NANYANG TECHNOLOGICAL UNIVERSITY

Modeling Heterogeneous Influences for Point-of-Interest Recommendation in Location-Based Social Networks

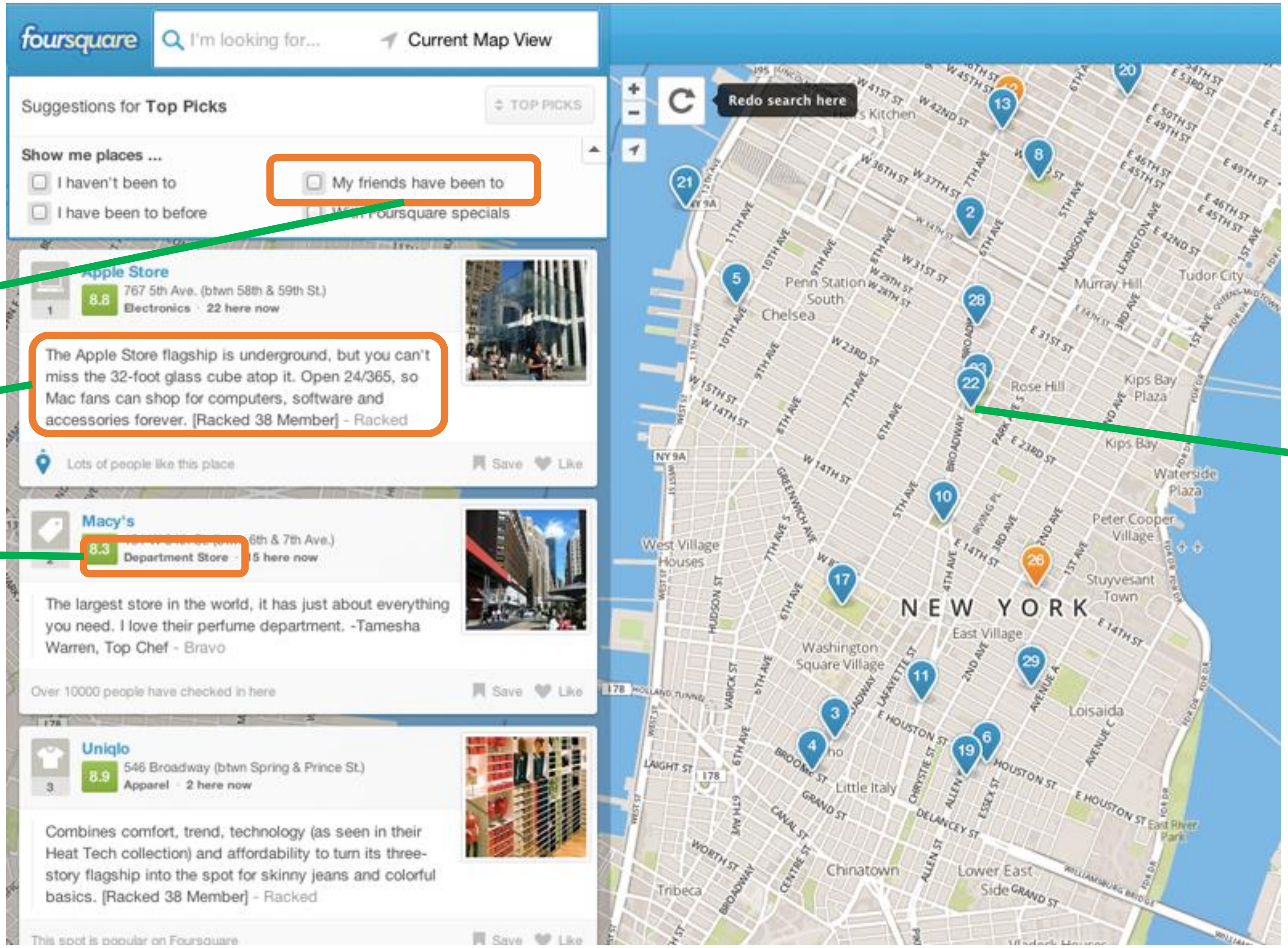
Qing Guo, Zhu Sun, Jie Zhang, and Yin-Leng Theng

Location-based Social Network

- Check-in

<user, POI, time>

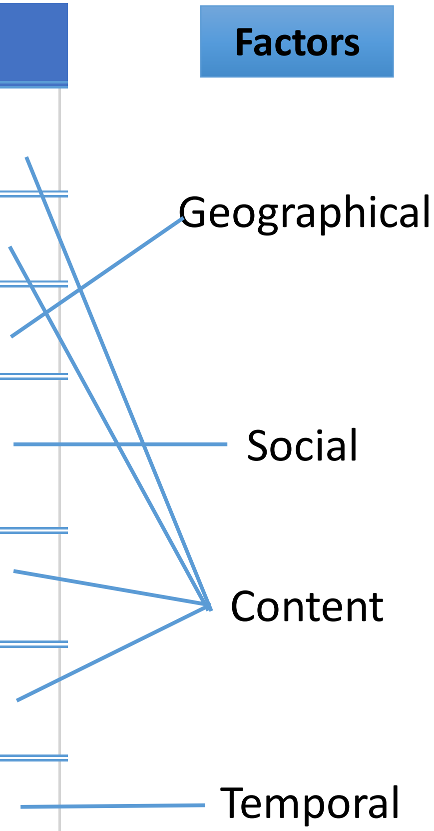
- Friendship
- Reviews
- Categories & tags



Latitude and longitude

POI Recommendation

Information Source	Information
POI	Categories
	Tags
	Lattitude, Longitude
User	Friendship (Social Links)
User-POI Interaction	Rating
	Reviews
	Time



- Motivation
  - Heterogeneous information
  - Various influence factors

Problem Statement

- Users' check-in decision is influenced by various factors. Although many POI recommender systems explore different influential factors to improve POI recommendation, they only partly exploit some of the factors. Therefore, there is a lack of research that attempt to incorporate these influential factors in a unified way.
- Two challenges
  - Different factor influence users' check-in decision in different way.
  - Information of different factors is represented in different way.

- Research Objectives
  - Identify appropriate information of content factor to model users’ preference more precisely.
  - Develop a knowledge graph to integrate various factors in a unified way.
  - Develop a latent factor model to generate POI recommendations with considering various influence factors as well as their personalized importance for each user and POI.

Overview

Research Objective 1

Heterogeneous Graph Construction

Research Objective 2

Neighbor Discovery

Incorporating meta paths into MF

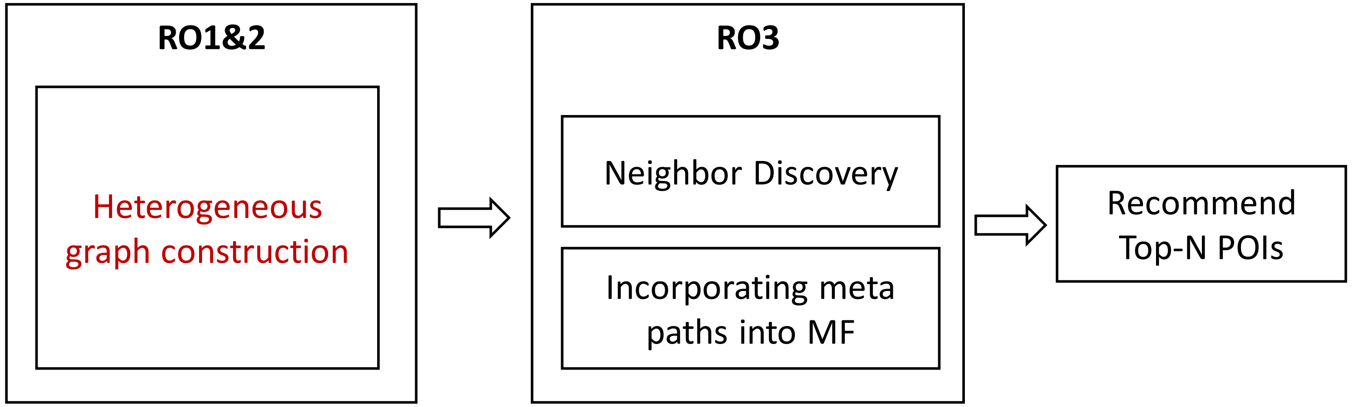
Recommend Top-N POIs

Dataset Description

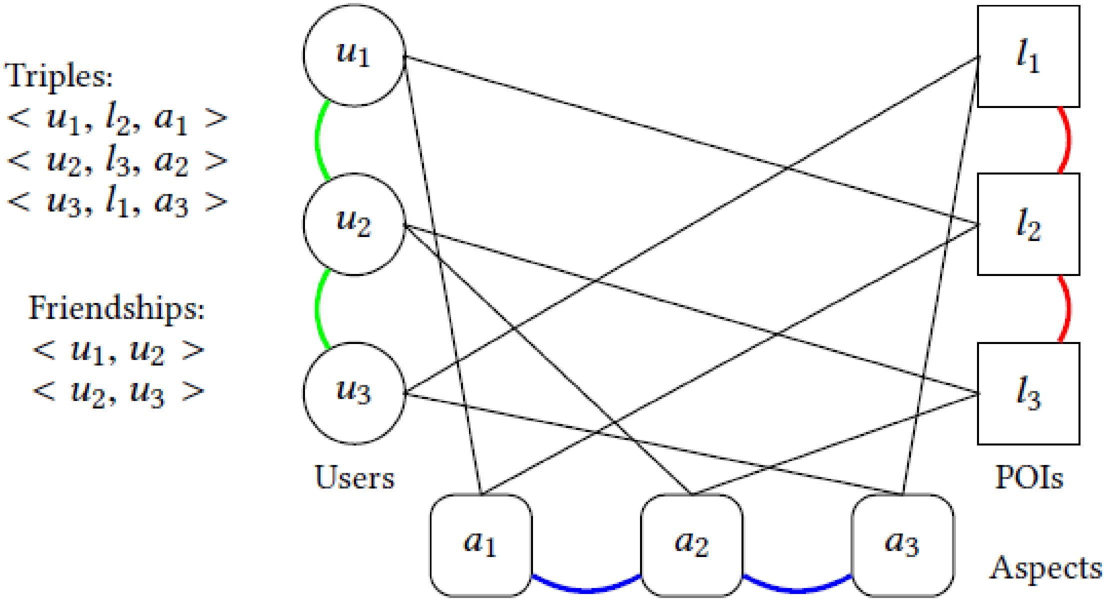
- Yelp dataset is utilized for experimental purpose.
- This study focuses on restaurants since aspects are domain specific.
- Phoenix, Las Vegas and Charlotte are chosen for this study.
- Aspect: "I like the *service* and *food* in this *Sushi* bar".

	Cities	CH	PH	LV
Dataset	Users	1,106	2,148	4,794
	POIs	2,724	2,603	4,274
	Aspects	1,355	1,952	4,180
	Reviews	33,966	61,596	150,080
	Categories	148	155	175
AGS-IG	User-POI	25,847	46,206	113,239
	User-Aspect	141,691	284,946	784,550
	User-User	5,803	23,766	46,407
	POI-Aspect	130,364	210,953	549,527
	POI-POI	79,530	112,248	405,310
	Aspect-Aspect	915,981	1,898,326	87,257,531
	Density	14.8%	11.5%	12.1%

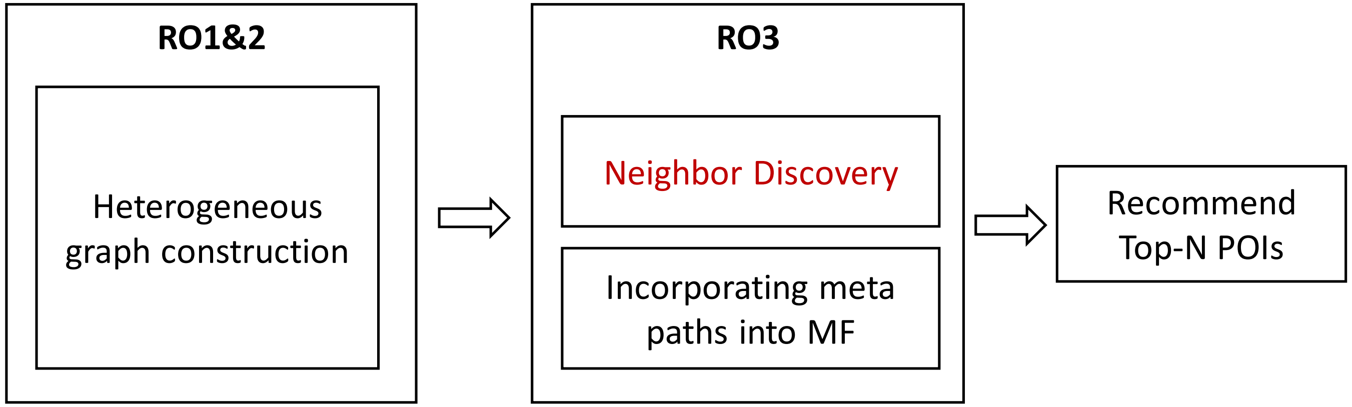
Intuitions for AGSG Construction



- **A**spect-aware **G**eo-**S**ocial Influence **G**raph (AGSG) is constructed based on the following intuitions:
  - **Intuition 1 (Geographical factor):** users prefer to visit place that are near to their usual activity areas (Ye et al., 2011; Yuan et al, 2014).
  - **Intuition 2 (Social factor):** friends share more interest that non-friends (Yang et al., 2013; Ye et al., 2010).
  - **Intuition 3 (Content factor):** a user prefer a POI because some aspects of this place attracts her, e.g., food or atmosphere (He et al., 2015).
  - **Intuition 4 (Content factor):** if user u_i express concerns about aspect a_p, u_i may also concern about aspect a_q if a_q is semantically similar to a_p.



Meta Path Selection



- Meta Path Selection
  - A meta path is a sequence of relations connecting different types of nodes (Sun et al., 2011), for example

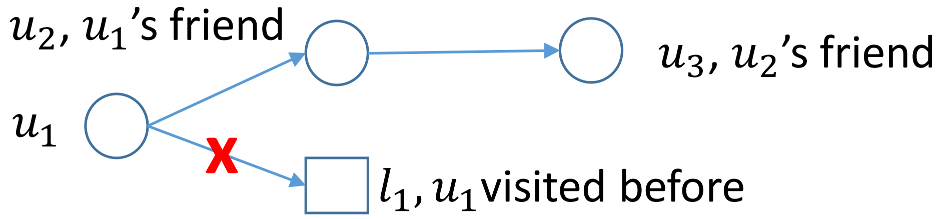
$$U \xrightarrow{mentioned} A \xrightarrow{isMentionedBy} U$$

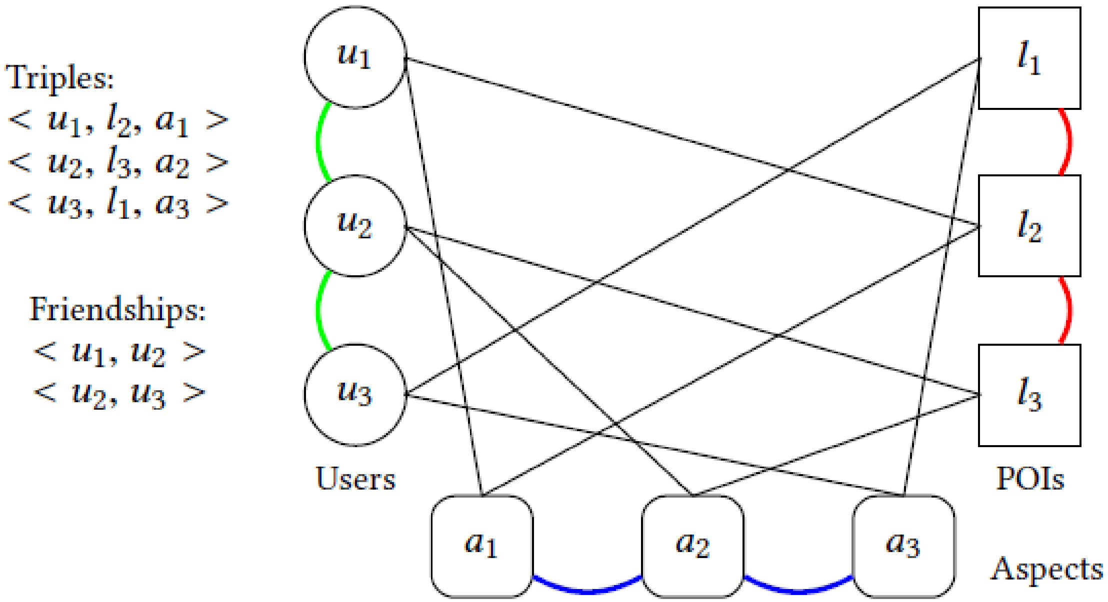
  - Different recommendation strategies can be modelled in a generic way.

Meta Path	Recommendation Strategy	Influence factor
ULU	User-based CF	User preference
UAU	User-based CF	Content factor
UU	User-based CF	Social factor
$ULALU$	User-based CF	User preference + Content factor
LUL	Item-based CF	Location preference
LL	Item-based CF	Geographical factor
LAL	Item-based CF	Content factor

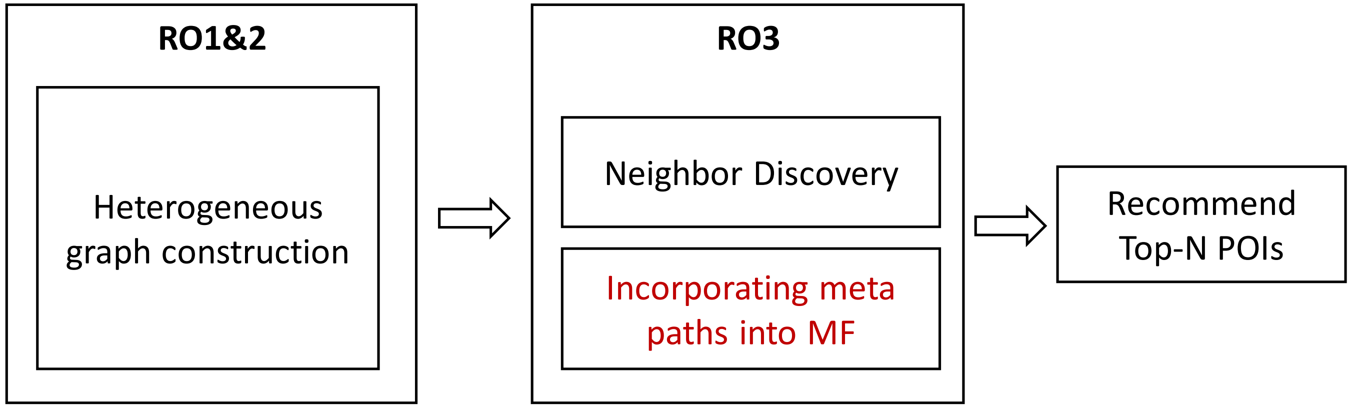
- Neighbor discovery by meta-path based random walk

For meta path UU,





Aspect-aware Geo-Social Matrix Factorization



- A unified latent factor model approach AGS-MF (**A**spect-aware **G**eo-**S**ocial Influence **M**atrix **F**actorization) is proposed by extending MF with the integration of the selected meta paths:

$$\begin{aligned}\mathcal{L} = {} & \frac{1}{2}\sum_i I_{i,j}(r_{i,j} - g(\mathcal{U}_i\mathcal{V}_j^\top))^2 + \frac{\lambda_u}{2}||\mathcal{U}||_F^2 + \frac{\lambda_l}{2}||\mathcal{V}||_F^2 \\ & + \frac{\alpha_u}{2}\sum_{p\in\mathcal{M}_u}\sum_i ||\Omega_{i,p}(\mathcal{U}_i - \sum_{u_k\in\mathcal{N}_p(u_i)} s_{i,k}\mathcal{U}_k)||_F^2 \\ & + \frac{\alpha_l}{2}\sum_{p\in\mathcal{M}_l}\sum_j ||\Theta_{j,p}(\mathcal{V}_j - \sum_{l_q\in\mathcal{N}_p(l_j)} s_{j,q}\mathcal{V}_q)||_F^2 \\ & + \frac{\lambda_\Omega}{2}||\Omega||_F^2 + \frac{\lambda_\Theta}{2}||\Theta||_F^2\end{aligned}$$

u_k is a neighbor of u_i based on meta path $p \in M_u$

l_q is a neighbor of l_j based on meta path $p \in M_l$

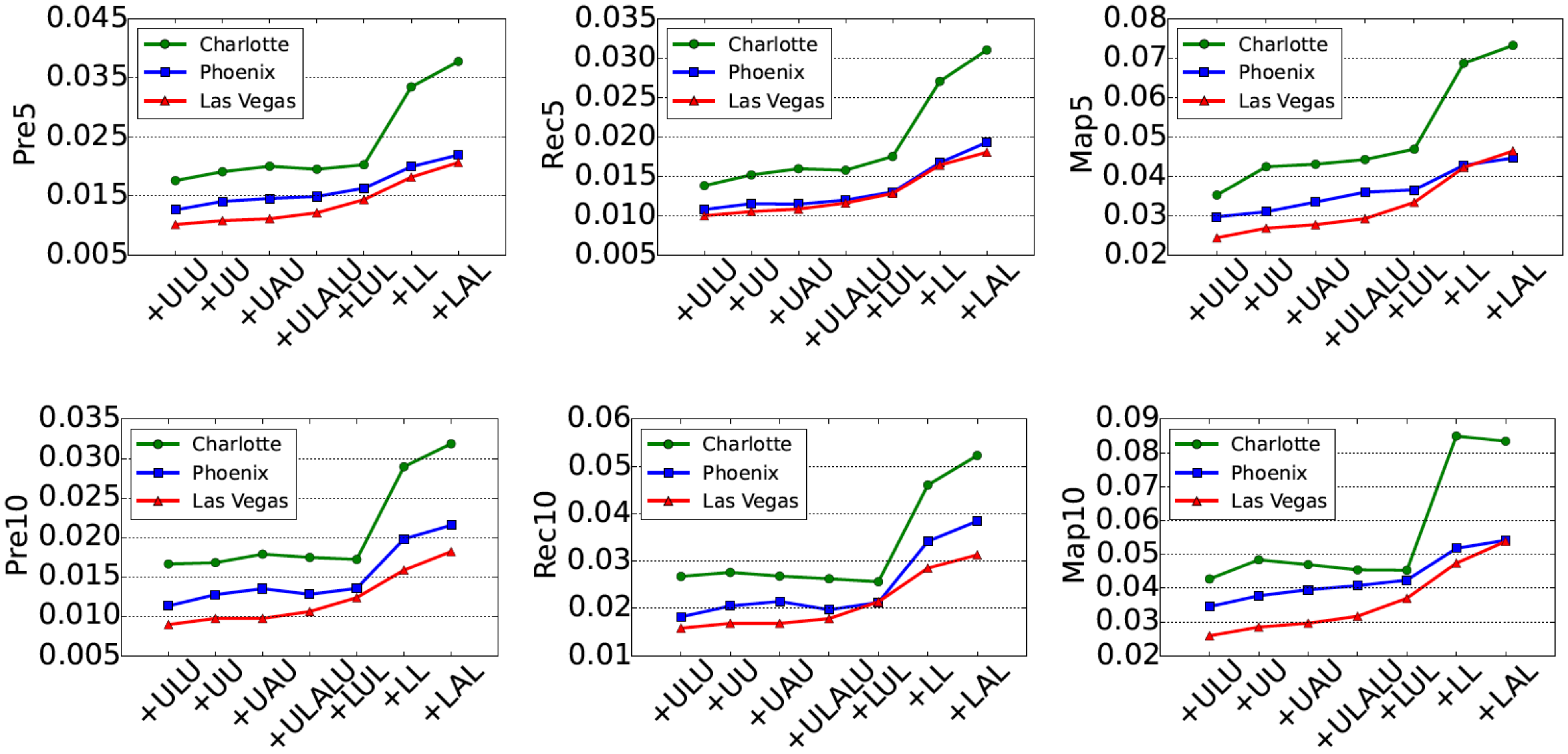
The importance of a meta path for a user or item

Experiments

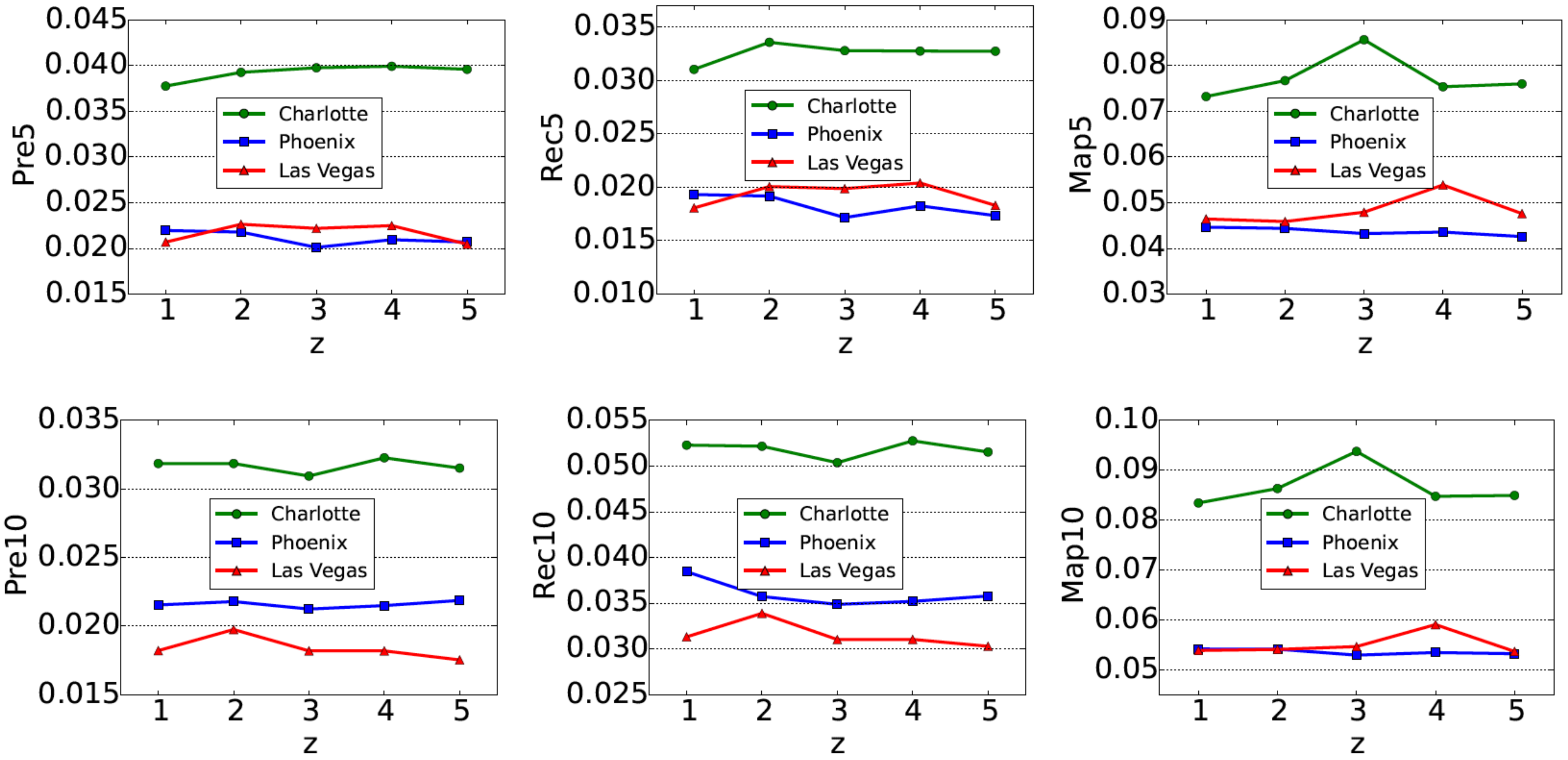
- Aspect Extraction
  - Toolkit provided by Zhang et al. is used in this study.
  - Aspect-opinion with sentiment polarity, e.g., (food, delicious, positive) is generated.
- Cross Validation
  - Users and POIs with less than 10 reviews are filtered out.
  - 5-fold cross validation is adopted for learning and testing.
- Evaluation metric
  - Precision@N
  - Recall@N

Results of Variants of AGS-MF

- Variants of AGSRec



- Effect of meta path length



- Comparison result

City	Metric	UCF	ICF	MF	SRMF	LFBCA	GeoMF	GeoSoCa	TriRank	AGSRec	AGS-MF	Improve(%)
Charlotte	Pre@5	1·483	0·976	1·817	2·586	2·007	2·901	2·586	2·821	3·018*	**3·776**	25·12
	Pre@10	1·456	0·922	1·798	2·134	1·881	2·562	2·134	2·532	2·633*	**3·186**	21·00
	Pre@20	1·433	0·945	1·581	1·917	1·939	2·372	1·917	2·360	**2·474**	2·411*	−2·54
	Rec@5	1·117	0·977	1·485	1·875	1·964	2·509	1·875	2·548	2·625*	**3·103**	18·21
	Rec@10	2·229	1·940	2·805	3·070	3·669	4·557	3·070	4·407	4·760*	**5·250**	9·87
	Rec@20	4·137	4·143	4·849	5·833	7·562	8·194	5·833	8·034	**8·558**	8·092*	−9·02
	MAP@5	3·086	2·039	3·659	5·328	4·188	5·702	5·328	5·586	6·033*	**7·323**	21·28
	MAP@10	3·788	2·566	4·369	5·922	4·939	6·807	5·922	6·747	6·949*	**8·345**	20·09
	MAP@20	4·255	3·061	4·751	6·281	5·589	7·528	6·281	7·203	7·946*	**8·441**	6·25
Phoenix	Pre@5	0·894	0·708	1·248	1·322	1·467	1·905	1·623	1·890	2·060*	**2·196**	6·60
	Pre@10	0·987	0·796	1·173	1·080	1·174	1·889	1·588	1·830	1·939*	**2·154**	11·09
	Pre@20	1·034	0·784	1·101	1·022	1·014	1·801	1·500	1·734	**1·869**	1·852*	−0·91
	Rec@5	0·693	0·771	1·028	1·022	1·536	1·723	1·600	1·751	1·821*	**1·932**	6·10
	Rec@10	1·604	1·742	1·880	1·628	3·592*	3·520	2·901	3·300	3·531	**3·845**	7·04
	Rec@20	3·370	3·489	3·603	3·092	6·130	5·850	5·700	6·126	**6·502**	6·163*	−5·21
	MAP@5	2·147	1·562	2·920	3·768	3·919	3·842	3·498	3·771	3·921*	**4·464**	13·85
	MAP@10	2·685	2·094	3·479	4·098	4·760	4·859	4·013	4·690	4·972*	**5·416**	8·93
	MAP@20	3·136	2·457	3·895	4·477	5·217	5·742	5·056	5·304	5·823*	**5·864**	0·70
Las Vegas	Pre@5	0·626	0·588	0·948	0·960	1·120	1·509	1·470	1·589	1·653*	**2·069**	25·17
	Pre@10	0·542	0·501	0·878	0·859	1·060	1·291	1·280	1·331	1·371*	**1·821**	32·82
	Pre@20	0·485	0·433	0·754	0·825	0·975	1·406	1·370	1·416	1·455*	**1·495**	2·75
	Rec@5	0·552	0·508	0·865	0·627	0·672	1·461	1·420	1·431	1·474*	**1·806**	22·52
	Rec@10	0·939	0·882	1·478	1·182	1·406	2·565	2·260	2·459	2·557*	**3·132**	22·49
	Rec@20	1·647	1·610	2·436	2·413	2·823	5·345	5·180	5·345	**5·502**	5·398*	−1·89
	MAP@5	1·340	1·310	2·191	1·714	2·222	3·428	3·250	3·306	3·571*	**4·644**	30·05
	MAP@10	1·612	1·586	2·530	2·074	2·685	3·860	3·820	3·882	4·154*	**5·385**	29·63
	MAP@20	1·819	1·613	2·251	2·393	3·063	4·282	4·100	4·349	4·883*	**5·580**	14·27

Contribution

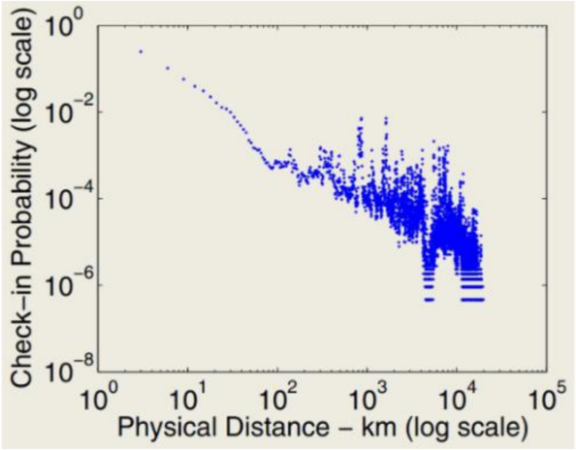
- Propose a novel heterogeneous graph model to incorporate various influence factors in a unified way.
- Propose a novel latent factor model that exploits semantics of the graph and can learn personalized importance of each influential factor for each user and POI.

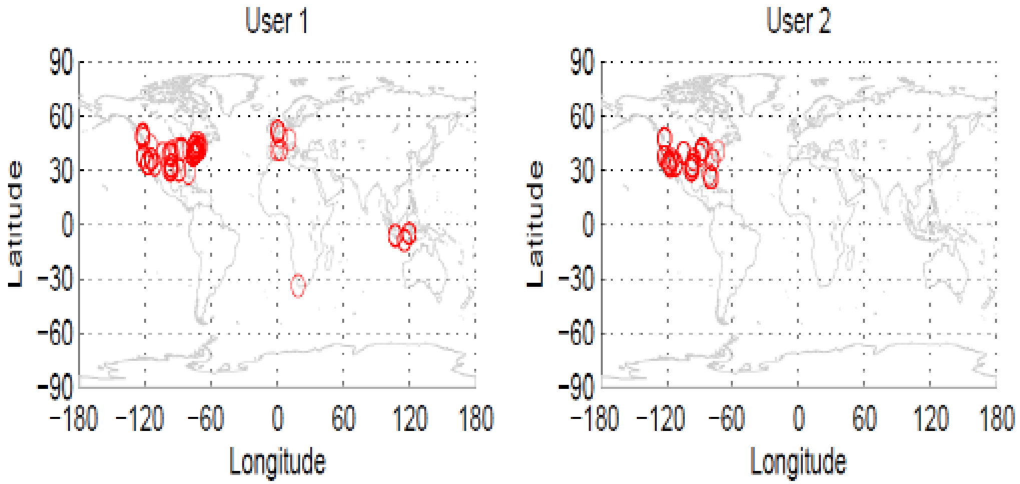
Thank you.

Guo Qing
qguo006@e.ntu.edu.sg

Geographical Factor

- Users' check-in behavior is geographically constrained.
- Representative methods that model geographical factor
  - Power law distribution model (Ye et al., 2011; Yuan et al., 2014)
  - Gaussian distribution (Cho et al., 2010; Cheng et al., 2012)
  - Kernel density estimation (Zhang et al., 2013)





Social Factor

- Friends in LBSNs share more common interests than non-friends.
- User-based Collaborative Filtering Framework.
  - Friend-based CF (Ye et al., 2010; Li et al., 2016)

$$r_{ij} = \frac{\sum_{u_k \in F_i} r_{kj} \cdot \omega_{ik}}{\sum_{u_k \in F_i} r_{kj}}$$

  - Random Walk (Noulas et al., 2012; Wang et al., 2013)
- Matrix Factorization Framework
  - The latent features of friends are similar in latent space (Cheng et al., 2011; Yang et al., 2013)

$$\arg\min_{U,L} \underbrace{\boxed{\sum_{i=1}^{|\mathcal{U}|}\sum_{j=1}^{|\mathcal{L}|} I_{ij}(g(c_{ij}) - g(U_i^T L_j))^2 + \lambda_1 ||U||_F^2 + \lambda_2 ||V||_F^2}}_{\text{Traditional MF}} + \beta \underbrace{\boxed{\sum_{i=1}^{N} \sum_{u_f \in F_i} sim(i,f) ||U_i - U_f||^2}}_{\text{Social regulation}}$$

Content Factor

- Inferring users' preference for categories
  - Categories provides information about the function of POIs.
  - Bao et al. (2012) exploit the hierarchy of categories.
- Inferring users' preference for aspects
  - Aspects can be used for modeling users' preference in a finer granularity (Yang et al., 2013; He et al., 2015).
- Geo-Topic model
  - A region is associated with certain topics which is hidden in reviews, categories and tags (Yuan et al., 2013; Yin et al., 2013; Zhao et al., 2015).
  - Geographical + Content