

## **DATA 620 Final Project Proposal**

**7/11/2021**

**Authors: John Mazon, LeTicia Cancel, Bharani Nitalla**

**Data Source:** <https://sites.google.com/site/yangdingqi/home/foursquare-dataset>

Location based social networks have attracted millions of users and massively contain their digital footprints. The dataset has crawled a part of these digital footprints from Foursquare in order to study the problems of personalized location recommendation and search. This dataset includes check-in and tip data of restaurant venues in NYC collected from Foursquare from 24 October 2011 to 20 February 2012. It contains 3112 users and 3298 venues with 27149 check-ins and 10377 tips.

checkin\_network has two columns. Each line represents a check-in event. The first column is user ID, while the second column is venue ID.

tips\_network has three columns. Each line represents a tip/comment a user left on a venue. The first and second columns are user ID and venue ID, respectively. The third column is tip text.

All user and venue IDs are anonymized, but they can be matched across the two files. In this dataset, tips may be observed even if a user didn't check-in at a venue, because check-ins are collected in a period of four months while tips are collected without time limitation.

### **Hypothesis**

Foursquare users explicitly express their interests in venues in two ways, including checking in and leaving tips to venues. We suspect a venue being frequently checked in indicates that the venue is popular in a sense that lots of people visit it and like to announce their visits to their friends. A venue being frequently tipped indicates that people are interested in the venue and would like to share their experience with all other users. It'd be interesting to know whether any community structures also exist, if two venues happen to belong to the same community, it means people often go from one to another, indicating these venues are popular among people with some similarities, possibly in taste, culture, occupation or lifestyle.

### **Preliminary Plan & Concerns**

The preliminary plan is to explore different networks since we have many data points that can be analyzed for relationships/communities. Once we have useful networks we can then further explore them based on the data such as comparison of communities of users who leave tips or comments.

### **Roles/Responsibilities**

John Mazon – Improve network analysis code, all centrality measure analysis and assist with final review

LeTicia Cancel – Final project proposal document, first draft of network analysis code including structuring data , assist with final review

Bharani Nittala – Sentiment analysis code, graphical text analysis and assist with final review