrates its capabilities to provide new valuable insights.\(^2\)

\(^2\)This work has been partially supported by the projects GRAISearch (FP7-PEOPLE-2013-IAPP) and VEL’INNOV (ANR INOV 2012).

---

\(1\) YoB \(\geq 1968, ZIP = 42400\)
\(2\) YoB \(\geq 1962, CAT = OURA\)
\(3\) YoB \(\geq 1980, TYP = \text{std}\)
\(4\) YoB \(\geq 1992, ZIP = 69003\)