

Social Analytics Project

A Graph Based Recommender System (Jiak*Bot 2.0)

**jiak - literal translation of the Hokkien verb “Eat”*

Chris Thng | Ong Han Ying | Tham Jun Quan | Wesley Chan Lun

Yang Cheng | Ye Zhibo



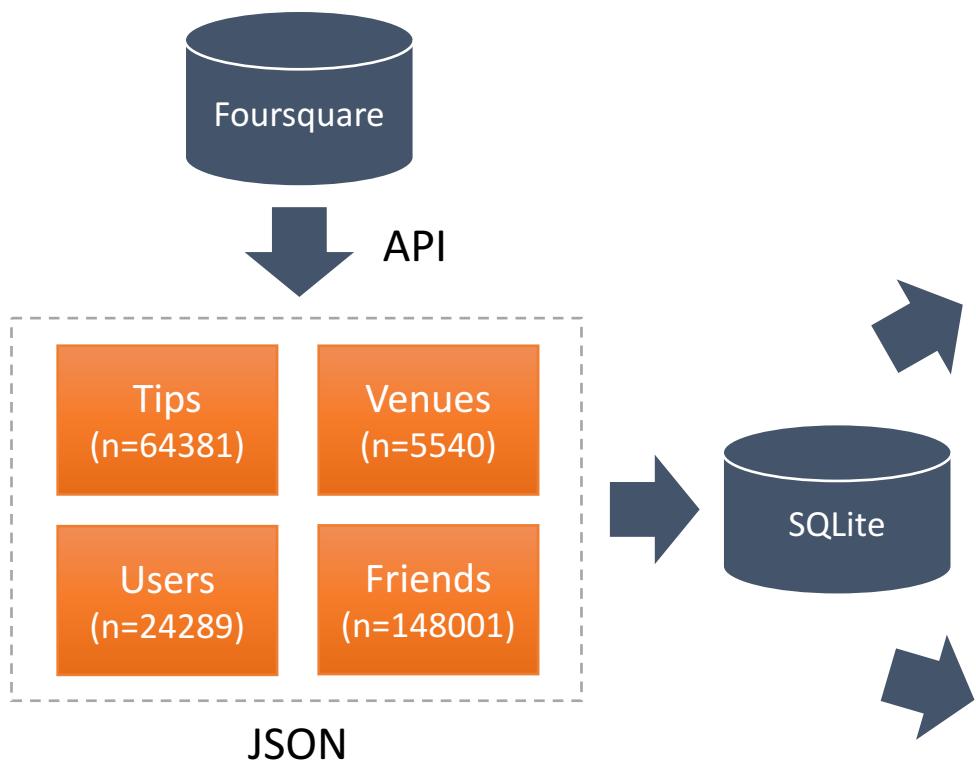
Objective

- Inspired from a common problem working class adults face Singapore - where and what to eat for lunch
- Aims to suggest places to eat through a chatbot through social graph mining and more specifically,
 - Detecting Community and
 - Providing Graph Based Recommendations



Methodology

Phase 1: Data Preparation



Phase 2: Analysis via Social Analytics

Clustering

Spectral Clustering /w Unipartite Projection

Gaussian Mixed Model

K-Means

Recommendations

New Users

Existing Users

Phase 3: Application

Web Visualization

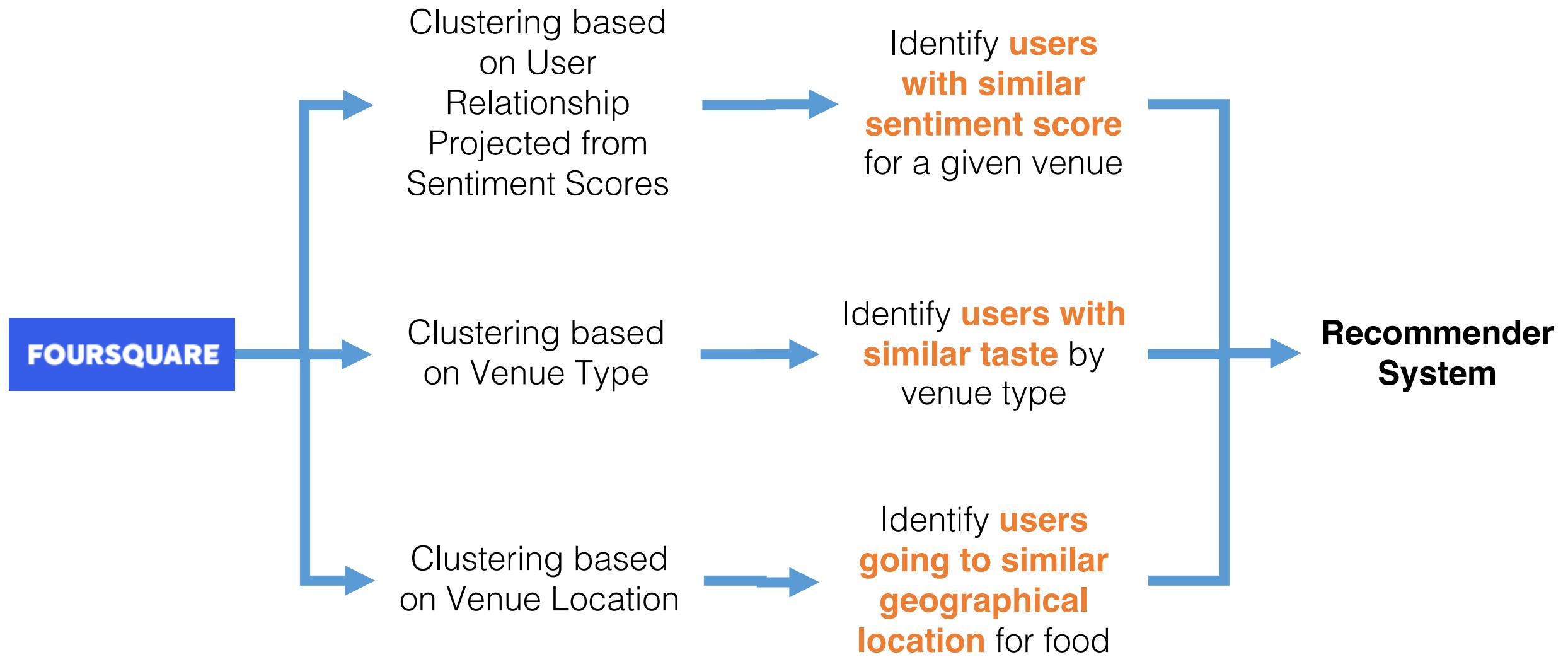
Community Detection

Restaurant Recommendation

Application

Jiakbot

Detecting Communities

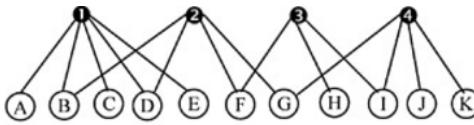


Clustering Based on Sentiment Scores

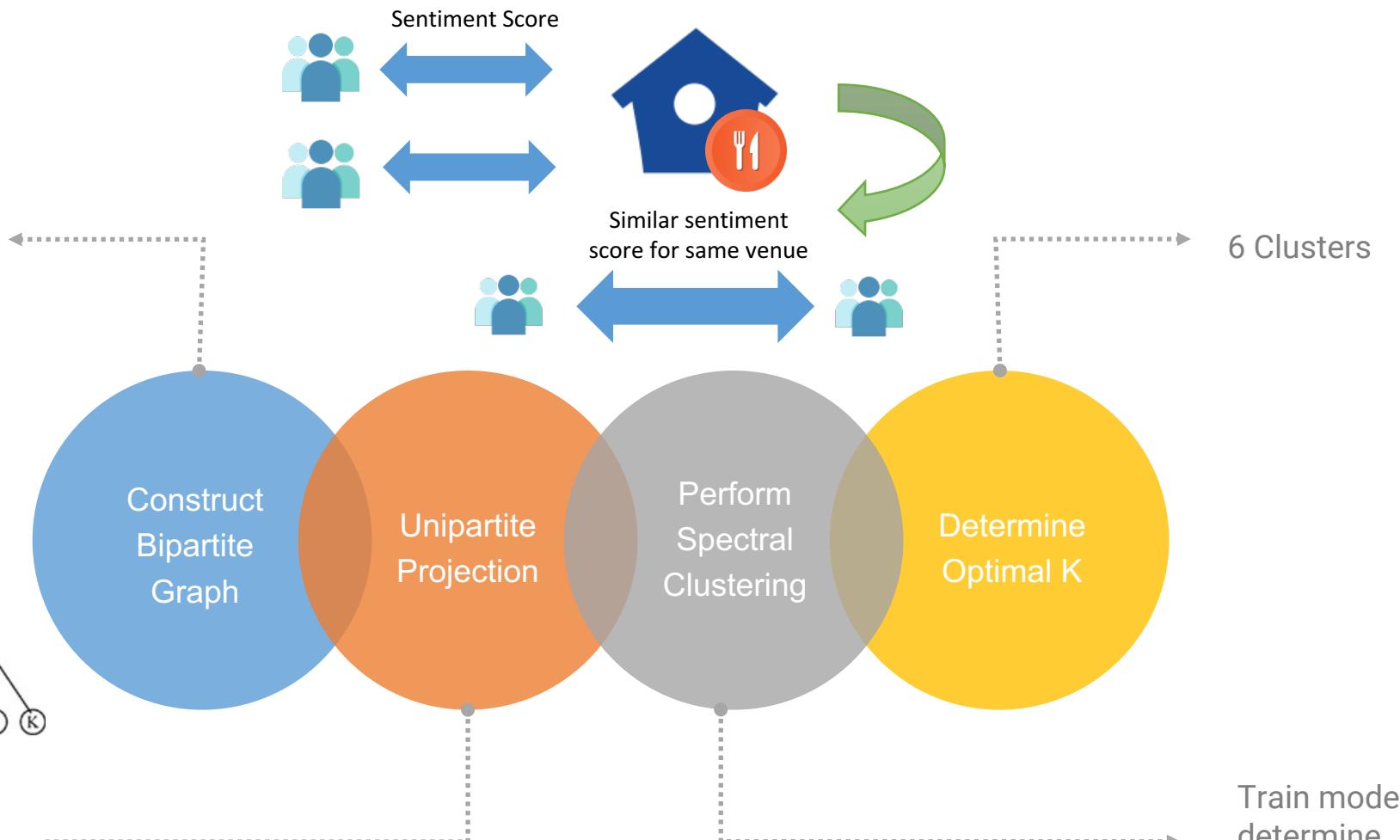
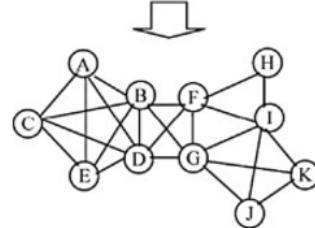
Bipartite Graph

Nodes: Users and Venues
Edges: Computed sentiment score derived from tips

user
venue

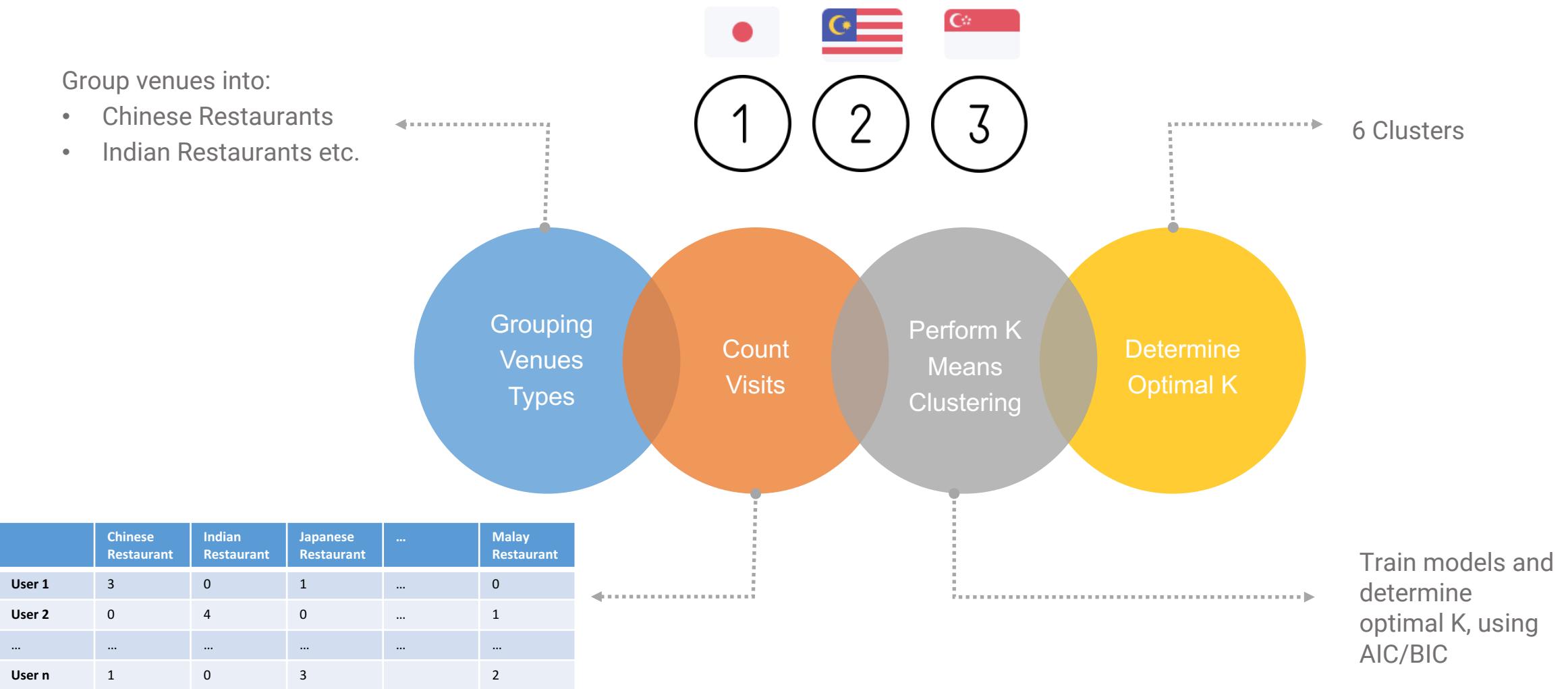


user - user



Train models and
determine
optimal K, using
AIC/BIC

Clustering Based on Venue Type



Clustering Based on Venue Type

Group venues based on their locations into 20 regions

- Central, North etc.



	Central (CBD)	Central (C1)	North East (NE1)	...	West (W1)
User 1	0.5	0	0	...	0.5
User 2	0.2	0.3	0.3	...	0.2
...
User n	0	0	0.5	...	0.5

Train models and determine optimal K, using AIC/BIC

How did we recommend?

A user is classified by:

- the type of restaurants they go to
- the location of the restaurants they go to
- the user-user similarity based on sentiment scores

Recommendations based on clusters identified earlier

Within each cluster, for each user,

- Identify users who visited the same restaurants
- Get the list of restaurants visited by these users
- Rank the restaurants by weighting sentiment and frequency of visits
- Return the top n restaurants by rank



How did we rank?

- Using a method proposed by Opsahl et al. (2010)
- A generalization for edge weights (sentiment score) and number of edges (frequency of visit).

$$C_D^{w\alpha} (i) = k_i \times \left(\frac{s_i}{k_i} \right)^\alpha = k_i^{(1-\alpha)} \times s_i^\alpha$$

Node	C_D	C_D^w	$C_D^{w\alpha}$ when $\alpha =$			
			0	0.5	1	1.5
A	2	8	2	4	8	16
B	4	8	4	5.7	8	11.3
C	2	6	2	3.5	6	10.4
D	1	1	1	1	1	1
E	2	8	2	4	8	16
F	1	7	1	2.6	7	18.5



In layman terms, we consider both sentiment and number of visits in ranking the venues



Evaluating the Recommendations

Evaluation is done by using the Mean Reciprocal Rank (MRR)

$$\text{MRR} = \frac{1}{|Q|} \sum_{i=1}^{|Q|} \frac{1}{\text{rank}_i}$$



In layman terms, it meant that on average, what is the rank position of the a relevant restaurant.



Evaluation Results

Mean Reciprocal Rank



Recent History



Full History



Venue



Rating



Location

0.0369

Relevant recommendation
ranked around #28



0.0519

Relevant recommendation
ranked around #20

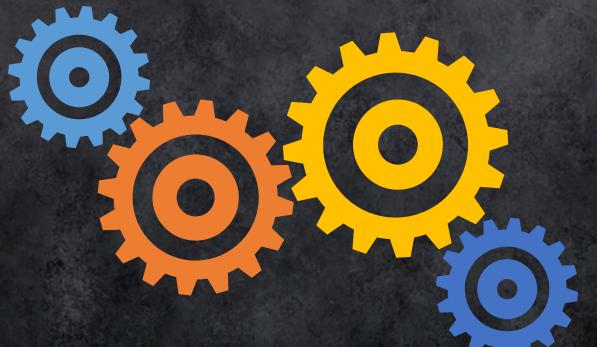


0.0672

Relevant recommendation
ranked around #15

Using the Results

- Web Visualisation
- Jiakbot 2.0 (@jiak_bot)

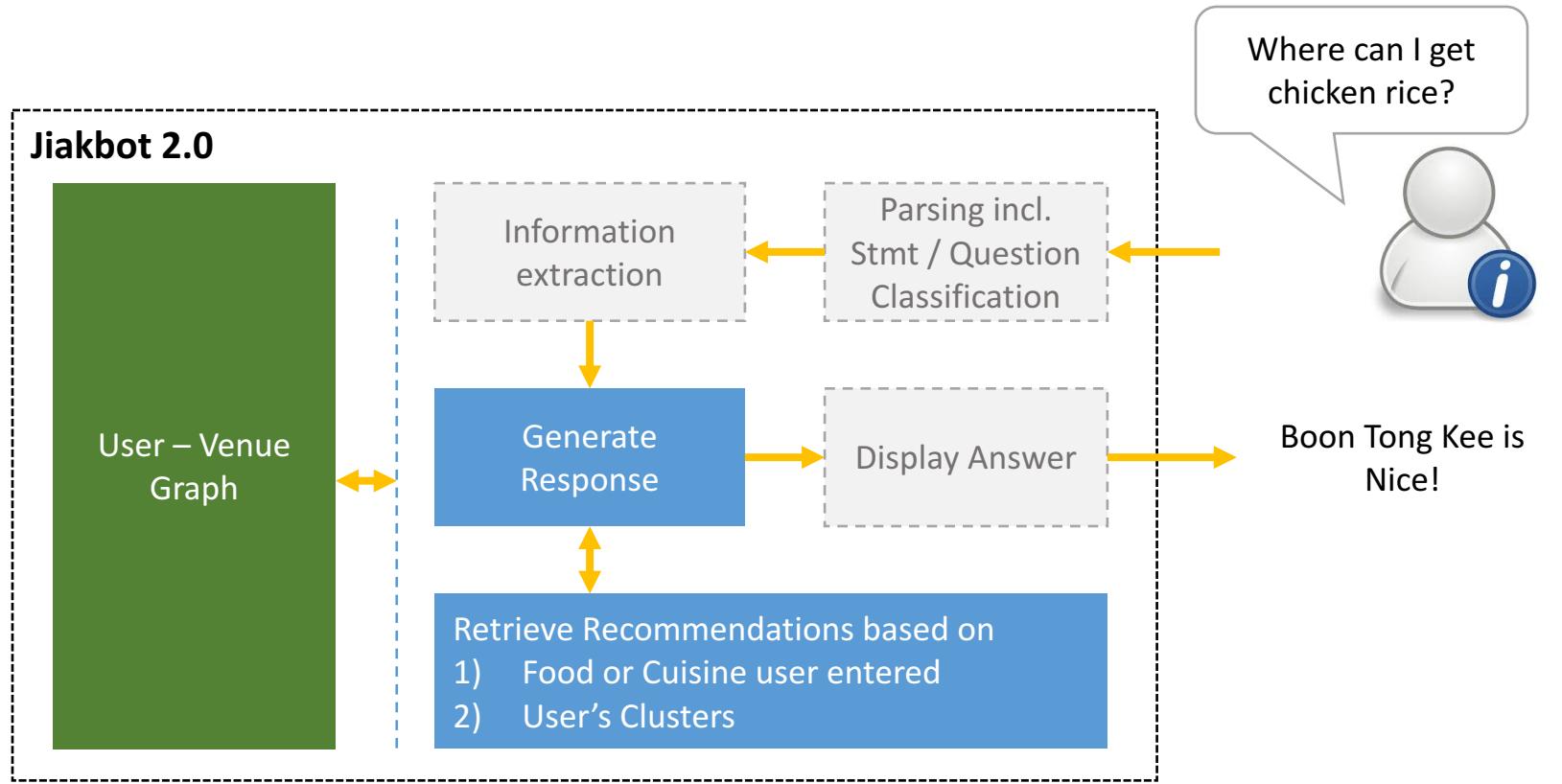


How JiakBot 2.0 Works

Data

Scrape businesses data from FourSquare including their reviews.

Using Graph Mining Techniques to make Recommendations



Q + A