INVESTIGATION OF HUMAN TRAVEL AND ACTIVITY PATTERNS BY LEVERAGING FOURSQUARE LOCATION-BASED SOCIAL NETWORK DATA

Student: ***Florian Cartuta***

1. INTRODUCTION: A description of the problem and a discussion of the background.

Researchers are interested to understand human activity patterns and travel behaviors. Conventionally, such information was in format of travel diaries, but the data is not highly available and difficult to gather by such methods. Also information about users gender, age, are available on Foursquare and can be correlated with the traffic and activity patterns in order to further understand gender differences in travel and activity which is a popular research topic. Diaries are used to record human daily movements, places visited, activities, time of day, duration of time spent at each place (venue). A diary is “a record of information in relation to the passage of time”

Information on the movement of people in the city can be used for: urban planning, location management, transportation management, etc thus being of interest also for local authorities and companies.

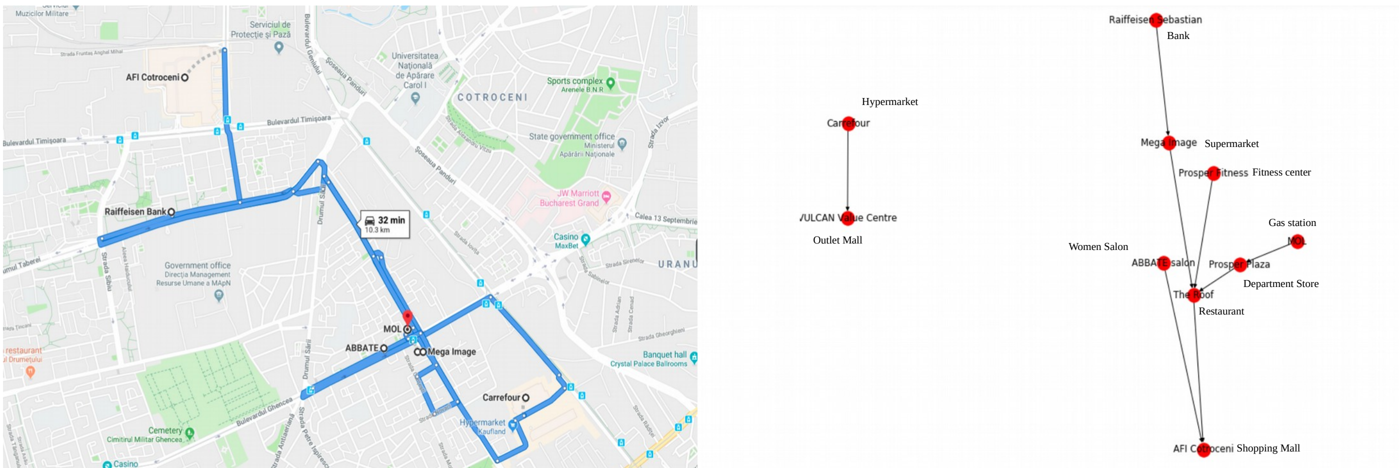
In my project, I will present an approach to understand the human activity and pattern behaviors in my town: *Bucharest, Romania*, using Foursquare’s location API (the GET NextVenue method).

Online social media, like geotagged travel photos, can be used as an alternative source of travel data. However, it is seriously limited because they cannot provide detailed contextual information on the activities for further analysis.

Suppose a group of people, report often visits to several locations in Bucharest, Romania. Figure 1a shows people visited several locations as pinpointed on the map. However *Figure 1b* provides much richer information such as traveling to gas station (‘MOL’), eating at the restaurant (‘The Roof’), shopping in the malls (‘Carrefour’, ‘AFI Cotroceni’), going to fitness, etc shown on a *directed graph*.

Travel Representation:

In LBSNs such as Foursquare, each venue represents a physical location (see Figure 1). Common types of venues include banks, restaurants, offices, apartments, hotels, bus stops, shops and gyms. In Figure 1, for instance, there is a pattern of users checks in at venue A (Prosper – Fitness Center) and venue B (‘The Roof’ restaurant) and venue C (‘AFI Cotroceni’ - shopping mall), consecutively. We can, therefore, deduce that people often moves from venue Fitness to venue Restaurant than to C Shopping Mall, irrespective of the specific route taken by the user between the two venues. A series of consecutive check-ins can record the user’s travel diary. Of note is that unlike GPS traces, check-ins seldom reflect the precise travel routes of users between distinct locations.



(a) Without Activity Information (b) With Activity Information (directed graph)

Figure 1: Visited locations in Bucharest Romania, neighborhood: ‘13 Septembrie’

The availability of location-aware mobile social applications like Foursquare changed the way people indicate their activity information. Users can now share their location in the form of venue check-in, with meta-data such as venue name, category, *next venue* from which they traveled from the current venue. Therefore, accurate identification of activity is possible.

From the Foursquare API, I used: *Get Next Venues* which returns venues that people often check in to after the current venue. Up to 5 venues are returned in each query, and results are sorted by how many people have visited that venue after the current one.

2. DATA COLLECTION : A description of the data and how it will be used to solve the problem

First step was to retrieve the list of Bucharest neighborhoods from Wikipedia: Bucharest_Romania_neighborhoods. Subsequently, I built a pandas dataframe containing neighborhood name as index and geographical coordinates: latitude and longitude as columns (fig 2). I obtained the geographical coordinates of the neighborhoods using the geopy Python client for geocoding web services.

Neighborhood	Latitude	Longitude
13 Septembrie	44.421519	26.065115
Aviatorilor	44.459504	26.081038
Aviației	44.482457	26.093999
Balta Albă	44.417151	26.181877
Baicului	44.444301	26.141879

Figure 2: Geographical coordinates of several neighborhoods in Bucharest, Romania

For the purpose of my project and due to the limitations of the Foursquare free account, I decided to restrict the identification of human travel and activity patterns to only one neighborhood, named: '*13 Septembrie*' (first on the previous list). But there is no limitation (except those of the Foursquare account) to apply same methodology for entire Bucharest city.

In the third step, I retrieved the information (venue name, location and category) for all the venues in the neighborhood (~100 venues), using Foursquare API SEARCH endpoint with a radius of 750 meters around the neighborhood central coordinates (figure 3)

	Neighborhood	Neighborhood Latitude	Neighborhood Longitude	Venue	Venue Latitude	Venue Longitude	Venue Category
0	13 Septembrie	44.421519	26.065115	ABBATE Salon	44.420100	26.063916	Salon / Barbershop
1	13 Septembrie	44.421519	26.065115	Prosper Fitness	44.420153	26.065467	Gym / Fitness Center
2	13 Septembrie	44.421519	26.065115	Green Day Spa	44.425332	26.062242	Cosmetics Shop
3	13 Septembrie	44.421519	26.065115	The Roof	44.420061	26.066116	Restaurant
4	13 Septembrie	44.421519	26.065115	LIDL	44.424773	26.063779	Supermarket
5	13 Septembrie	44.421519	26.065115	Casa Bia	44.416991	26.069137	Eastern European Restaurant
6	13 Septembrie	44.421519	26.065115	TAJ	44.422938	26.073669	Indian Restaurant
7	13 Septembrie	44.421519	26.065115	Casa Brândușa Restaurant & Garden	44.416964	26.065128	Restaurant
8	13 Septembrie	44.421519	26.065115	ABBATE salon	44.420111	26.063879	Salon / Barbershop
9	13 Septembrie	44.421519	26.065115	Snatch Pub	44.425732	26.070151	Pub
10	13 Septembrie	44.421519	26.065115	Wellness Cuisine	44.427574	26.061519	Restaurant
11	13 Septembrie	44.421519	26.065115	Mesopotamia	44.417890	26.072145	Doner Restaurant
12	13 Septembrie	44.421519	26.065115	Carrefour	44.418093	26.072153	Department Store
13	13 Septembrie	44.421519	26.065115	Anturaj	44.423165	26.063959	Bar
14	13 Septembrie	44.421519	26.065115	Westlife Fitness	44.427573	26.066398	Gym

Figure 3: Geographical coordinates, name and category of all venues in '13 Septembrie' neighborhood

For all venues of the '13 Septembrie' neighborhood, I registered the Next (visited) Venues in a pandas dataframe from which I will later build the directed graphs showing the travel and activity patterns like in Figure 1b.

	Source_name	Source_categ	Source_id	Source_lat	Source_lng	Source_addr	Target_name	Target_categ	Target_id
0	MOL	Gas Station	4d14cf30401db60cdf4ddba4	44.420797	26.065114	Calea 13 Septembrie nr. 204-206A	Prosper Plaza	Department Store	4c7543892db5236a2f74bc79
0	ABBATE salon	Salon / Barbershop	4c4716cd417b20a11f68dca9	44.420111	26.063879	Calea 13 septembrie nr. 206	AFI Cotroceni	Shopping Mall	4ba79045f964a520c79d39e3
0	Mega Image	Supermarket	4c94d3ed94a0236abef68f12	44.420332	26.066003	Calea 13 Septembrie nr. 221-225	The Roof	Restaurant	4e9870e1775bdedf44426987
0	Prosper Plaza	Department Store	4c7543892db5236a2f74bc79	44.420046	26.065664	Calea 13 Septembrie 221-225	The Roof	Restaurant	4e9870e1775bdedf44426987
0	The Roof	Restaurant	4e9870e1775bdedf44426987	44.420061	26.066116	Calea 13 Septembrie nr. 221-223	AFI Cotroceni	Shopping Mall	4ba79045f964a520c79d39e3
0	Carrefour	Department Store	5407ec17498e5458a8a4d1bf	44.418093	26.072153	Str. Mihail Sebastian nr. 88	VULCAN Value Centre	Outlet Mall	5407d878498eef5e68d11aee
0	Prosper Fitness	Gym / Fitness Center	4e528116ae6054e936229adf	44.420153	26.065467	Calea 13 Septembrie	The Roof	Restaurant	4e9870e1775bdedf44426987

3. Methodology:

In this section, the approach used to investigate frequent activity and travel patterns of LBSN users within ‘13 Septembrie’ Bucharest neighborhood is presented. In order to retrieve the venues of the neighborhood, the Foursquare API ‘Explore’ end-point was used. In the next step, for each venue id, was tried to get the venues where people often check-in after the current venue, using the ‘Get Next Venues’ end-point. The above process was continued until no more “next venues” were available.

The Get Next Venues request was used to form relationships between venues showing the predominant travel and activity patterns in the neighborhood.

Results were stored in a pandas dataframe and were presented in a directed graph using NetworkX python library (Figure 4).

4. Results:

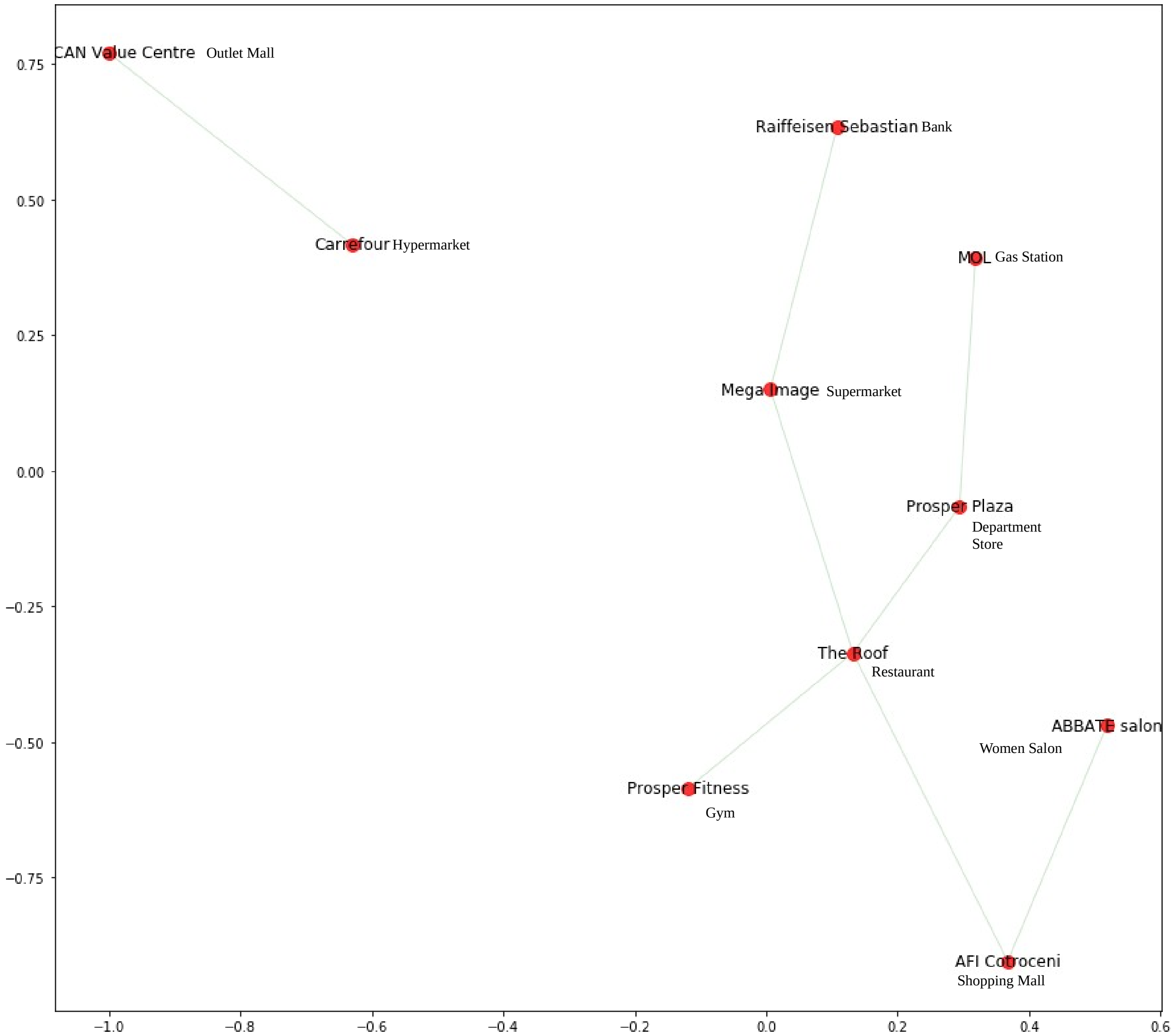


Figure 4: Frequent travel and activity patterns in ‘13 Septembrie’ Bucharest neighborhood

Looking at the directed graph in figure 4, it is observed that the most often activities in the neighborhood were visits to the bank, shopping centers, hair styling saloons and fitness centers.
It also can be observed that there were not identified frequent check-ins into public transportation facilities which is consistent with the fact that this neighborhood is not covered by the Bucharest metro line. Also there are no frequent check-ins to other type of cultural venue categories like museums, theaters, etc which shows another attribute of this neighborhood which lacks such important cultural venues.

5. Discussion

This project only uncovers a small part of what can be done using machine learning and LBSN data. There are interesting opportunities which can be followed as next step, for example to explore how gender influences travel and activity patterns.
There are many directions which can be followed using this methodology and the full extent of the LBSN data. I am interested to explore some of them in the future and would recommend to other machine learning enthusiasts.

6. Conclusion

This paper reports on the empirical investigation, using Foursquare check-in data, of how neighborhood planning influences travel and activity patterns of active social media users. The empirical results reveal that main travel and activity patterns of active local users in ‘13 Septembrie’ neighborhood focuses around shopping, banking and restaurants but there is a lack of important socio-cultural venues like: theaters, museums, etc.
Also it emphasizes the lack of metro line coverage in this neighborhood.
These findings can be taken into consideration by the local authorities to improve the urban planning.