Cross-Domain Recommendation
via Clustering on Multi-Layer Graphs

Goncharov Alexey
MIPT november 2017

Доклад основан на двух работах:

http://farseev.azurewebsites.net/slides/SIGIR17_Farseev.pdf
http://nusmultisource.azurewebsites.net/slides.pdf

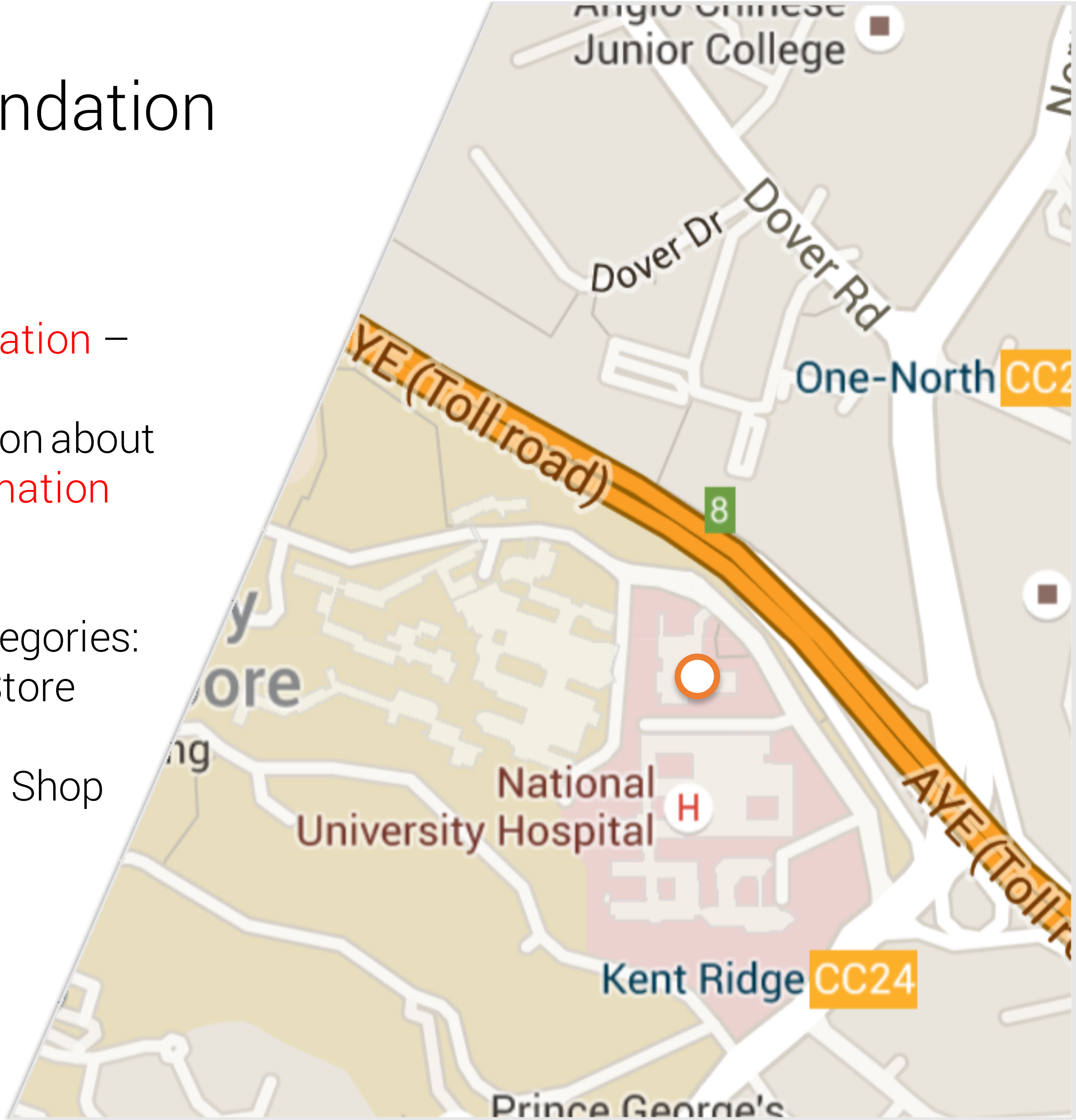
Venue Category Recommendation

Collaborative Venue Category Recommendation – recommendation of venue categories (i.e. restaurant, cinema) to user using information about his/her profile (i.e. past visits) and/or information about users from the same domain.

Venue categories:

Venue categories:
Clothing Store
Hotel
Ice Cream Shop

Total 764 different categories



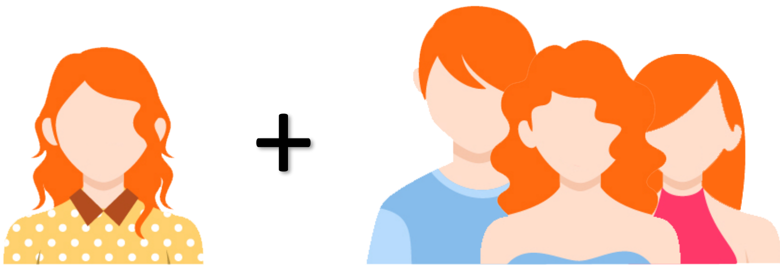
Idea 1: Utilization of Individual And Group Knowledge for Better Recommendation

User Community-Based Collaborative Recommendation

We perform venue category recommendation based on both individual and group knowledge =>

naturally models the impact of society on an individual's behavior during the selection of a new place to go:

$$rec(u) = sort\left(\gamma \cdot vec_u + \theta \frac{\sum_{v \in C_u} vec_v}{|C_u|}\right)$$



What do we need user communities for?

+ Users from the same community (extracted from multi-source data) may have similar location preferences

+ Search within user community significantly reduces search space during the recommendation process

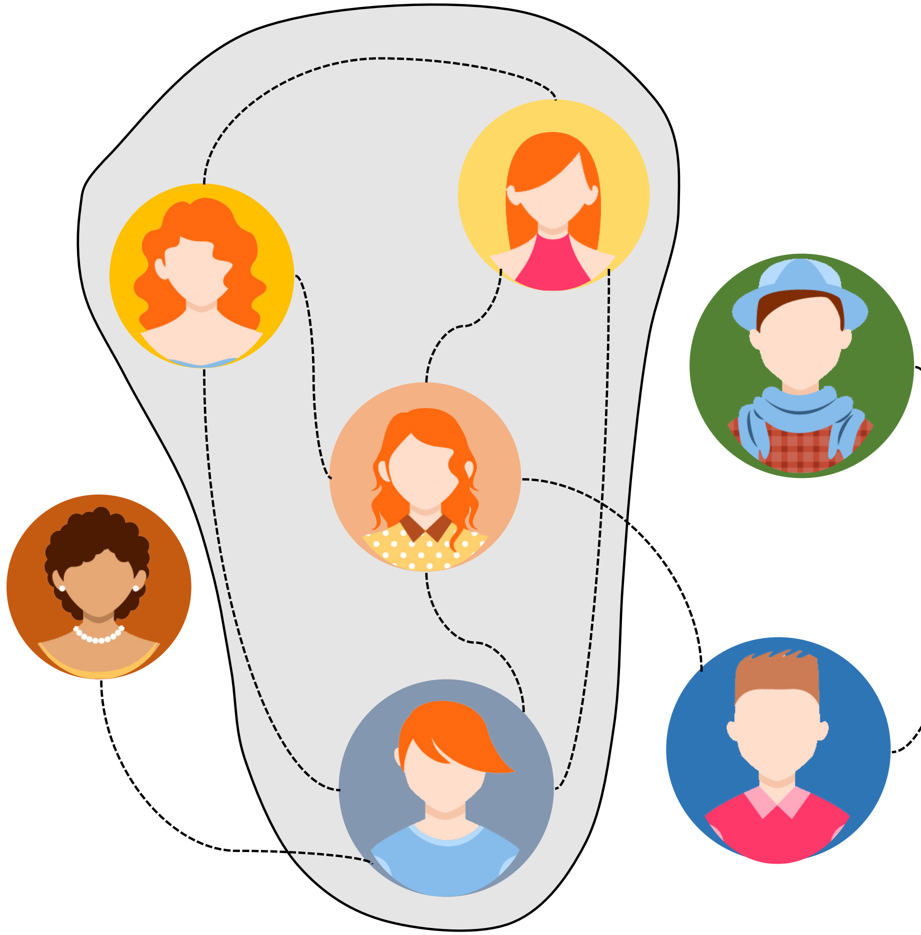
Example of User Communities (1)

Community 1: Gingers

Community K: Darker Hair

User Relation and Community Representations

One way to find user communities is to model users' relationships in the form of a graph so that dense subgraphs are considered to be user communities.

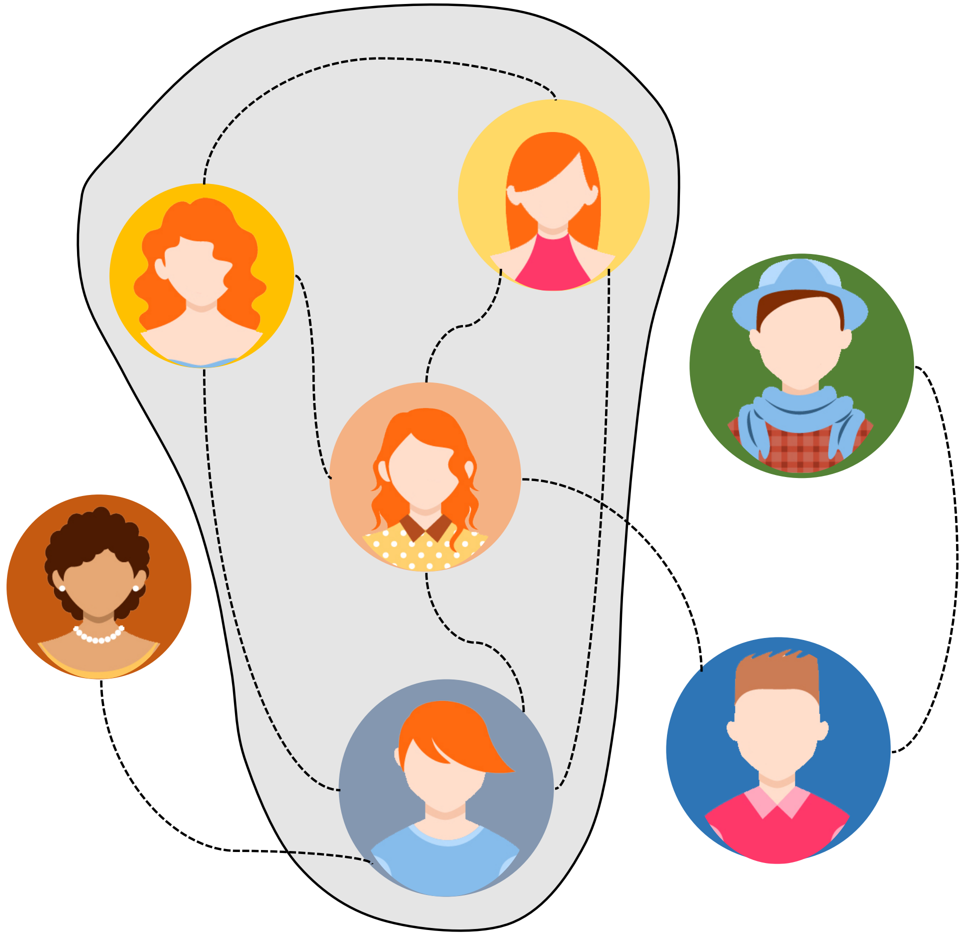


Community Detection based on a single data source

One of the commonly formulations is **MinCut** problem.

For a given number k of subsets, the **MinCut** involves choosing a partition $C_1, \dots, C_k$ such that it minimizes the expression:

$$cut(C_1, \dots, C_k) = \sum_{i=1}^{k} W(C_i, \bar{C}_i)$$



*W is the sum of weights of edges attached to vertices in C_i

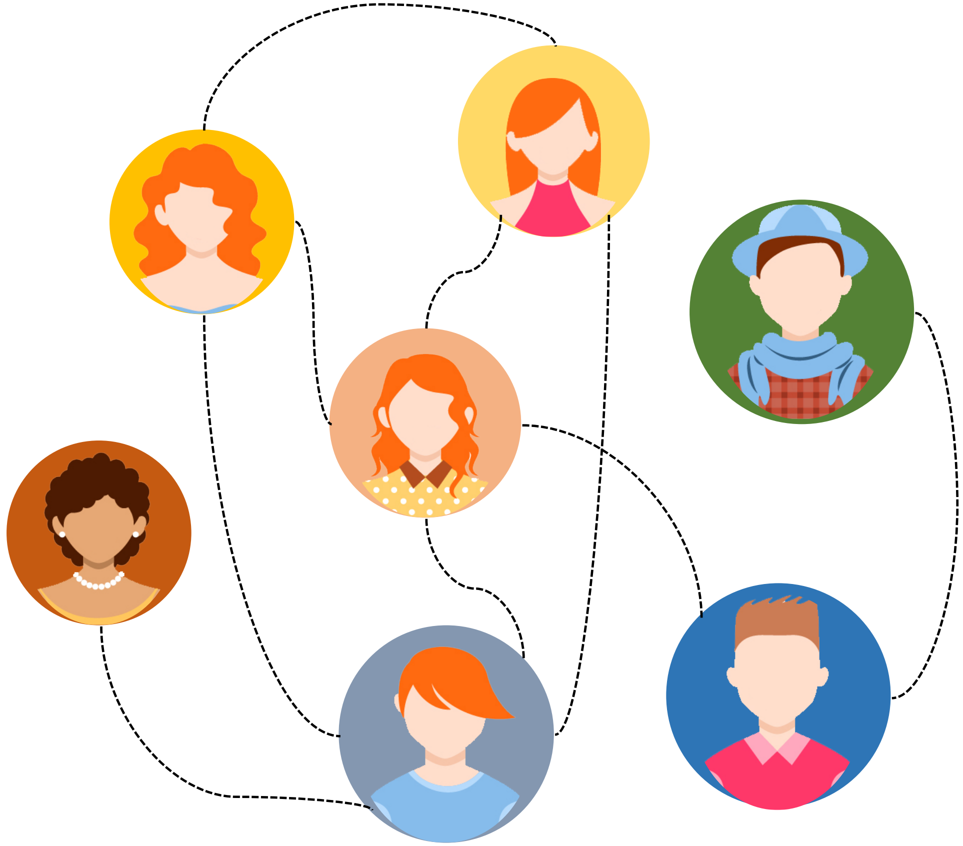
How to solve MinCut problem?

Approximation of MinCut as
standard trace minimization problem:

$$\min_{U \in R^{n \times k}} \mathrm{tr}(U^T L U), \mathrm{s.t.}\, U^T U = I$$

which can be solved by Spectral Clustering:

1. Calculates Laplacian matrix $L \in R^{n \times n}$
2. Builds matrix of the first k eigenvectors $U \in R^{n \times k}$ correspond to the smallest eigenvalues of L
3. Clusters data in a new space U using i.e. k-means algorithm

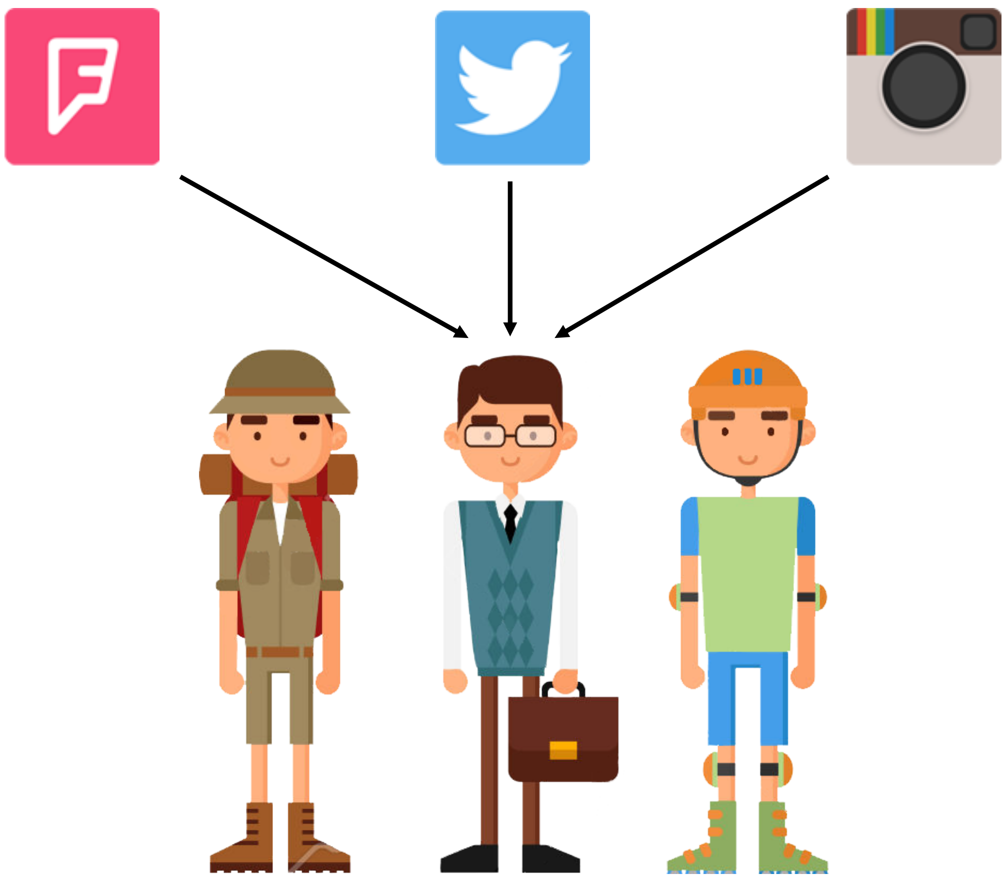


Idea 2: Utilization of Multi-Source Data

Most of user actively use ≈3 social networks

* GlobalWebIndex. 2016. GWI Social report. http://www.globalwebindex.net/blog/internet-users-have-average-of-5-social-media-accounts

Multi-source data describe user from multiple views

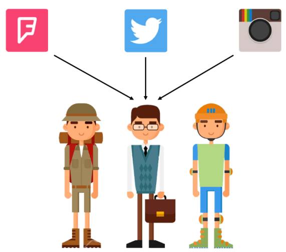


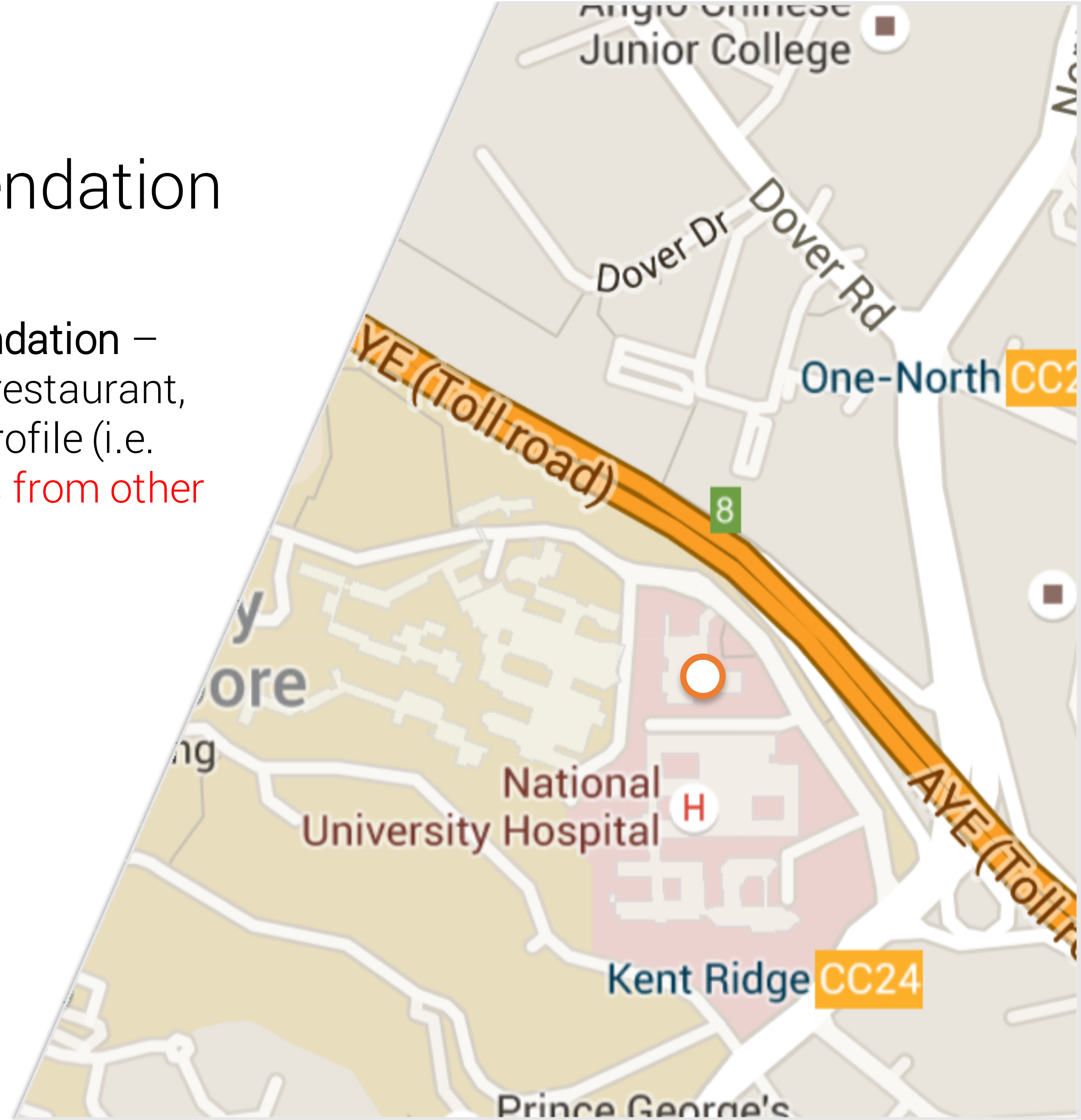
Cross-Domain
Venue Category Recommendation

Cross Domain - Venue category recommendation – recommendation of venue categories (i.e. restaurant, cinema) using information about his/her profile (i.e. past visits) and/or information about users from other sources (i.e. images, texts, location types).

Venue categories:
- Clothing Store
- Hotel
- Ice Cream Shop

Multi-Source Data:



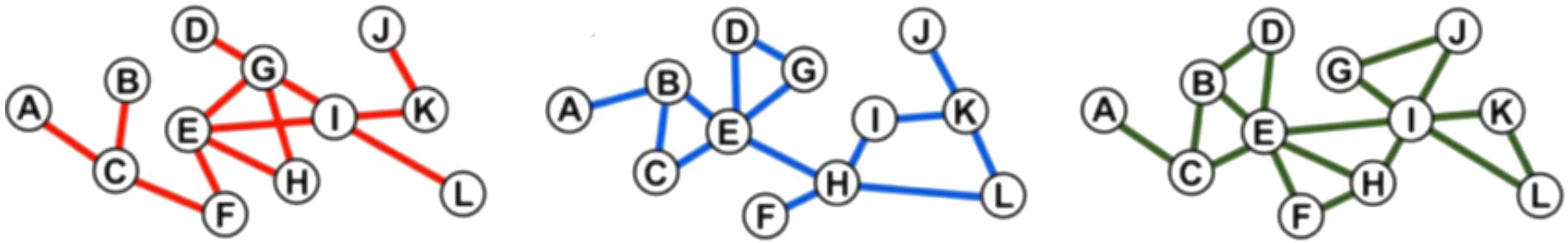


Community Detection must performed in a Cross-Source Manner…

Problems:

- Data source integration
- Community detection

How to represent multi-source data?



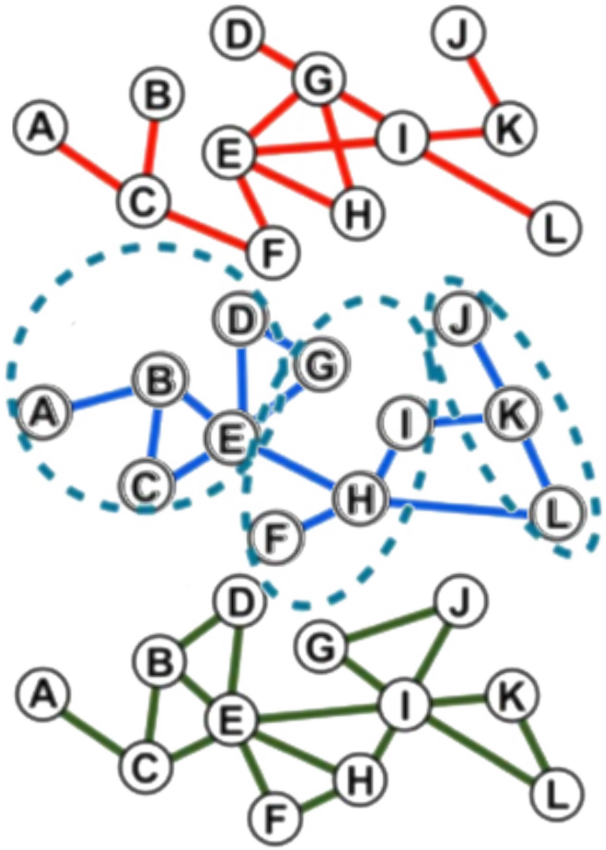
Multi-layer graph – graph G, where $G = \{G_i\}$, $G_i = (V, E_i)$

Extending definition of spectral clustering

$$\min_{U \in R^{n \times k}} \sum_{i=1}^{M} \mathrm{tr}(U^T L_i U), \mathrm{s.t.}\, U^T U = I$$

$$\min_{U \in R^{n \times k}} \mathrm{tr}(U^T L_{sum} U), \text{where } L_{sum} = \sum_{i=1}^{M} L_i$$

Such approximation could suffer from poor generalization ability.

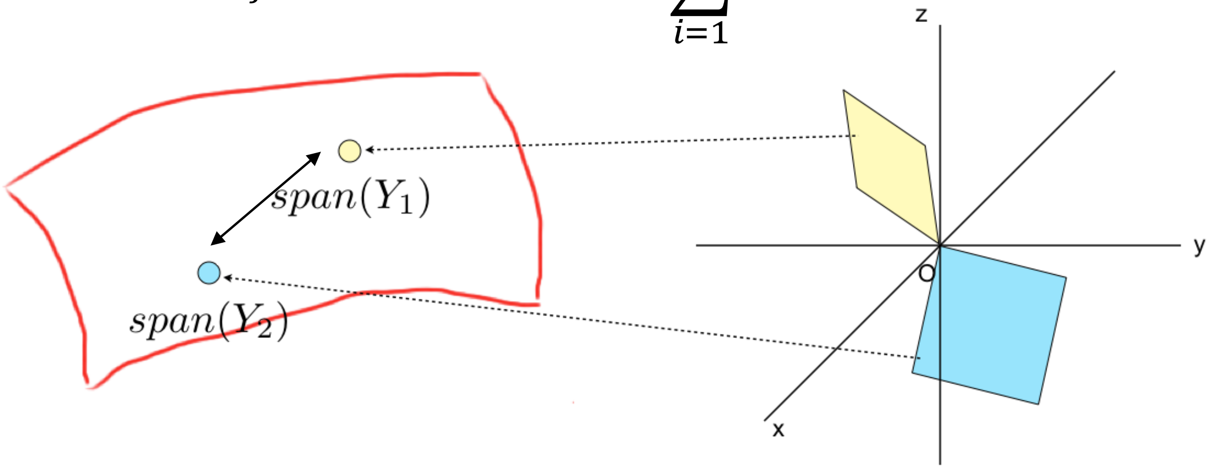


Regularized Clustering on Multi-layer Graph -1

Use Grassman Manifolds to keep final latent representation “close” to all layers of multi-layer graph*. Where projected distance between two spaces Y_1 and Y_2:

$$d^2_{Proj}(Y_1, Y_2) = \frac{1}{2}\left\|Y_1 Y_1^T - Y_2 Y_2^T\right\|_F^2, \text{where } \|A\|_F \text{ is the Frobenius norm}$$

$$d^2_{Proj}\left(S, \{S_i\}_{i=1}^M\right) = kM - \sum_{i=1}^{M} \mathrm{tr}\left(SS^T - S_i S_i^T\right)$$



* X. Dong, P. Frossard, P. Vandergheynst, and N. Nefedov. Clustering on multi-layer graphs via subspace analysis on grassmann manifolds. IEEE Transactions on Signal Processing, 2014.

Regularized Clustering on Multi-layer Graph -2

Extends the objective function to introduce the subspace analysis regularization

$$\min_{U\in\mathbb{R}^{n\times k}} \sum_{i=1}^{M} \mathrm{tr}\left(U^T L_i U\right) + \alpha\left(kM - \sum_{i=1}^{M} \mathrm{tr}\left(UU^T U_i U_i^T\right)\right), \mathrm{s.t.}\ U^T U = I$$

$$\min_{U\in\mathbb{R}^{n\times k}} \mathrm{tr}(U^T \boxed{L_{mod}} U)$$

$$L_{mod} = \sum_{i=1}^{M} (L_i - \alpha U_i U_i^T)$$

Idea 4: Making use of
Inter-Layer (Inter-Source) Relations

Incorporating inter-layer relationship (1)

By using distance on Grassman Manifolds, we present the new objective function for the i^{th} layer:

$$\min_{\hat{U}_i \in \mathbb{R}^{n\times k}} \operatorname{tr}\left(\hat{U}_i^T L_i \hat{U}_i\right) + \beta_i \left(kM - \sum_{j=1, j\neq i}^{M} w_{i,j} \operatorname{tr}\left(\hat{U}_i \hat{U}_i^T U_j U_j^T\right) \right)$$

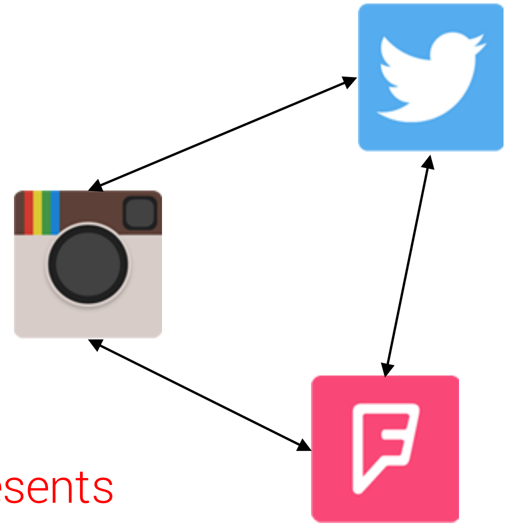
$$\min_{\hat{U}_i \in \mathbb{R}^{n\times k}} \operatorname{tr}\left(\hat{U}_i^{\,T} \boxed{\hat{L}_i} \hat{U}_i\right)$$

$$\boxed{\hat{L}_i = L_i - \beta_i \sum_{j=1, j\neq i}^{M} w_{i,j} \operatorname{tr}\left(U_j U_j^T\right)}$$

But how can we determine $\mathbf{w_{i,j}}$ when computing *i-th* layer ?

$$\min_{\widehat{U}_i \in \mathbb{R}^{n \times k}} \operatorname{tr}\left(\widehat{U}_i{}^T \hat{L}_i \widehat{U}_i\right)$$

$$\hat{L}_i = L_i - \beta_i \sum_{j=1, j \neq i}^{M} \boxed{w_{i,j}} \operatorname{tr}\left(U_j U_j^T\right)$$



Inter-layer relationship graph $\boldsymbol{R}(\boldsymbol{V}, \boldsymbol{E})$ – weighted graph which represents the similarity between layers.

$$\forall (i,j) \in E, w_{i,j} = \frac{\sum_{k=2}^{K} \left(1 - \frac{\left\| M_{i,k} - M_{j,k} \right\|}{\sqrt{N(N-1)}}\right)}{K - 1}$$

where $M_{i,k}$ is clustering co-occurrence matrix of layer i, $m_{a,b} = 1$, if users a and b assigned to the same cluster, and 0 otherwise.

Final objective function

Let's combine equations from previous slides to define the final objective function:

$$\min_{U \in \mathbb{R}^{n\times k}} \sum_{i=1}^{M} \mathrm{tr}\left(U^T \hat{L}_i U\right) + \alpha\left(kM - \sum_{i=1}^{M} \mathrm{tr}\left(UU^T \hat{U}_i \hat{U}_i^T\right)\right) =$$

$$= \min_{U \in \mathbb{R}^{n\times k}} \mathrm{tr}\left(U^T \boxed{\sum_{i=1}^{M} (\hat{L}_i - \alpha \hat{U}_i \hat{U}_i^T)} U\right)$$

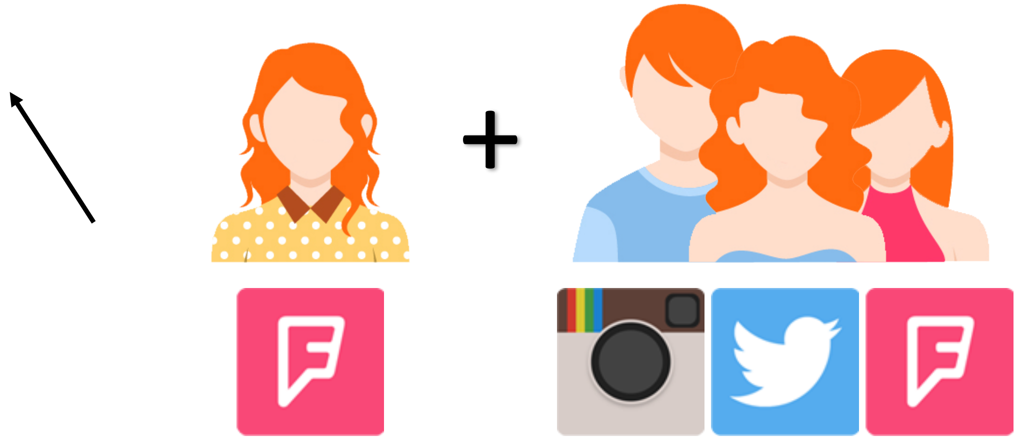
Problems

- ~~Community detection~~
- ~~Data source integration~~

Recall: Community-Based Cross-Domain Recommendation

We perform venue category recommendation based on both individual and group knowledge, where group knowledge is obtained from multiple sources:

$$rec(u) = sort\left(\gamma \cdot vec_u + \theta \frac{\sum_{v \in C_u} vec_v}{|C_u|}\right)$$



Foursquare

Instagram

NUS-MSS Dataset

Dataset* is presented as a set of features, extracted from user-generated data in three social networks:

- text based from Twitter (LDA, LIWC, text features)
- image based from Instagram (concepts)
- location based from Foursquare (LDA, categories, Mobility Features)

Foursquare categories is splited into two parts: 3 months data (train) and 2 months (test).

* A. Farseev, N. Liqiang, M. Akbari, and T.-S. Chua. **Harvesting multiple sources for user profile learning: a Big data study.** ACM International Conference on Multimedia Retrieval (ICMR). China. June 23-26, 2015.

Twitter

Data Sources

Text Features:

Linguistic features: LIWC; Latent Topics

Heuristic features: Writing behavior

LIWC LDA

Location Features:

Location Semantics: Venue Category Distribution

Mobility Features: Areas of Interest (AOI)

Mobility

Location Type Preferences

Image Features

Image Concept Distribution (Image Net)

Image Concepts

Images

Data representation

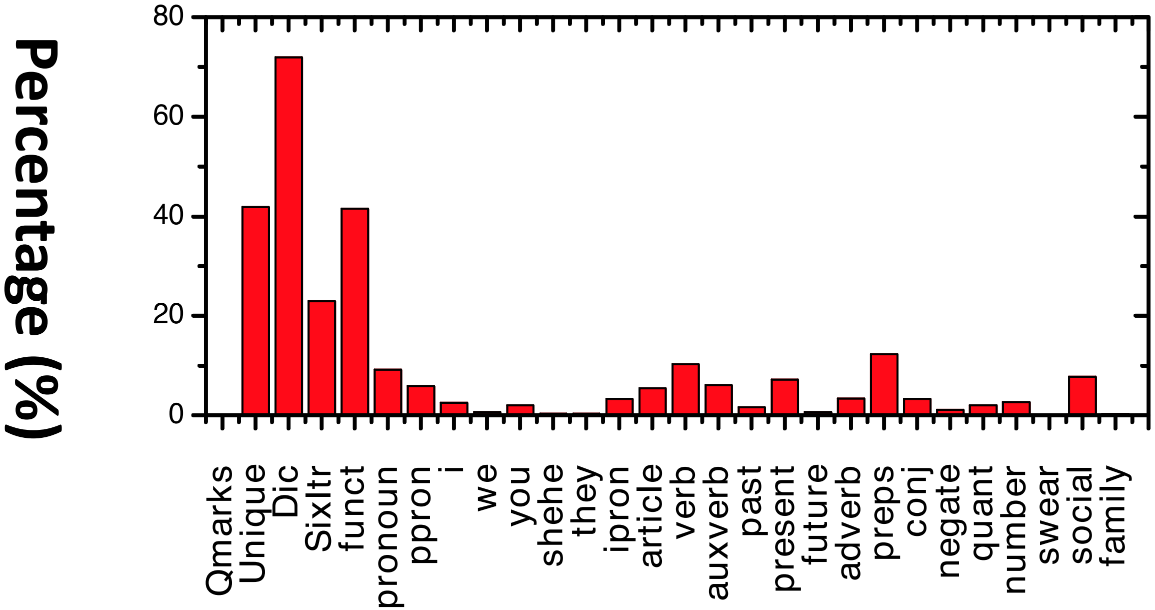
- Linguistic features
  - **LIWC**
  - User Topics
- Heuristic features
  - Writing behavior

An efficient and effective method for studying the various emotional, cognitive, structural, and process components present in individuals' verbal and written speech samples. Can be highly related to one's demography.

A text analysis software.

Dictionary

Word category

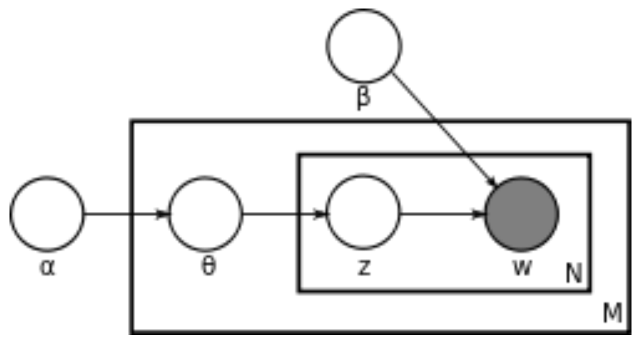


Data representation

- Linguistic features
  - LIWC
  - **User Topics**
- Behavioral features
  - Writing behavior

Users of **similar gender** and **age** may talk about **similar topics** e.g. **female** users – about **shopping, male** – about **cars**; **youth** – about **school** while **elderly** – about **health**.

LDA word distribution over **50 topics** for collected Twitter timeline.



Data representation

- Linguistic features
  - LIWC
  - User Topics
- Heuristic features
  - **Writing behavior**

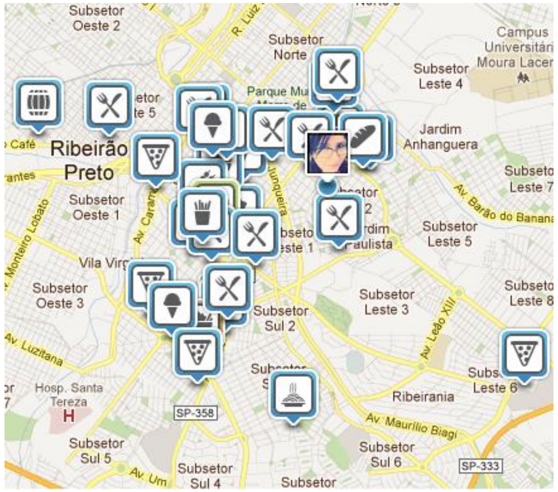
As we mention from our research – user's writing behavioral patterns are highly correlated with e.g. age (individuals from 10 – 20 years old are making two times less grammatical errors than 20 -30 years old individuals)

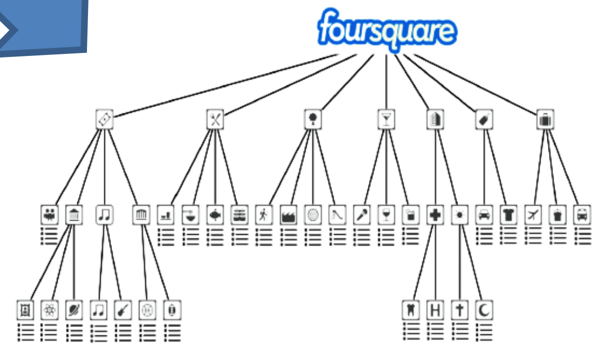
Feature name	Description
Number of hash tags	Number of hash tags mentioned in message
Number of slang words	Number of slang words one use in his tweets. We calculate number of slang words / tweet and compute average slang usage
Number of URLs	Number of URL's one usually use in his/her tweets
Number of user mentions	Number of user mentions – may represent one's social activity
Number of repeated chars	Number of repeated characters in one tweets (e.g. noooooooo, wahhhhhhh)
Number of emotion words	Number of words that are marked with not – neutral emotion score in Sentiment WordNet
Number of emoticons	Number of common emoticons from Wikipedia article
Average sentiment level	Module of average sentiment level of tweet obtained from Sentiment WordNet
Average sentiment score	Average sentiment level of tweet obtained from Sentiment WordNet
Number of misspellings	Number of misspellings fixed by Microsoft Word spell checker
Number Of Mistakes	Number of words that contains mistake but cannot be fixed by Microsoft Word spell checker
Number of rejected tweets	Number of tweets where 70% of words either not in English or cannot be fixed by Microsoft Word spell checker
Number of terms average	Average number of terms per / tweet
Number of Foursquare check-ins	Number of Foursquare check-ins performed by user
Number of Instagram medias	Number of Instagram medias posted by user
Number of Foursquare tips	Number of Foursquare Tips that user post in a venue
Average time between check-ins min	Average time between two sequential check-ins - represents Foursquare user activity frequency

Data representation

- Location features
  - **Location semantics**
  - Location topics

We map all Foursquare check – ins to Foursquare categories from category hierarchy.





Venue semantics such as venue categories can be related to users demography. E.g. individuals who tent to visit night clubs are usually belong to 10 – 20 or 20 – 30 years old age groups.

For case when user performed check-ins in two restaurants and airport but did not perform check-ins in other venues:

	$Category_1$	…	$Category_{restaurant}$	…	$Category_{airport}$	…	$Category_n$
U_1	0	0	2	0	1	0	0
…	*	*	*	*	*	*	*
U_n	*	*	*	*	*	*	*

Data representation

- Image features
  - **Image concept learning**

Extracted **image concepts** may represents **user interests** and be related to one's demography. For example **female** user may take pictures of **flowers, food**, while **male** – of **cars or buildings**.

SIFT, LBP, CH
FEATURES*

IMAGENET

IMAGE CONCEPTS
VECTOR

*The concept learning Tool was provided by Lab of Media Search LMS.
It was evaluated based on ILSVRC2012 competition dataset and performed with average accuracy @10 - 0.637

Evaluation Baselines

Recommender Systems

Popular (POP) —recommendation based on user's past experience

Popular All (POP All) —recommendation based on experience of all users

Multi-Source Re-Ranking (MSRR) — linearly combines recommendation results from all data modalities

Nearest Neighbor Collaborative Filtering (CF) — recommendation based on top k most similar Foursquare users

Early Fusion (EF) — fuses multi-source data into a single feature vector

SVD++ — makes use of the "implicit feedback" information

FM— brings together the advantages of different factorization-based models via regularization.

Community Detection Approaches

$\mathbf{C^3R} - \hat{\mathbf{L}}_{\mathbf{i}}$ — $\mathrm{C^3R}$ recommendation without inter-layer regularization

$\mathbf{C^3R} - \hat{\mathbf{L}}_{\mathbf{i}} - \hat{\mathbf{L}}_{\mathbf{Mod}}$ — $\mathrm{C^3R}$ recommendation without inter-layer regularization and sub-space regularization

$\mathbf{C^3R}-\boldsymbol{Comm}$ — $\mathrm{C^3R}$ recommendation without user community extraction

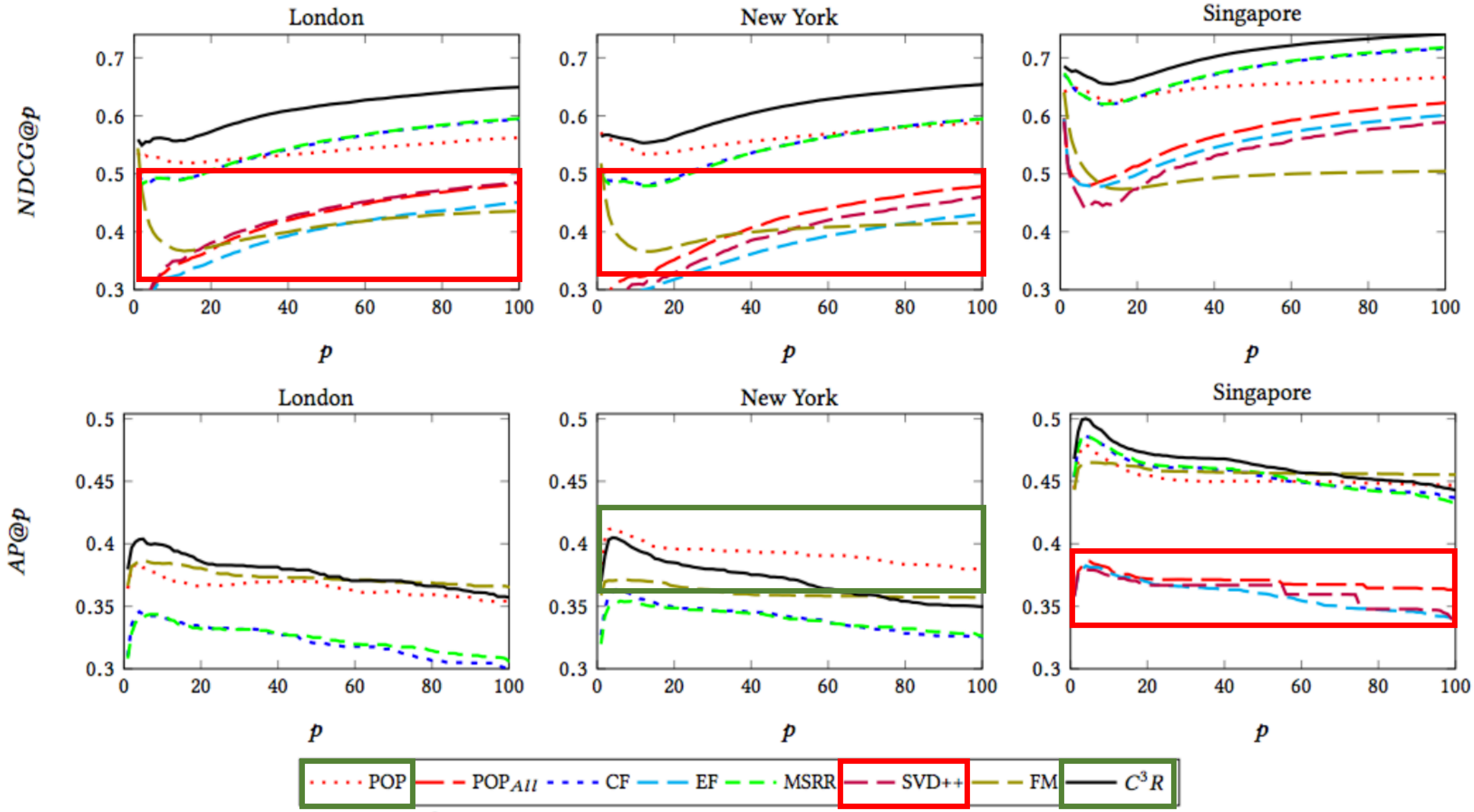
$\mathbf{C^3R}$ (DBScan) — $\mathrm{C^3R}$ recommendation, where user communities are detected by Density-Based clustering (DBScan)

$\mathbf{C^3R}$ (x-means) — $\mathrm{C^3R}$ recommendation, where user communities are detected by x-means clustering

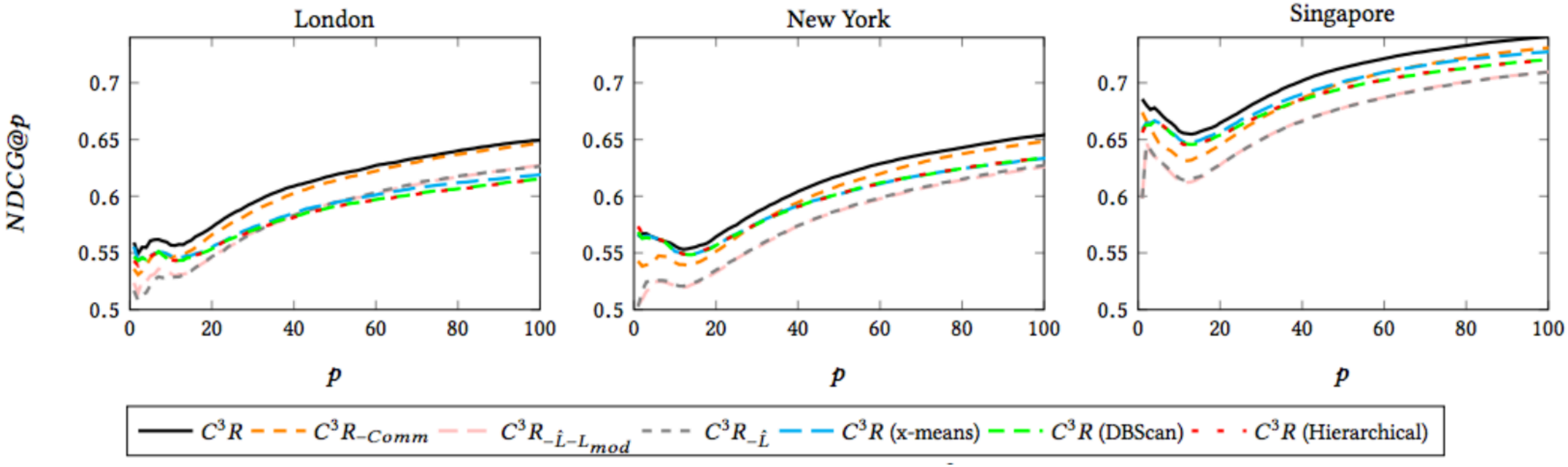
$\mathbf{C^3R}$ (Hierarchical) — $\mathrm{C^3R}$ recommendation, where user communities are detected by Hierarchical Clustering

$\mathbf{C^3R}$ — Our Approach

Evaluation against other recommender systems



Evaluation against other community detection approaches

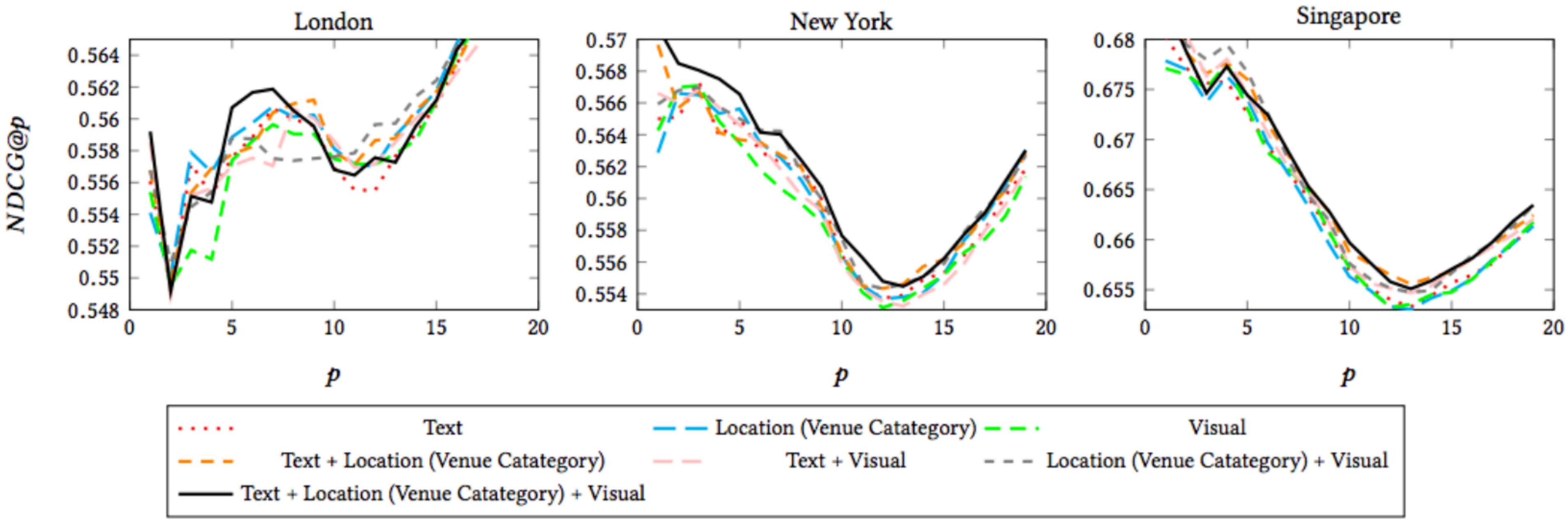


+ Incorporation of group knowledge is is important

+ Multi-modal clustering performs better than single-source clustering

+ Incorporation of Inter-Source relationship is crucial.

Evaluation against source combinations



+ In different geo regions, different data sources are of different importance

+ Location data is more powerful than other data modalities

$$
W_R = \left(\begin{array}{c|ccccc}
 & \textbf{tw} & \textbf{4sq} & \textbf{inst} & \textbf{tmp} & \textbf{mob} \\
\hline
\textbf{tw} & 1 & 0.632 & 0.621 & 0.643 & 0.561 \\
\textbf{4sq} & 0.632 & 1 & 0.614 & 0.631 & 0.570 \\
\textbf{inst} & 0.621 & 0.614 & 1 & 0.621 & 0.551 \\
\textbf{tmp} & 0.643 & 0.631 & 0.621 & 1 & 0.560 \\
\textbf{mob} & 0.561 & 0.570 & 0.551 & 0.560 & 1
\end{array}\right)
$$

Examples of detected user communities

Name	Bag of Words for different modalities		
	Text	**Visual**	**Location**
Gadgets 832 users	device, launcher, android	mouse, digital clock, hard disc	electronics store, tech startup, technology building
Arts 538 users	painting, landscape, reflection	obelisk, paint-brush, pencil box	arts & crafts store, arts & entertainment, museum
Food 446 users	dining, cof-fee, cook-ing	pineapple, mi-crowave, fry-ing pan	italian restaurant, pizzeria, macanese restaurant

Future Work

Community Detection is more useful when it is Source-Dependent

=> Introduce Supervision Into Clustering

How?

- Graph Construction Level – reweight edges according to prior knowledge about existing user communities
- Model Level – introduce community-related constraints into clustering

Summary

+ Multi-View Data is crucial for User Community Detection

+ For the task of venue category recommendation, both Group And Individual Knowledge are Important

+ Venue Category Recommendation is not a conventional recommendation task: users visit many venue types from the past. (items from the train set often occur in test set)

The Released Datasets

http://nusmss.azurewebsites.net

NUS-MSS

http://nussense.azurewebsites.net

NUS-SENSE

Our Tutorial on Multi-View Learning @ WST WSSS'17

http://tutorial.farseev.com

Normalized Discounted Cumulative Gain (*NDCG*) measure, which is defined as:

user in i category

$$NDCG@p = \frac{DCG@p}{IDCG@p}, DCG@p = \sum_{i=1}^{p} \frac{2^{rel_i}}{\log_2(i+1)}, rel_i = \frac{Cat_i}{N_{Cat}},$$

Average Precision (*AP*), which is defined as:

$$AP@p = \frac{1}{\sum_{i=1}^{p} r_i} \sum_{i=1}^{p} r_i \left(\frac{\sum_{j=1}^{i} r_j}{i} \right), r_i = \begin{cases} 1, & \text{i is in top p visited cat.)} \\ 0, & \text{otherwise.} \end{cases}$$