Meta-Graph Based Recommendation Fusion over Heterogeneous Information Networks

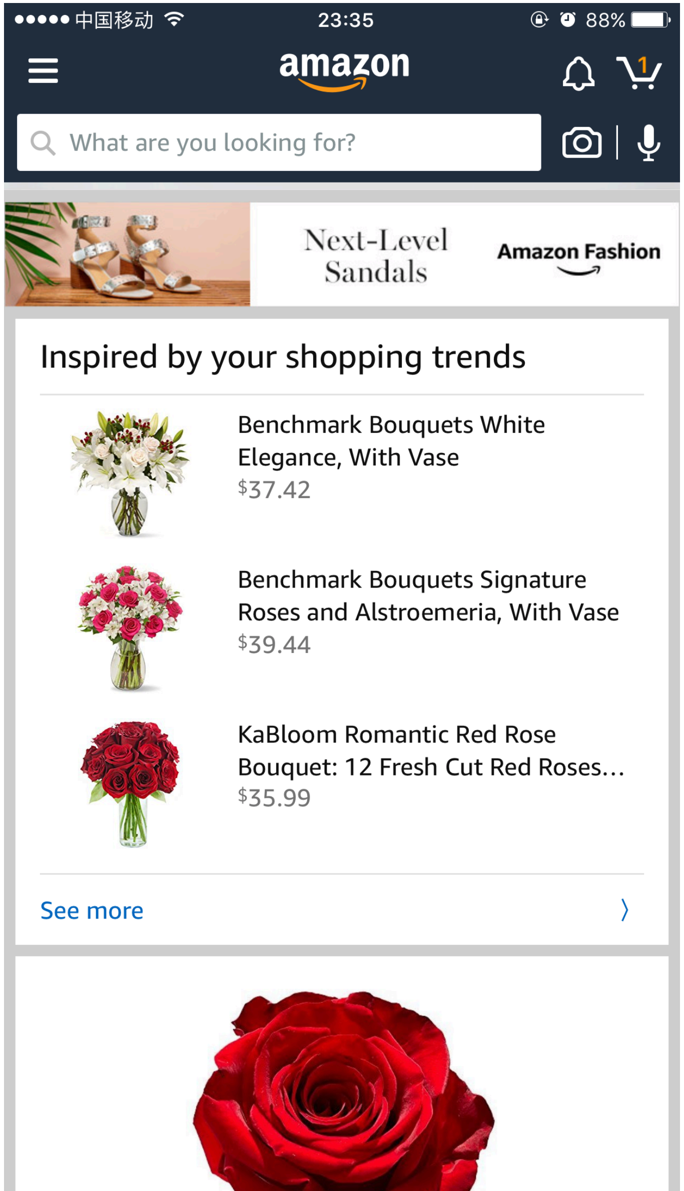
Huan Zhao

Joint work with Quanming Yao, Jianda Li, Yangqiu Song and Dik Lun Lee

Department of CSE, HKUST, Hong Kong

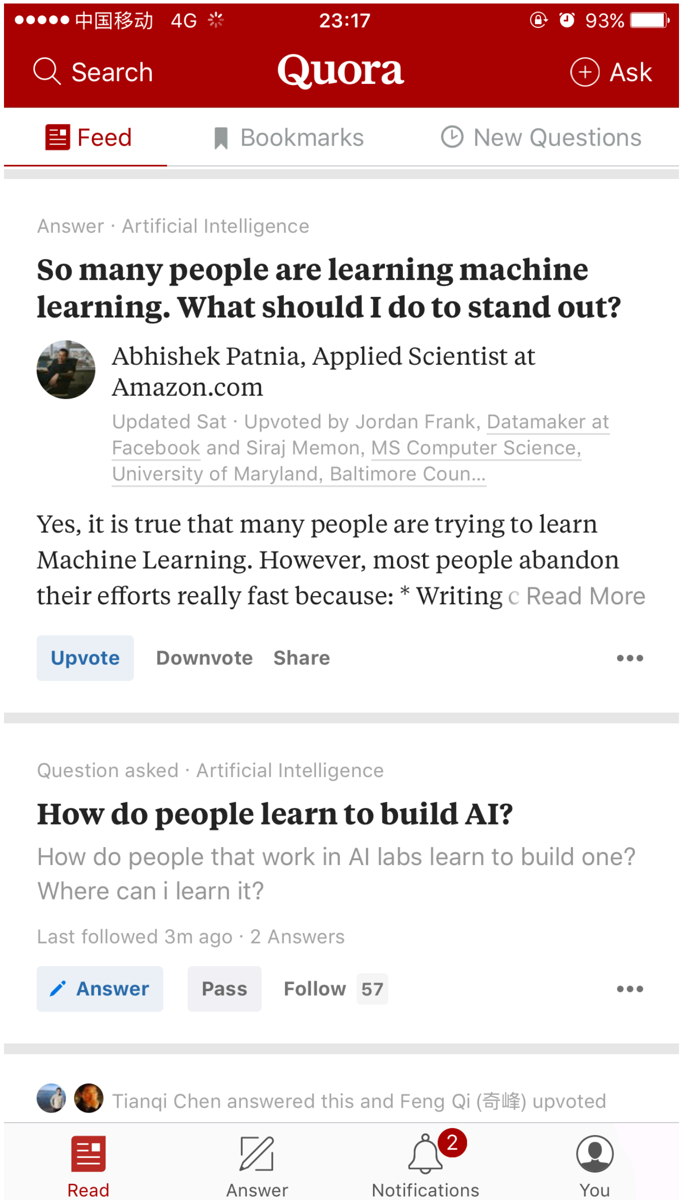
香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

Recommender System



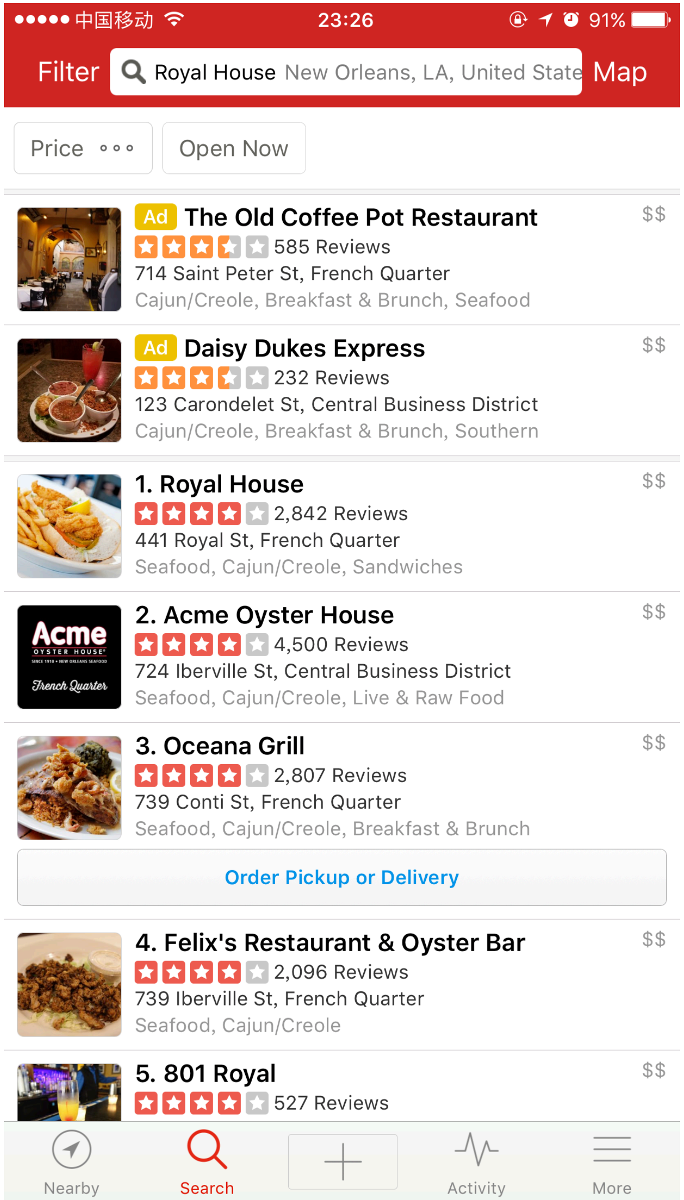
Products Recommendation

Recommender System



Questions and Answers Recommendation

Recommender System



Restaurants Recommendation

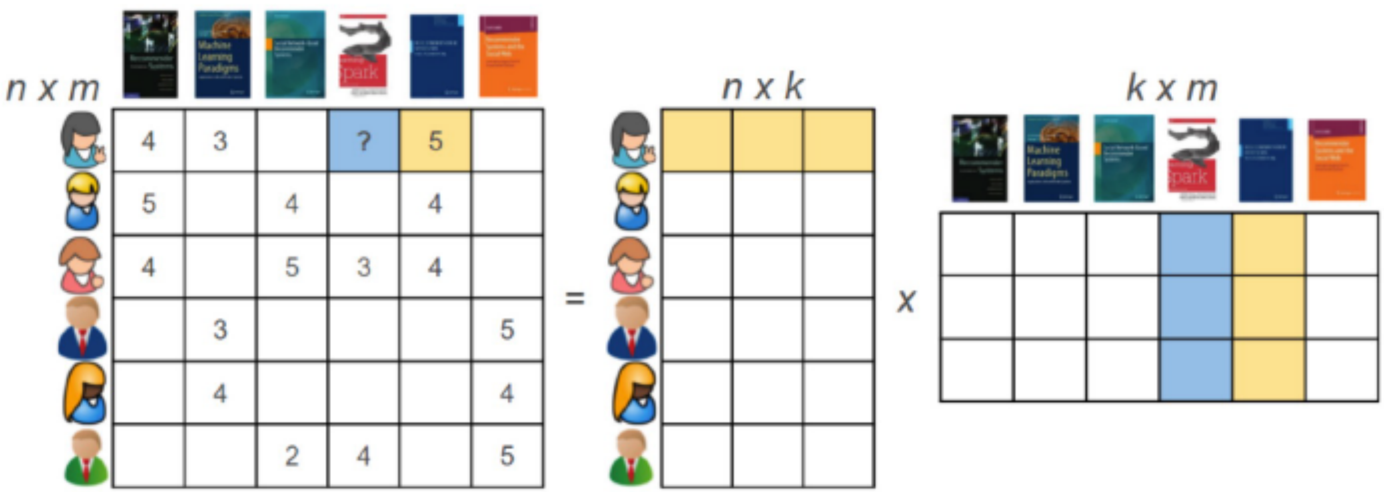
Recommender System

- Recommender systems (RS) are everywhere.
- They are not only useful for people, but also create huge revenues for companies.

- The most popular RS method is collaborative filtering (CF).
  - User-based CF
  - Item-based CF
  - Matrix Factorization (MF)

Matrix Factorization

- Matrix Factorization is one of the most popular methods for collaborative filtering



$$\min_{U,B} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} \left(R_{ij} - u_i b_j \right)^2 + \frac{\lambda_1}{2} ||U||_F^2 + \frac{\lambda_2}{2} ||B||_F^2$$

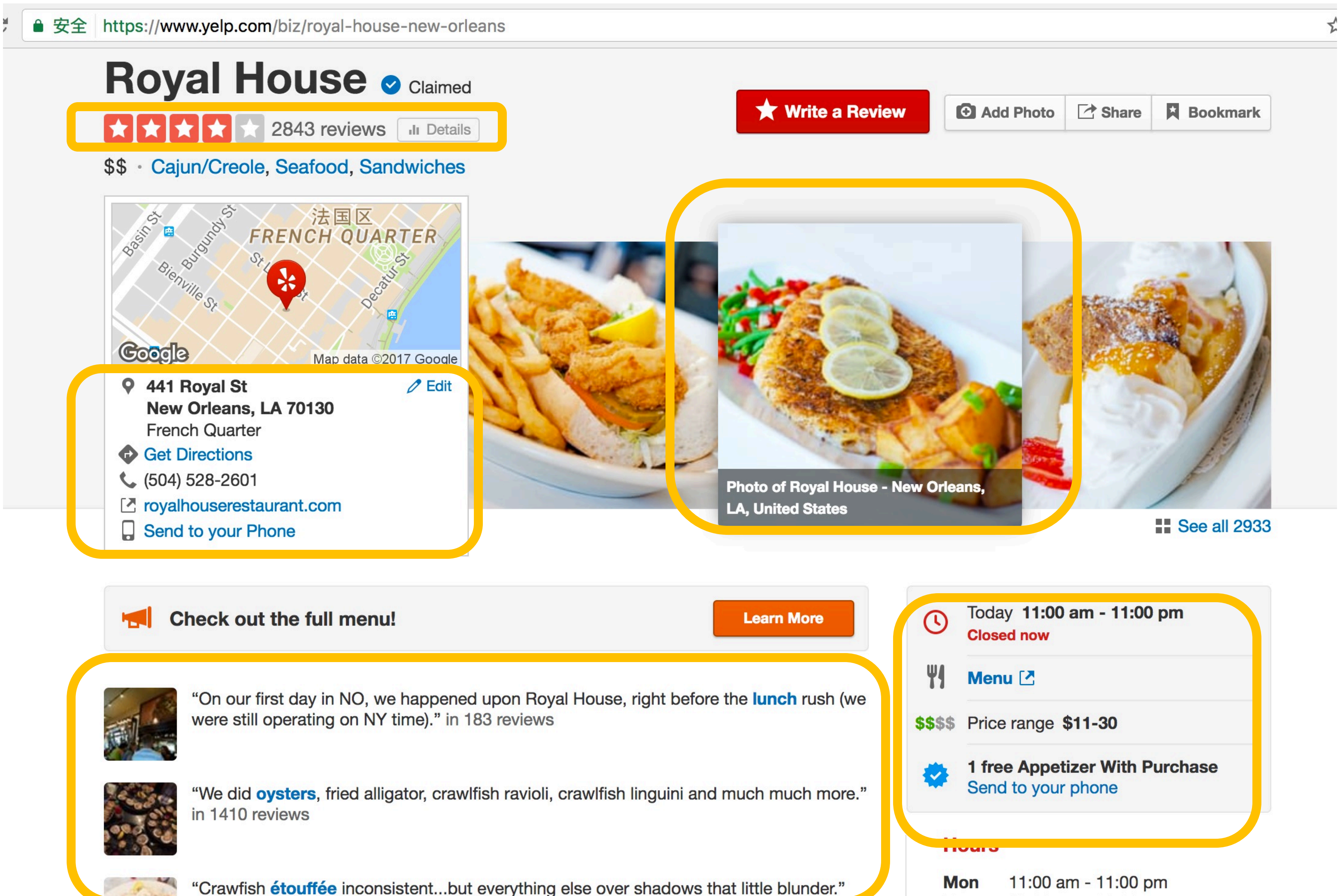
https://buildingrecommenders.wordpress.com/2015/11/18/overview-of-recommender-algorithms-part-2/

Problems of MF

- Sparsity of the rating matrix
  - More than 99% entries are missing.
- Cold Start
  - Some users or items have no ratings.

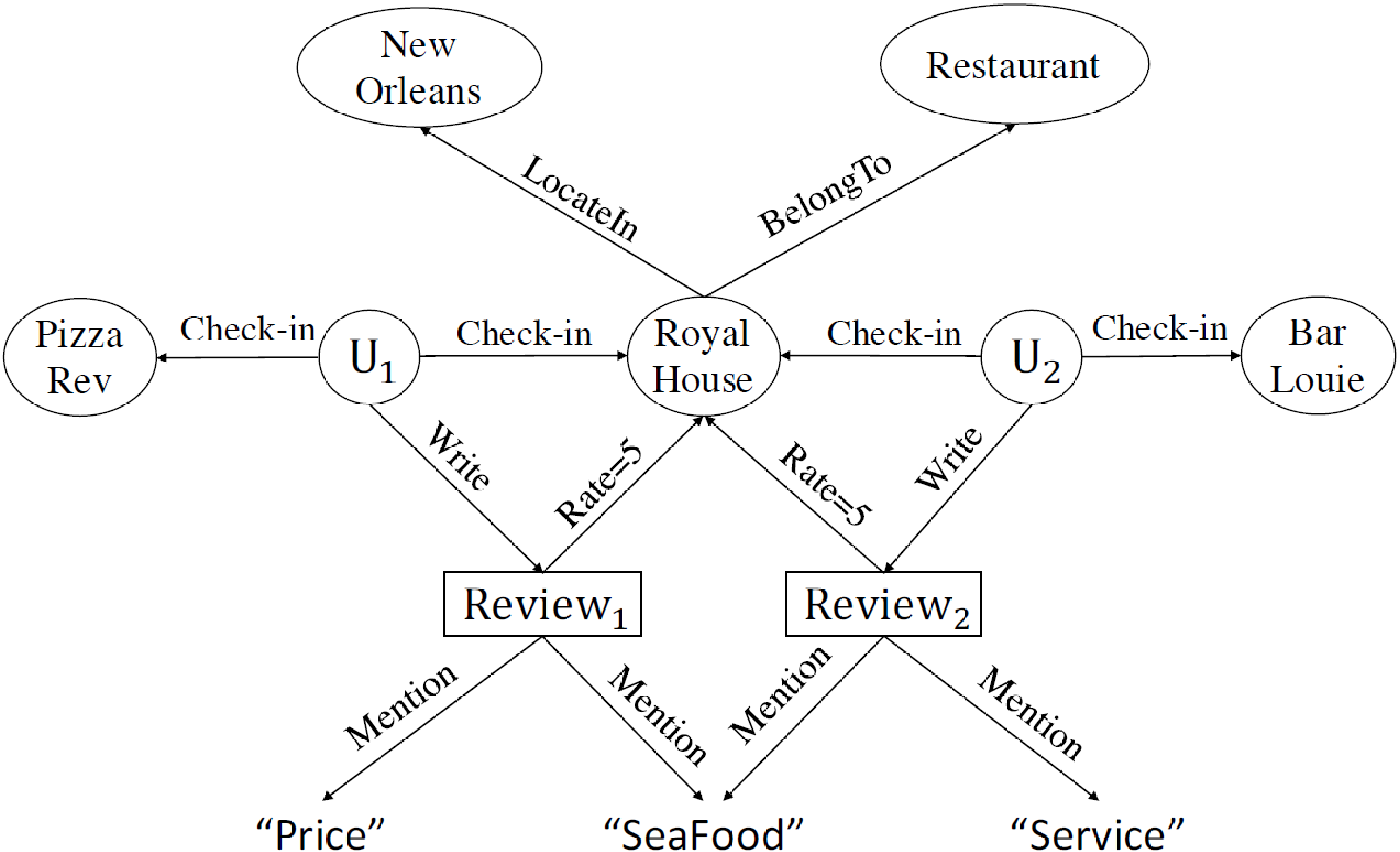
- More importantly, MF is not applicable for nowadays RS.
  - The rating matrix is not sufficient for rich side information.

Example



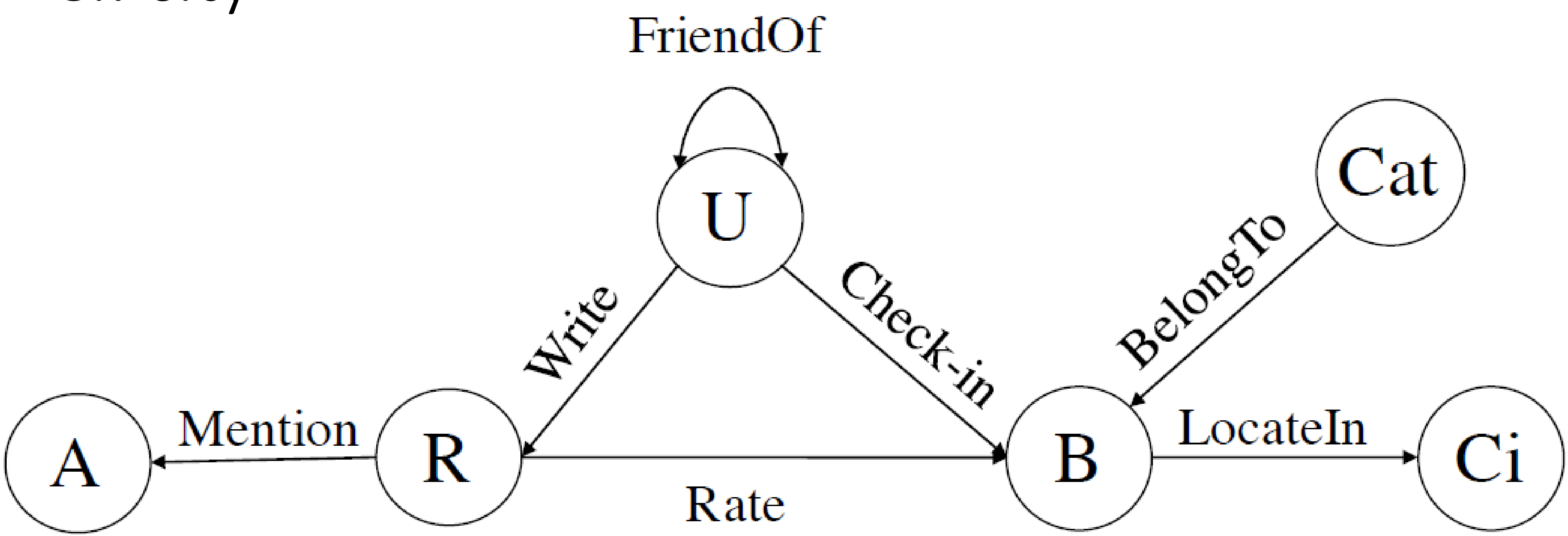
安全 https://www.yelp.com/biz/royal-house-new-orleans
Royal House
Claimed
2843 reviews
Details
Write a Review
Add Photo
Share
Bookmark
$$ · Cajun/Creole, Seafood, Sandwiches
法国区
FRENCH QUARTER
Basin St
Burgundy St
Bienville St
Decatur St
Google
Map data ©2017 Google
441 Royal St
New Orleans, LA 70130
French Quarter
Edit
Get Directions
(504) 528-2601
royalhouserestaurant.com
Send to your Phone
Photo of Royal House - New Orleans, LA, United States
See all 2933
Check out the full menu!
Learn More
Today 11:00 am - 11:00 pm
Closed now
Menu
$$$$ Price range $11-30
1 free Appetizer With Purchase
Send to your phone
Hours
Mon 11:00 am - 11:00 pm
"On our first day in NO, we happened upon Royal House, right before the lunch rush (we were still operating on NY time)." in 183 reviews
"We did oysters, fried alligator, crawlfish ravioli, crawlfish linguini and much much more." in 1410 reviews
"Crawfish étouffée inconsistent...but everything else over shadows that little blunder."

It's a Heterogeneous Information Network!



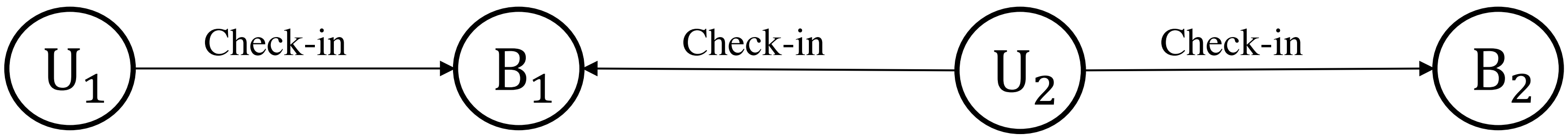
A Typical Network Schema of Yelp

- R: reviews;
- U: users;
- B: business;
- Cat: category of item;
- Ci: city

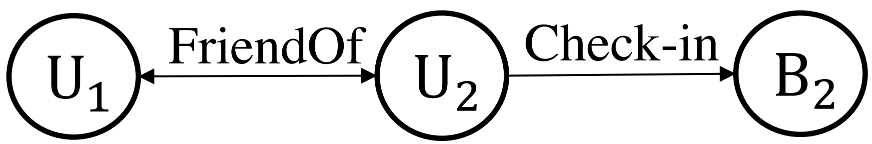


Meta-path based RS

- Recommending strategies can be modeled by meta-paths



User-based CF



Social Recommendation

Content-based Recommendation

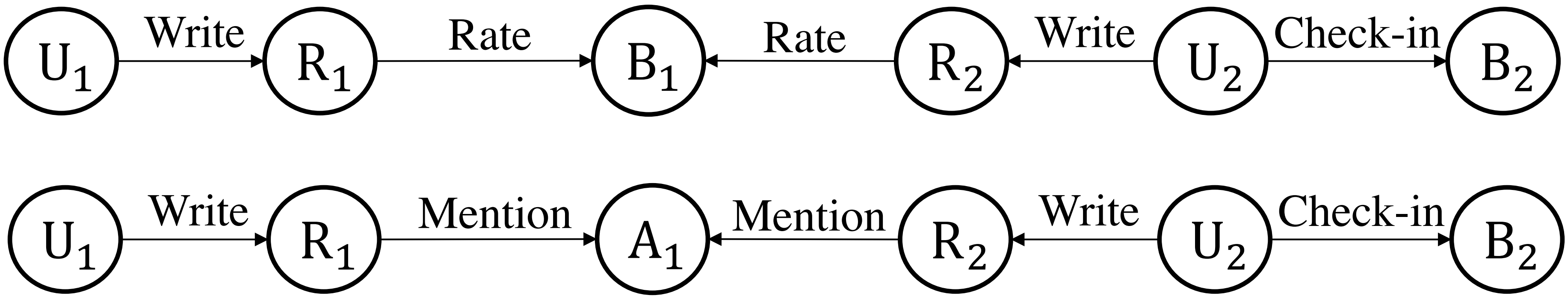
Meta-path based RS

- Similarity is the key factor.

- Embeddings of users and items are learnt from similarities, which are used in a linear model for rating prediction. [Yu et. al., WSDM'14]

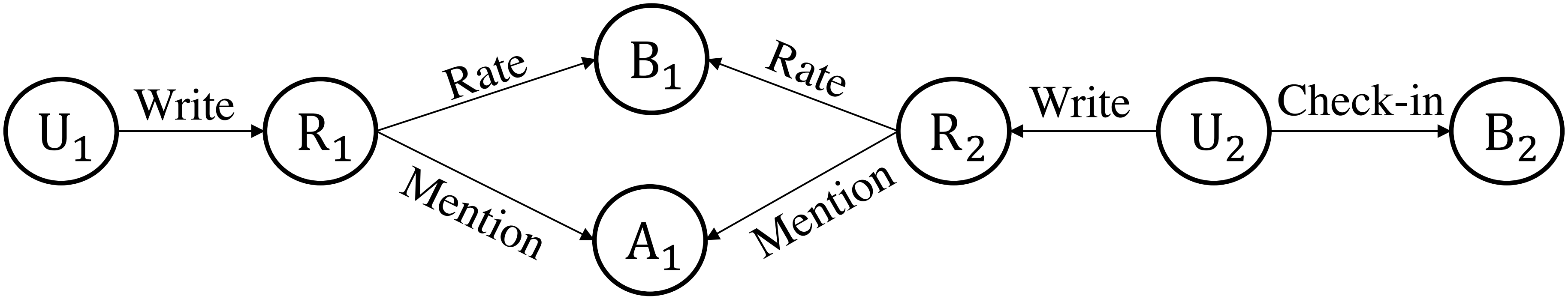
- Ratings are predicted by weighted ensemble of similar users' ratings, where similarities are learnt from different meta-paths. [Shi et. al., CIKM' 15]

Problems of meta-path



What if R_1 and R_2 mention the same aspect for the same business?

Problems of meta-path



What if R_1 and R_2 mention the same aspect for the same business?

Meta-path fails for such complex relations, which are very common in nowadays recommending scenarios.

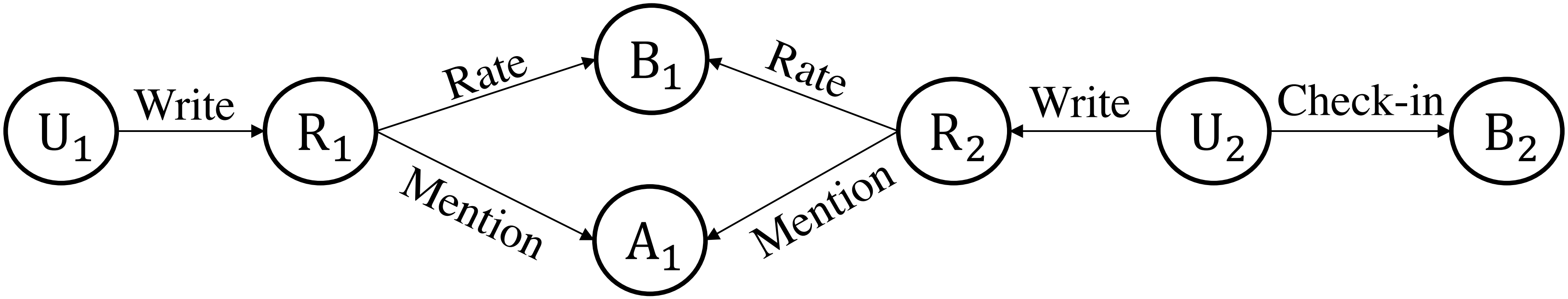
Our Work

- Meta-graph based recommendation.

- Powerful prediction model by “MF + FM”.

- Automatic selection of important meta-graphs.

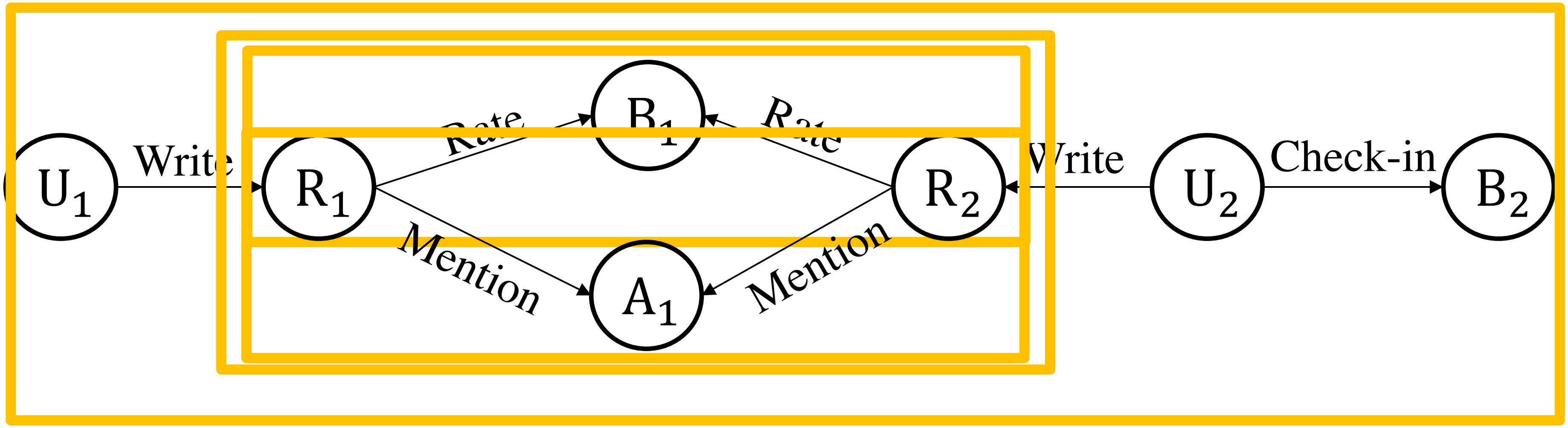
Meta-graph based RS



A meta-graph is a directed acyclic graph (DAG) with a single source node and a single sink (target) node. [Huang et. al., KDD’16, Fang et. al., ICDE’16]

Meta-path is a special case of meta-graph.

Meta-graph based RS



Compute C_{P_1}: $C_{P_1} = W_{RB} \cdot W_{RB}^T$

Compute C_{P_2}: $C_{P_2} = W_{RA} \cdot W_{RA}^T$

Compute C_{S_r}: $C_{S_r} = C_{P_1} \odot C_{P_2}$

Compute C_M: $C_M = W_{UR} \cdot C_{S_r} \cdot W_{UR}^T \cdot W_{UB}$

Assemble multiple meta-graphs

- In previous work, linear ensemble methods are used. [Yu et. al., WSDM'14, Shi et. al., CIKM'15]
- We argue in this work non-linear relations also needed to be captured.

- Factorization Machine (FM) is chosen.
  - Capture non-linear interactions among features.
  - Good ability of prediction in recommending scenario.

Assemble multiple meta-graphs

- Factorization Machine [Rendle ICDM'10, TIST'12]

$$\hat{y}(w,V) = \boxed{w_0 + \sum_{i=1}^{d} w_i x_i^n} + \boxed{\sum_{i=1}^{d} \sum_{j=i+1}^{d} \langle v_i, v_j \rangle x_i^n x_j^n}$$

  - Linear model.
  - Non-linear, second order interactions.

"MF + FM" framework

- For each meta-graph, do MF:

$$\min_{U,B} \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{n} I_{ij} (R_{ij} - u_i b_j)^2 + \frac{\lambda_1}{2} ||U||_F^2 + \frac{\lambda_2}{2} ||B||_F^2$$

- Given all MF latent features:
  - *L* meta-graphs
  - *F* dimension of MF

$$\mathbf{x}^n = \underbrace{\mathbf{u}_i^{(1)}, \ldots, \mathbf{u}_i^{(l)}, \ldots, \mathbf{u}_i^{(L)}}_{L \times F} \underbrace{\mathbf{b}_j^{(1)}, \ldots, \mathbf{b}_j^{(l)}, \ldots, \mathbf{b}_j^{(L)}}_{L \times F}$$

- Do FM:

$$\hat{y}(w, V) = w_0 + \sum_{i=1}^{d} w_i x_i^n + \sum_{i=1}^{d} \sum_{j=i+1}^{d} \langle v_i, v_j \rangle x_i^n x_j^n$$

Meta-graph Selection

- The original cost function of FM

$$\min_{w,V} \sum_{n=1}^{N} \left(y^n - \hat{y}^n(w,V)\right)^2$$

$$\hat{y}(w,V) = w_0 + \sum_{i=1}^{d} w_i x_i^n + \sum_{i=1}^{d} \sum_{j=i+1}^{d} \langle v_i, v_j \rangle x_i^n x_j^n$$

- + group lasso:

$$\Phi_w(w) = \sum_{l=1}^{2L} ||w_l||_2 \qquad \Phi_V(V) = \sum_{l=1}^{2L} ||V||_2$$

L meta-graphs

- In side meta-graph: L2 norm
- Between meta-graphs: L1 norm

nonmonotonous accelerated proximal gradient (nmAPG) algorithm [Li and Lin, NIPS'15]

Experiments

Yelp-200k				
Relations(A-B)	Number of A	Number of B	Number of (A-B)	Avg Degrees of A/B
User-Business	36,105	22,496	191,506	5.3/8.5
User-Review	36,105	191,506	191,506	5.3/1
User-User	17,065	17,065	140,344	8.2/8.2
Business-Category	22,496	869	67,940	3/78.2
Business-Star	22,496	9	22,496	1/2,499.6
Business-State	22,496	18	22496	1/1,249.8
Business-City	22,496	215	22,496	1/104.6
Review-Business	191,506	22,496	191,506	1/8.5
Review-Aspect	191,506	10	955,041	5/95,504.1

Amazon-200k				
Relations(A-B)	Number of A	Number of B	Number of (A-B)	Avg Degrees of A/B
User-Business	59,297	20,216	183,807	3.1/9.1
User-Review	59,297	183,807	183,807	3.1/1
Business-Category	20,216	682	87,587	4.3/128.4
Business-Brand	95,33	2,015	9,533	1/4.7
Review-Business	183,807	20,216	183,807	1/9.1
Review-Aspect	183,807	10	796,392	4.3/79,639.2

Experiments

- Two more datasets are used, provided by [Shi et. al., CIKM' 15].
  - denoted as CIKM-Yelp and CIKM-Douban.
- Density

	Amazon-200k	Yelp-200k	CIKM-Yelp	CIKM-Douban
Density	0.015%	0.024%	0.086%	0.630%

- Evaluation Metric

$$RMSE = \sqrt{\frac{\sum_{(i,j)\in R_{test}}(R_{ij}-\hat{R}_{ij})}{|R_{test}|}}$$

Smaller means better performance.

Comparison Results

Traditional Approaches

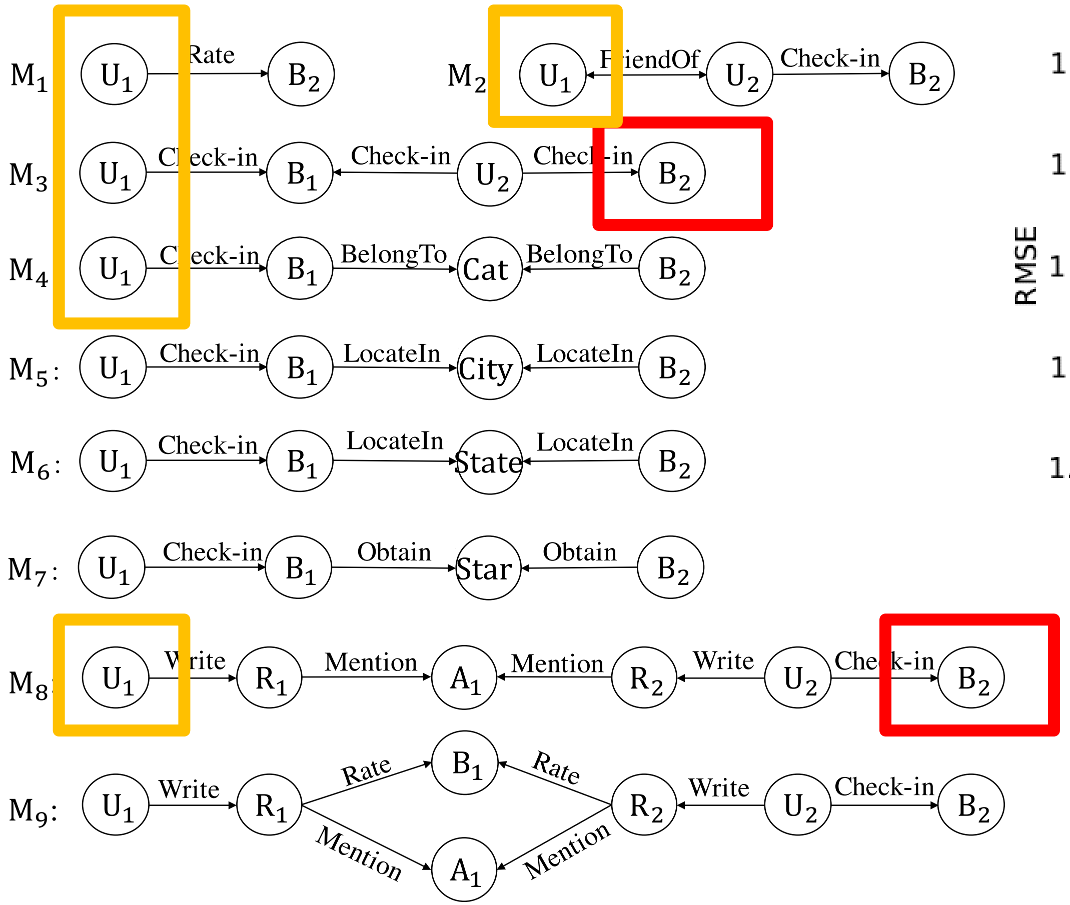
HIN Based Approaches

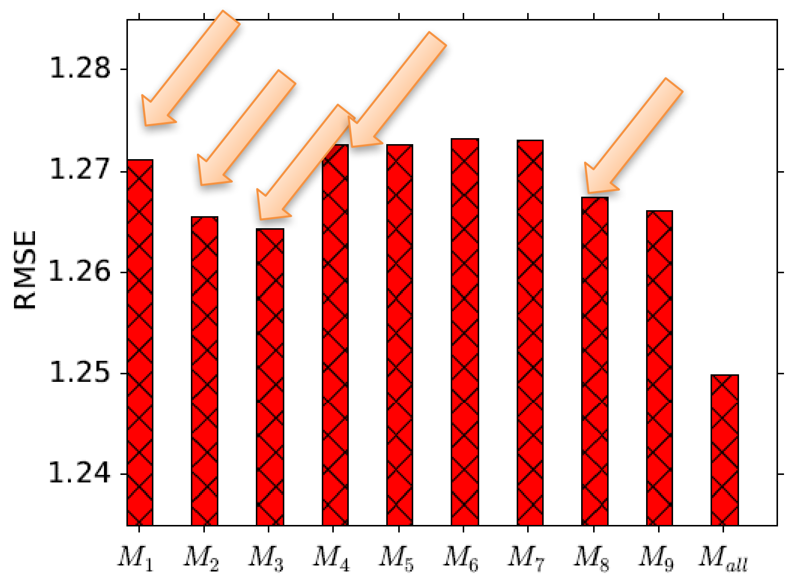
	Amazon-200k	Yelp-200k	CIKM-Yelp	CIKM-Douban
RegSVD	2.9656 (+60.0%)	2.5141 (+49.9%)	1.5323 (+27.7%)	0.7673 (+9.0%)
FMR	1.3462 (+11.9%)	1.7637 (+28.6%)	1.4342 (+22.8%)	0.7524 (+7.2%)
HeteRec	2.5368 (+53.2%)	2.3475 (+47.0%)	1.4891 (+25.6%)	0.7671 (+9.0%)
SemRec	- -	1.4603 (+13.8%)	1.1559 (+4.2%)	0.7216 (+3.2%)
FMG	**1.1864**	**1.2588**	**1.1074**	**0.6985**

- FMG(Our model) consistently beats all baselines on all datasets, demonstrating the effectiveness of our method.
- FMG beats SemRec [Shi et. al., CIKM' 15] on the same datasets with the same meta-paths, demonstrating the effectiveness of the adoption of FM.
- SemRec is better than HeteRec [Yu et. al., WSDM'2014] mainly because it uses meta-path in style "U-*-U-B", while HeteRec uses "U-B-*-B".

Selected Meta-graphs for Yelp

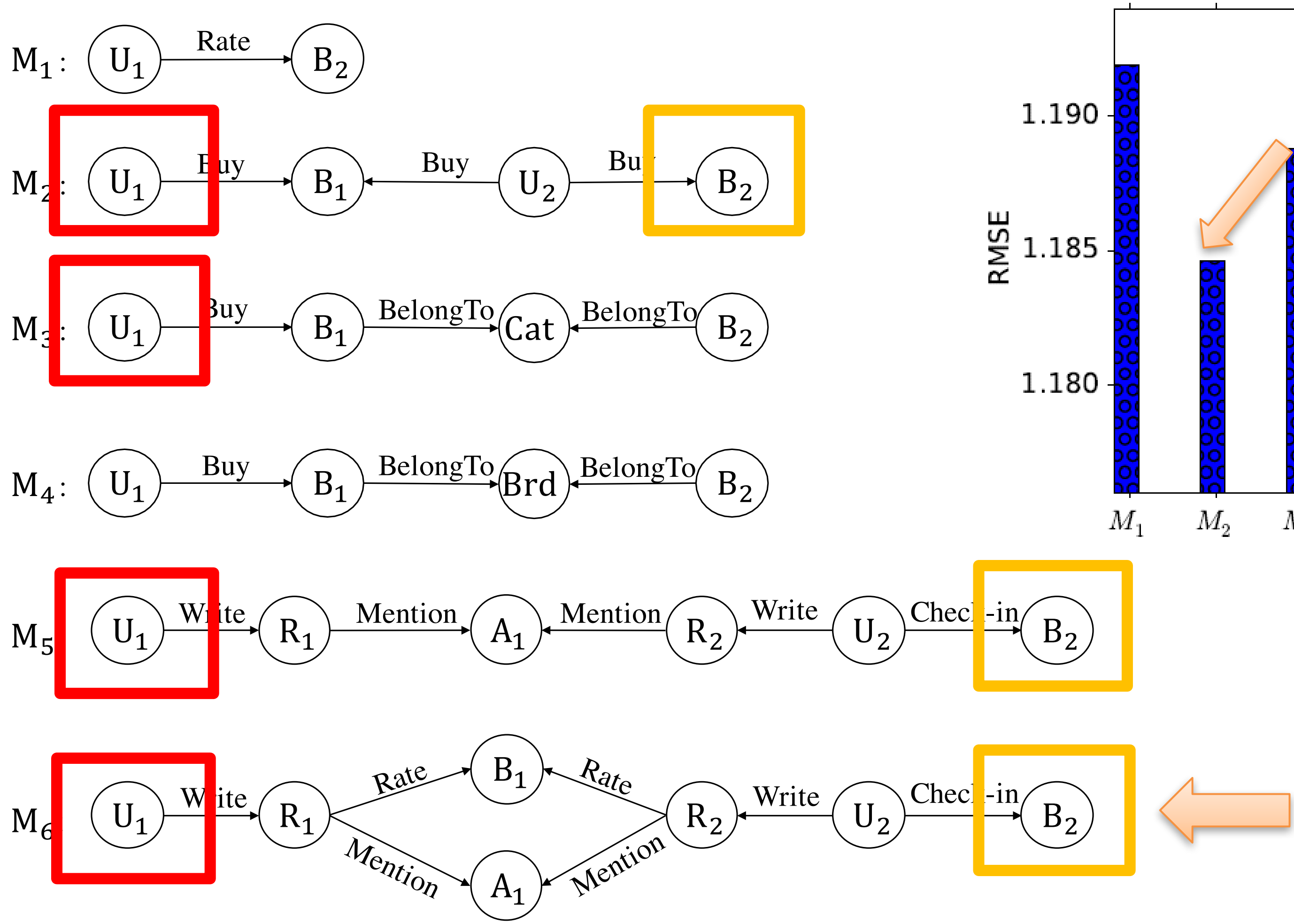
		User-Part		Item-Part	
		w	**V**	**w**	**V**
Yelp	Important	$M_1 - M_4, M_6, M_8$	$M_1 - M_3, M_5, M_8$	$M_1 - M_5, M_8, M_9$	M_3, M_8
	Useless	M_5, M_7, M_9	M_4, M_6, M_7, M_9	M_6, M_7	$M_1, M_2, M_4 - M_7, M_9$

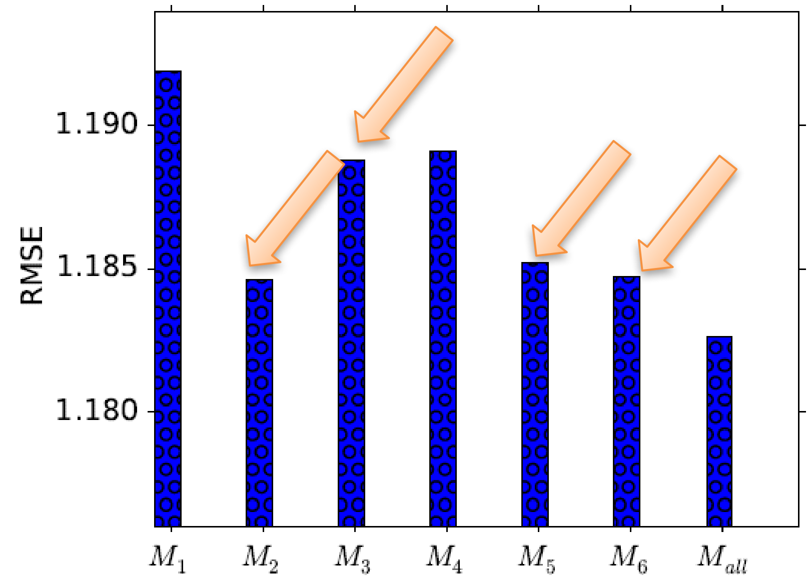




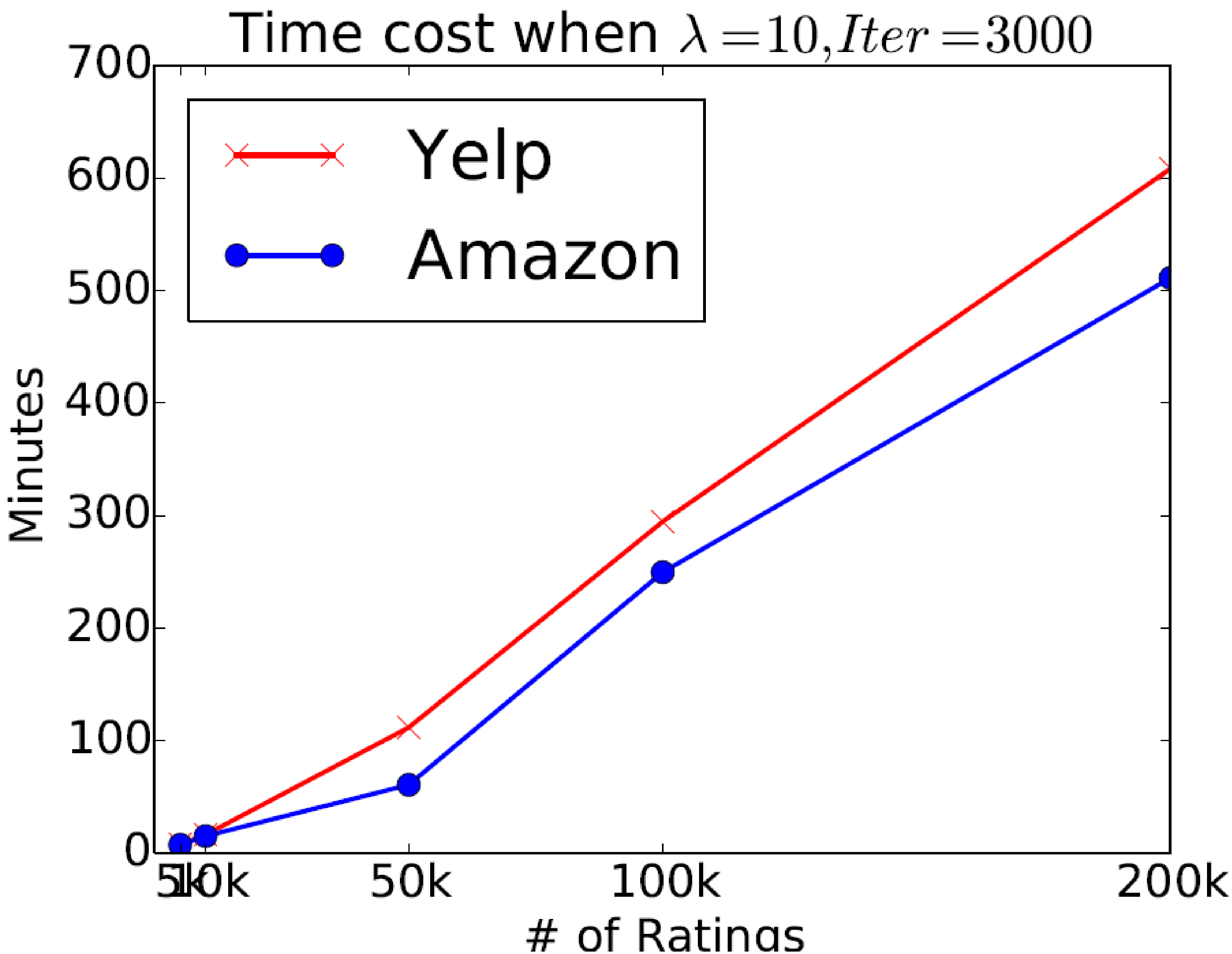
Selected Meta-graphs for Amazon

		User-Part		Item-Part	
		w	**V**	**w**	**V**
Amazon	Important	$M_1 - M_3, M_5$	$M_1 - M_6$	M_2, M_3, M_5, M_6	M_2, M_5, M_6
	Useless	M_4, M_6	-	M_1, M_4	M_1, M_3, M_4





Scalability of Algorithm



Summary

- The first work to use meta-graph on HIN to unify rich side information for recommender system.
- “MF+FM” is used to effectively assemble all the meta-graphs.
- Group lasso is used to select important meta-graphs.