wsdm 2014
New York City
February 24-28

Personalized Entity Recommendation: A Heterogeneous Information Network Approach

Xiao Yu, Xiang Ren, Yizhou Sun†, Quanquan Gu,
Bradley Sturt, Urvashi Khandelwal, Brandon Norick, Jiawei Han
University of Illinois, at Urbana Champaign
†Northeastern University

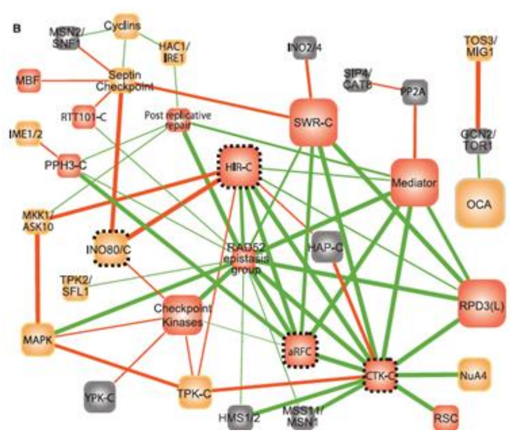
02/26/2014

Booming Age of
Heterogeneous Information Networks

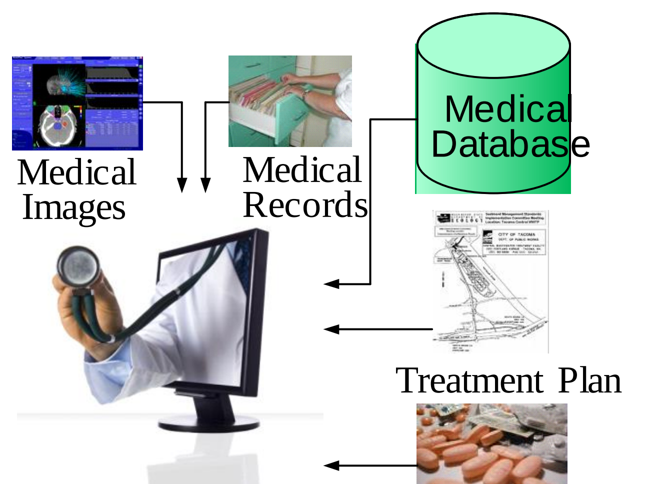
Information network with multi-typed entities and relationships

Social Media

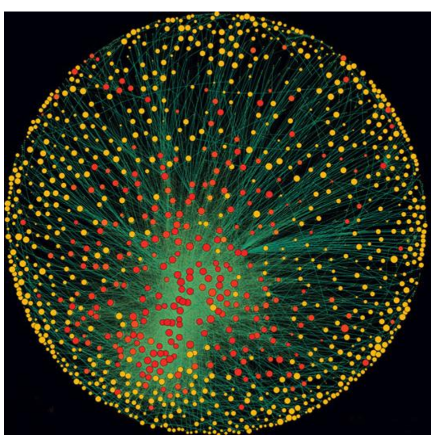
E-Commerce



Biology



Healthcare

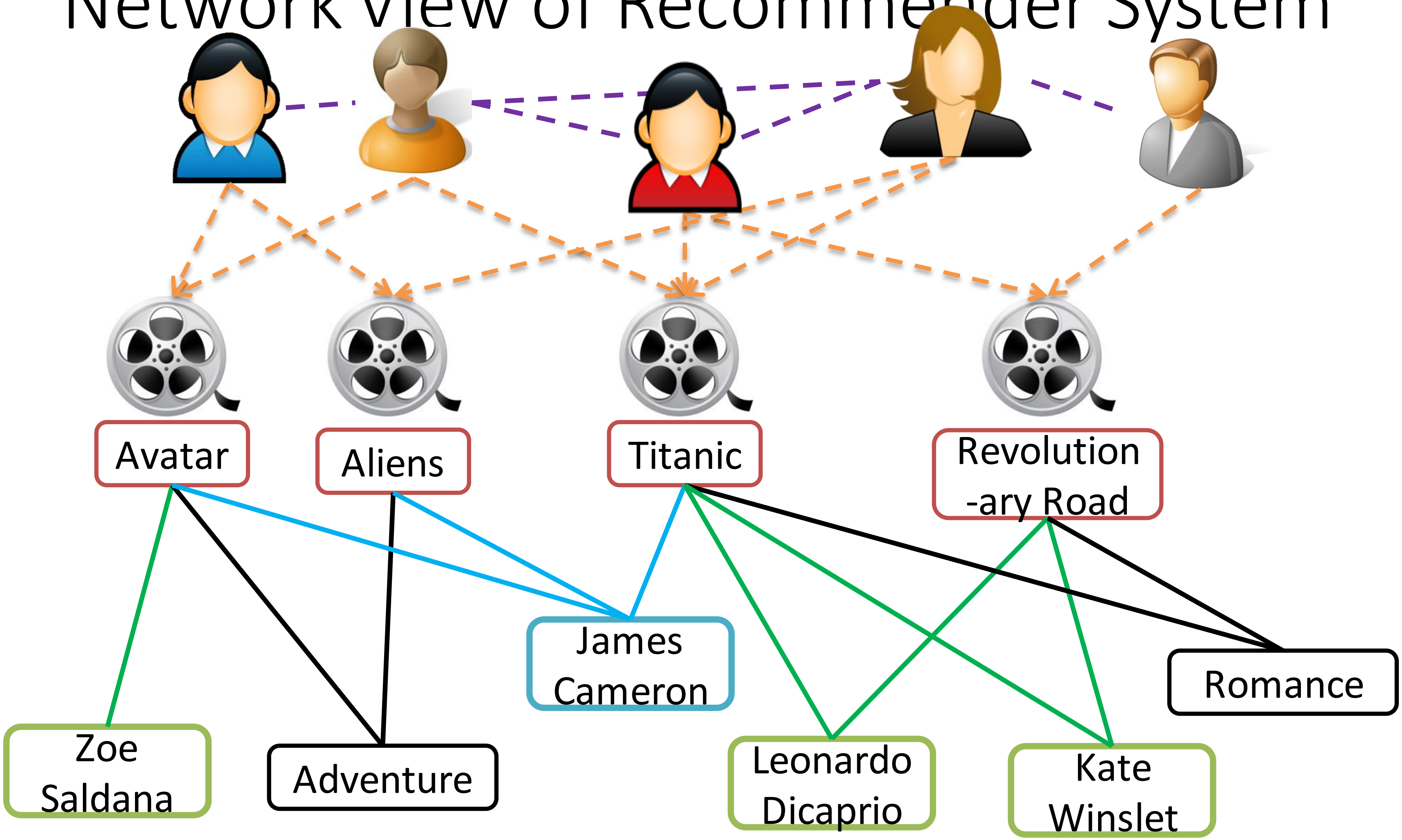


Global Economy

Hybrid Collaborative Filtering with Networks

- Utilizing network relationship information can enhance the recommendation quality
- However, most of the previous studies only use single type of relationship between users or items (e.g., social network Ma,WSDM'11, trust relationship Ester, KDD'10, service membership Yuan, RecSys'11)

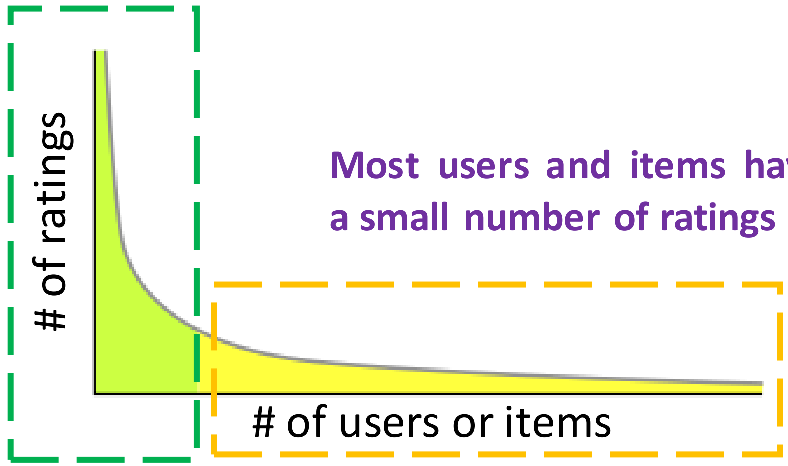
The Heterogeneous Information Network View of Recommender System



Relationship heterogeneity alleviates data sparsity

Collaborative filtering methods suffer from data sparsity issue

A small number of users and items have a large number of ratings

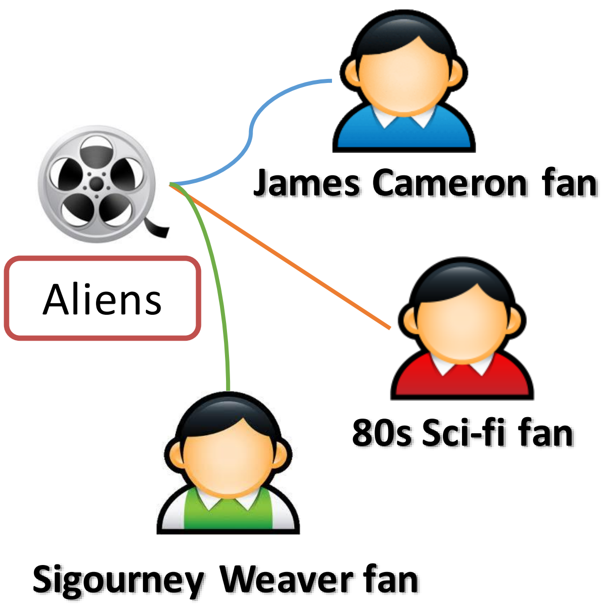


Most users and items have a small number of ratings

- Heterogeneous relationships complement each other
- Users and items with limited feedback can be connected to the network by different types of paths
  - Connect new users or items (cold start) in the information network

Relationship heterogeneity based personalized recommendation models

Different users may have different behaviors or preferences



Different users may be interested in the same movie for different reasons

Two levels of personalization

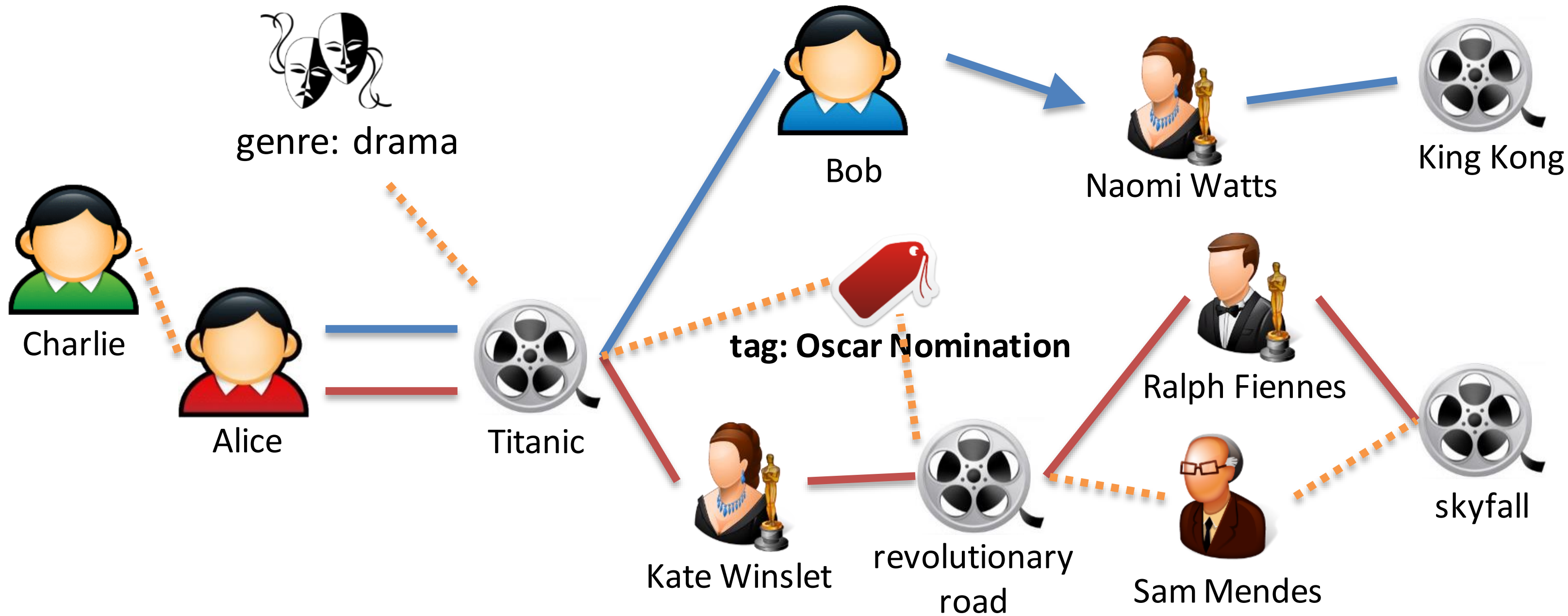
Data level

- Most recommendation methods use one model for all users and rely on personal feedback to achieve personalization

Model level

- With different entity relationships, we can learn personalized models for different users to further distinguish their differences

Preference Propagation-Based Latent Features



Generate *L* different meta-path (path types) connecting users and items

→

Propagate user implicit feedback along each meta-path

→

Calculate latent-features for users and items for each meta-path with NMF related method

Recommendation Models

Observation 1: Different meta-paths may have different importance

Global Recommendation Model

ranking score

features for user *i* and item *j*

$$\boxed{\hat{r}(u_i, e_j)} = \sum_{q=1}^{L} \theta_q \cdot \boxed{\hat{U}_i^{(q)} \hat{V}_j^{(q)T}} \qquad (1)$$

the q-th meta-path

Observation 2: Different users may require different models

Personalized Recommendation Model

user-cluster similarity

$$\hat{r}_p(u_i, e_j) = \sum_{k=1}^{c} \boxed{sim(C_k, u_i)} \sum_{q=1}^{L} \theta_q^{\{k\}} \cdot \hat{U}_i^{(q)} \hat{V}_j^{(q)T} \qquad (2)$$

c total soft user clusters

Parameter Estimation

- Bayesian personalized ranking (Rendle UAI'09)
- Objective function

sigmoid function $\sigma(x) = \frac{1}{1+e^{-x}}.$

$$\min_{\Theta} \; -\sum_{u_i \in \mathcal{U}} \boxed{\sum_{(e_a > e_b) \in R_i}} \ln \sigma(\hat{r}(u_i, e_a) - \hat{r}(u_i, e_b)) + \frac{\lambda}{2} \|\Theta\|_2^2 \qquad (3)$$

for each correctly ranked item pair
i.e., u_i gave feedback to e_a but not e_b

Soft cluster users with NMF + k-means → For each user cluster, learn one model with Eq. (3) → Generate personalized model for each user on the fly with Eq. (2)

Learning Personalized Recommendation Model

Experiment Setup

- Datasets

Name	#Items	#Users	#Ratings	#Entities	#Links
IM100K	943	1360	89,626	60,905	146,013
Yelp	11,537	43,873	229,907	285,317	570,634

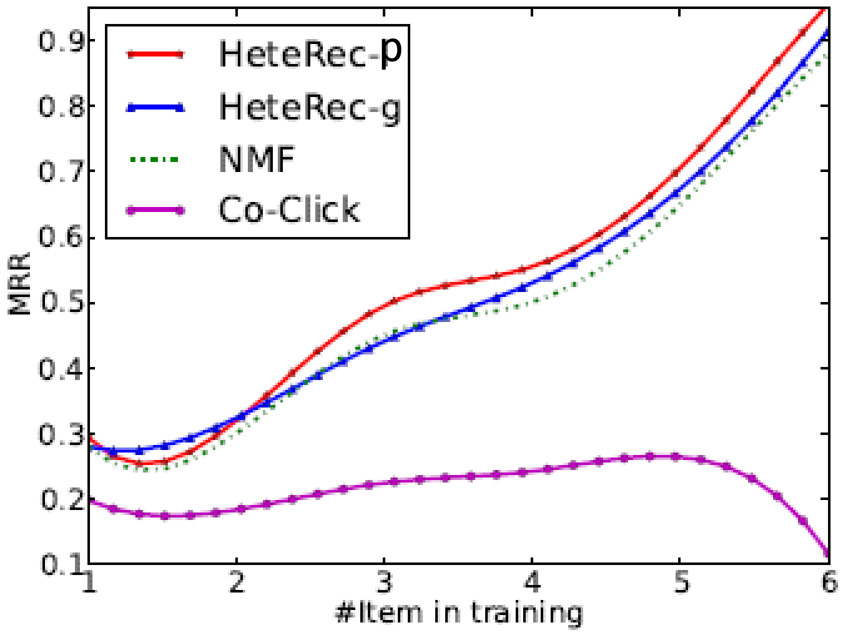
- Comparison methods:
  - Popularity: recommend the most popular items to users
  - Co-click: conditional probabilities between items
  - NMF: non-negative matrix factorization on user feedback
  - Hybrid-SVM: use Rank-SVM with plain features (utilize both user feedback and information network)

Performance Comparison

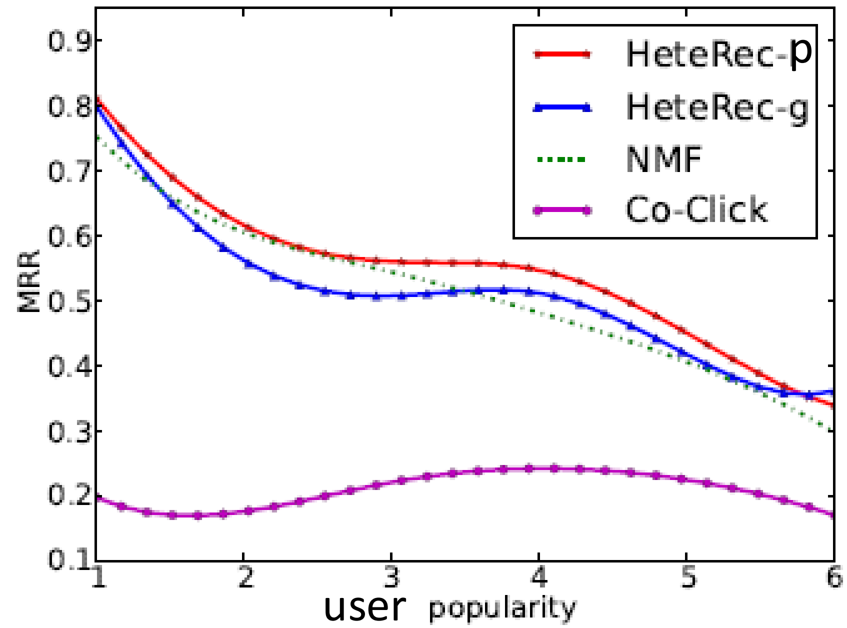
Method	IM100K				Yelp			
	Prec1	Prec5	Prec10	MRR	Prec1	Prec5	Prec10	MRR
Popularity	0.0731	0.0513	0.0489	0.1923	0.00747	0.00825	0.00780	0.0228
Co-Click	0.0668	0.0558	0.0538	0.2041	0.0147	0.0126	0.01132	0.0371
NMF	0.2064	0.1661	0.1491	0.4938	0.0162	0.0131	0.0110	0.0382
Hybrid-SVM	0.2087	0.1441	0.1241	0.4493	0.0122	0.0121	0.0110	0.0337
HeteRec-g	0.2094	0.1791	0.1614	0.5249	0.0165	0.0144	0.0129	0.0422
HeteRec-l	**0.2121**	**0.1932**	**0.1681**	**0.5530**	**0.0213**	**0.0171**	**0.0150**	**0.0513**

HeteRec personalized recommendation (HeteRec-p) provides the best recommendation results

Performance under Different Scenarios



(a) Performance Change with User Feedback Number



(b) Performance Change with User Feedback Popularity

HeteRec–p consistently outperform other methods in different scenarios

- better recommendation results if users provide more feedback
- better recommendation for users who like less popular items

Contributions

- Propose latent representations for users and items by propagating user preferences along different meta-paths
- Employ Bayesian ranking optimization technique to correctly evaluate recommendation models
- Further improve recommendation quality by considering user differences at model level and define personalized recommendation models
    - Two levels of personalization