Based on Soho of Rome

Measures and Algorithms for finding similar neighborhoods in Cities.  
By – G´eraud Le Falhery, Aristides Gionis and Michael Mathioudakis

Finding Similar Neighborhood

Karan Jeet Singh – A20327513

Kanwal Preet Singh – A20327416

**Introduction**

This project deals with a simple yet intriguing problem of a traveler. Suppose you travelling from City of New York in United States to London in UK and you are fairly new to London. In order to feel more like home you would be interested in visiting the neighborhood which are similar to the neighborhood back in your hometown. As visiting a similar neighborhood not only gives you a sense of being at home, it also make your stay more pleasant and enjoyable. With this in mind, we are trying to solve this problem by finding out a neighborhood in one city similar to a neighborhood in another city.

**Data**

Quality information regarding a place can be obtained from today’s location aware social media platforms. For our project we are using the data provided by the FourSquare Application, it is based on the approach developed by G´eraud Le Falhery, Aristides Gionis and Michael Mathioudakis. By default the information shared on FourSqaure is private. We were only able to collect the data published by the user on twitter. As the data from twitter is publically available. When a chooses to share his/her FourSquare checkin via a tweet, partial information from his/her FourSquare checkin is also embedded in it. We use the information to find out about the venues where users have checkedin. This helps us to find out various venues present in a city. We are only using the top rated venues in FourSquare to reduce our search domain as including all venues in a city requires a lot of computation power. When venues are gathered, we identify individual venue with their features respectively. Then we calculate the dissimilarity between the two neighborhoods of the different cities. That is the set of venues that are geographically close to each other. The neighborhoods then with the least dissimilarity are selected. We gathered data using the twitters live stream. Also we used the historical data provided to establish ground truth and which further helped in identify the neighborhoods and also, helped in finding the similarities as well.

INFORMATION REGARDING THE DATA

**Methods**

**Experiments**

**Conclusion and Future work**

We can further improve our approach by the following approaches:

* Including more data from FourSquare because as of now were are only using the data published on twitter.
* It fails to detect neighborhoods in cities with low or no FourSquare checkins on twitter like in citi of Chicago.
* Including information from various photo sharing platforms as we can also retrieve the information regarding a place using geo tagged images.
* Including information regarding the transportation and weather as well.
* Environmental information like the air quality can also be included.