

# TIME 2 OUT NYC

- WITH MACHINE LEARNING

---

Yuxiang Zhang, Shi Fan, Xiaoge Wu, Tianyi Gu



# Introduction

- ❑ A vast database of reviews and ratings available in social media
- ❑ Reading all the reviews of each business alone is time consuming
- ❑ As a result, we intended to develop a venue recommendation system to help New Yorkers find the best places to visit in New York City
- ❑ Utilize geospatial social media data to calculate similarities between users, according to their historical venue check-ins
- ❑ Compare their historical check-ins with those users similar to them and those other check-in users they follow

## Where to hang out?





# Data

- ❑ Indirect Foursquare check-in records shared on Twitter
- ❑ 113 GB geotagged Tweets around NYC
- ❑ From February 2014 to February 2015
- ❑ Features include User ID, Name and Bios; Check-in message, time and location (latitude, longitude & city name)
- ❑ Remove irrelevant venues (i.e. airport, bus stops, offices...) and non-frequent users (less than 5 check-ins in 13 months)
- ❑ Final dataset: 4,365 users, 838 venues, and 54,017 check-ins



**FOURSQUARE**





# Data

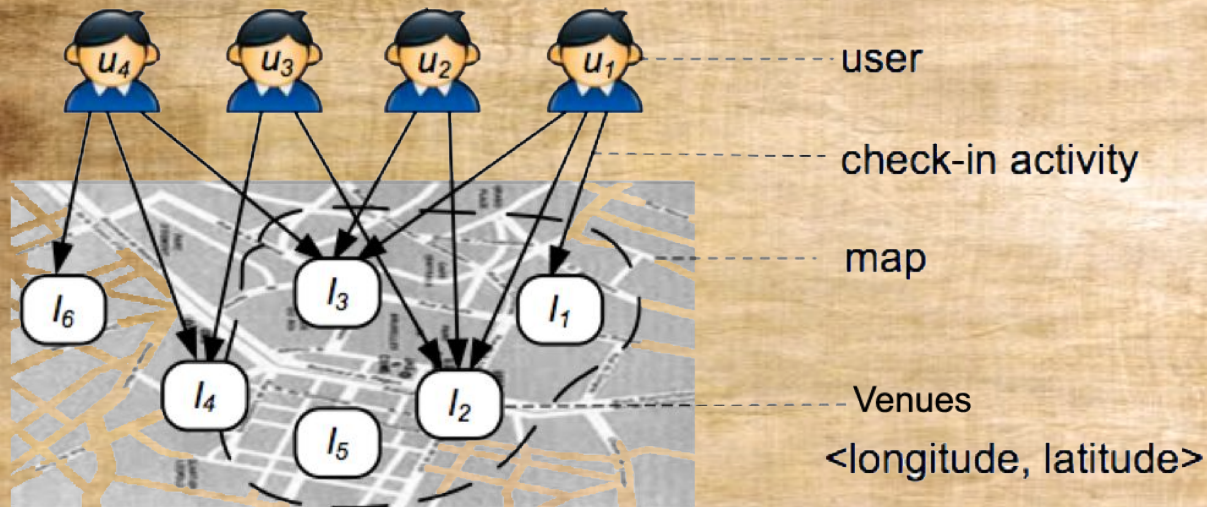
- ❑ The Twitter data were collected from the Twitter's API by the Visualization and Data Analysis Lab at New York University (VIDA-NYU).
- ❑ Special thanks to Professor Huy T. Vo for making the data available to us.





# Methodology User-Based Collaborative Filtering

WHAT IS THAT?



Check-In Record:  
 $C(i,j)$  in  $[0,1]$

User Similarity:

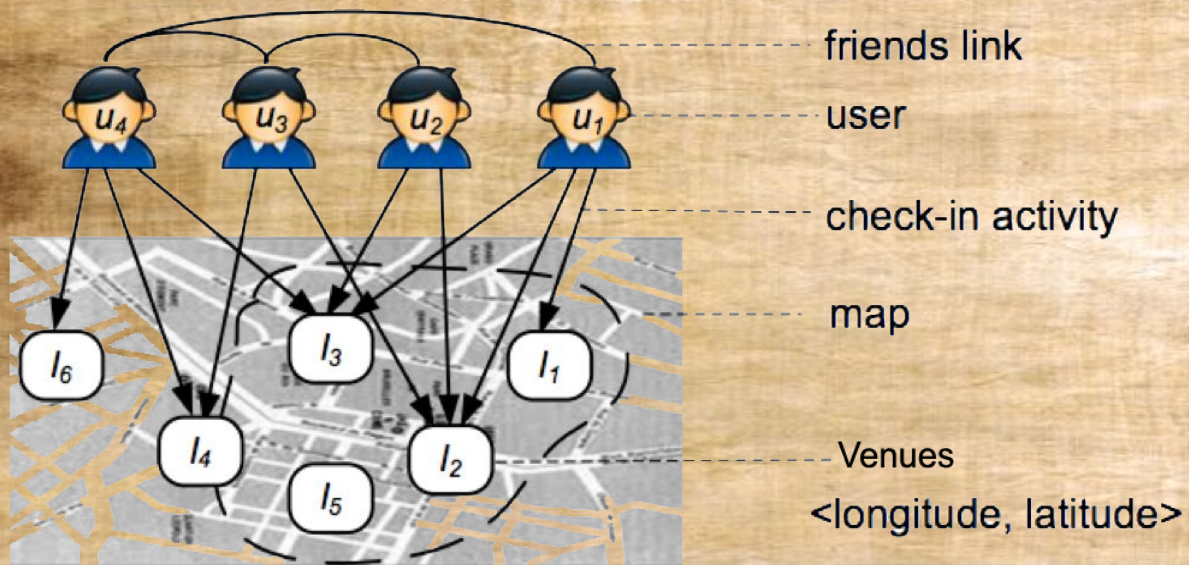
$$\text{Similarity} = \frac{\sum_{k \in V} c_{i,k} \times c_{j,k}}{\sqrt{\sum_{k \in V} c_{i,k}^2} \times \sqrt{\sum_{k \in V} c_{j,k}^2}}$$

Check-In Probability:  
 $\hat{C}(i,j)$  in  $[0,1]$



# Methodology Friend-Based Collaborative Filtering

WHAT IS DIFFERENT?



Following certain users may indicate a tendency of referring to their online activities



# Methodology    Spectral Clustering

- ❑ Based on user similarity matrix, we can view the set of user as a connected graph. Every user node is connected with others according to their similarity.
- ❑ Cluster labels ~ friendship, for users who do not have friend information.



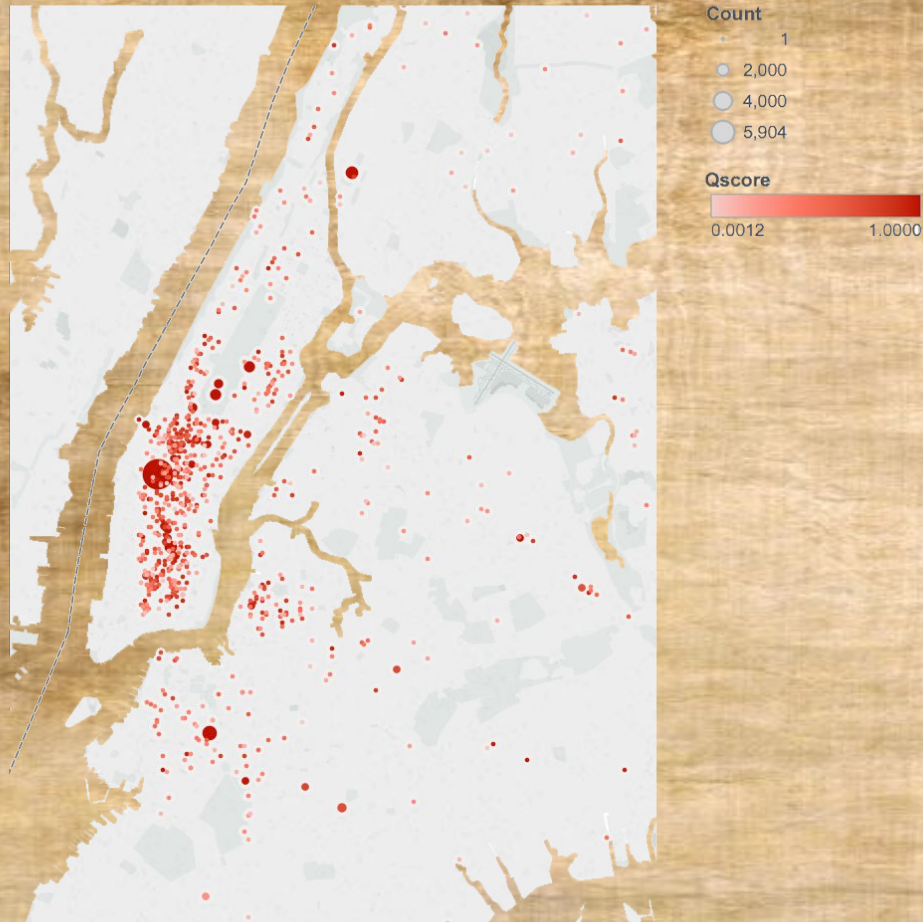


# Result

Place Visited	Category	Place Recommended	Category
Studio 450	Event Space	The 13th Step	Harlem Tavern
Mercury Bar	Bar	Delicatessen	American Restaurant
Sinigual Contemporary Mexican Cuisine	Mexican Restaurant	Cafeteria	New American Restaurant
Oasis	Falafel Restaurant	Harlem Tavern	Music Venue
Cascabel Taqueria	Mexican Restaurant	Pacha NYC	Nightclub
The DL	Lounge	Mother's Ruin	Cocktail Bar
Sweet and Vicious	Lounge	Radegast Hall & Biergarten	Beer Garden







- ❑ Each point stands for a venue
- ❑ The darker the color, the higher the preference level
- ❑ The bigger the point, the more the check-ins





# Future Work

- ❑ **Temporal Dynamics**
  - ❑ **Capture temporal trends in user activities**
- ❑ **Venue Categorization**
  - ❑ **Provide contextualized information on venues**
- ❑ **Matrix Factorization**
  - ❑ **Discover latent features between users and venues**

GitHub Repo: <https://github.com/fanshi118/Time-Out-New-York-MLC>





# Thank You!

