

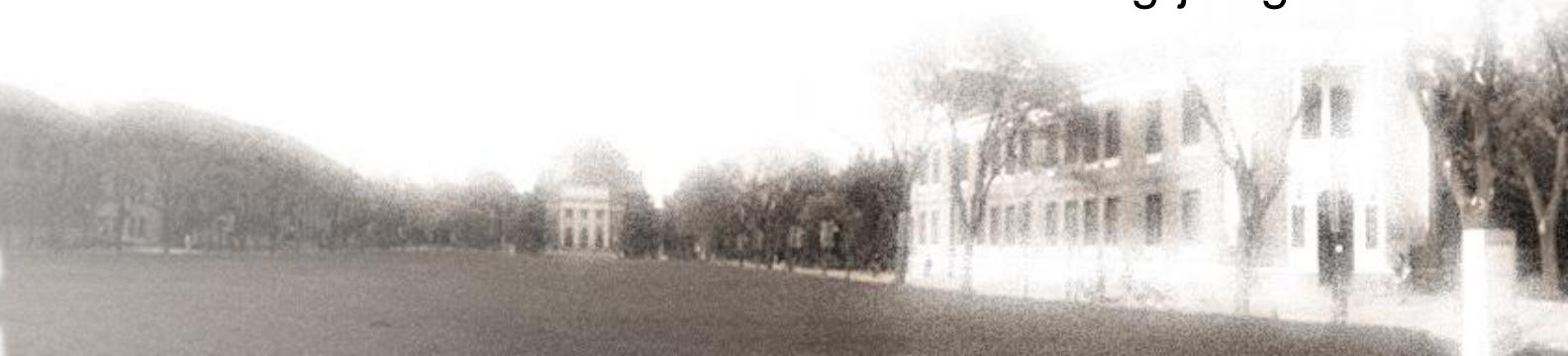


清华大学
Tsinghua University



User Behavior Analysis in Social Networks: Pattern Discovery, Prediction and Anomaly Detection

Meng Jiang 蒋 朦
www.meng-jiang.com





My Group in Tsinghua

Professor Jianping Wu (Network)
Professor Maosong Sun (NLP)

Professor Yuanchun Shi (HCI)
Professor Shimin Hu (Graphics)

Professor Shiqiang Yang (Multimedia)
Professor Jiangtao Wen (Net-of-Things)
Professor Lifeng Sun (Multimedia)

Professor Wenwu Zhu (Multimedia)
Assistant Professor Peng Cui (DM & Multimedia)

Department of Computer
Science and Technology

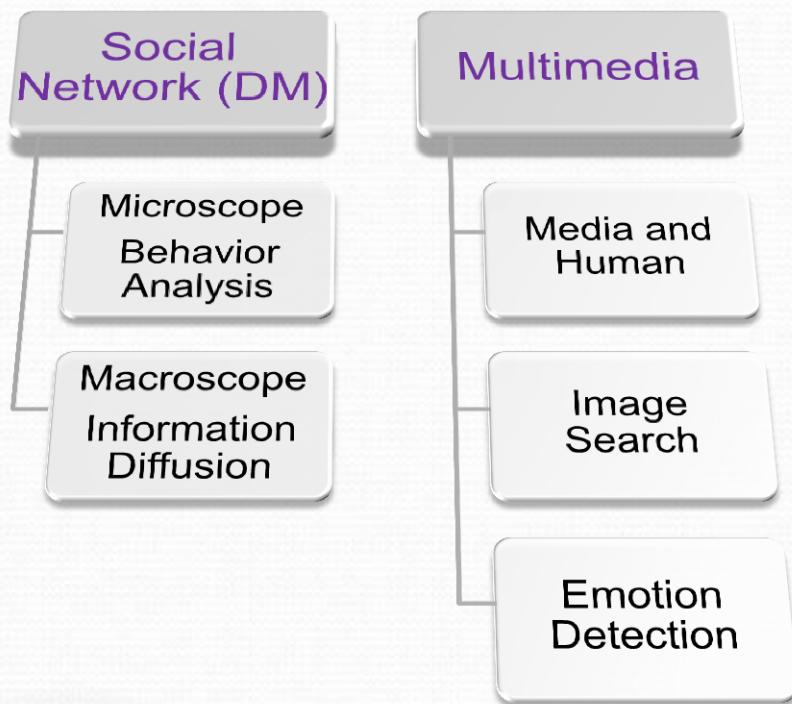
Institute of HCI and
Media Integration

Lab of Multimedia and
Networking

Social Media Lab



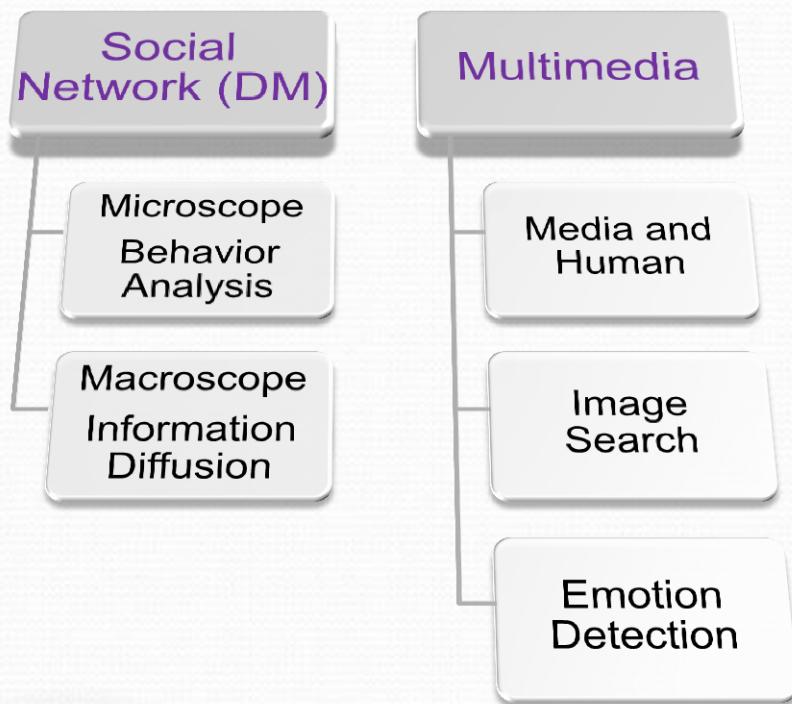
Social-sensed Multimedia Computing



- ❖ Behavior analysis
 - Pattern discovery and prediction [KDD'14]
 - Social recommendation [CIKM'12, CIKM'12, TKDE'14]
 - Anomaly detection [PAKDD'14, WWW'14 Poster, KDD'14]
- ❖ Information diffusion
 - Item-level influence [AAAI'11, SIGIR'11]
 - Outbreak detection [KDD'13]
- ❖ Big social data
 - Heterogeneous hashing [KDD'13]
- ❖ Multimedia computing [TMM'09, DMKD'11, TIP'12, TMM'12, MM'12 Best, TOMCCAP'13, TOMCCAP'14]
- ❖ Image search [MMM'13 Best, TOIS'14, CVIU'14]
- ❖ Emotion detection [MM'13, ICMR'14]



Social-sensed Multimedia Computing



❖ Behavior analysis

- Pattern discovery and prediction [KDD'14]
- Social recommendation [CIKM'12, CIKM'12, TKDE'14]
- Anomaly detection [PAKDD'14, WWW'14 Poster, KDD'14]

❖ Information diffusion

- Item-level influence [AAAI'11, SIGIR'11]
- Outbreak detection [KDD'13]

❖ Big social data

- Heterogeneous hashing [KDD'13]

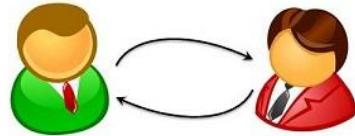
❖ Multimedia computing [TMM'09, DMKD'11, TIP'12, TMM'12, MM'12 Best, TOMCCAP'13, TOMCCAP'14]

❖ Image search [MMM'13 Best, TOIS'14, CVIU'14]

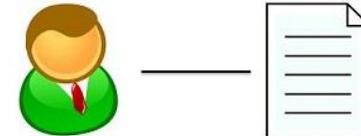
❖ Emotion detection [MM'13, ICMR'14]



My Research Interests



User-user link



User-item link



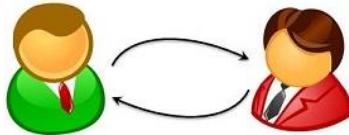
Normal use



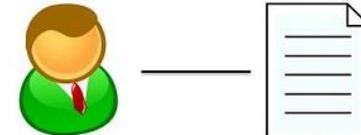
Anomalous use



My Research Interests



User-user link



User-item link



Normal use

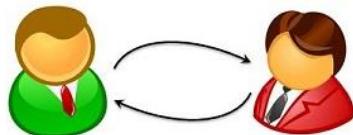
Influence mining
Community detection
Friend recommendation



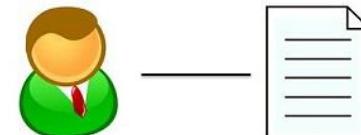
Anomalous use



My Research Interests



User-user link

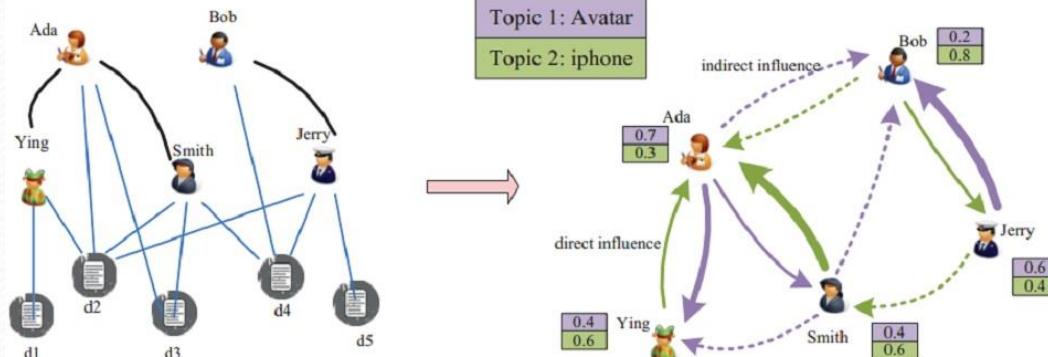


User-item link



Normal use

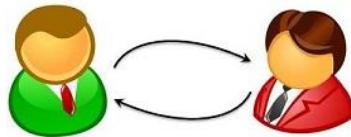
Topic-level influence [Liu et al. CIKM 2010]



Anomalous use



My Research Interests



User-user link



User-item link

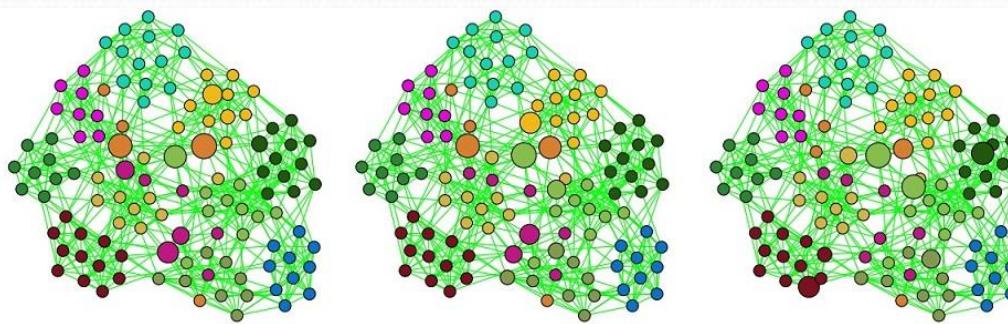


Normal use



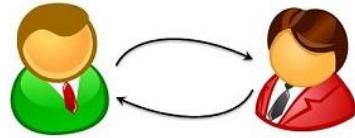
Anomalous use

User diversity [Lu et al. Multimed Tools Appl 2012]

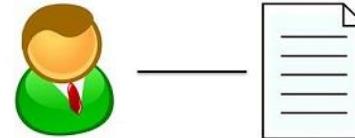




My Research Interests



User-user link



User-item link



Normal use

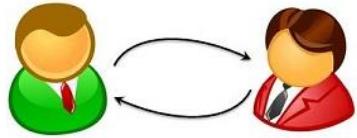
Behavior prediction
Social recommendation
Behavior pattern discovery



Anomalous use



My Research Interests



User-user link



User-item link

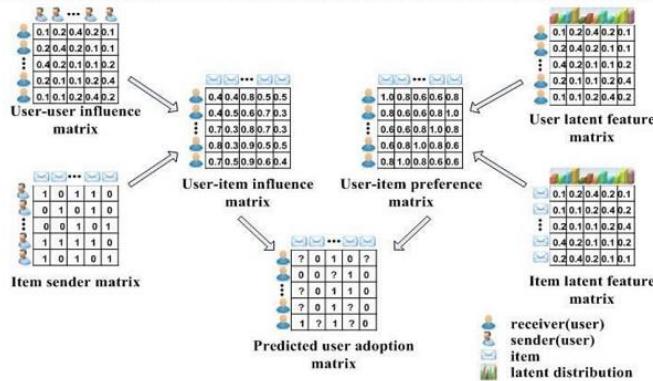


Normal use



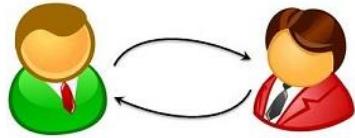
Anomalous use

Social contextual recommendation [Jiang et al. CIKM 2012]

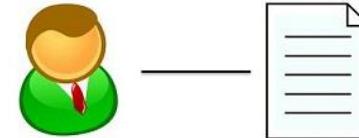




My Research Interests



User-user link



User-item link

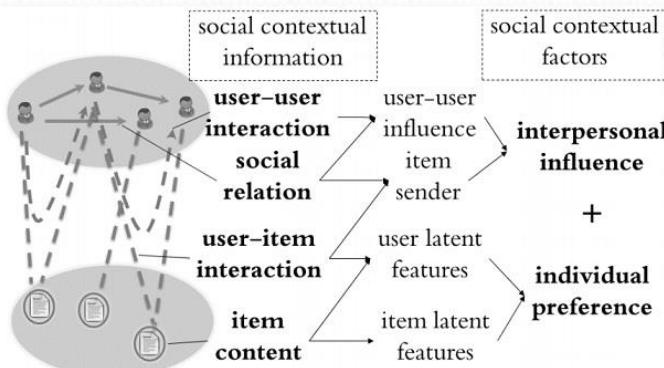


Normal use



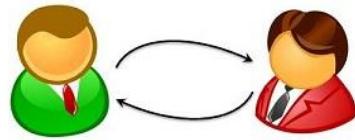
Anomalous use

Scalable recommendation with social context
[Jiang et al. TKDE 2014]





My Research Interests



User-user link



User-item link

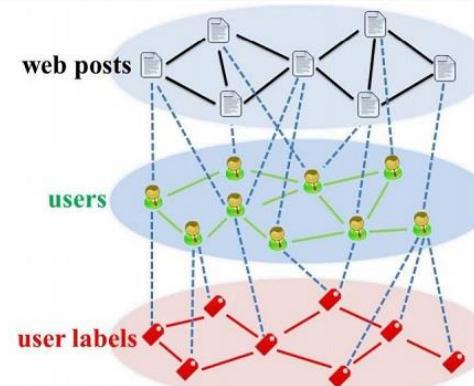
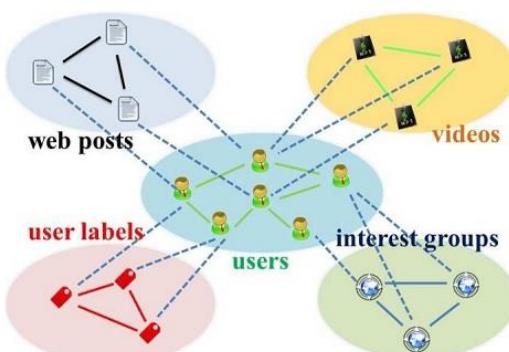


Normal use



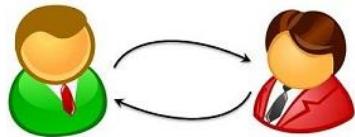
Anomalous use

Cross-domain social recommendation [Jiang et al. CIKM 2012]





My Research Interests



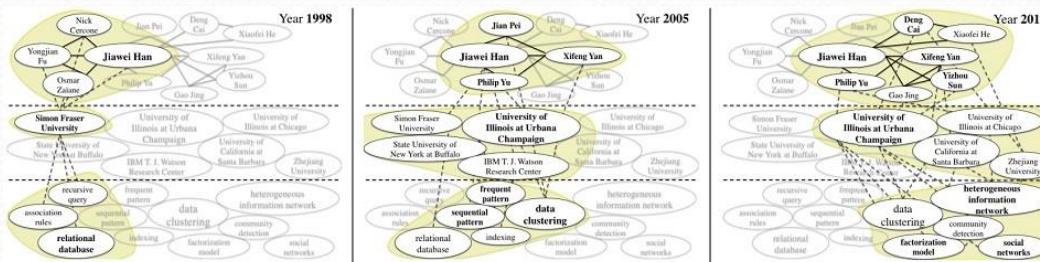
User-user link



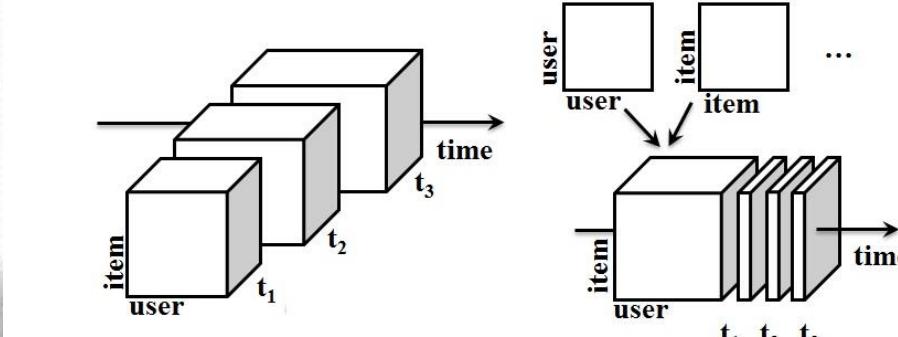
User-item link



Normal use

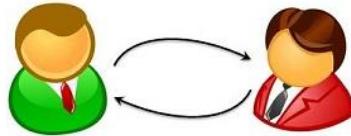


Anomalous use

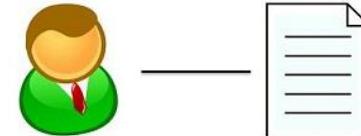




My Research Interests



User-user link



User-item link



Normal use

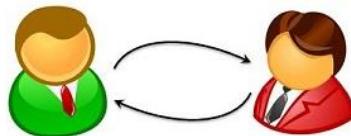


Anomalous use

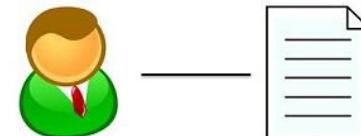
Zombie follower detection
Suspicious behavior patterns



My Research Interests



User-user link



User-item link

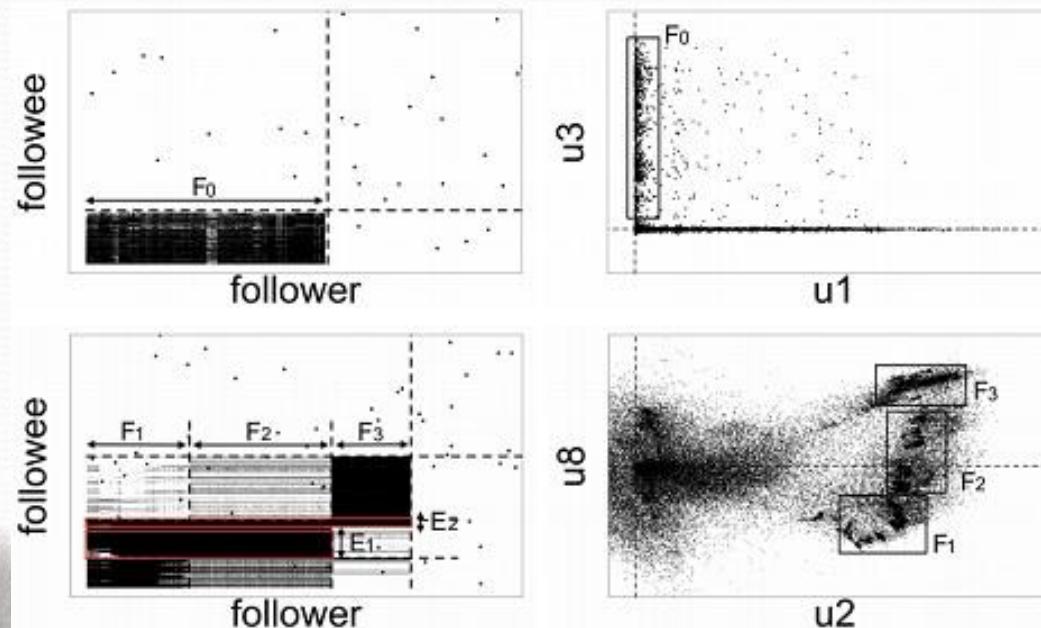


Normal use



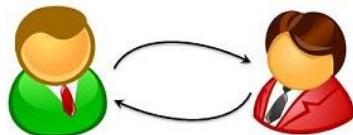
Anomalous use

Dense bipartite core detection [Jiang et al. PAKDD 2014]

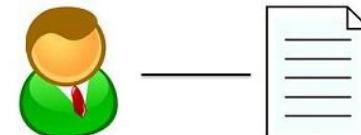




My Research Interests



User-user link



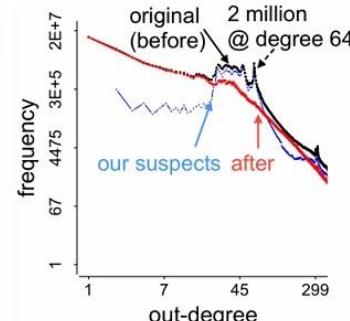
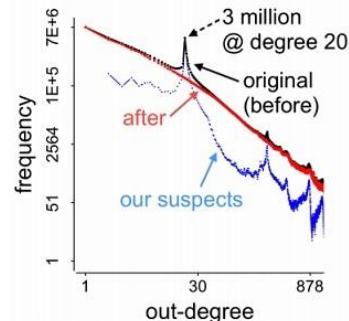
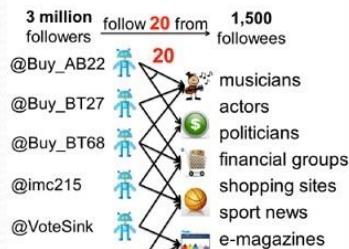
User-item link



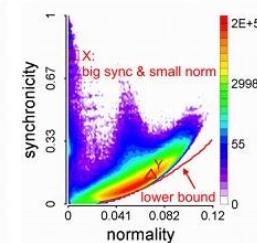
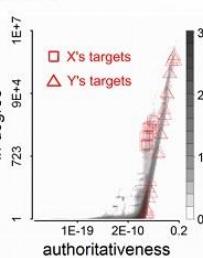
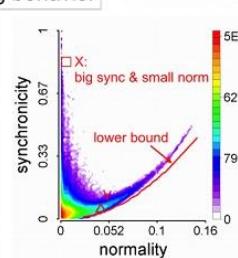
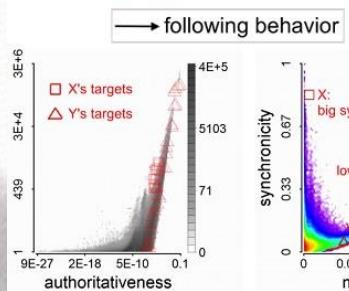
Normal use

Zombie follower detection

[Jiang et al. WWW 2014 Poster, KDD 2014]

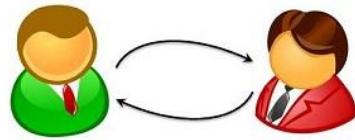


Anomalous use

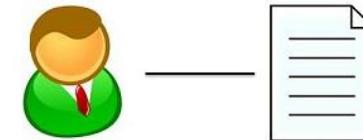




My Research Interests



User-user link



User-item link



Normal use

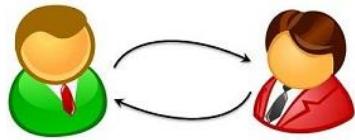


Anomalous use

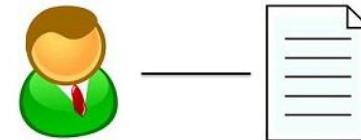
Spam/fraud detection
Promoting behavior patterns
[On-going]



My Research Interests



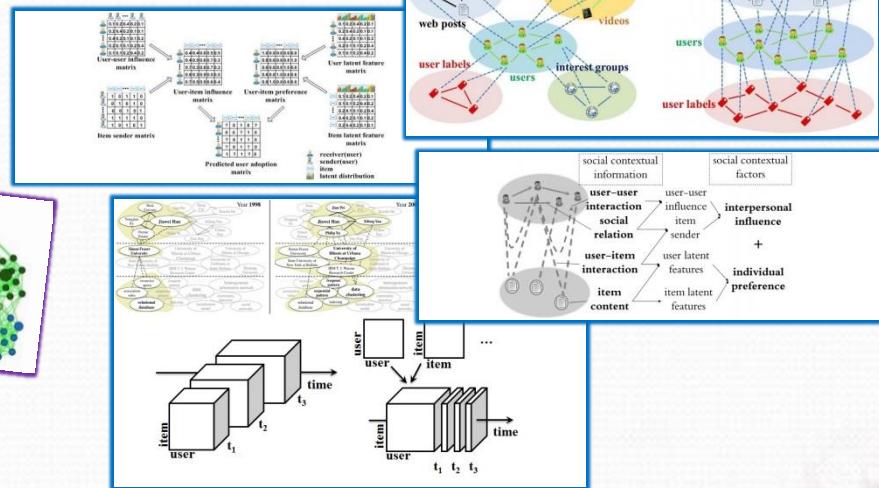
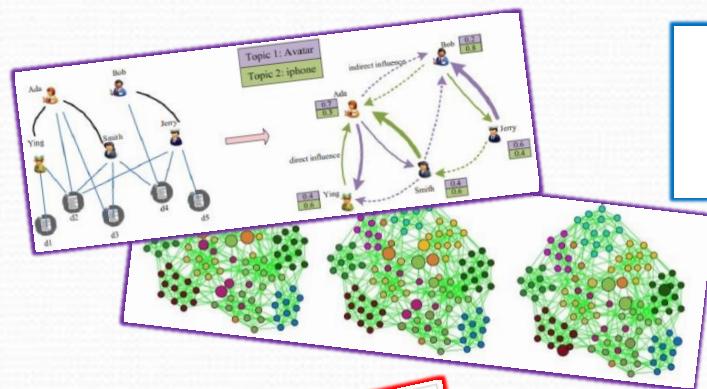
User-user link



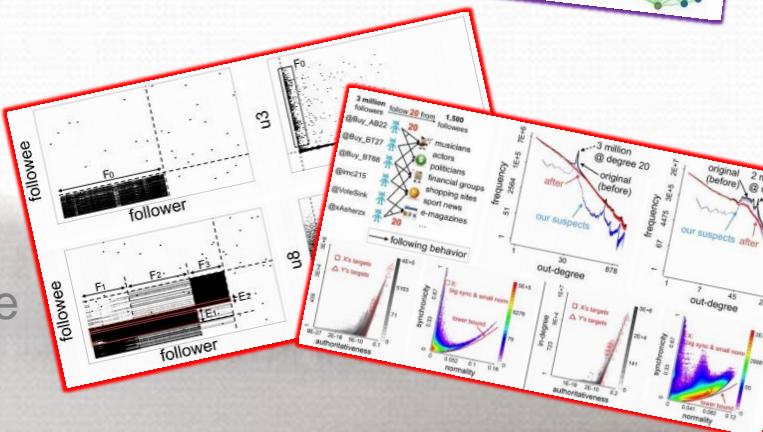
User-item link



Normal use



Anomalous use



Behavior analysis

Pattern discovery and prediction [KDD'14]

Social recommendation [CIKM'12, CIKM'12, TKDE'14]

Anomaly detection [PAKDD'14, WWW'14 Poster, KDD'14]



Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



Bad user-user links



Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



Bad user-user links



Social Recommendation

❖ Problem

- Too many messages are generated and received every minute.
- How to recommend posts/rank feeds in social networks?

❖ Challenges

- **High sparsity** for user-item matrix factorization/completion

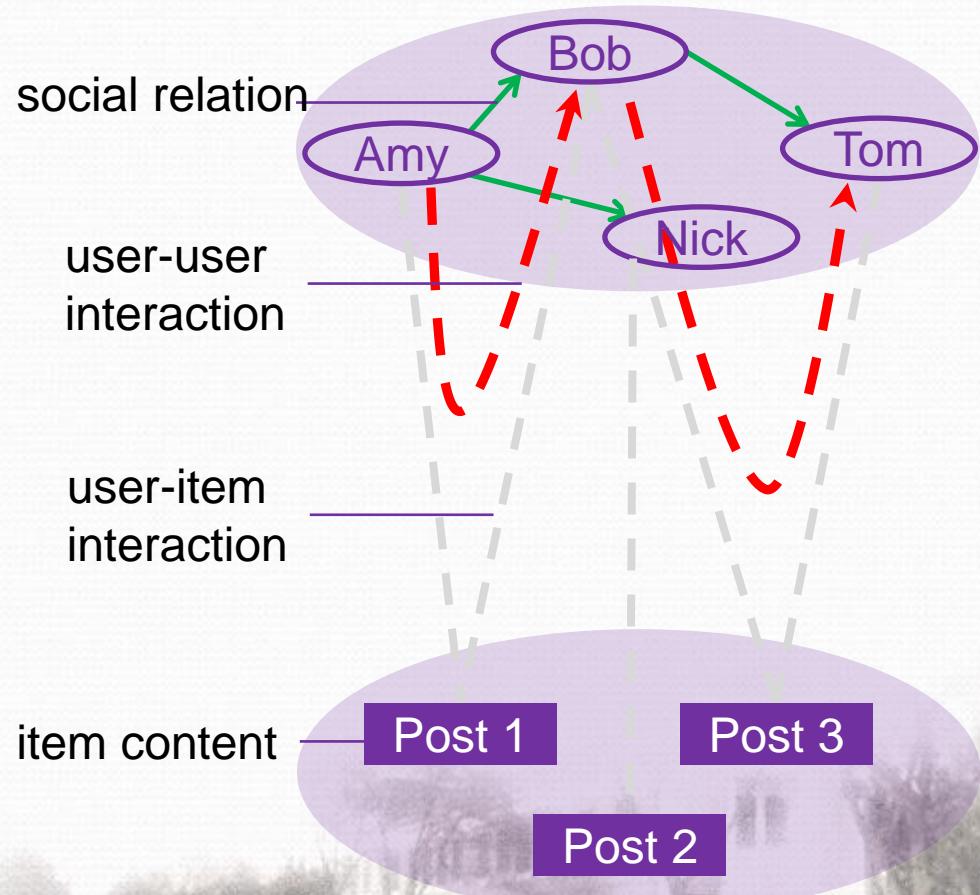


	User 1	User 2	...	User M
Post 1	1	1	...	1
Post 2	0	?	...	?
...
Post N	1	?	...	1



Social Contextual Information

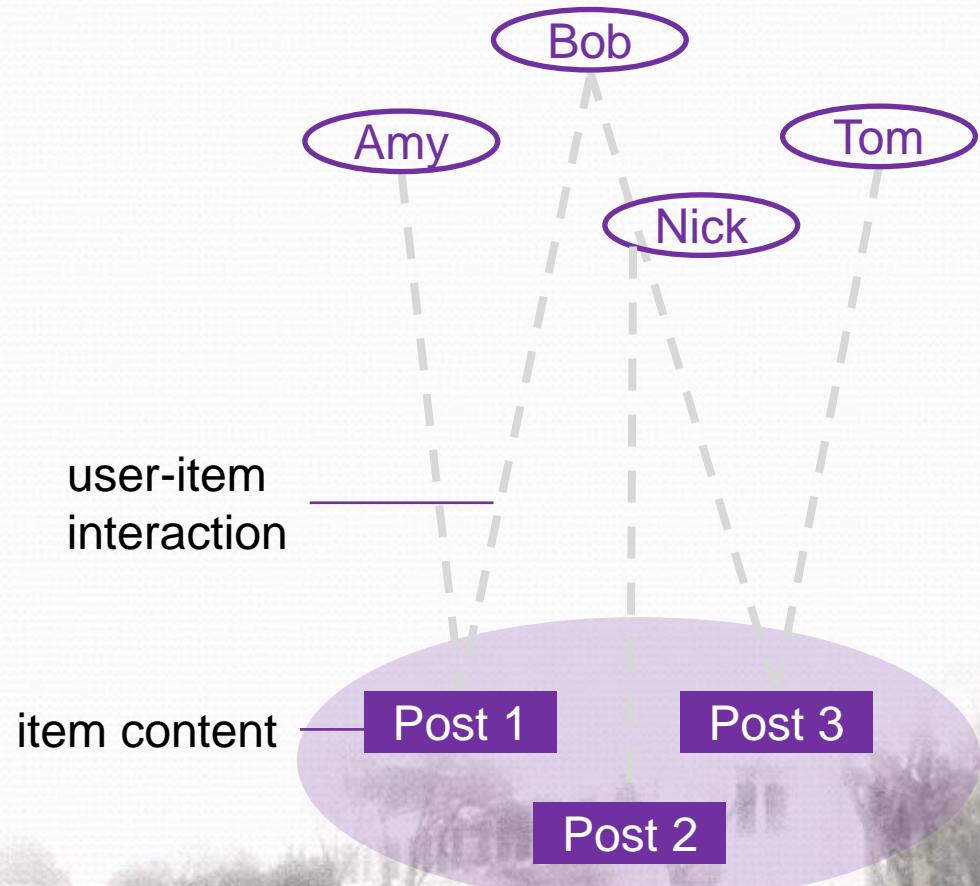
- ❖ More knowledge! But ...
- ❖ Can we use social relations (user-user links) to help infer missing user-item links?
- ❖ Can we fully use social contextual information on the networks?





Related Works

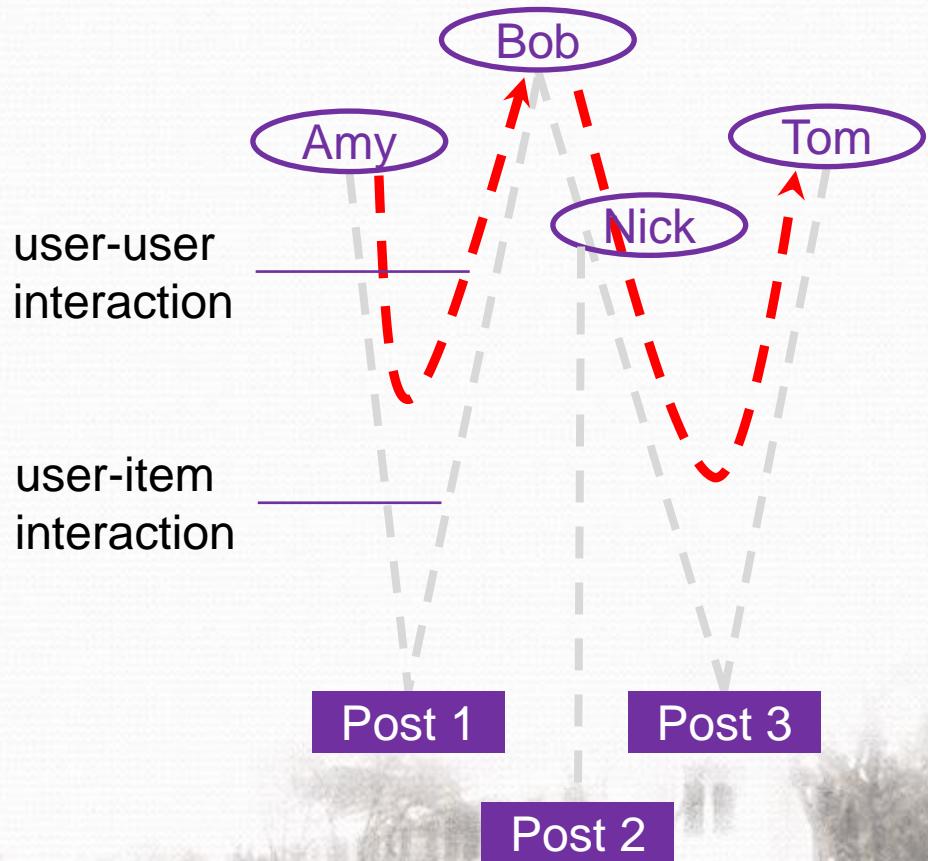
- ❖ Content-based filtering
- ❖ Collaborative filtering





Related Works

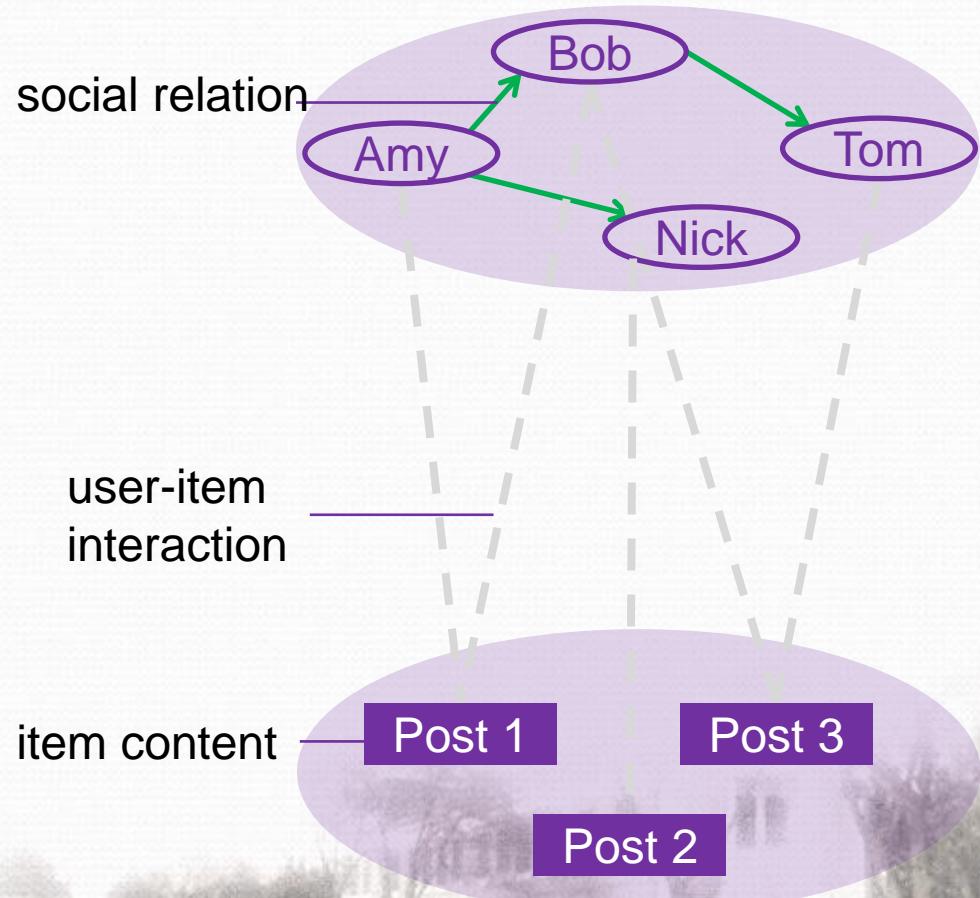
- ❖ Content-based filtering
- ❖ Collaborative filtering
- ❖ Trust-based recommendation
- ❖ Influenced-based recommendation





Related Works

- ❖ Content-based filtering
- ❖ Collaborative filtering
- ❖ Trust-based recommendation
- ❖ Influenced-based recommendation
- ❖ Social recommendation with MF/Social Regularization





Related Works

	Social relation	User-user interaction	User-item interaction	Item content
Content-based & CF	✗	✗	✓	✓
Trust & Influence	✗	✓	✓	✗
SoRec & SoReg	✓	✗	✓	✓
?	✓	✓	✓	✓

- ❖ How to fully use contextual information for social recommendation?
- ❖ Answer: Understand **user intention of adopting messages**.



User Intention of Adopting Messages



Peng Cui : Is there anyone who call for paper via Renren? Hah!
http://media.cs.tsinghua.edu.cn/~multimedia/cuipeng/IR_SI_SocialMedia.htm

2011-01-05 13:47

[Reply](#) | [Share](#)

Call for paper? About social media? Wow!

Personal Preference

This is my best friend and co-author!

Interpersonal influence



Meng Jiang : Support! //Peng Cui : Is there anyone who call for paper via Renren? Hah!
http://media.cs.tsinghua.edu.cn/~multimedia/cuipeng/IR_SI_SocialMedia.htm

2011-01-05 14:05

[Reply](#) | [Share](#)



User Intention of Adopting Messages



Maosong Sun : KDD Summer School on Mining the Big Data will be held in Tsinghua. This is the first time for KDD to hold Summer School. Dean Xiaoyong Du, Dr. Hang Li and me are the Chairs. Today Jiawei Han(UIUC), Christos Faloutsos(CMU) and Bing Liu(UIC) gave lectures for 2 hours each.

2012-08-10 21:47

[Retweet](#) | [Save](#) | [Reply](#)

Amazing! Summer school! It is KDD!
Jiawei Han! Christos Faloutsos! Bin Liu!

Personal Preference

This is the Dean of my Department!

Interpersonal influence

His research area is Artificial Intelligence and Data Mining!



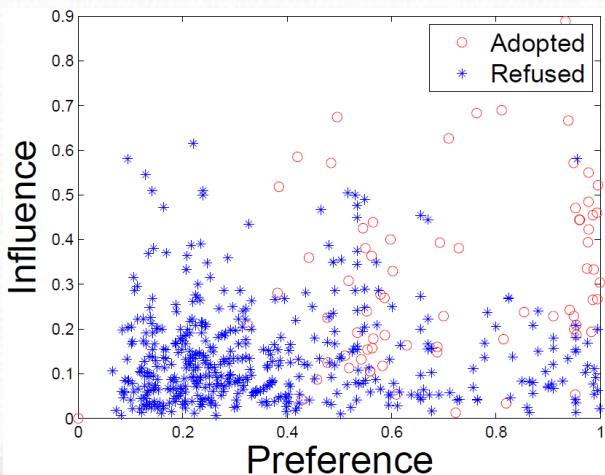
Meng Jiang : Amazing! //**Maosong Sun** : KDD Summer School on Mining the Big Data will be held in Tsinghua. This is the first time for KDD to hold Summer School. Dean Xiaoyong Du, Dr. Hang Li and me are the Chairs. Today Jiawei Han(UIUC), Christos Faloutsos(CMU) and Bing Liu(UIC) gave lectures for 2 hours each.

2012-08-11 09:35

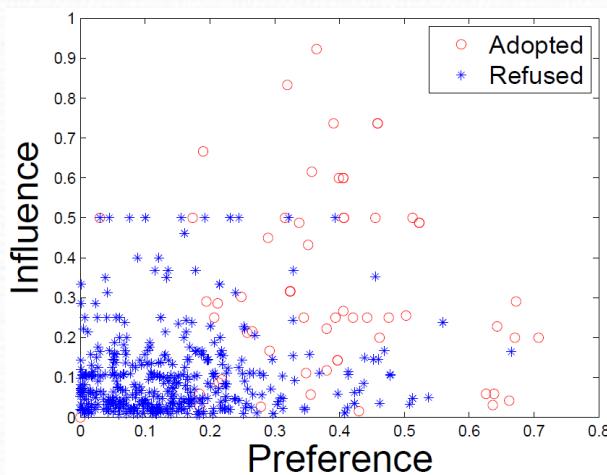
[Retweet](#) | [Save](#) | [Reply](#)

User Intention of Adopting Messages

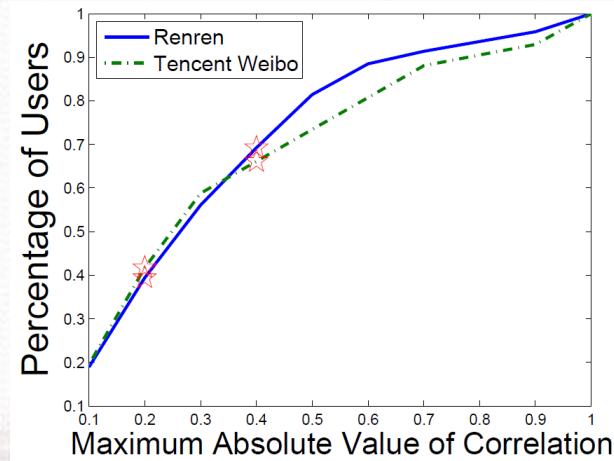
- ❖ Do we care the item content? Do we care the sender?
- ❖ Preference: topic-level user-item similarity
- ❖ Influence: user-sender interaction frequency



Renren

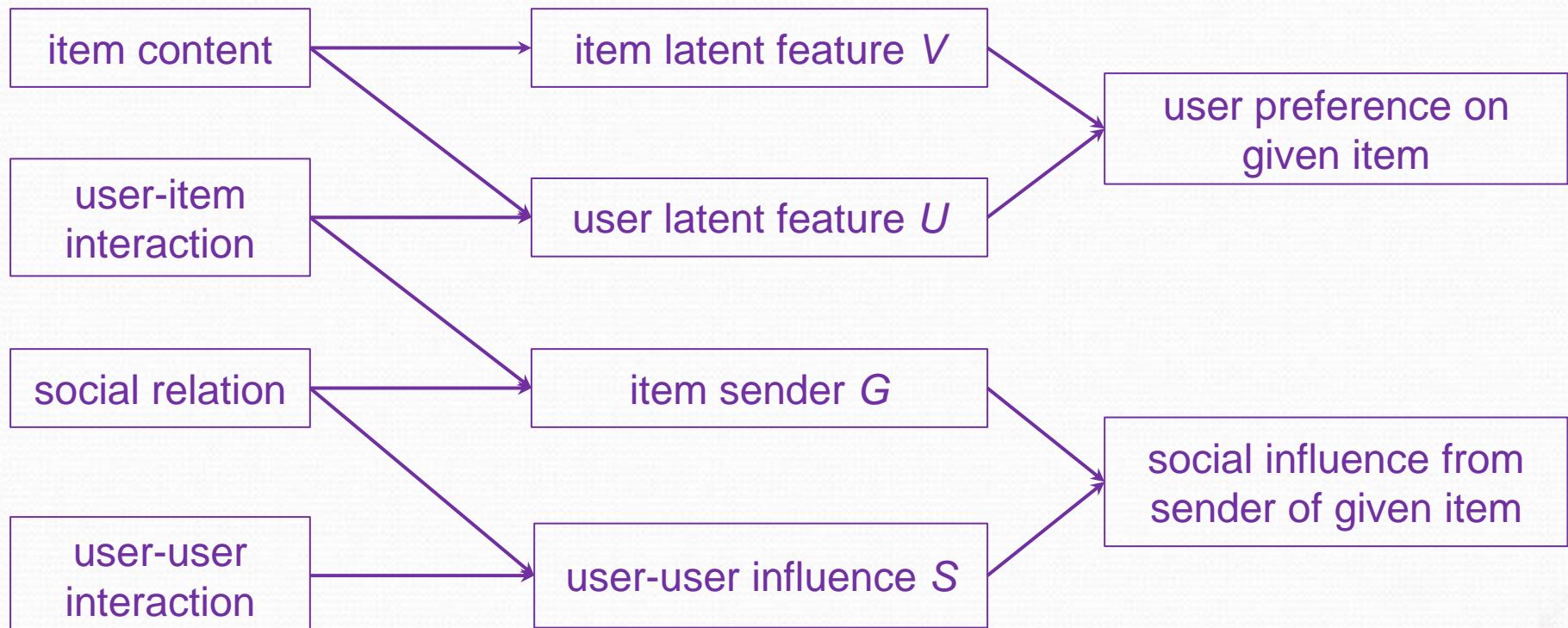


Tencent Weibo



Correlation(Preference, Influence) is small.

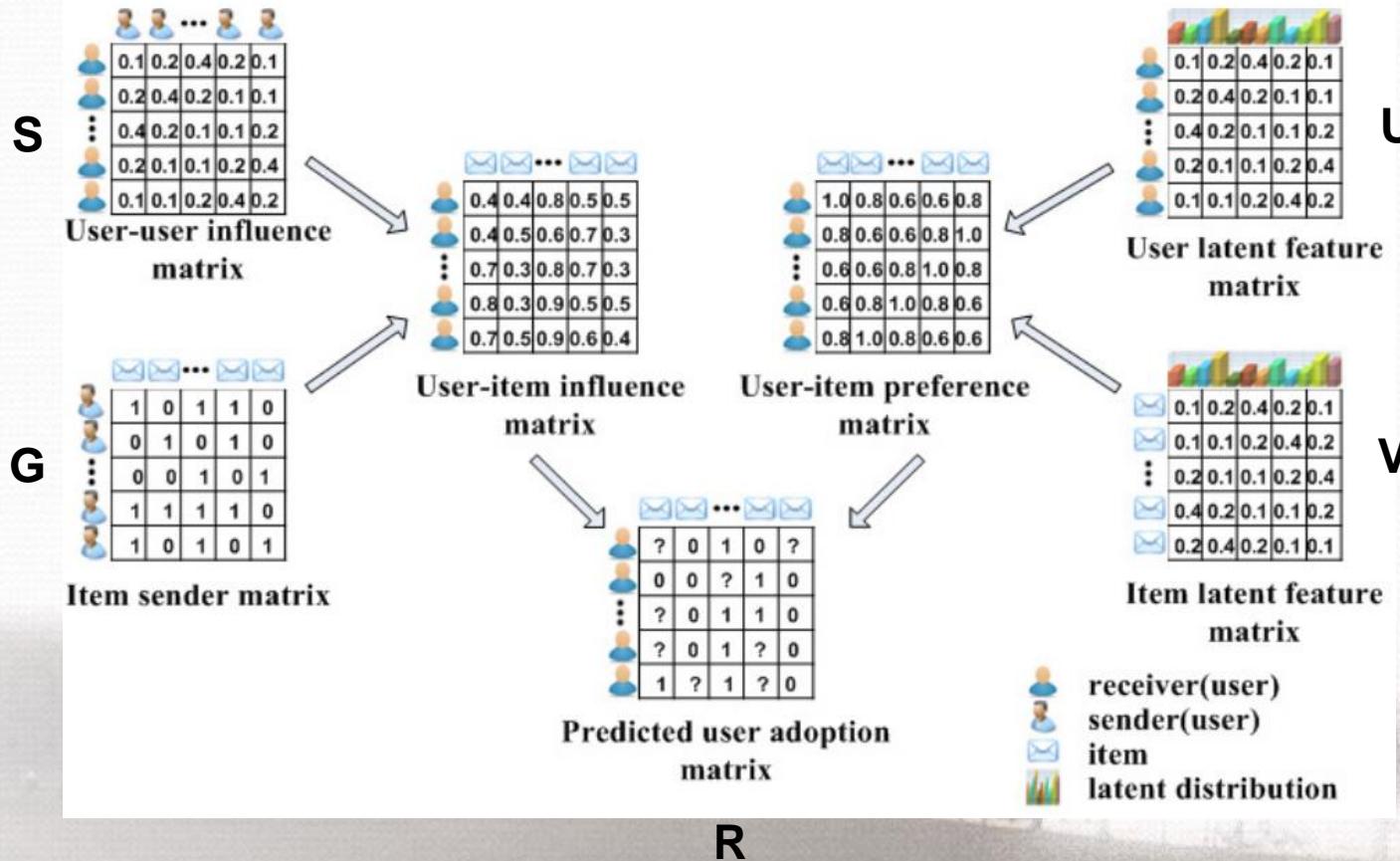
Framework: Social Contextual Factors





Framework: Social Contextual Factors

$$P(\mathbf{R}|\mathbf{S}, \mathbf{U}, \mathbf{V}, \sigma_R^2) = \prod_{i=1}^M \prod_{j=1}^N \mathcal{N}(\mathbf{R}_{ij} | \mathbf{S}_i \mathbf{G}_j^\top \odot \mathbf{U}_i^\top \mathbf{V}_j, \sigma_R^2)$$





ContextMF Algorithm

- ❖ Minimize sum-of-squared errors function

$$\begin{aligned}\mathcal{J} = & \|\mathbf{R} - \mathbf{S}\mathbf{G}^\top \odot \mathbf{U}^\top \mathbf{V}\|_F + \alpha \|\mathbf{W} - \mathbf{U}^\top \mathbf{U}\|_F \\ & + \beta \|\mathbf{C} - \mathbf{V}^\top \mathbf{V}\|_F + \gamma \|\mathbf{S} - \mathbf{F}\|_F \\ & + \delta \|\mathbf{S}\|_F + \eta \|\mathbf{U}\|_F + \lambda \|\mathbf{V}\|_F\end{aligned}$$

- ❖ Block coordinate descent scheme with gradients.

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{S}} = & 2 \left(-\mathbf{R}(\mathbf{G} \odot \mathbf{V}^\top \mathbf{U}) + (\mathbf{S}\mathbf{G}^\top \odot \mathbf{U}^\top \mathbf{V})\mathbf{G} \right. \\ & \left. + \gamma(\mathbf{S} - \mathbf{F}) + \delta\mathbf{S} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{U}} = & 2 \left(-\mathbf{V}\mathbf{R}^\top + \mathbf{V}(\mathbf{G}\mathbf{S}^\top \odot \mathbf{V}^\top \mathbf{U}) - 2\alpha\mathbf{U}\mathbf{W} \right. \\ & \left. + 2\alpha\mathbf{U}\mathbf{U}^\top \mathbf{U} + \eta\mathbf{U} \right)\end{aligned}$$

$$\begin{aligned}\frac{\partial \mathcal{J}}{\partial \mathbf{V}} = & 2 \left(-\mathbf{U}\mathbf{R} + \mathbf{U}(\mathbf{S}\mathbf{G}^\top \odot \mathbf{U}^\top \mathbf{V}) - 2\beta\mathbf{V}\mathbf{C} \right. \\ & \left. + 2\beta\mathbf{V}\mathbf{V}^\top \mathbf{V} + \lambda\mathbf{V} \right)\end{aligned}$$



New Users and New Items Coming

New data = Entries between...	Old N items	New ΔN items
Old M users	Done! $O(k^2 L(M+N)^2)$	$\mathcal{J}_{\Delta V} = \ \Delta C - \Delta V^\top V\ _F^2, \frac{\partial \mathcal{J}}{\partial \Delta V} = -2V\Delta C^\top + O(\Delta V)$ $O(k^2 L \Delta NN) \ll O(k^2 L N(M+N))$
New ΔM users	$\mathcal{J}_{\Delta S} = \ \Delta F - \Delta S\ _F^2, \frac{\partial \mathcal{J}}{\partial \Delta S} = -2\Delta F + O(\Delta S)$ $\mathcal{J}_{\Delta U} = \ \Delta W - \Delta U^\top U\ _F^2, \frac{\partial \mathcal{J}}{\partial \Delta U} = -2U\Delta W^\top + O(\Delta U)$ $O(k^2 L \Delta MM) \ll O(k^2 L M(M+N))$	Sorry...

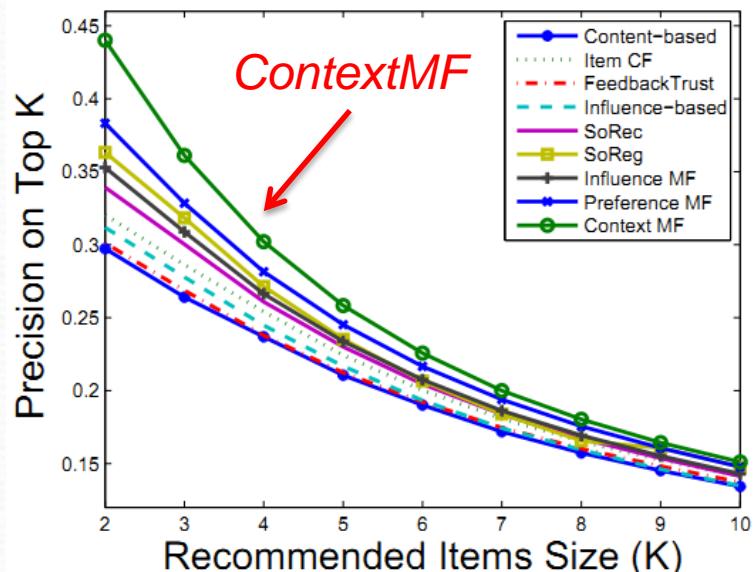
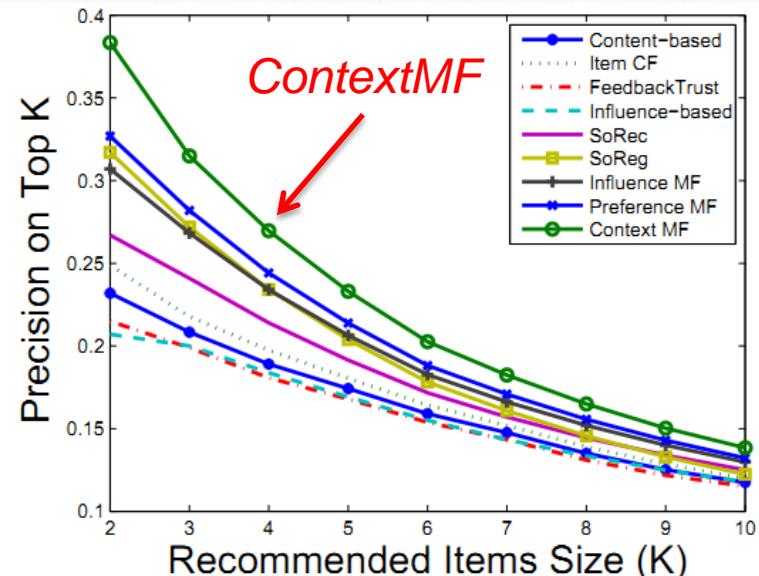


Effectiveness in Predicting Adoptions

Method	MAE	RMSE	$\hat{\tau}$	$\hat{\rho}$
Renren Dataset				
Content-based [1]	0.3842	0.4769	0.5409	0.5404
Item CF [25]	0.3601	0.4513	0.5896	0.5988
FeedbackTrust [22]	0.3764	0.4684	0.5433	0.5469
Influence-based [9]	0.3859	0.4686	0.5394	0.5446
SoRec [19]	0.3276	0.4127	0.6168	0.6204
SoReg [20]	0.2985	0.3537	0.7086	0.7140
Influence MF	0.3102	0.3771	0.6861	0.7006
Preference MF	0.3032	0.3762	0.6937	0.7036
Context MF	0.2416	0.3086	0.7782	0.7896
Tencent Weibo Dataset				
Content-based [1]	0.2576	0.3643	0.7728	0.7777
Item CF [25]	0.2375	0.3372	0.7867	0.8049
FeedbackTrust [22]	0.2830	0.3887	0.7094	0.7115
Influence-based [9]	0.2651	0.3813	0.7163	0.7275
SoRec [19]	0.2256	0.3325	0.7973	0.8064
SoReg [20]	0.1997	0.2962	0.8390	0.8423
Influence MF	0.2183	0.3206	0.8179	0.8258
Preference MF	0.2111	0.3088	0.8384	0.8453
Context MF	0.1514	0.2348	0.8570	0.8685

	Renren	Tencent Weibo
MAE	-19.1%	-24.2%
RMSE	-12.8%	-20.7%
Kendall's	+9.82%	+2.1%
Spearman's	+10.6%	+3.1%

Effectiveness in Predicting Adoptions


Renren

Tencent Weibo

	Renren	Tencent Weibo
Top-5 Precision	+21.7%	+12.3%
Top-10 Precision	+10.8%	+6.85%



Efficiency in Incremental Data

- ❖ Better than SoReg [Ma et al. WSDM 2011]
- ❖ A little bit worse than offline learning (re-training)
- ❖ Save time from hours to minutes when $M, N \sim \text{million}$ and $\Delta M, \Delta N \sim \text{thousand}$

Dataset	RMSE (smaller is better)			ERR (bigger is better)			Time cost	
	<i>SoReg</i>	$\Delta ContextMF$	$ContextMF^\Delta$	<i>SoReg</i>	$\Delta ContextMF$	$ContextMF^\Delta$	$\Delta ContextMF$	$ContextMF^\Delta$
RΔM1000	0.342	0.263	0.257	0.555	0.610	0.636	172s	41.7h
RΔM10000	0.502	0.464	0.444	0.481	0.542	0.559	1610s	41.7h
TΔM1000	0.168	0.122	0.105	0.652	0.764	0.783	54.2s	2.42h
TΔM10000	0.342	0.333	0.317	0.534	0.611	0.651	531s	2.42h
RΔN1000	0.335	0.276	0.276	0.570	0.663	0.680	97.3s	41.7h
RΔN10000	0.546	0.478	0.465	0.514	0.587	0.609	941s	41.7h
TΔN1000	0.218	0.192	0.173	0.726	0.824	0.864	17.8s	2.42h
TΔN10000	0.427	0.376	0.355	0.658	0.720	0.751	160s	2.42h

Meng Jiang, Peng Cui, Rui Liu, Qiang Yang, Fei Wang, Wenwu Zhu and Shiqiang Yang. Social Contextual Recommendation. *The 21st ACM International Conference on Information and Knowledge Management (CIKM)*, 2012.

Meng Jiang, Peng Cui, Fei Wang, Wenwu Zhu and Shiqiang Yang. Scalable Recommendation with Social Contextual Information. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2014.



Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



Bad user-user links



Social Recommendation

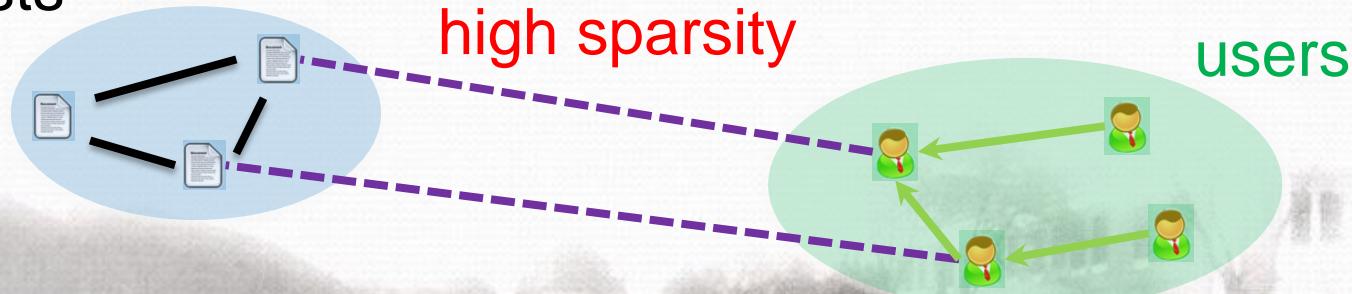
❖ Problem

- Usually we train historical behavioral data (user-post links) to understand their interests. However, if history ~ 0 ?
- How can we recommend posts for new(-registered) users?

❖ Challenges

- **Cold-start and extremely high sparsity** in user-item domain

web posts





Social Recommendation

- ❖ We have social relation domain.
- ❖ We also have auxiliary knowledge in other domains.
- ❖ **User label domain**



Peng Cui
Haidian, Beijing
Company: Tsinghua

Choose < 10 from 200+ labels like ‘iPhone fan’

User labels (5)
Tsinghua, Ph.D., **World Wide Web**,
Social Network, Social Media



Meng Jiang
Haidian, Beijing
University: Tsinghua

User labels (9)
Chinese food, **World Wide Web**,
Social Network, Data Mining, Liverpool
Football Club, NBA, Humors, Sports,
Ph.D. Candidates



Social Recommendation

- ❖ We have social relation domain.
- ❖ We also have auxiliary knowledge in other domains.
- ❖ **Interest group domain**



Interest Groups (2)



Tsinghua
University

I love sing!



Interest Groups (3)



Tsinghua
University

Social
Media &
Reputation

World Wide
Web Team



Social Recommendation

- ❖ How to construct social network with multiple domains?
 - We have user-post, user-label and user-group links.
 - No relations between item domains. No post-label links in nature.
 - Stronger social relations can help collaborate user-item links.
 - More collaborating in user-item links can strengthen the social relations.

web posts



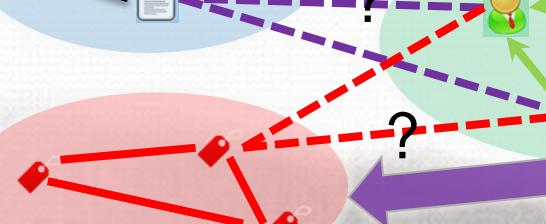
users

web posts



users

user labels



user labels

Social Recommendation

❖ Answer: A star-structured graph.

❖ Properties

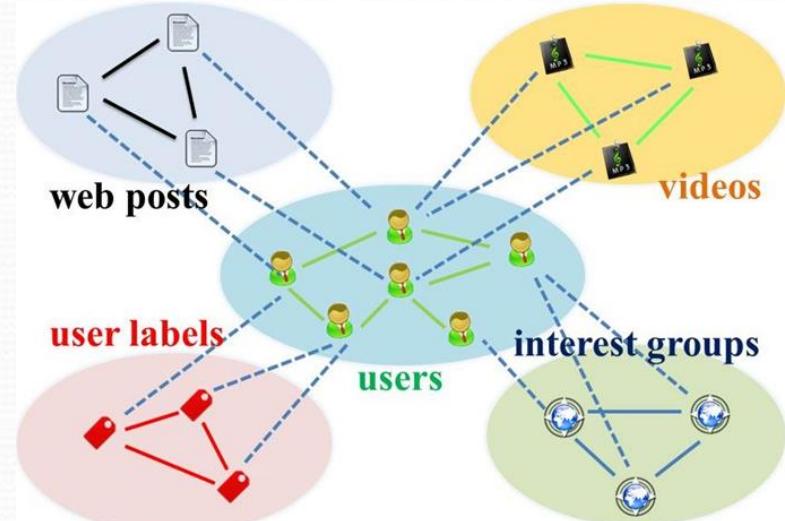
- Relational
 - Within-domain and cross-domain
- Heterogeneous
 - Different types of item domains
- Sparse
 - Different sparsity levels

❖ Key idea

- Use social relation domain as a bridge

❖ Method

- Transfer learning + Random walk with restarts

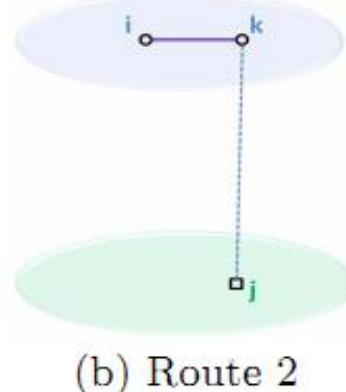
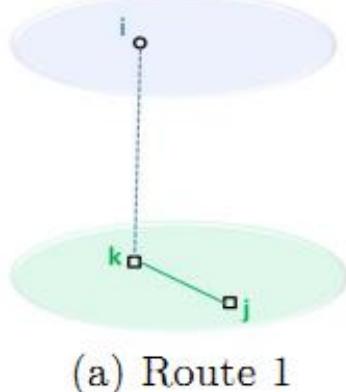




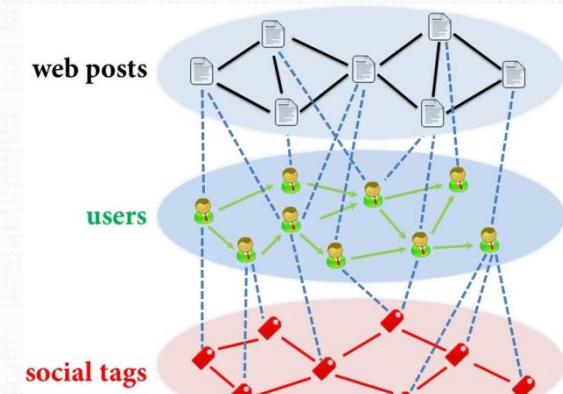
Hybrid Random Walk

❖ On second-order star-structured graph

❖ Update cross-domain links



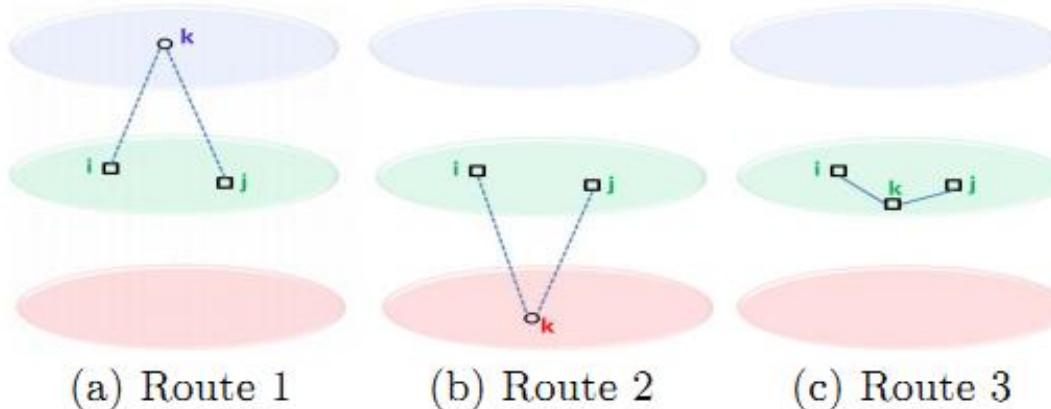
$$p_{ij}^{(\mathcal{U}\mathcal{P})^+} = \delta \sum_{u_k \in \mathcal{U}} r_{ik}^{(\mathcal{U})} p_{kj}^{(\mathcal{U}\mathcal{P})^+} + (1 - \delta) \sum_{p_k \in \mathcal{P}} p_{ik}^{(\mathcal{U}\mathcal{P})^+} r_{kj}^{(\mathcal{P})}$$
$$p_{ij}^{(\mathcal{U}\mathcal{P})^-} = \delta \sum_{u_k \in \mathcal{U}} r_{ik}^{(\mathcal{U})} p_{kj}^{(\mathcal{U}\mathcal{P})^-} + (1 - \delta) \sum_{p_k \in \mathcal{P}} p_{ik}^{(\mathcal{U}\mathcal{P})^-} r_{kj}^{(\mathcal{P})}$$
$$\mathbf{P}^{(\mathcal{U}\mathcal{P})^+}(t+1) = \delta \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{U}\mathcal{P})^+}(t) + (1 - \delta) \mathbf{P}^{(\mathcal{U}\mathcal{P})^+}(t) \mathbf{R}^{(\mathcal{P})}$$
$$\mathbf{P}^{(\mathcal{U}\mathcal{P})^-}(t+1) = \delta \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{U}\mathcal{P})^-}(t) + (1 - \delta) \mathbf{P}^{(\mathcal{U}\mathcal{P})^-}(t) \mathbf{R}^{(\mathcal{P})}$$





Social Recommendation

❖ Update within-domain links



$$r_{ij}^{(\mathcal{U})} = \tau^{(\mathcal{P})} \left(\mu \sum_{p_k \in \mathcal{P}} p_{ik}^{(\mathcal{U}\mathcal{P})+} p_{jk}^{(\mathcal{U}\mathcal{P})+} + (1 - \mu) \sum_{p_k \in \mathcal{P}} p_{ik}^{(\mathcal{U}\mathcal{P})-} p_{jk}^{(\mathcal{U}\mathcal{P})-} \right)$$

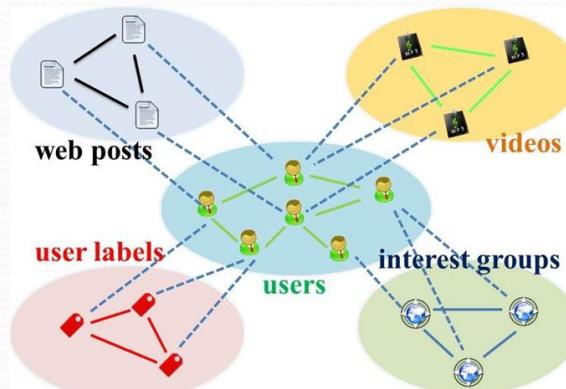
$$+ \tau^{(\mathcal{T})} \sum_{t_k \in \mathcal{T}} p_{ik}^{(\mathcal{U}\mathcal{T})+} p_{jk}^{(\mathcal{U}\mathcal{T})+} + \tau^{(\mathcal{U})} \sum_{u_k \in \mathcal{U}} r_{ik}^{(\mathcal{U})} r_{kj}^{(\mathcal{U})}$$

$$\begin{aligned} \mathbf{R}^{(\mathcal{U})}(t+1) = \\ \tau^{(\mathcal{P})} (\mu \mathbf{P}^{(\mathcal{U}\mathcal{P})+}(t) \mathbf{P}^{(\mathcal{U}\mathcal{P})+}(t)^T + (1 - \mu) \mathbf{P}^{(\mathcal{U}\mathcal{P})-}(t) \mathbf{P}^{(\mathcal{U}\mathcal{P})-}(t)^T) \\ + \tau^{(\mathcal{T})} \mathbf{P}^{(\mathcal{U}\mathcal{T})+}(t) \mathbf{P}^{(\mathcal{U}\mathcal{T})+}(t)^T + \tau^{(\mathcal{U})} \mathbf{R}^{(\mathcal{U})}(t) \mathbf{R}^{(\mathcal{U})}(t)^T \end{aligned}$$



Social Recommendation

- ❖ On high-order star-structured graph



$$\mathbf{P}^{(\mathcal{UD}_i)^+}(t+1) = \delta_i \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{UD}_i)^+}(t) + (1 - \delta_i) \mathbf{P}^{(\mathcal{UD}_i)^+}(t) \mathbf{R}^{(\mathcal{D}_i)}$$

$$\mathbf{P}^{(\mathcal{UD}_i)^-}(t+1) = \delta_i \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{UD}_i)^-}(t) + (1 - \delta_i) \mathbf{P}^{(\mathcal{UD}_i)^-}(t) \mathbf{R}^{(\mathcal{D}_i)}$$

$$\begin{aligned} \mathbf{R}^{(\mathcal{U})}(t+1) &= \sum_{\mathcal{D}_i \in \mathcal{D}} \tau_i \mu_i \mathbf{P}^{(\mathcal{UD}_i)^+}(t) \mathbf{P}^{(\mathcal{UD}_i)^+}(t)^T \\ &\quad + \sum_{\mathcal{D}_i \in \mathcal{D}} \tau_i (1 - \mu_i) \mathbf{P}^{(\mathcal{UD}_i)^-}(t) \mathbf{P}^{(\mathcal{UD}_i)^-}(t)^T \\ &\quad + \tau^{(\mathcal{U})} \mathbf{R}^{(\mathcal{U})}(t) \mathbf{R}^{(\mathcal{U})}(t)^T \end{aligned}$$



Data Set

- ❖ Tencent Weibo (January 2011)

Domain	Size	Cross-domain links	
		Accept	Refuse
User	53.4K	—	—
Web post	142K	1.47M (0.02%)	3.40M (0.04%)
User label	111	330K (5.57%)	—



Good to Transfer?

❖ Comparative Algorithms (RWR)

- $\mathbf{W}^{(P)}$: Use web post similarity?
- $\mathbf{W}^{(U)}$: Use social relation?
- $\mathbf{R}^{(U)}$: Update tie strength?
- $\mathbf{W}^{(T)}$: Use user label similarity?

Algorithm	$\mathbf{R}^{(U)}$	$\mathbf{W}^{(U)}$	$\mathbf{W}^{(P)}$	$\mathbf{W}^{(T)}$
HRW	✓	✓	✓	✓
BRW- R_U -P (TrustWalker)	✓	✓	✓	✗
BRW- R_U	✓	✓	✗	✗
BRW- W_U -P	✗	✓	✓	✗
BRW- W_U (ItemRank)	✗	✓	✗	✗
BRW-P	✗	✗	✓	✗



Good to Transfer?

❖ Compare with RWR models

Algorithm	MAE	Precision	Recall	F1	Kendall's $\hat{\tau}$
HRW	0.227±1.5e-3	0.711±1.3e-3	0.921±1.4e-3	0.802±1.1e-3	0.792±2.5e-3
BRW- R_U -P (TrustWalker)	0.276±1.1e-3	0.657±7.6e-4	0.935±9.8e-4	0.772±7.6e-4	0.774±1.6e-3
BRW- R_U	0.282±5.3e-3	0.655±4.0e-3	0.921±1.2e-2	0.765±7.7e-3	0.725±2.8e-3
BRW- W_U -P	0.292±1.1e-3	0.666±7.0e-4	0.900±5.2e-4	0.765±6.6e-4	0.725±8.5e-4
BRW- W_U (ItemRank)	0.318±1.4e-3	0.671±1.5e-3	0.713±2.4e-3	0.691±1.2e-3	0.661±2.2e-3
BRW-P	0.438±2.6e-4	0.571±3.4e-4	0.499±4.2e-4	0.532±3.2e-4	0.606±2.3e-4

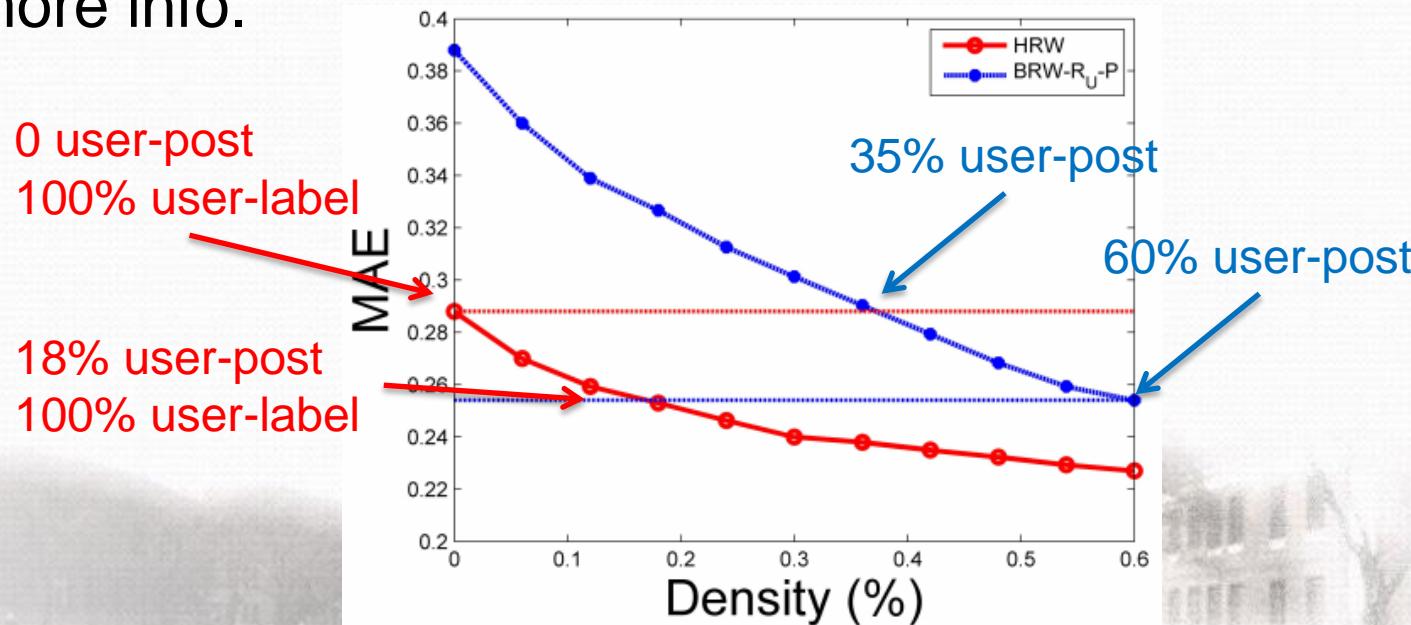
❖ Compare with Baselines

Algorithm	MAE	Precision	Recall	F1	Kendall's $\hat{\tau}$
HRW	0.227±1.5e-3	0.711±1.3e-3	0.921±1.4e-3	0.802±1.1e-3	0.792±2.5e-3
BRW- R_U -P (TrustWalker) [10]	0.276±1.1e-3	0.657±7.6e-4	0.935±9.8e-4	0.772±7.6e-4	0.774±1.6e-3
BRW- W_U (ItemRank) [8]	0.318±1.4e-3	0.671±1.5e-3	0.713±2.4e-3	0.691±1.2e-3	0.661±2.2e-3
MCF [5]	0.352±2.3e-4	0.592±1.8e-3	0.951±6.0e-4	0.730±1.3e-3	0.582±4.3e-4
CF [22]	0.506±3.4e-4	0.552±1.5e-3	0.589±7.2e-4	0.570±1.0e-3	0.540±5.2e-4



Good to Transfer? More Insights!

- ❖ If we do transfer (from user-label domain), we need only ~30% to reach the same performance.
- ❖ Advice: build more features to let new users provide more info.





Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



Bad user-user links



Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- **Multi-faceted**

Write a paper

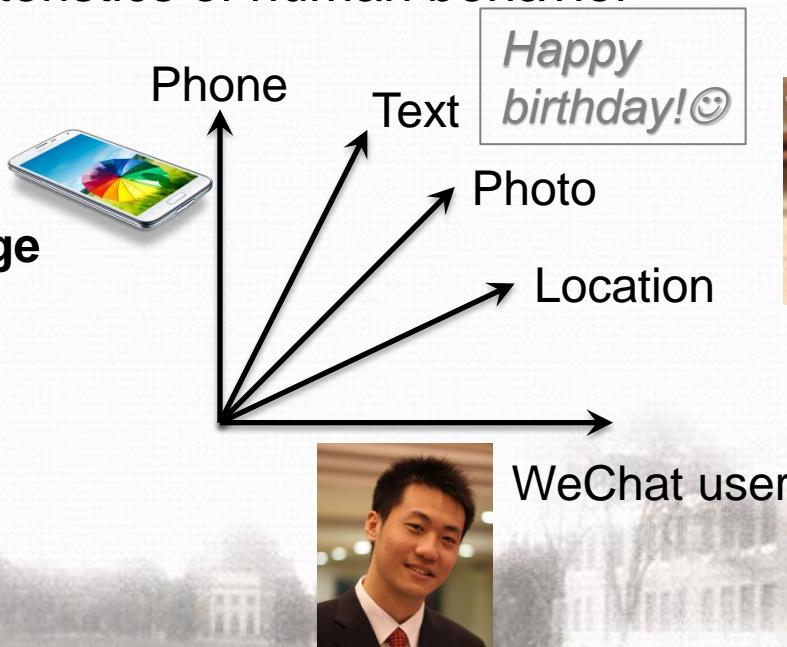


Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- **Multi-faceted**

Post a WeChat message

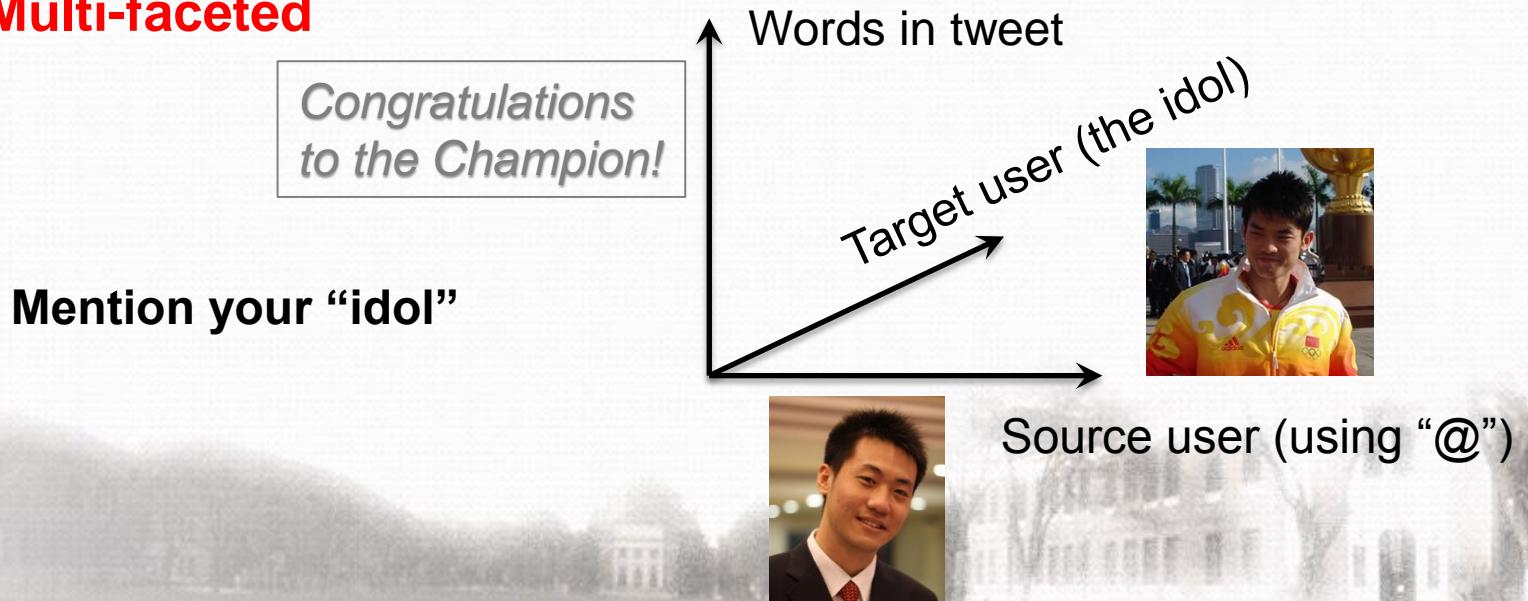




Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- **Multi-faceted**

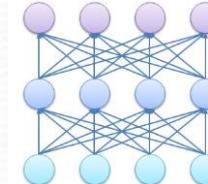
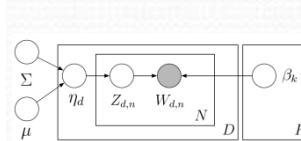
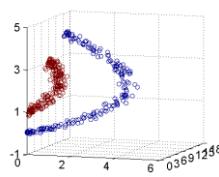




Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- Multi-faceted
- **Evolutionary**



time

Write a paper

Stanford
University



Baidu 百度



Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- Multi-faceted
- **Evolutionary**

Post a WeChat message

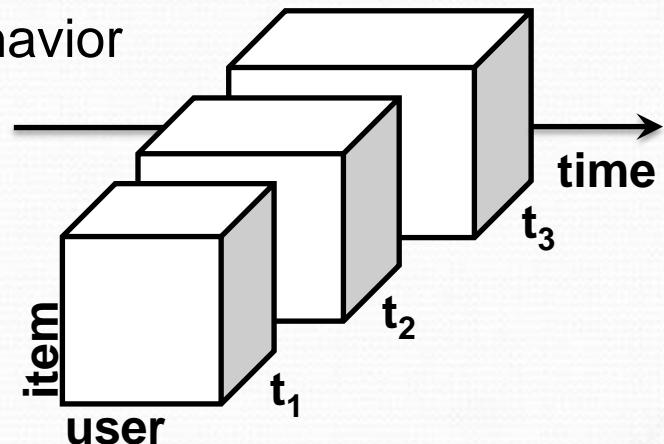




Human Behavior Pattern

❖ Problem

- Human behavior pattern discovery and prediction: significant for industry, national security and life experience.
- Two basic characteristics of human behavior
- **Multi-faceted**
- **Evolutionary**
- How to model human behavior?
- **Tensor sequence.**
- How to do pattern discovery and prediction?
- **Tensor decomposition and completion.**



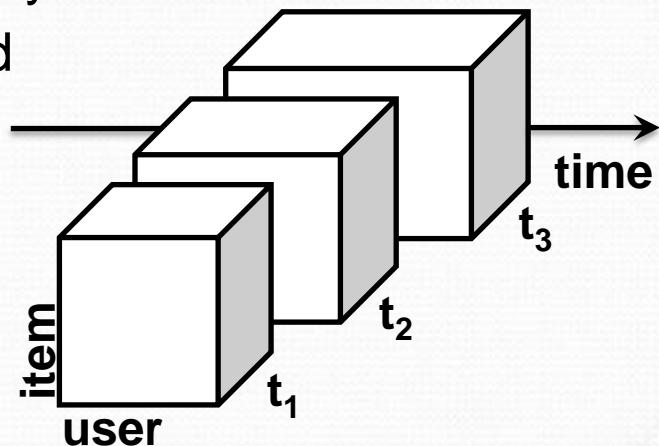


Human Behavior Pattern

❖ Challenges in high-order tensor decomposition

- **High sparsity**
- Write a paper: #author * #affiliation * #keyword
- “@idol”: #(src) user * #(tgt) user * #word

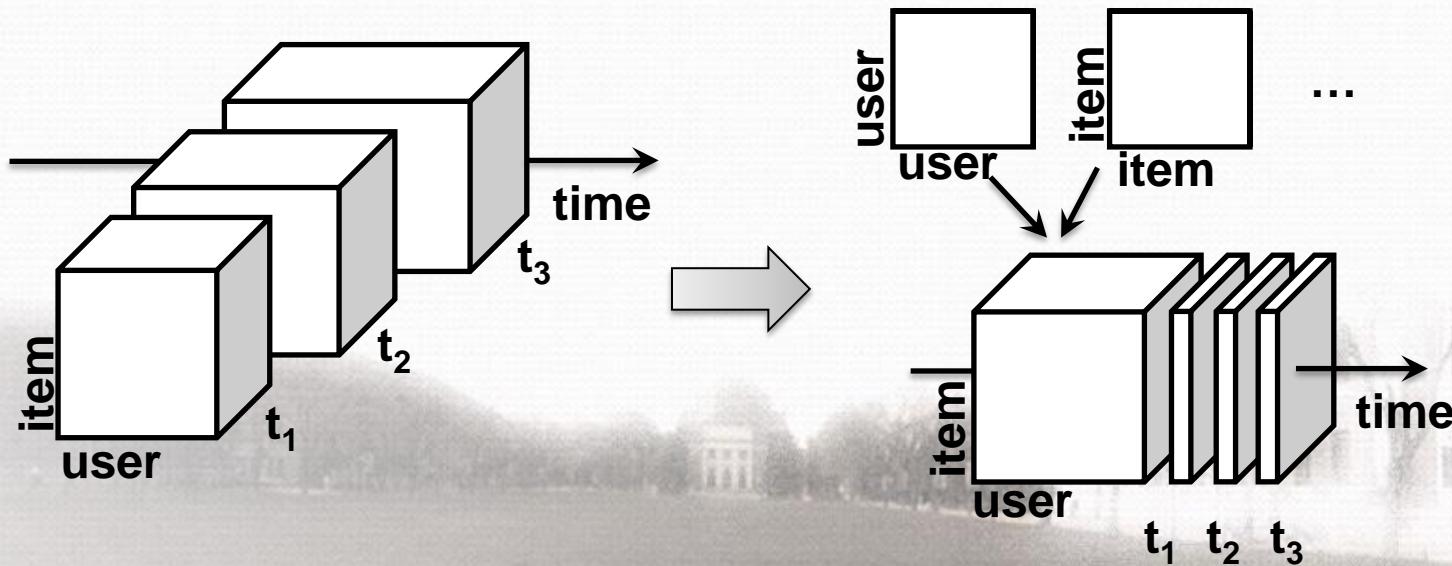
- **High complexity**
- Long sequence of large tensors
- Slow: decomposition at each time





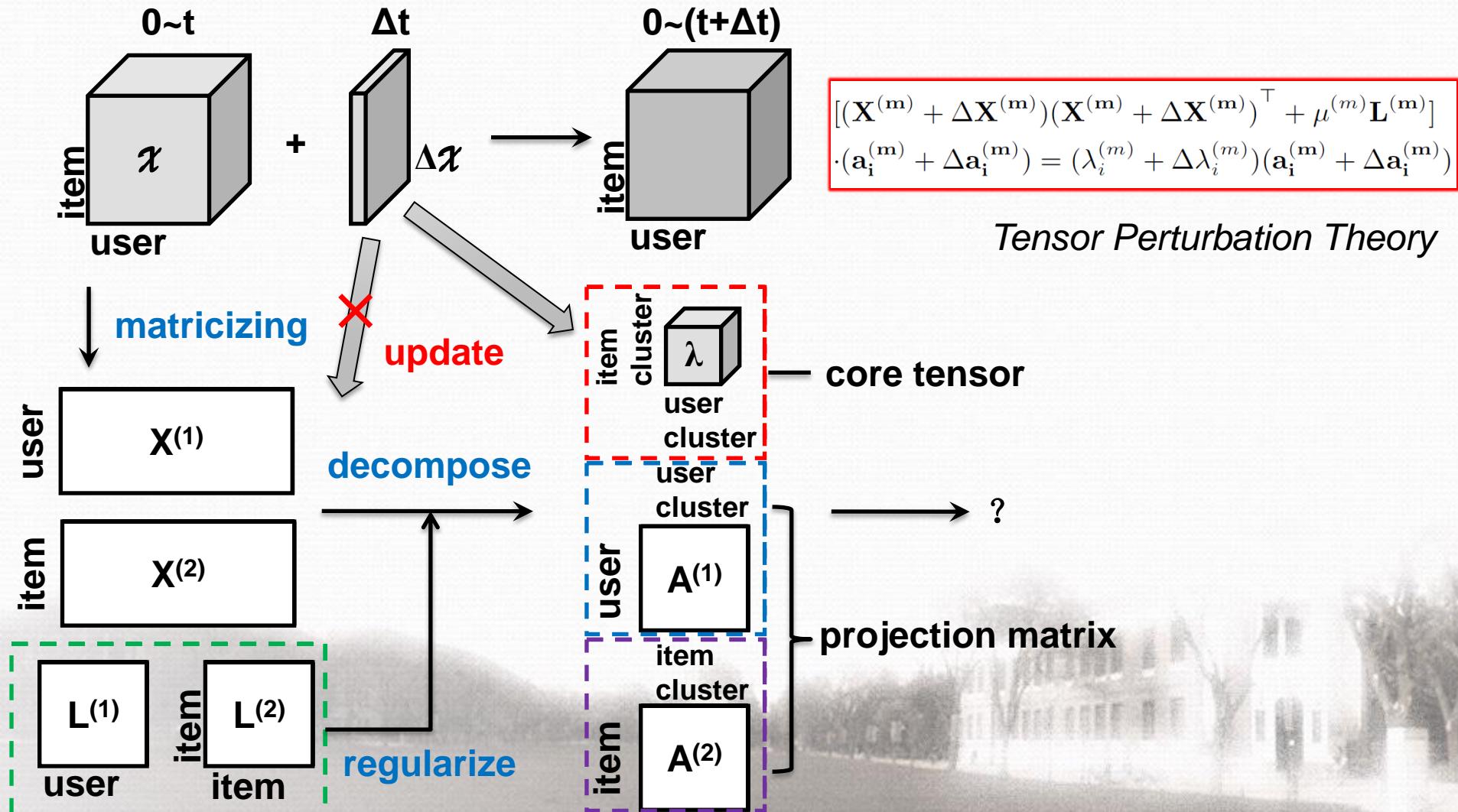
Human Behavior Pattern

- ❖ High sparsity: auxiliary knowledge as regularizers
 - Author - affiliation - keyword + co-authorship (author-author)
 - Src user - tgt user - word + social (user-user)
- ❖ High complexity: update decomposition results (projection matrix) with new coming piece of data





FEMA: Flexible Evolutionary Multi-faceted Analysis





FEMA Algorithm

Approximation

Require: $\mathcal{X}_t, \Delta\mathcal{X}_t, \mathbf{A}_t^{(m)}|_{m=1}^M, \lambda_t^{(m)}|_{m=1}^M$

for $m = 1, \dots, M$ **do**

for $i = 1, \dots, r^{(m)}$ **do**

 Compute $\Delta\lambda_{t,i}^{(m)}$ using

$$\Delta\lambda_i^{(m)} = \mathbf{a}_i^{(m)\top} (\mathbf{X}^{(m)} \Delta\mathbf{X}^{(m)\top} + \Delta\mathbf{X}^{(m)} \mathbf{X}^{(m)\top}) \mathbf{a}_i^{(m)}$$

 and compute

$$\lambda_{t+1,i}^{(m)} = \lambda_{t,i}^{(m)} + \Delta\lambda_{t,i}^{(m)};$$

 Compute $\Delta\mathbf{a}_{t,i}^{(m)}$ using

$$\Delta\mathbf{a}_i^{(m)} = \sum_{j \neq i} \frac{\mathbf{a}_j^{(m)\top} (\mathbf{X}^{(m)} \Delta\mathbf{X}^{(m)\top} + \Delta\mathbf{X}^{(m)} \mathbf{X}^{(m)\top}) \mathbf{a}_i^{(m)}}{\lambda_i^{(m)} - \lambda_j^{(m)}} \mathbf{a}_j^{(m)}$$

 and compute

$$\mathbf{a}_{t+1,i}^{(m)} = \mathbf{a}_{t,i}^{(m)} + \Delta\mathbf{a}_{t,i}^{(m)} \text{ and } \mathbf{A}_{t+1}^{(m)} = \{\mathbf{a}_{t+1,i}^{(m)}\};$$

end for

end for

$$\mathcal{Y}_{t+1} = (\mathcal{X}_t + \Delta\mathcal{X}_t) \prod_{m=1}^M \times_{(m)} \mathbf{A}_{t+1}^{(m)\top};$$

return $\mathbf{A}_{t+1}^{(m)}|_{m=1}^M, \lambda_{t+1}^{(m)}|_{m=1}^M, \mathcal{Y}_{t+1}$

Bound Guarantee

$$|\Delta\lambda_i^{(m)}| \leq 2(\lambda_{\mathbf{X}^{(m)\top} \mathbf{X}^{(m)}}^{\max})^{\frac{1}{2}} \|\Delta\mathbf{X}^{(m)}\|_2$$

$$|\Delta\mathbf{a}_i^{(m)}| \leq 2\|\Delta\mathbf{X}^{(m)}\|_2 \sum_{j \neq i} \frac{(\lambda_{\mathbf{X}^{(m)\top} \mathbf{X}^{(m)}}^{\max})^{\frac{1}{2}}}{|\lambda_i^{(m)} - \lambda_j^{(m)}|}$$





Leveraging Multi-faceted Information

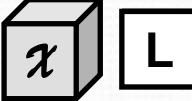
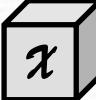
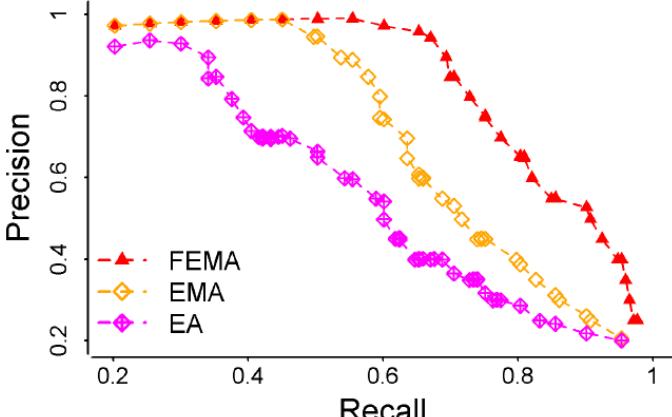
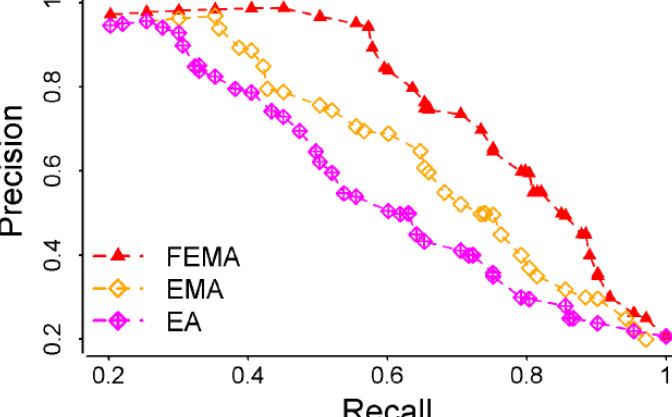
“Who”-“What keyword”?
Use “Where” (affiliation)?

“Who”-“@Whom”?
Use “What” (tweet word)?

	MAS (Data mining subset)		Tencent Weibo (January, 2011)	
	MAE	RMSE	MAE	RMSE
FEMA	0.735	0.944	0.894	1.312
EMA	0.794	1.130	0.932	1.556
EA	0.979	1.364	1.120	1.873
Cross Validation				



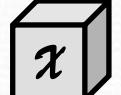
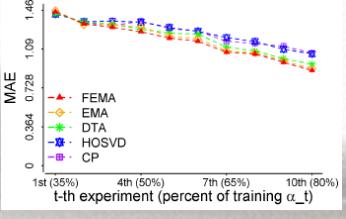
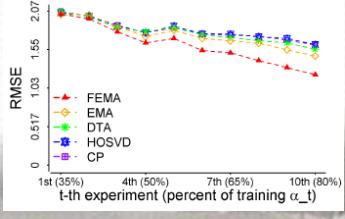
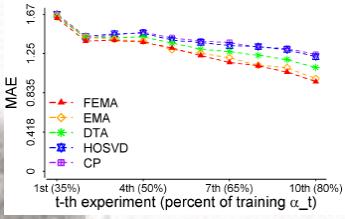
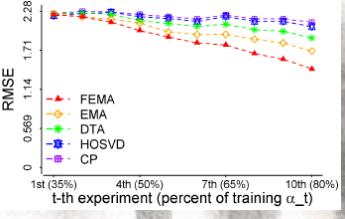
Leveraging Multi-faceted Information

	MAS (Data mining subset)		Tencent Weibo (January, 2011)	
	MAE	RMSE	MAE	RMSE
FEMA 	0.735	0.944	0.894	1.312
EMA 	0.794	1.130	0.932	1.556
EA 	0.979	1.364	1.120	1.873
				



Leveraging Flexible Regularizations

“Who”-“Where”-“What keyword”? “Who”-“@Whom”-“What”?

	MAS (Data mining subset)		Tencent Weibo (January, 2011)	
	MAE	RMSE	MAE	RMSE
FEMA 	0.893	1.215	0.954	1.437
EMA 	0.909	1.466	0.986	1.698
DTA	0.950	1.556	1.105	1.889
HOSVD	1.047	1.618	1.220	2.054
CP	1.055	1.612	1.243	2.117
Cross Validation				



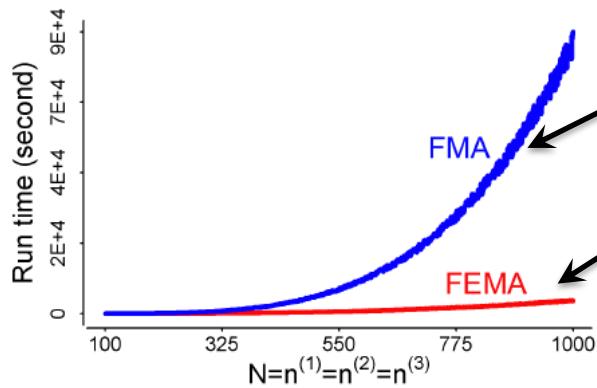
Leveraging Flexible Regularizations

“Who”-“Where”-“What keyword”? “Who”-“@Whom”-“What”?

	MAS (Data mining subset)		Tencent Weibo (January, 2011)	
	MAE	RMSE	MAE	RMSE
FEMA 	0.893	1.215	0.954	1.437
EMA 	0.909	1.466	0.986	1.698



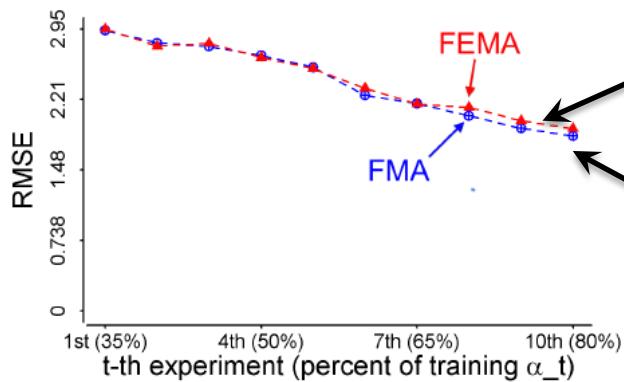
Efficiency, Loss and Parameters



Re-decompose updated matrices

Evolutionary analysis:
update λ and a with ΔX

Time vs Num. objects N

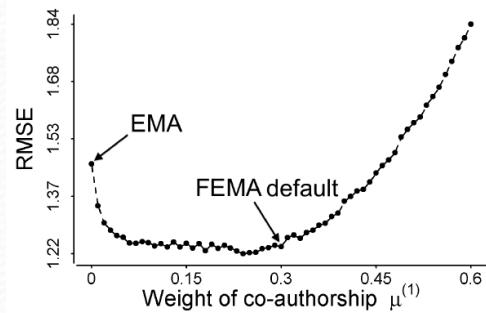


Evolutionary analysis:
update λ and a with ΔX

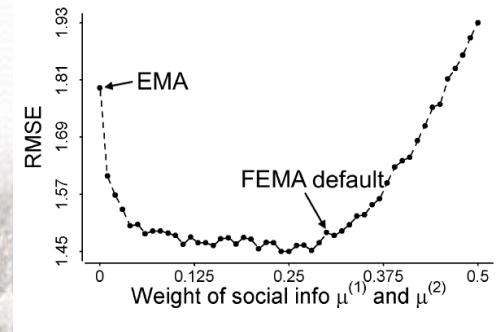
Re-decompose updated matrices

The loss is small.

Insensitive to regularization weight



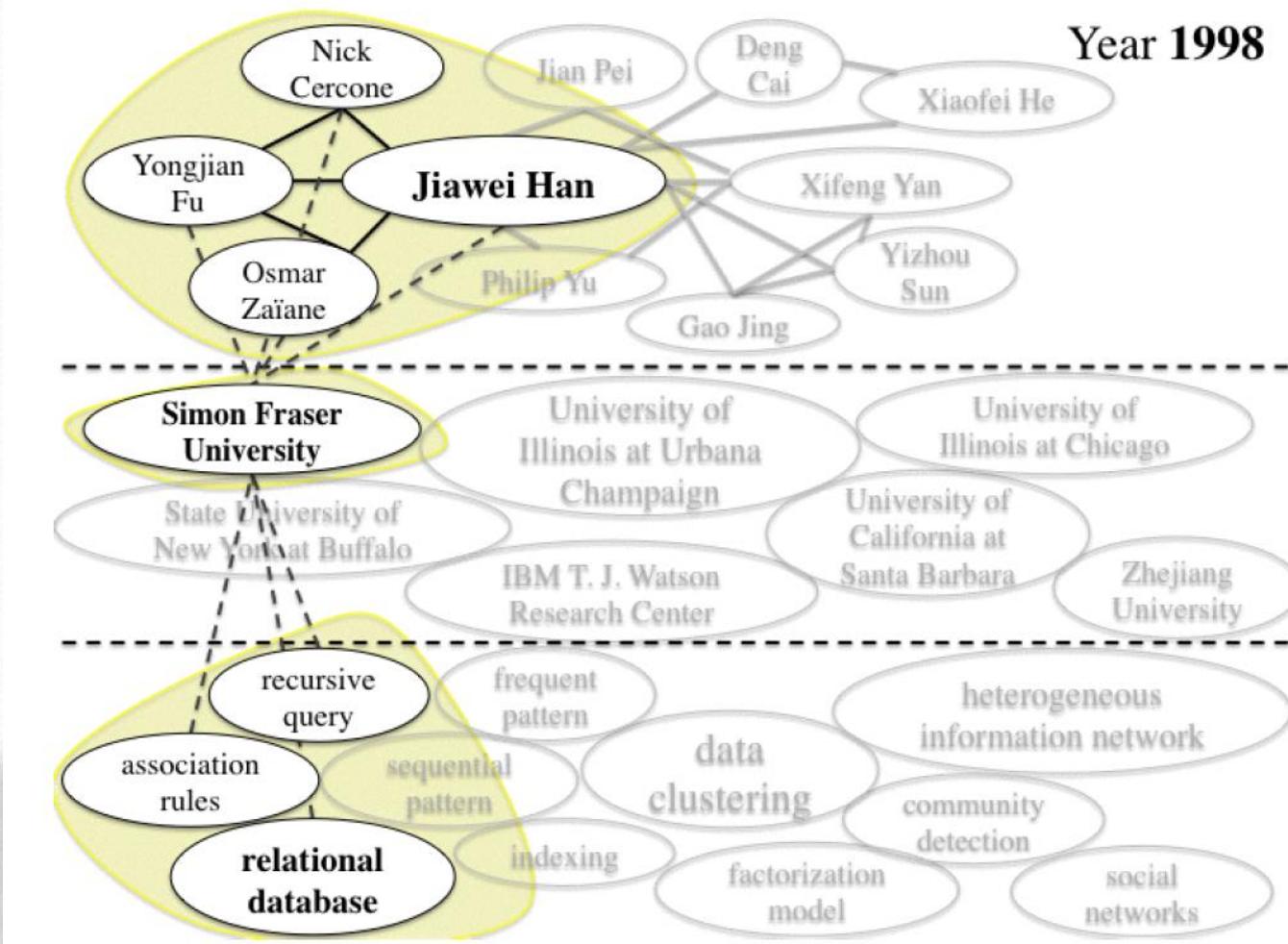
(a) Academic data MAS



(b) Tweet data WEIBO

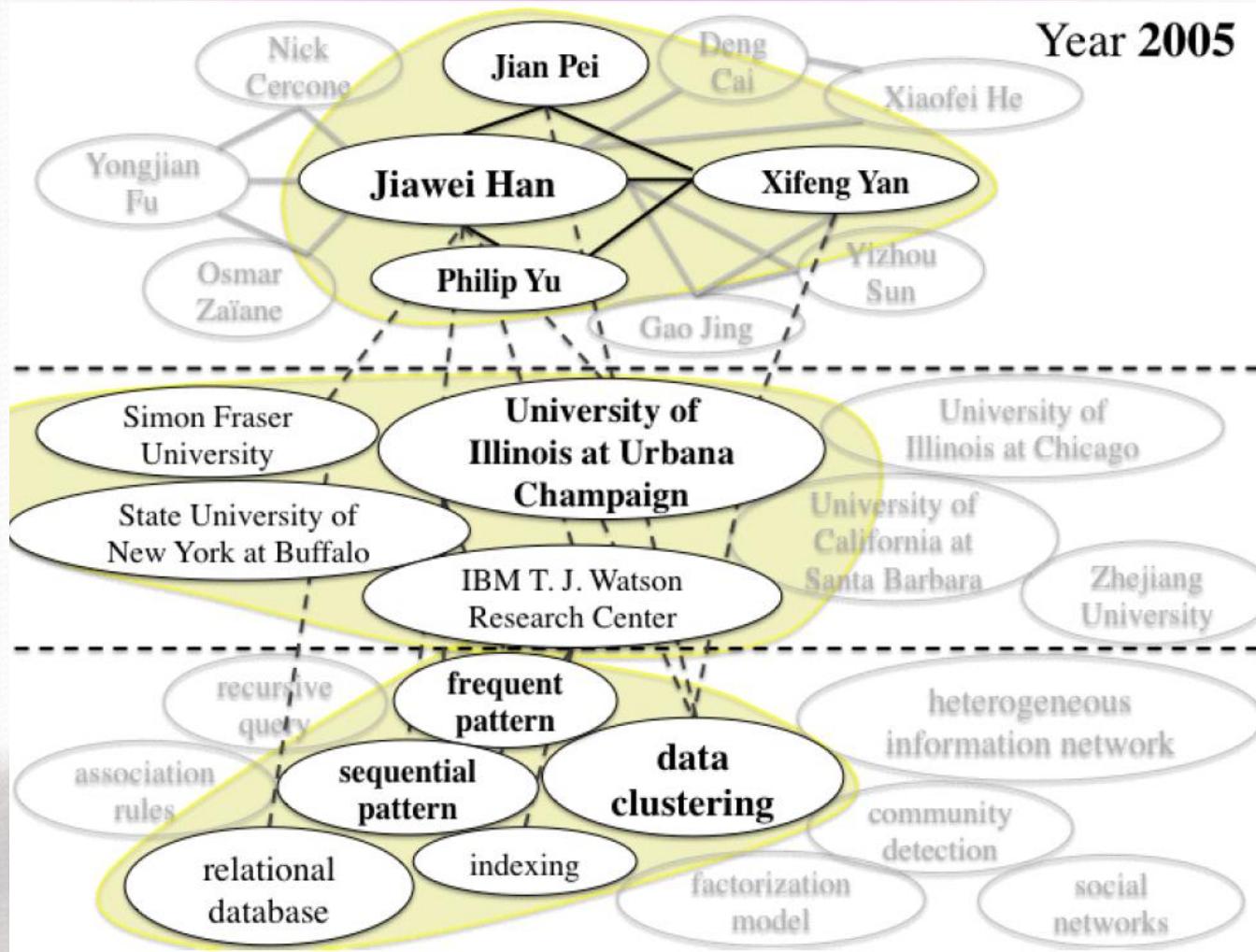


First Cluster in MAS DM set



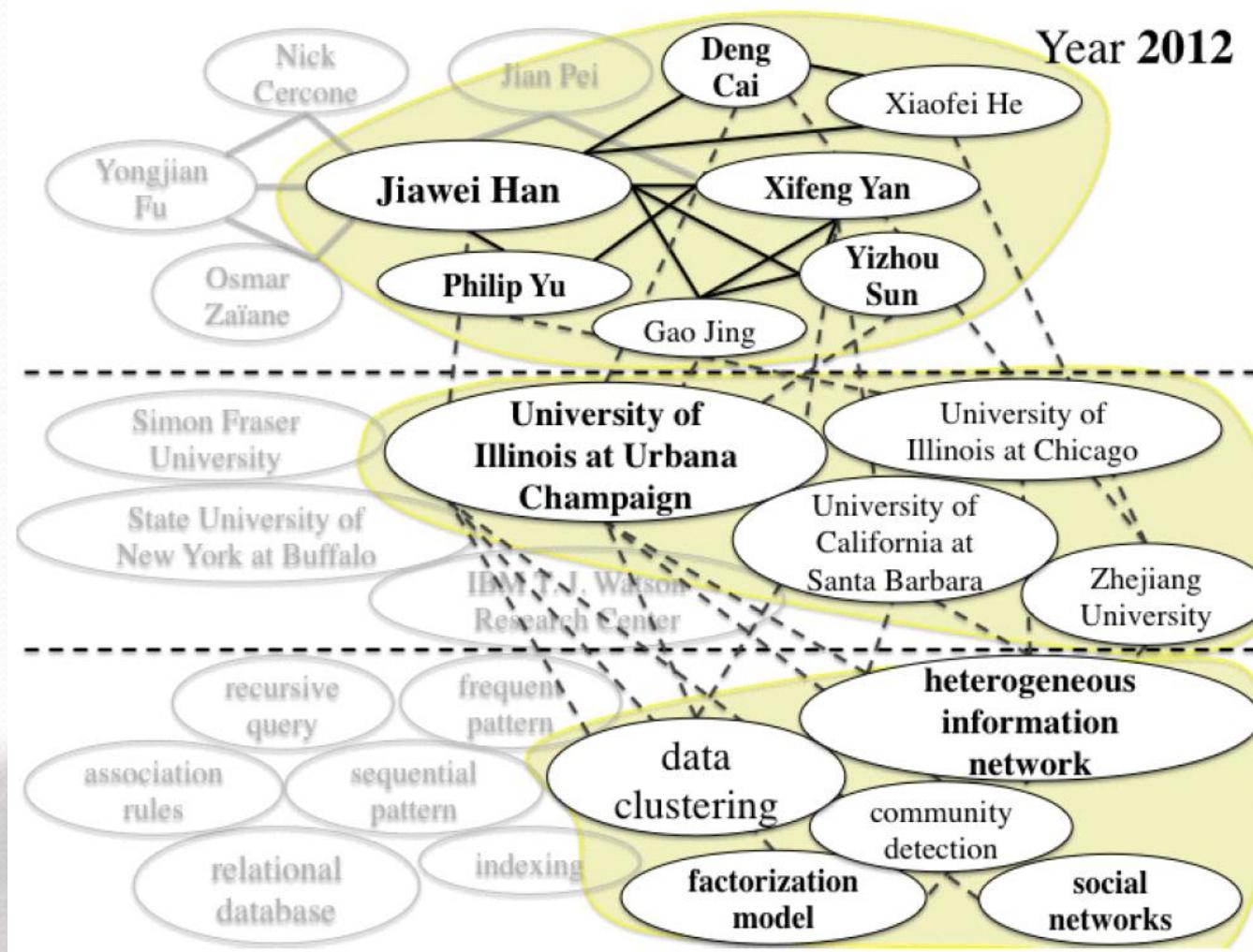


First Cluster in MAS DM set



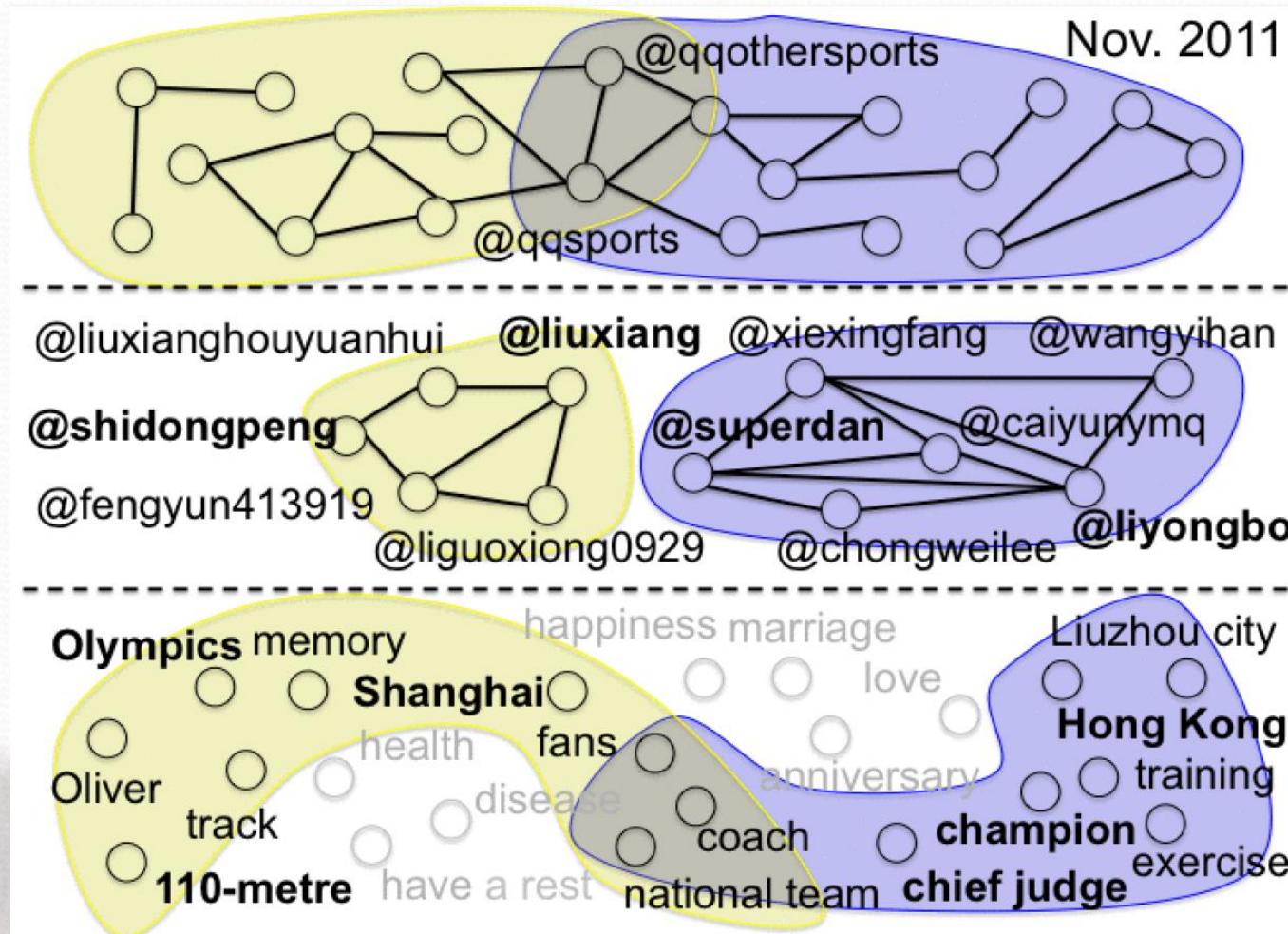


First Cluster in MAS DM set



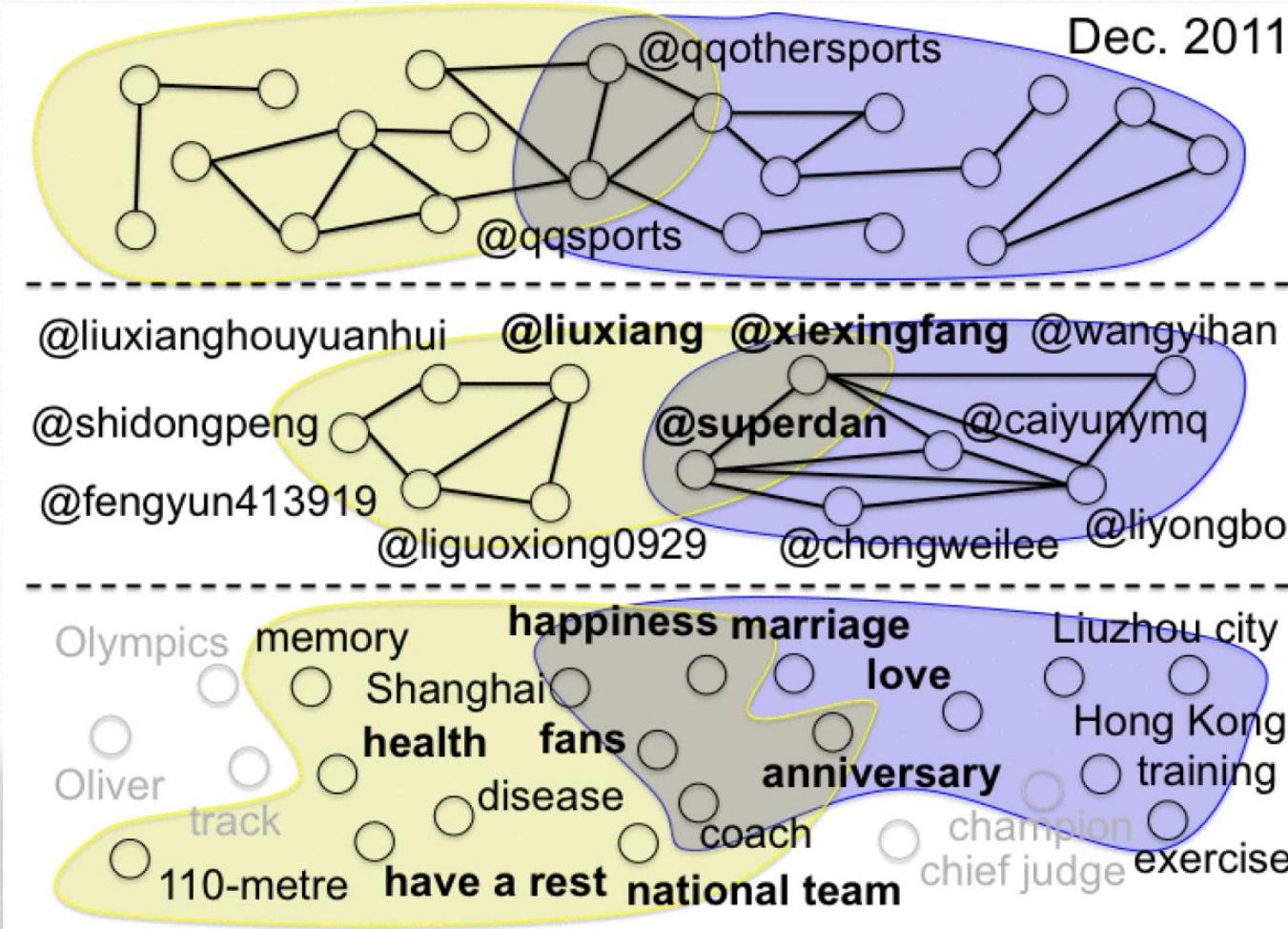


First Two Clusters in Tencent Weibo set





First Two Clusters in Tencent Weibo set





Demo: www.meng-jiang.com

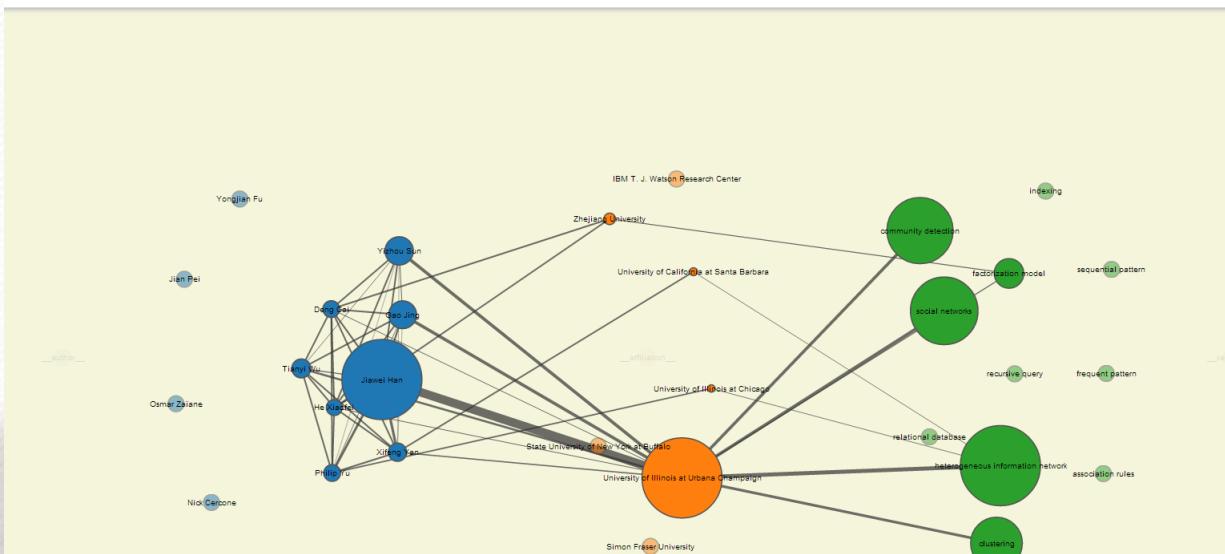
FEMA: Flexible Evolutionary Multi-faceted Analysis for Dynamic Behavioral Pattern Discovery. (KDD 2014)

Meng Jiang, Peng Cui, Fei Wang, Xinran Xu, Wenwu Zhu and Shiqiang Yang.

[How do research topics of Professor Jiawei Han's group change? What happened in the data mining communities and why?](#)

[How do Weibo users mention \(using "@"\) their idols \(@superdan, etc.\) in their tweets? What are the words in their tweets?](#)

A demo for dynamic behavioral pattern discovery using FEMA



Year 2012

Press <- / -> or drag the bar to change year.

Node Detail

Category	keyword
Name	recursive query
Weight	0

Link Detail

People always change. Professor Jiawei Han moved from Simon Fraser University to University of Illinois at Urbana Champaign. His excellent students graduated and young students became rising stars. Social networks emerged. How do research topics of his group change and why?

Blue: authors; Orange: affiliations;
Green: keywords.

Meng Jiang, Peng Cui, Fei Wang, Xinran Xu, Wenwu Zhu and Shiqiang Yang. FEMA: Flexible Evolutionary Multi-faceted Analysis for Dynamic Behavioral Pattern Discovery. *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.



Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



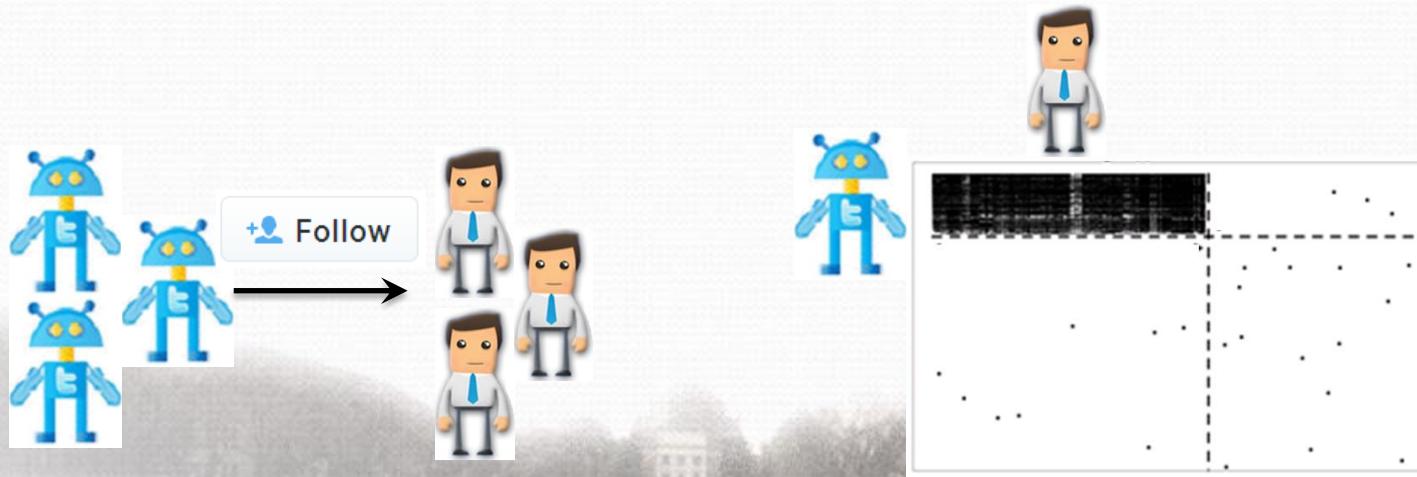
Bad user-user links



Strange Behavior/Connectivity Patterns

❖ Problem

- Strange behavior on large graph: dense bipartite cores
- **Twitter/Weibo “who-follows-whom”**





Strange Behavior/Connectivity Patterns

❖ Problem

- Strange behavior on large graph: dense bipartite cores
- Twitter/Weibo “who-follows-whom”
- **eBay “who-reviews-what”**

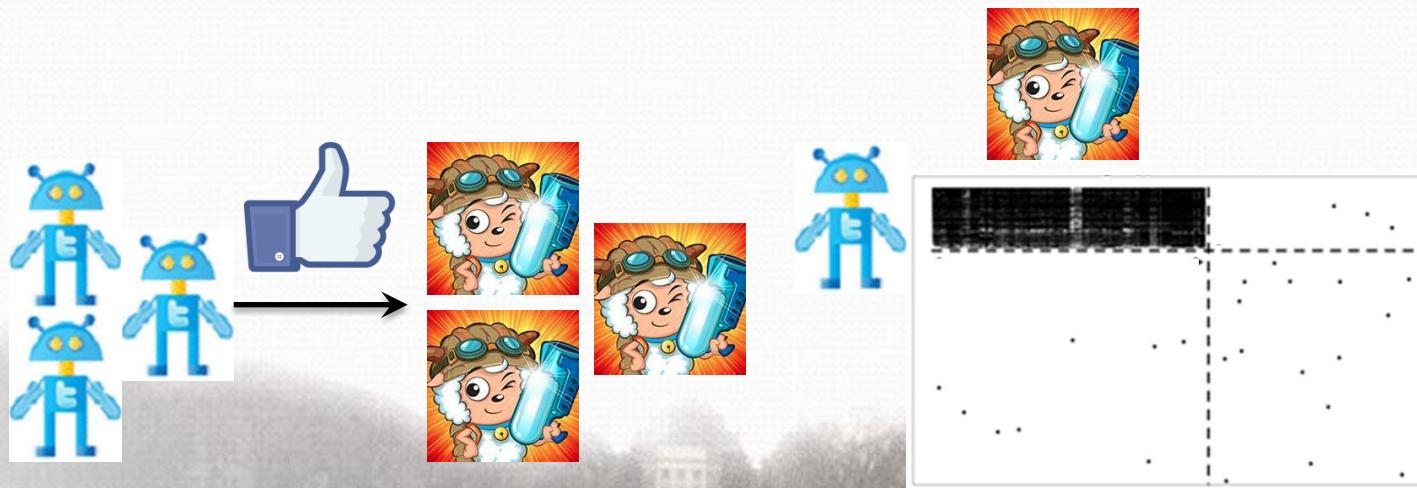




Strange Behavior/Connectivity Patterns

❖ Problem

- Strange behavior on large graph: dense bipartite cores
- Twitter/Weibo “who-follows-whom”
- eBay “who-reviews-what”
- **Facebook “who-likes-page”**



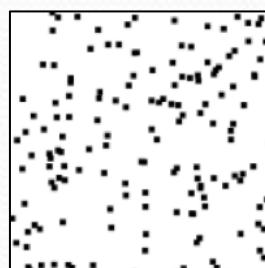


Strange Behavior/Connectivity Patterns

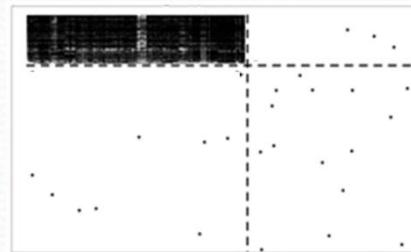
❖ Problem

- Strange behavior on large graph: dense bipartite cores
- Twitter/Weibo “who-follows-whom”
- eBay “who-reviews-what”
- Facebook “who-likes-page”
- **“Lockstep” behavior**

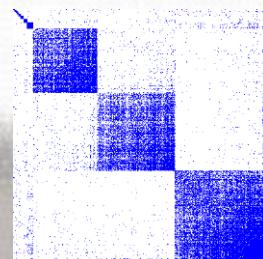
- Given



reordering →



- Usually we have

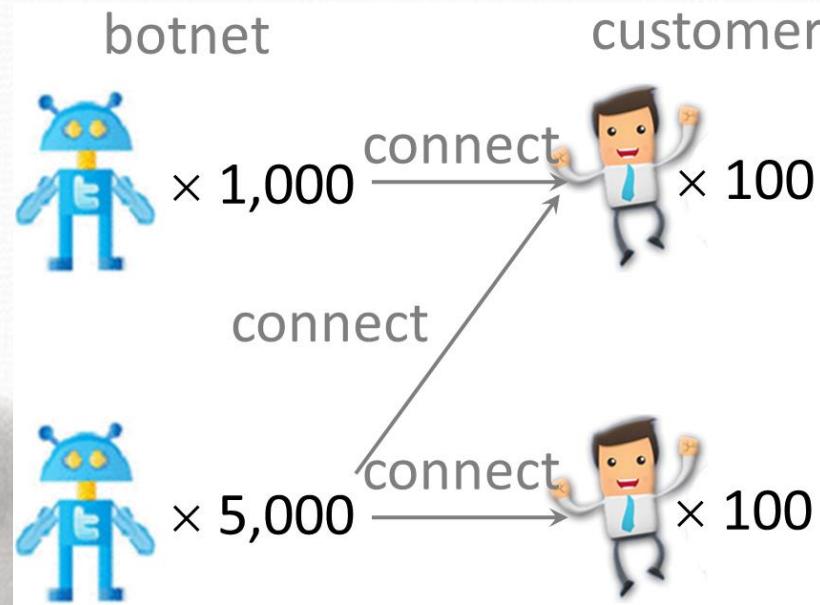




Strange Behavior/Connectivity Patterns

❖ Challenges

- “Lockstep” behavior
- Overlapping

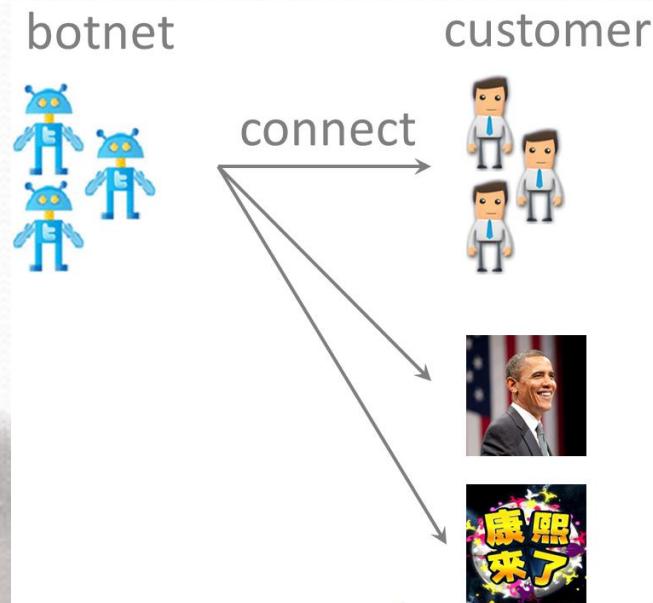




Strange Behavior/Connectivity Patterns

❖ Challenges

- “**Lockstep**” behavior
- Overlapping
- **Camouflage**

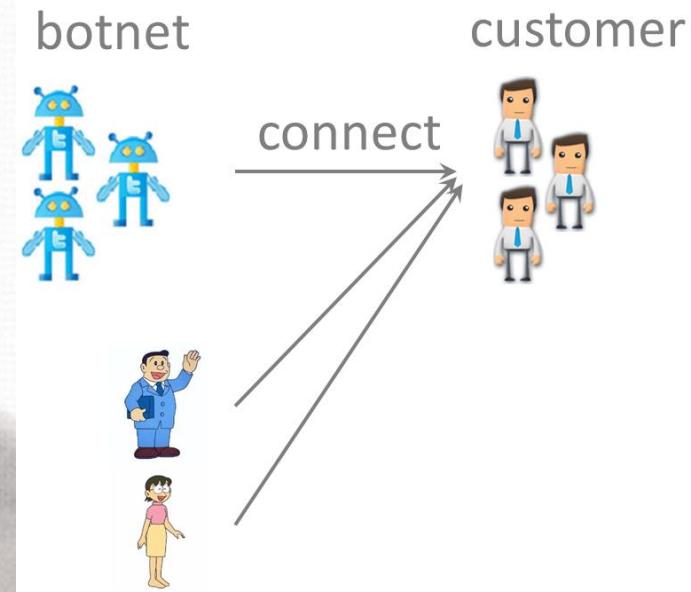




Strange Behavior/Connectivity Patterns

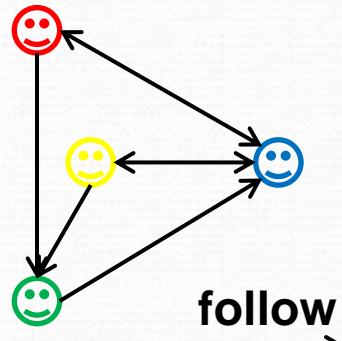
❖ Challenges

- “Lockstep” behavior
- Overlapping
- Camouflage
- **Fame**

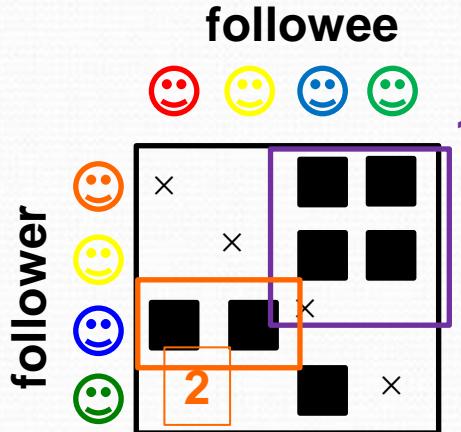




SVD Reminder

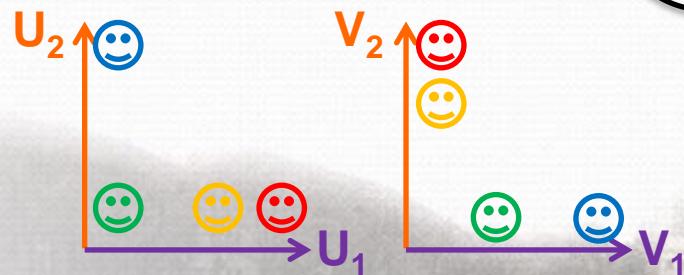


Graph Structure

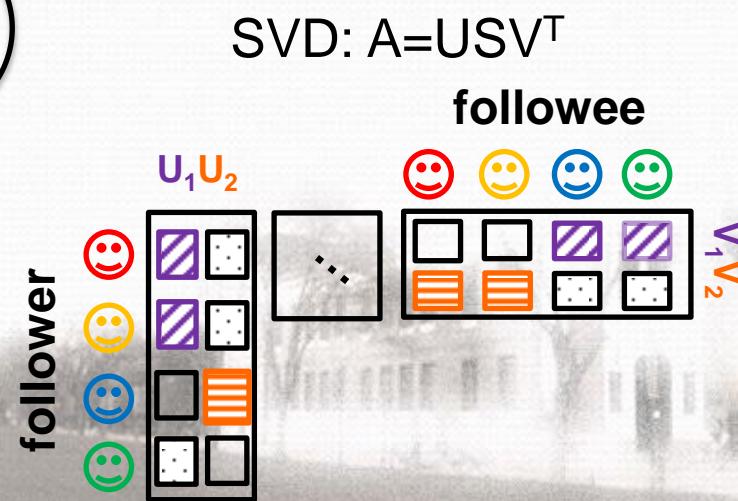


Adjacency Matrix

Pairs of singular vectors:



“Spectral Subspace Plot”

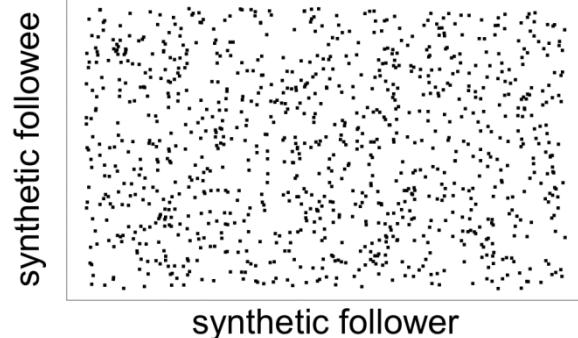




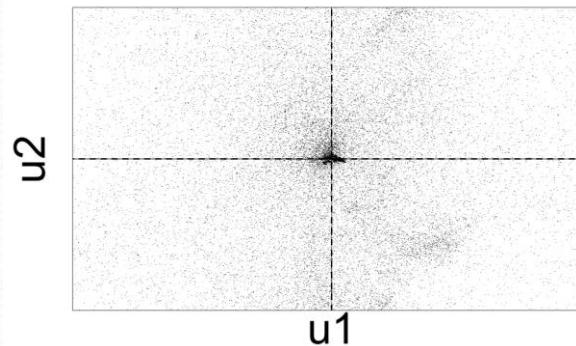
Lockstep and Spectral Subspace Plot

- ❖ Case #0: No lockstep behavior in random power law graph of 1M nodes, 3M edges
- ❖ Random \longleftrightarrow “Scatter”

Adjacency Matrix



Spectral Subspace Plot

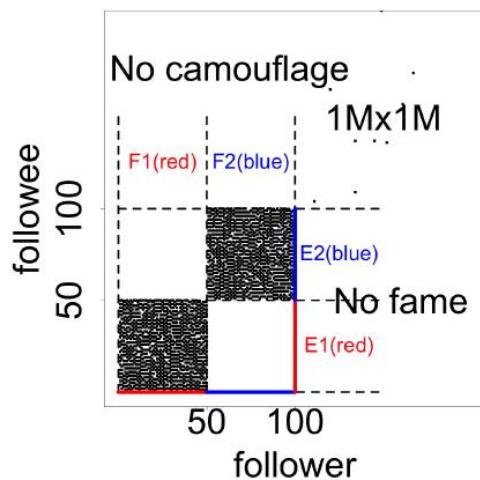




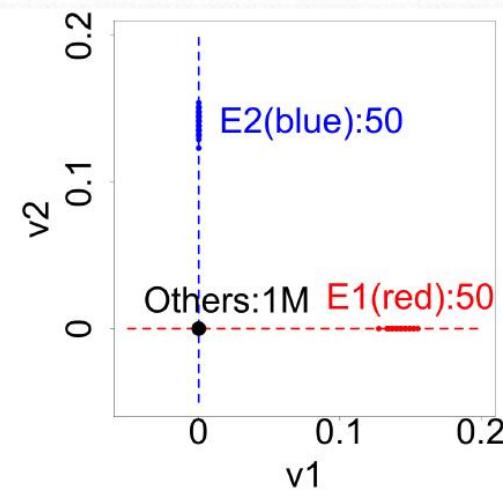
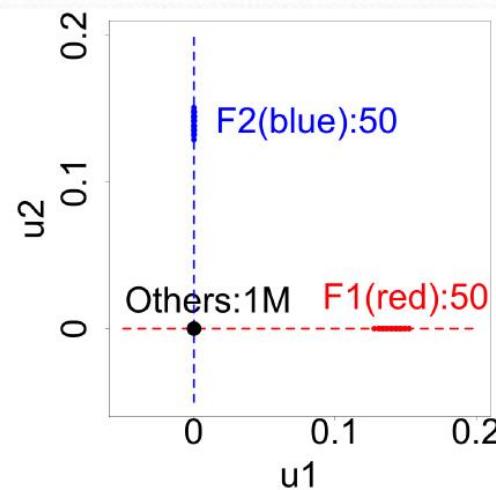
Lockstep and Spectral Subspace Plot

- ❖ Case #1: non-overlapping lockstep
- ❖ “Blocks” \longleftrightarrow “Rays”

Adjacency Matrix



Spectral Subspace Plot



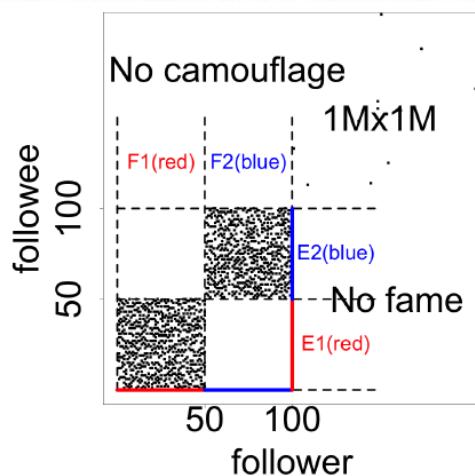
Rule 1 (short “rays”): two blocks, high density (90%), no “camouflage”, no “fame”



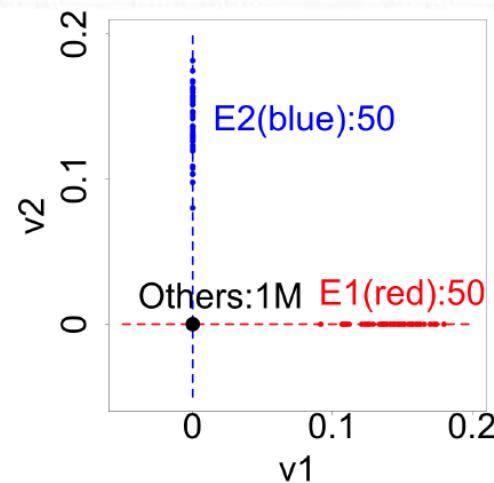
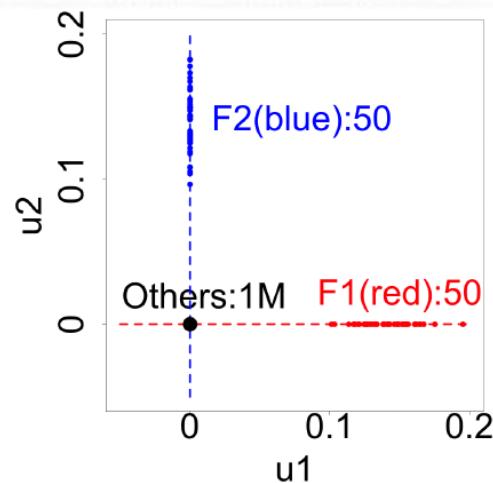
Lockstep and Spectral Subspace Plot

- ❖ Case #2: non-overlapping lockstep
- ❖ “Blocks; low density” \longleftrightarrow Elongation

Adjacency Matrix



Spectral Subspace Plot

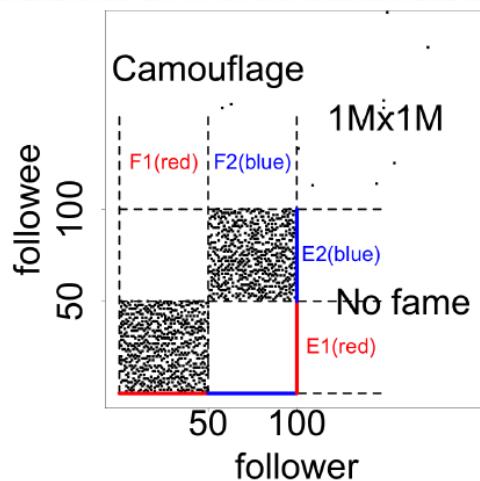


Rule 2 (long “rays”): two blocks, low density (50%), no “camouflage”, no “fame”

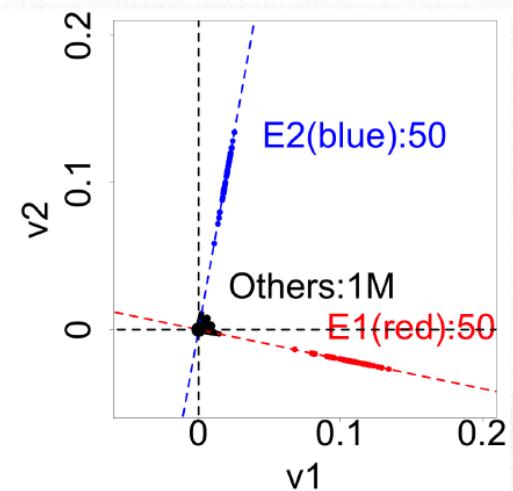
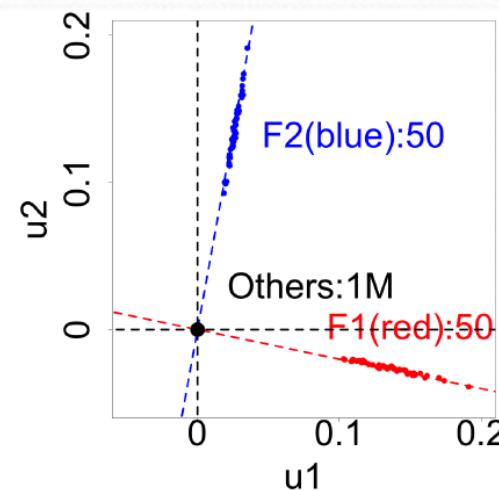
Lockstep and Spectral Subspace Plot

- ❖ Case #3: non-overlapping lockstep
- ❖ “**Camouflage**” (or “Fame”) \longleftrightarrow Tilting “Rays”

Adjacency Matrix



Spectral Subspace Plot

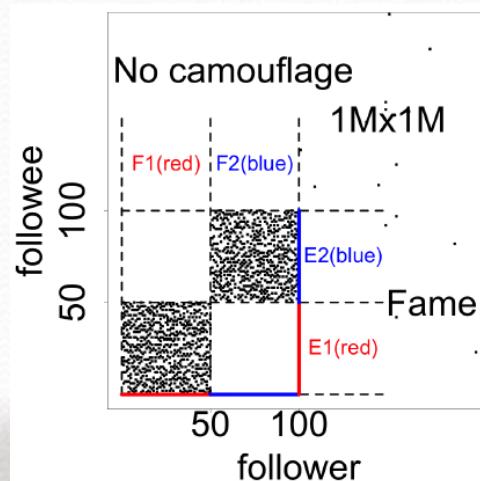


Rule 3 (tilting “rays”): two blocks, with “camouflage”, no “fame”

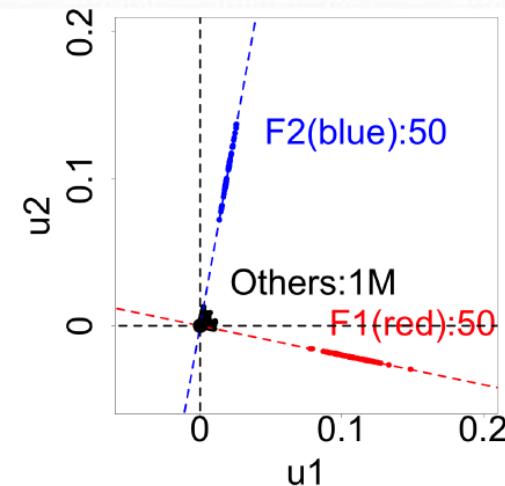
Lockstep and Spectral Subspace Plot

- ❖ Case #3: non-overlapping lockstep
 - ❖ “Camouflage” (or “Fame”) \longleftrightarrow Tilting “Rays”

Adjacency Matrix



Spectral Subspace Plot



Rule 3 (tilting “rays”): two blocks, no “camouflage”, with “fame”

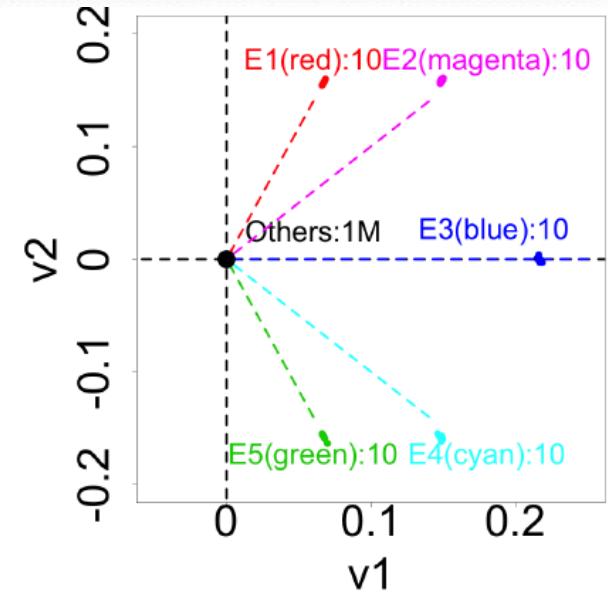
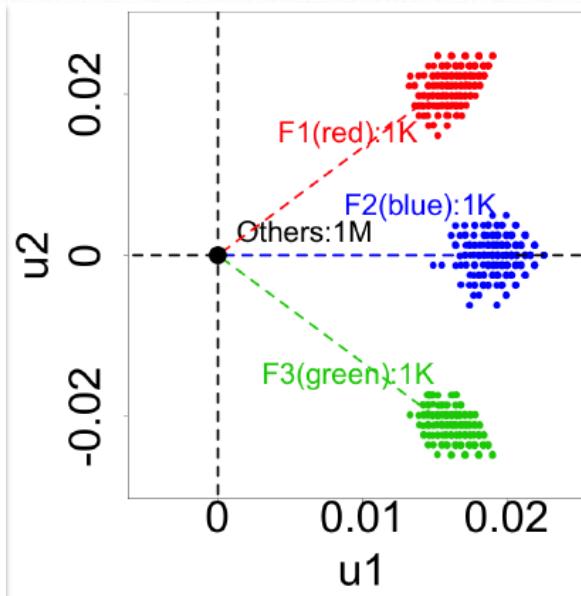
Lockstep and Spectral Subspace Plot

- ❖ Case #4: ? lockstep
- ❖ “?” \longleftrightarrow “Pearls”

Adjacency Matrix

?

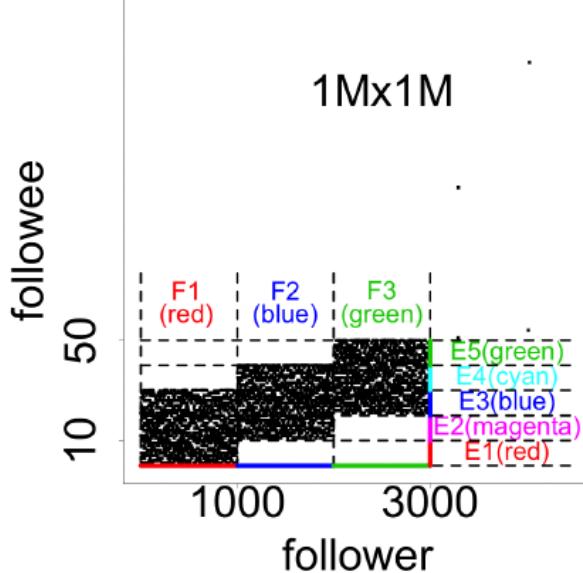
Spectral Subspace Plot



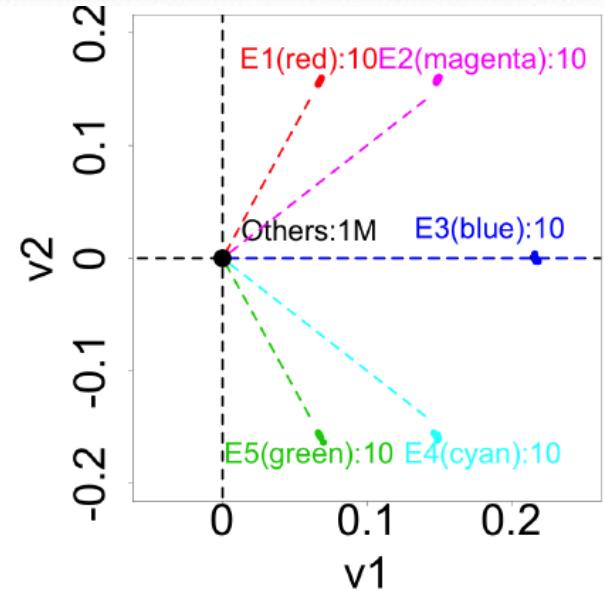
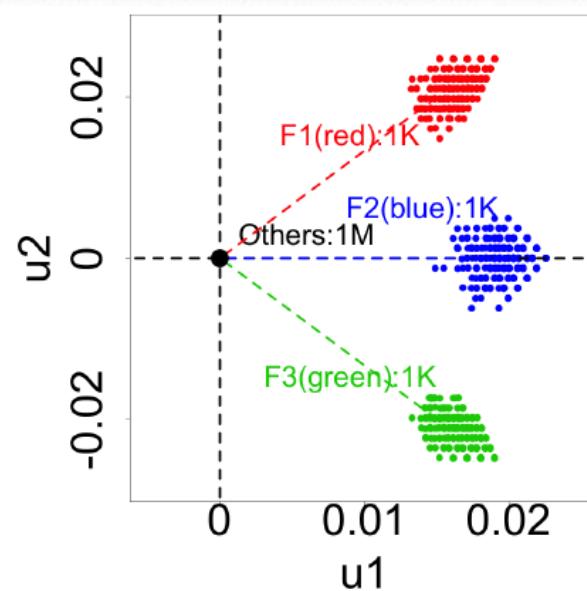
Lockstep and Spectral Subspace Plot

- ❖ Case #4: **overlapping lockstep**
- ❖ “**Staircase**” \longleftrightarrow “Pearls”

Adjacency Matrix



Spectral Subspace Plot



Rule 4 (“pearls”): a “staircase” of three partially overlapping blocks.



Algorithm

❖ Step 1: Seed selection

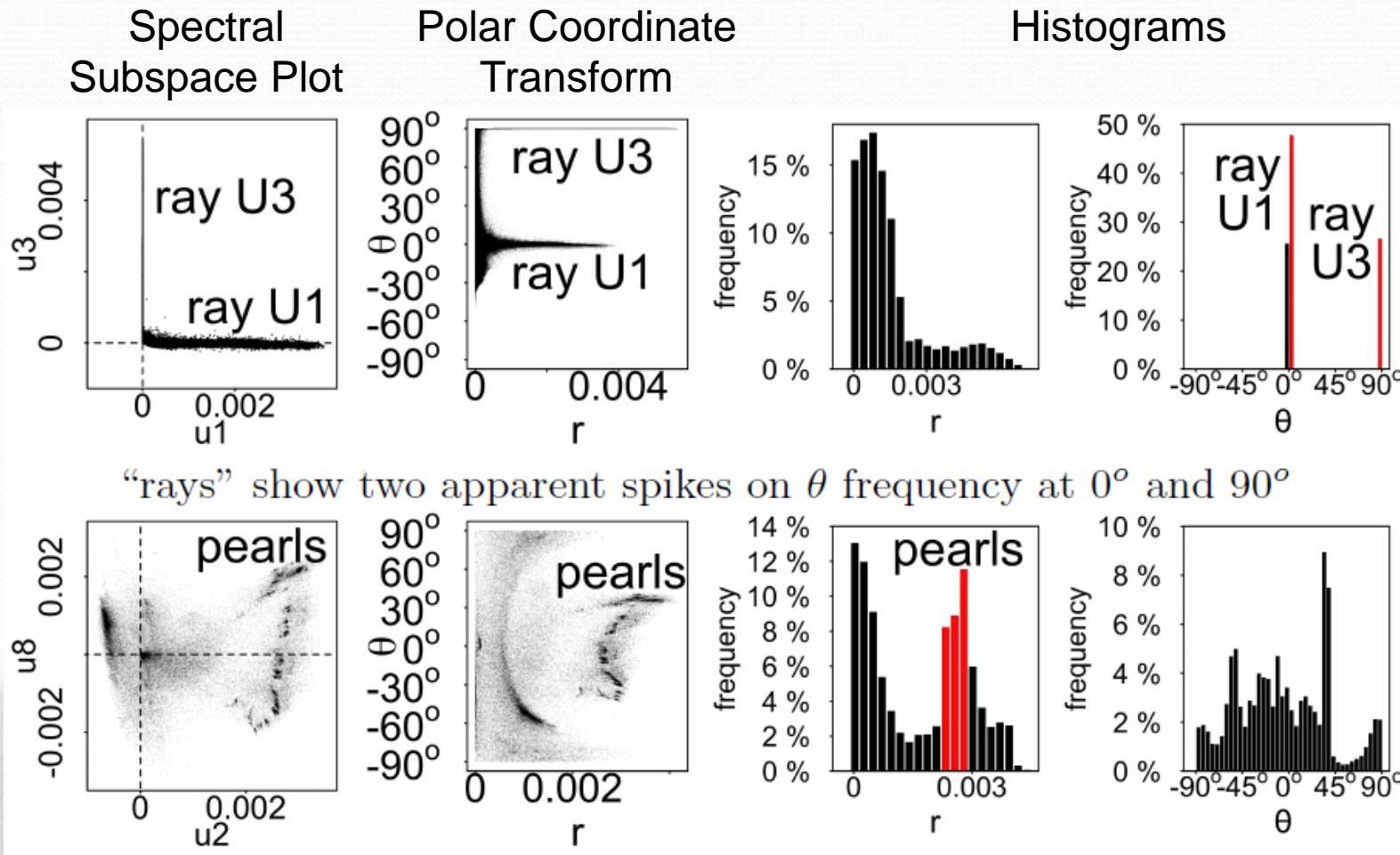
- Spot “Rays” and “Pearls”
- Catch seed followers

❖ Step 2: Belief Propagation

- Blame followees with strange followers
- Blame followers with strange followees



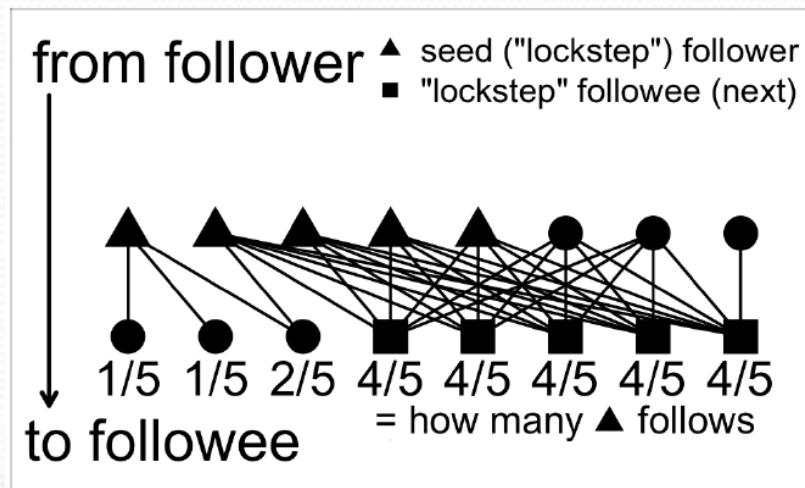
Automatically Spot “Rays” and “Pearls”



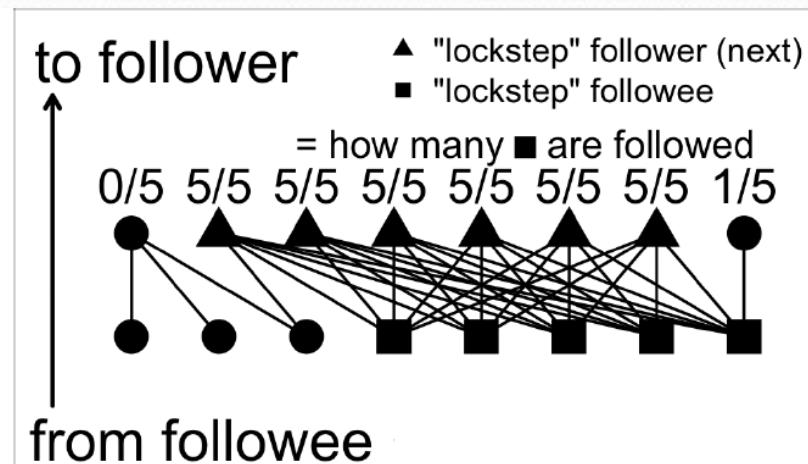


BP-based Algorithm

- ❖ Blame followees with strange followers
- ❖ Blame followers with strange followees



(a) select “lockstep” followees:
from (seed) followers to followees

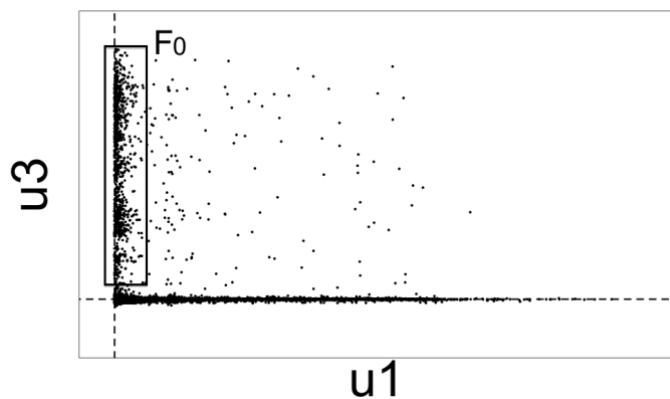


(b) select “lockstep” followers:
from followees to followers

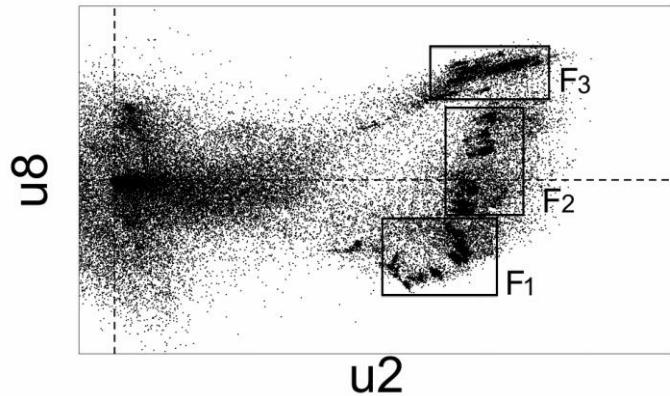


Real Data

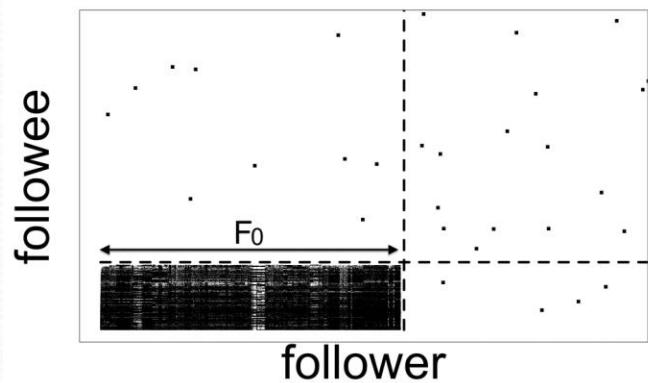
“Rays”



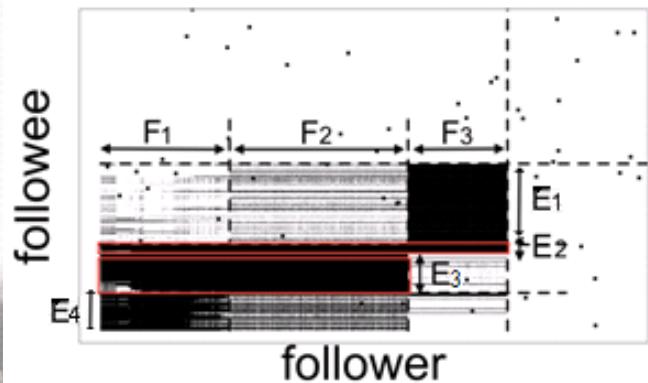
“Pearls”



“Block”



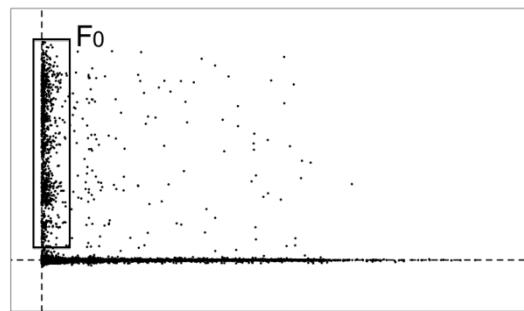
“Staircase”



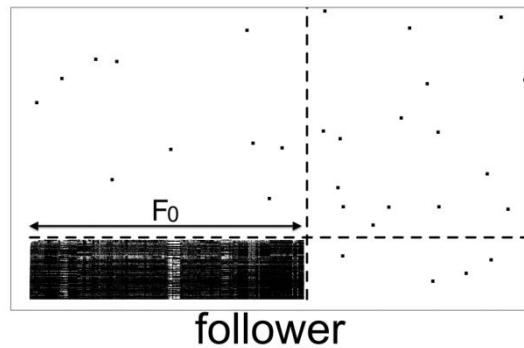


Real Data

“Rays”



“Block”

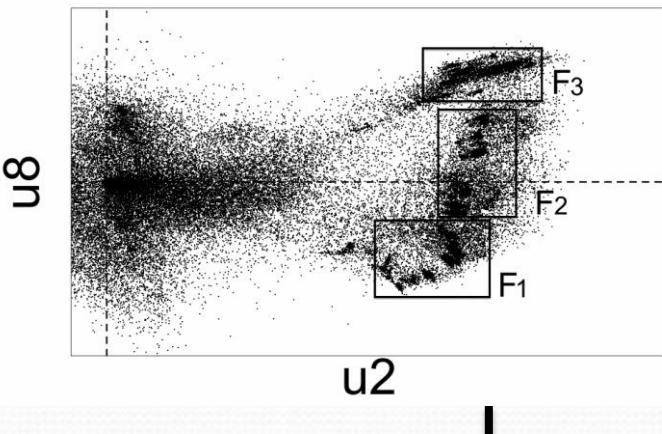


	“ray” F_0	“p”
Num. seeds	100	1,000
Size of block	$83, 208 \times 30$	$3, 10 \times 10$
Density	81.3%	9.1%
Camouflage	0.14%	0.01%
Fame	0.05%	0.01%
Out-degree	231 ± 109	31 ± 10
In-degree	2.0 ± 1.4	1.0 ± 0.5

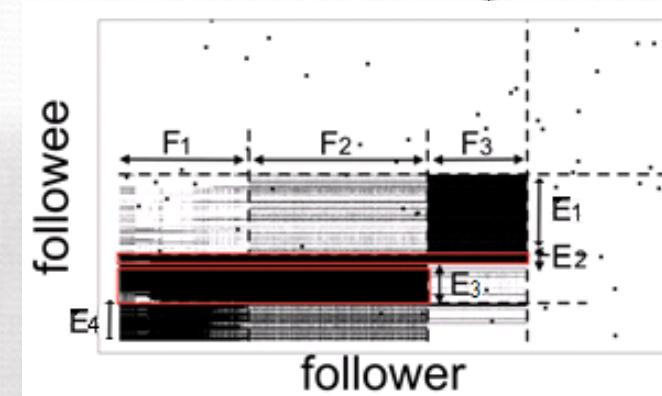


Real Data

“Pearls”



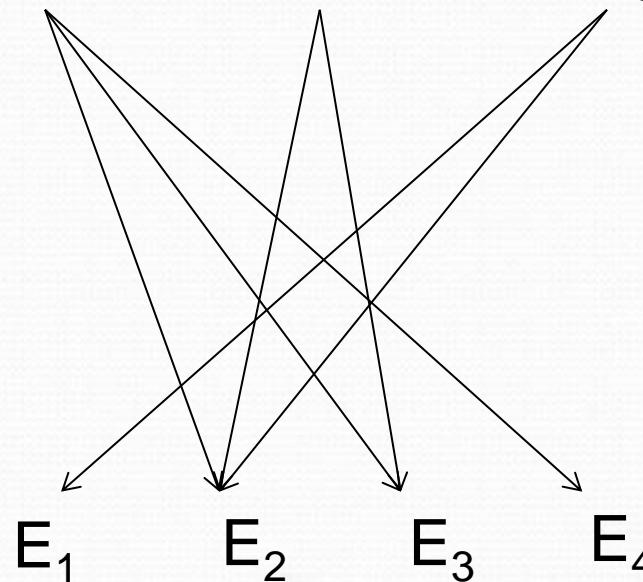
“Staircase”



3,188
in F_1

7,210
in F_2

2,457
in F_3

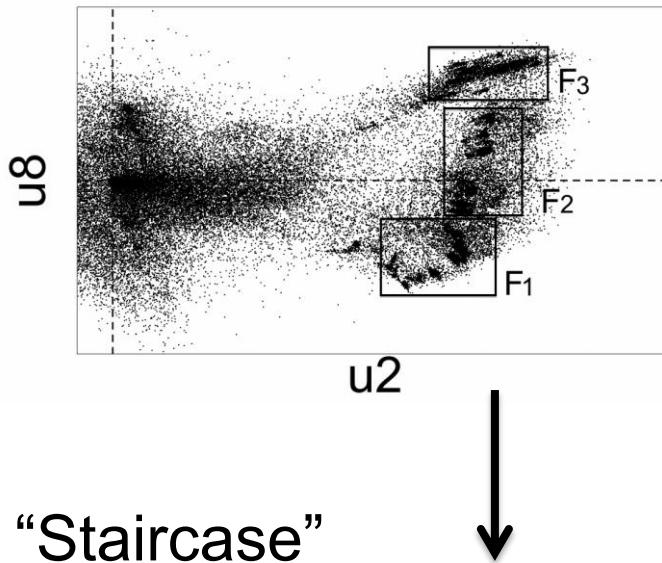


“F-E”	F_1 -...	F_2 -...	F_3 -...
Density	91.3%	92.6%	89.1%

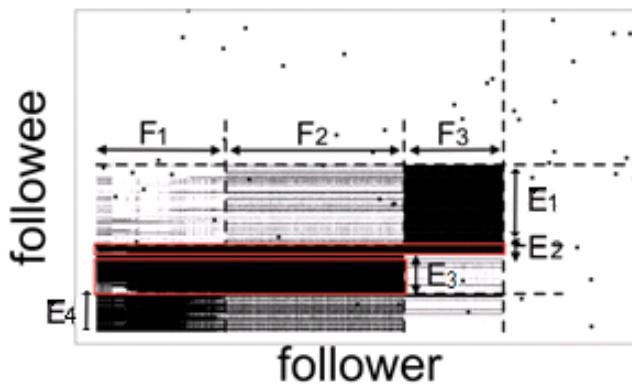


Real Data

“Pearls”



“Staircase”

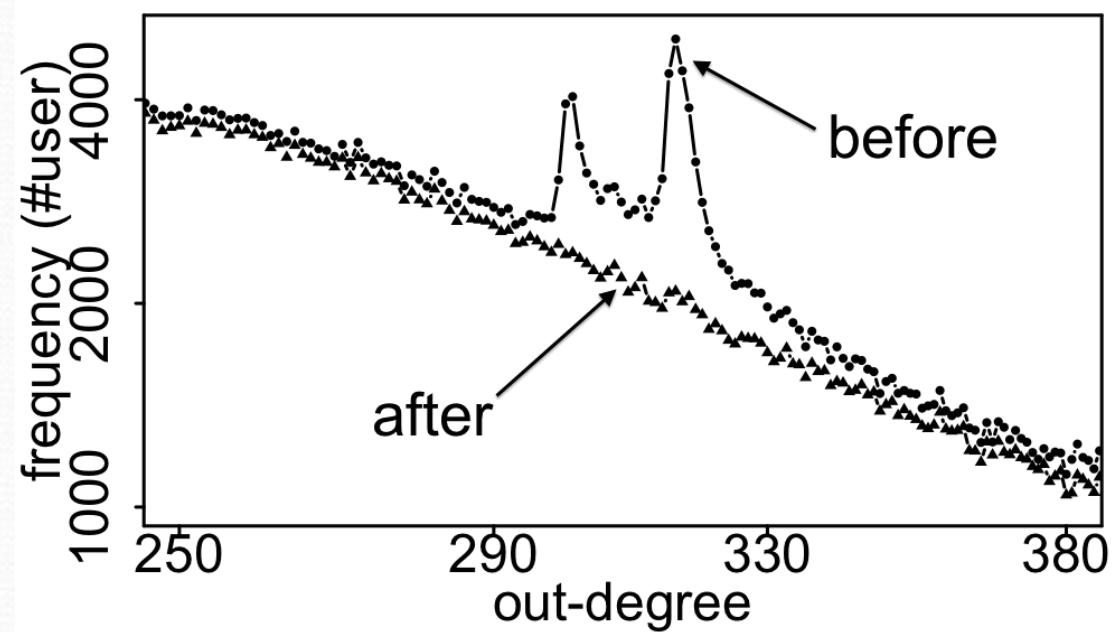
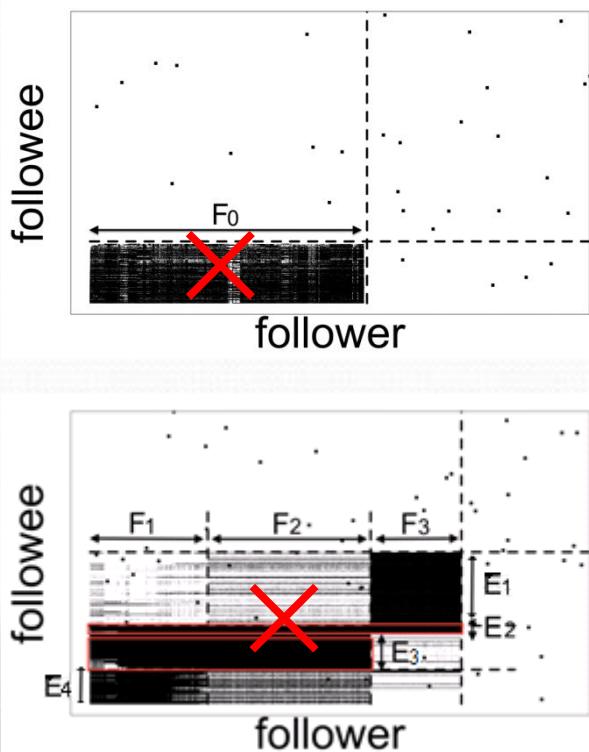


	“pearl” F_1	“pearl” F_2	“pearl” F_3	“pearl” Total
	1,239	107	990	—
	$3,188 \times 135$	$7,210 \times 79$	$2,457 \times 148$	$10,052 \times 270$
	91.3%	92.6%	89.1%	43.1%
	0.06%	0.10%	0.05%	0.07%
	1.93%	1.94%	1.72%	1.73%
	310 ± 7	312 ± 7	304 ± 5	310 ± 7
	9 ± 6	10 ± 6	17 ± 13	12 ± 9



Real Data

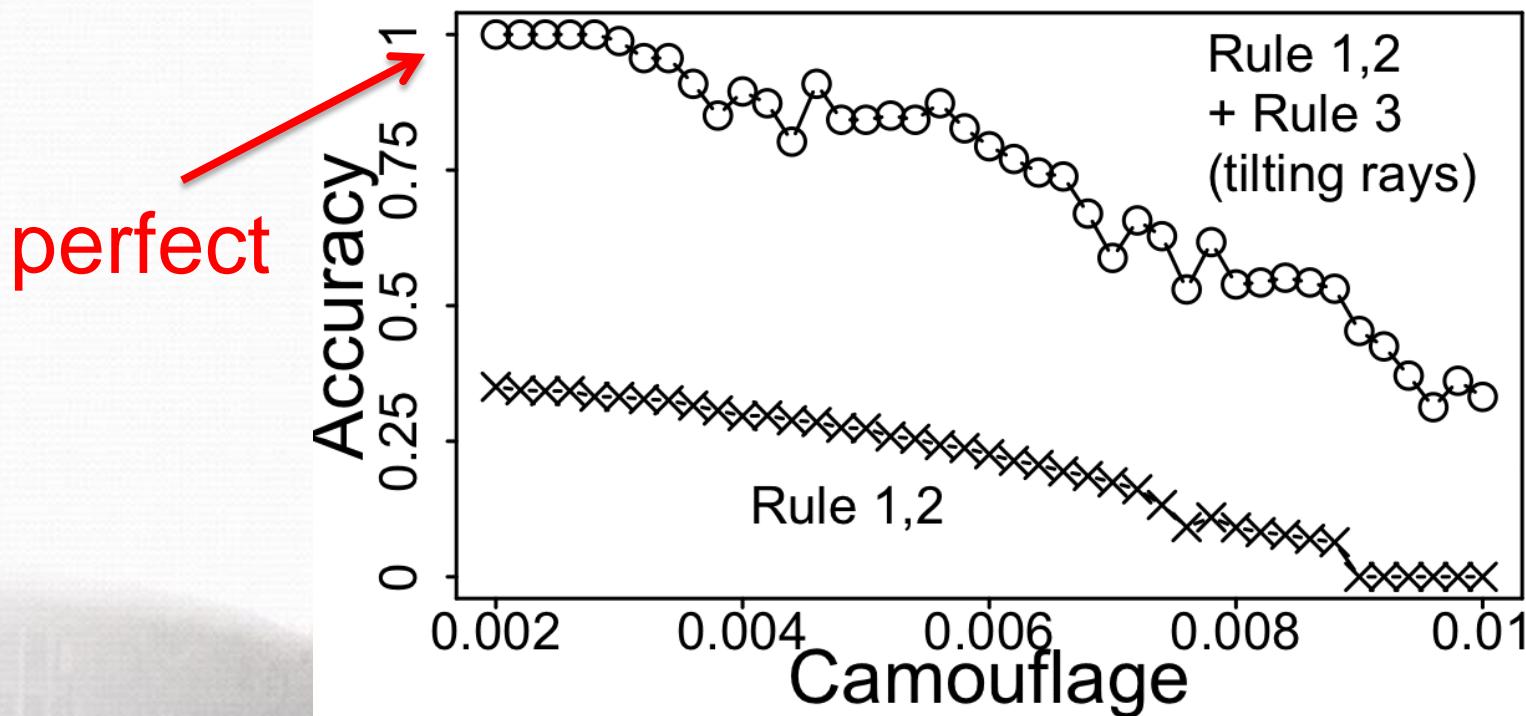
❖ Spikes on the out-degree distribution





Synthetic Data

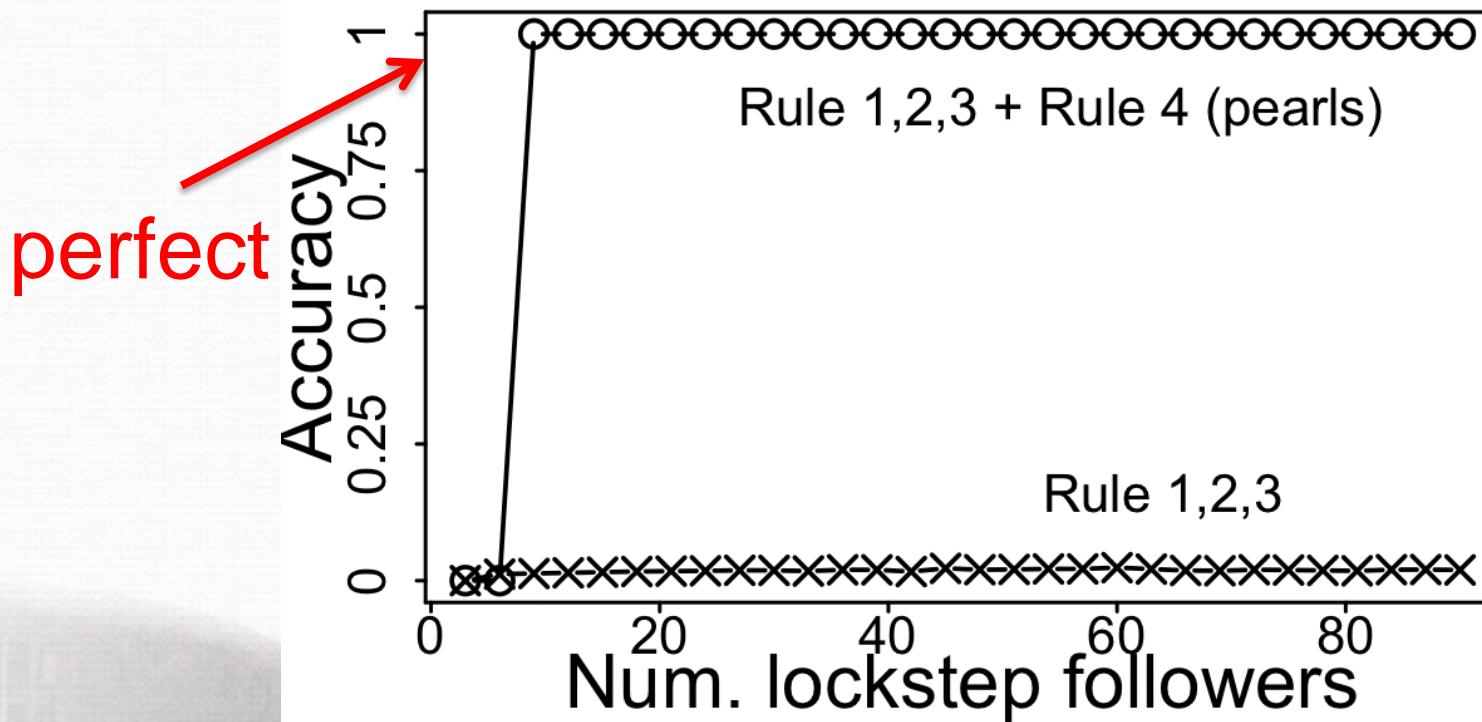
- ❖ Inject lockstep behavior with “camouflage”





Synthetic Data

- ❖ Inject overlapping lockstep behavior





Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links



Bad user-user links



Buy Twitter Followers

- ❖ **(Fake) friends with (Real) benefits**
- ❖ **I paid \$5 for 4,000 Twitter followers, and here's what I found**
- ❖ <https://medium.com/i-data/fake-friends-with-real-benefits-eec8c4693bd3>

The screenshot shows a Google search results page with the query "buy twitter followers" entered into the search bar. Below the search bar, there is a dropdown menu with suggestions: "buy twitter followers", "buy twitter followers cheap", "buy twitter followers review", and "buy twitter accounts".

The main search results area contains several links:

- Buy Twitter Followers**
www.followersale.com/
Buy Twitter Followers for cheap. Buy Twitter Followers in 48 Hours, Get Real Followers, cheapest place to buy twitter followers.
- Fake Twitter Followers: An Easy and Dangerous Game**
theweek.com/twitter.../fake-followers-an-easy-game-but-n...
Dec 15, 2012 – I felt pretty comfortably smug about myself, especially considering I'd gotten this far by dumbly typing in "How to buy Twitter followers" into my ...
- If You Buy Twitter Followers, You'll Regret It - AllTwitter**
www.mediabistro.com/alltwitter/buy-followers_b254...
by Mary Long – in 2,279 Google+ circles - More by Mary Long
Jul 17, 2012 – We've told you before that buying Twitter followers does not work, but it appears to be time for a refresher. Why? Thanks to the website Fiverr, ...
- Buy Twitter Followers 1000 Quality Followers For Only \$14 Save Now!**
intertwitter.com/
Buy Twitter followers on the cheap without following. World's Largest Supplier of Quality Twitter followers online! 100% Money Back Guarantee!
Contact Us - Facebook - Re-Tweets - Buy Instagram Followers
- Twitter Followers For Sale - NYTimes.com**
www.nytimes.com/2012/08/23/.../twitter-followers-for-sale.ht...
Aug 22, 2012 – A Google search for "buy Twitter followers" turns up dozens of Web sites like USocial.net, InterTwitter.com, and FanMeNow.com that sell Twitter ...

On the right side of the search results, there is a sponsored shopping result for "Shop for buy twitter followers on Google". It features a blue Twitter bird icon, the text "1,000 Real Twitter Followers", "\$5.00 - eBay", and the message "Find great deals on eBay!". Below this, there is a link "See more shopping results on Google".



Buy Twitter Followers

- ❖ **(Fake) friends with (Real) benefits**
- ❖ **I paid \$5 for 4,000 Twitter followers, and here's what I found**
- ❖ <https://medium.com/i-data/fake-friends-with-real-benefits-eec8c4693bd3>

I will send you more than 27000 TWITTER followers on your account within 12 hours without needing your... for \$5
I will send 27000 twitter followers on your account... (by twitter_magic)

I will send you followers on twitter,no password,no following, 100 followers in 5 days for \$5
I will send you 100 twitter followers, no password... (by commentsofcredit)

I will teach You step by step How To Get Unlimited Real Twitter Followers in 24hrs for \$5
24hrs service..unlimited access to followers... (by joycenica)

I will teach you how to get unlimited UK real twitter followers to your acc in 24hrs for \$5
This is a simple and short guild on how to add Unlimited... (by joycenica)

I will teach you how to get unlimited US real twitter followers to your acc in 24hrs for \$5
This is a simple and short guild on how to add Unlimited... (by joycenica)

I will give you 1000+ twitter followers without your password for \$5
I will give you 1000+ SAFE twitter followers to... (by thebooster)

I will add 500 US real twitter followers in 20 hours for \$5
500 real US twitter followers, best quality in... (by twinkles)

I will send you a list of over 300 thought provoking questions to ask your Twitter followers for \$5
No need to think of these questions on your own... (by shagam)

I will get you 27,500+ TWITTER followers on your Twitters profile to skyrocket your follower... for \$5
Order NOW to receive 27,500+ NEW Twitter Followers... (by twitter_magic)

I will get you 200+ real INSTAGRAM followers for \$5
★☆ Without following others - Without using fake... (by hamsey87)



Buy Twitter Followers

- ❖ **(Fake) friends with (Real) benefits**
- ❖ **I paid \$5 for 4,000 Twitter followers, and here's what I found**
- ❖ <https://medium.com/i-data/fake-friends-with-real-benefits-eec8c4693bd3>

Annalisa Monsod
@AnnalisaMonsodz FOLLOW YOU

By: Dan Schawbel on November 30th, 2010 at 11:19 am
South Cambridgeshire, UK

17 TWEETS 2,000 FOLLOWING 12 FOLLOWERS [Follow](#)

Tweets

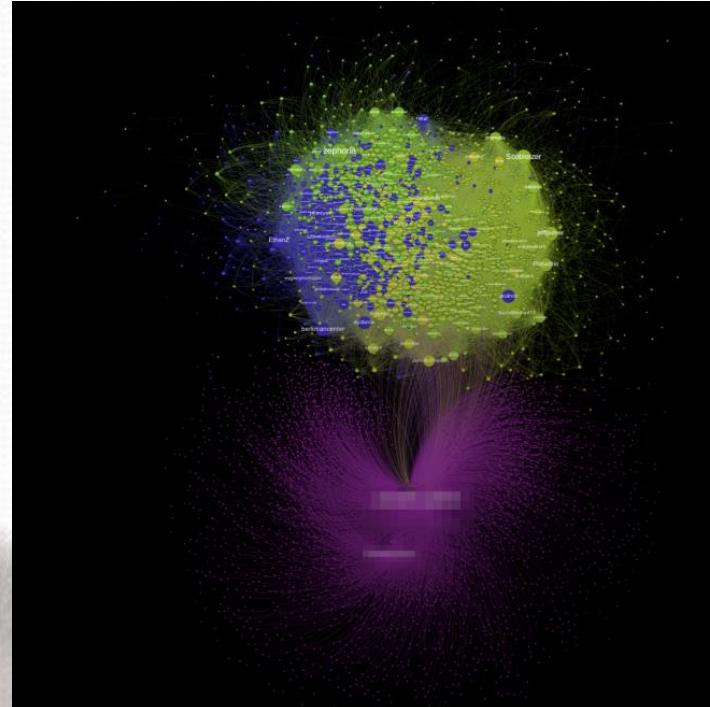
Annalisa Monsod @AnnalisaMonsodz 6 Jul
Entrepreneurial Spotlight: Daniel Brusilovsky - Grasshopper Group
[Expand](#)

Annalisa Monsod @AnnalisaMonsodz 6 Jul
The only place I see that he's typed it as ?MAC? is in the titles, in
which case all of the letters are capitalized, as it's a title
[Expand](#)



Buy Twitter Followers

- ❖ **(Fake) friends with (Real) benefits**
- ❖ **I paid \$5 for 4,000 Twitter followers, and here's what I found**
- ❖ <https://medium.com/i-data/fake-friends-with-real-benefits-eec8c4693bd3>





Zombie Follower Detection

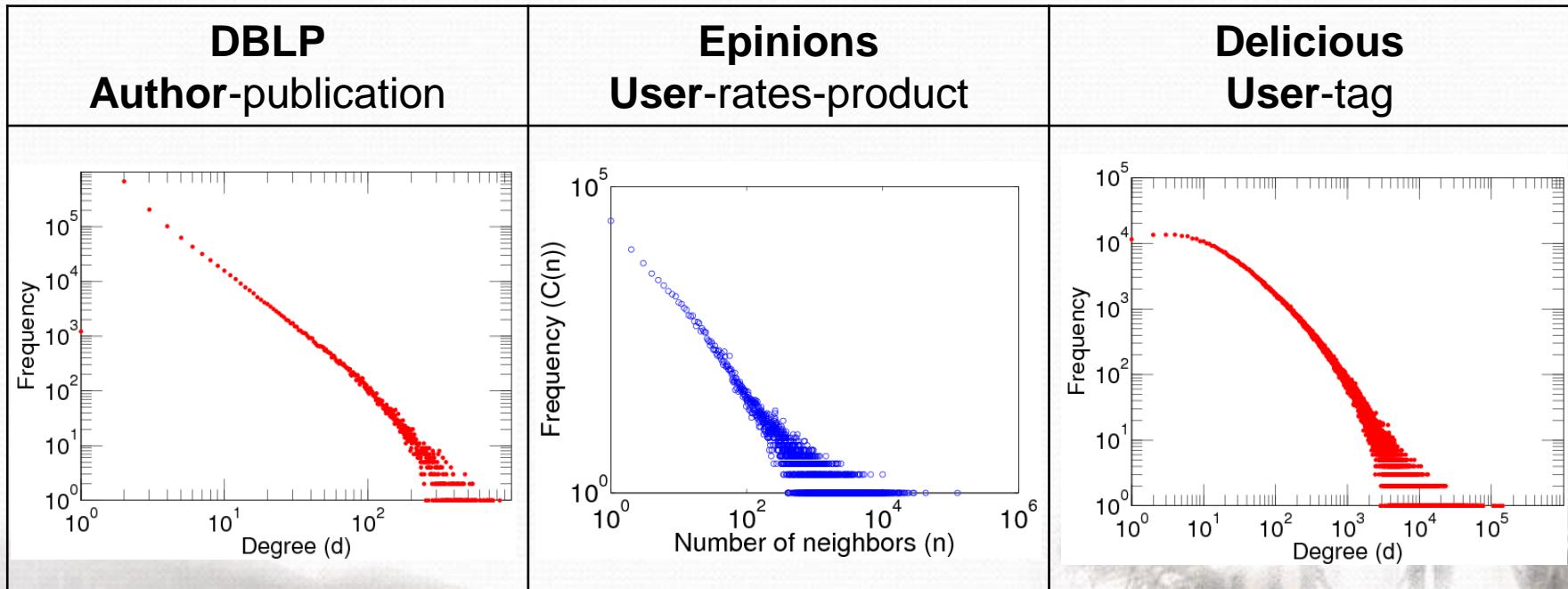
❖ Challenges

- **Scalability:** How to catch zombie followers from large graphs of millions of nodes and billions of edges?



Out-degree Distribution

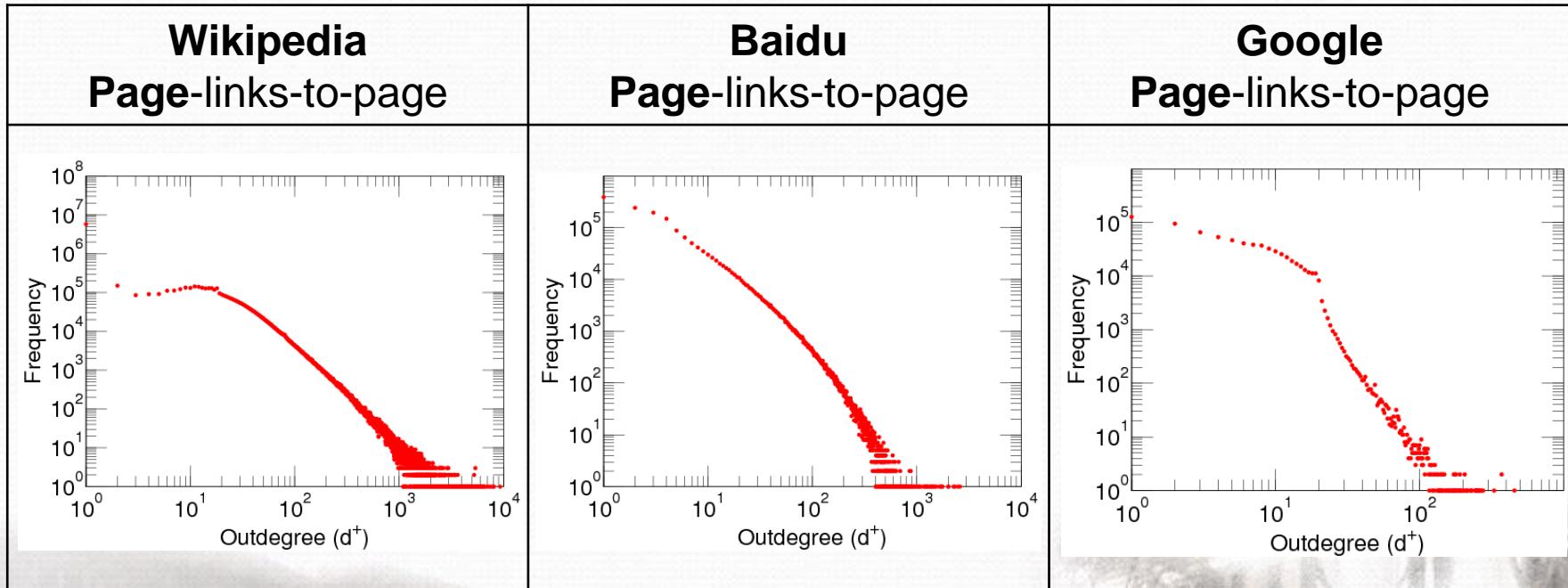
- ❖ Power-law distribution (bipartite graph)





Out-degree Distribution

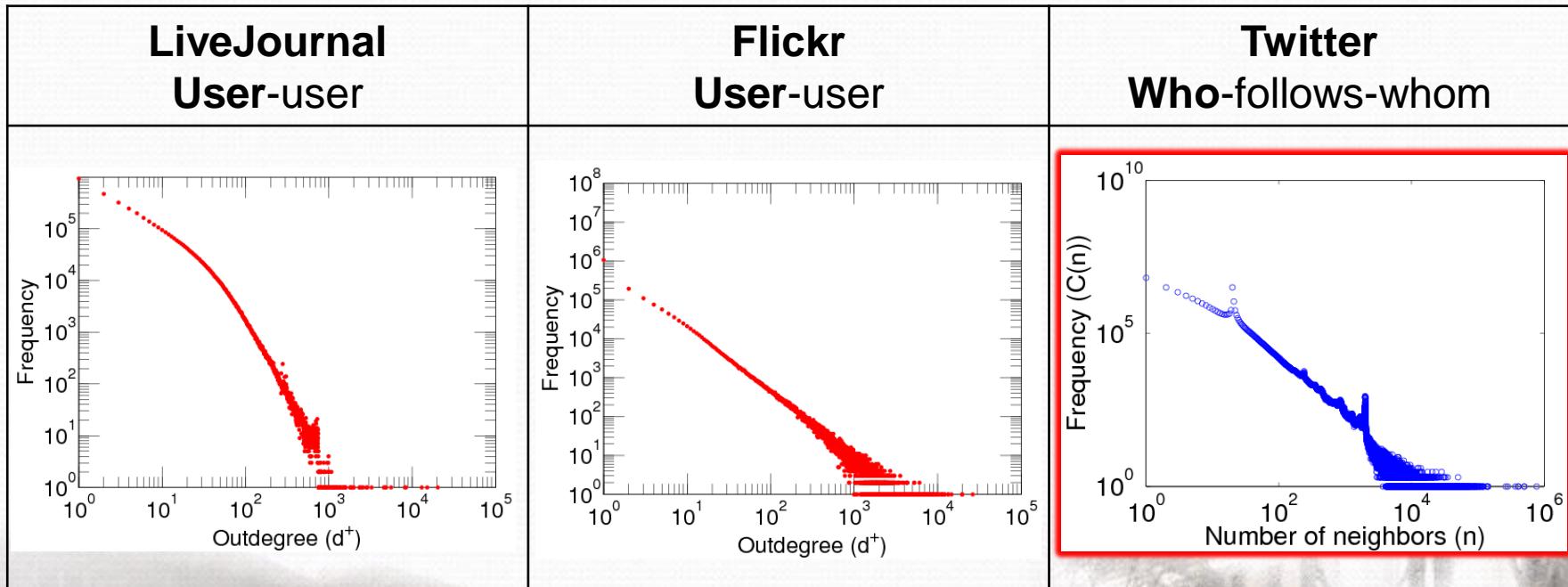
- ❖ Power-law distribution (directed graph)





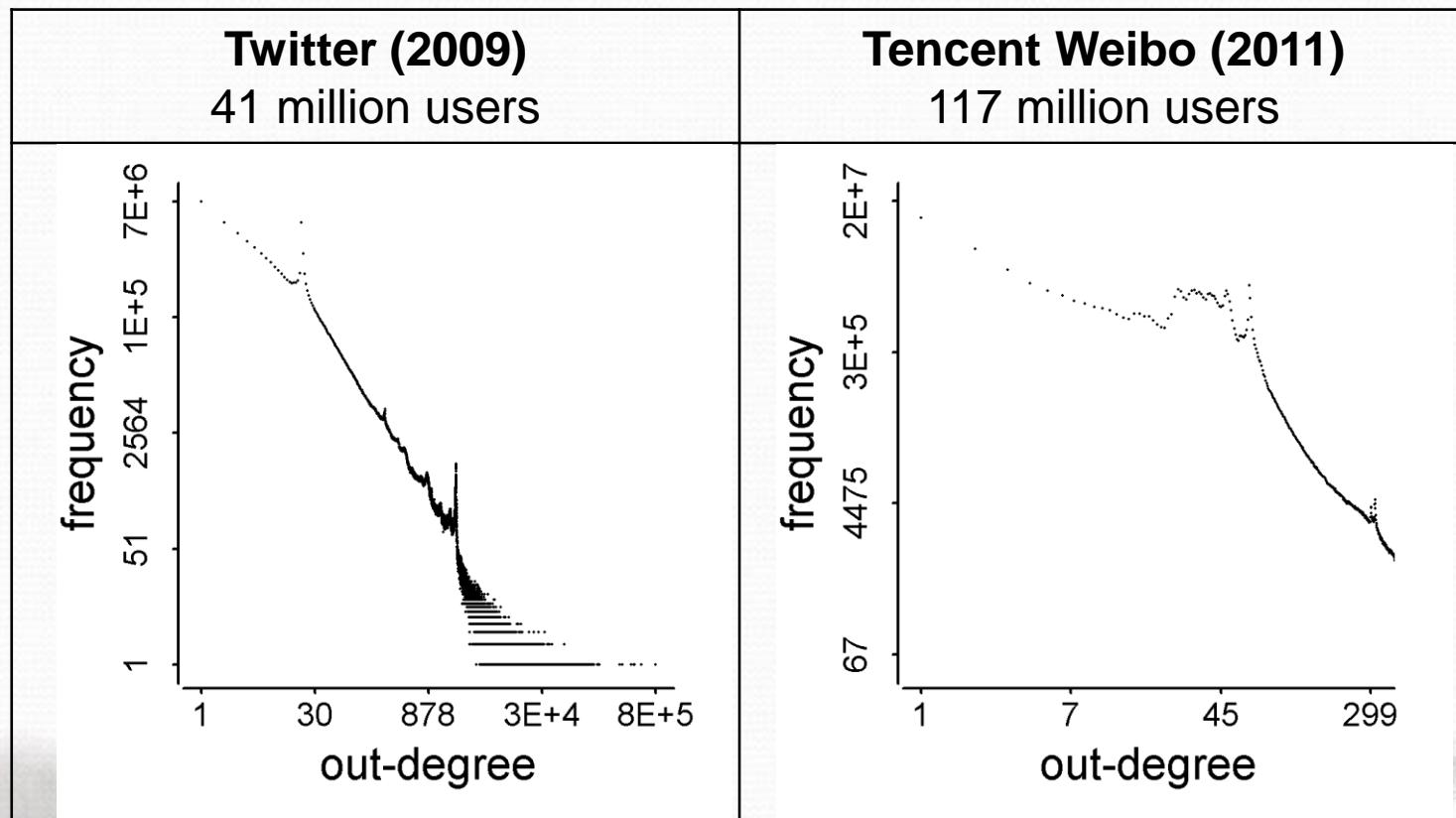
Out-degree Distribution

- ❖ Power-law distribution (directed graph - social network)





What We Have...





Zombie Follower Detection

❖ Challenges

- **Scalability:** How to catch zombie followers from large graphs of millions of nodes and billions of edges? **Can we explain the spikes on out-degree distributions?**



Zombie Follower Detection

❖ Challenges

- Scalability: How to catch zombie followers from large graphs of millions of nodes and billions of edges? Can we explain the spikes on out-degree distributions?
- **Camouflage:** fake profile, no or little content, extra performance



Annalisa Monsod
@AnnalisaMonsodz FOLLOWS YOU
By: Dan Schawbel on November 30th, 2010 at 11:19 am
South Cambridgeshire, UK

17 TWEETS 2,000 FOLLOWING 12 FOLLOWERS

Tweets

Annalisa Monsod @AnnalisaMonsodz 6 Jul
Entrepreneurial Spotlight: Daniel Brusilovsky - Grasshopper Group
Expand

Annalisa Monsod @AnnalisaMonsodz 6 Jul
The only place I see that he's typed it as ?MAC? is in the titles, in
which case all of the letters are capitalized, as it's a title
Expand



Zombie Follower Detection

❖ Challenges

- Scalability: How to catch zombie followers from large graphs of millions of nodes and billions of edges? Can we explain the spikes on out-degree distributions?
- **Camouflage:** fake profile, no or little content, extra performance

The image shows a user profile page for a user named "AjaQwX1Z3". The profile has 0 tweets, 741 following, and 344 followers. The bio reads: "Joined June 2009". Below the bio is a "Tweet to AjaQwX1Z3" button. A "Who to follow" section lists several accounts, including "UPMC Health Plan" (Followed), "Microsoft Store" (Followed), and "Li Zhaonian" (Followed). The main feed displays several tweets from other users. One tweet from "nikhil_parekh" (@nikhil_parekh) discusses being an Internet Marketer, Affiliate Marketer, and Public Speaker. Another tweet from "Anthony Gemma" (@AnthonyGemma) discusses being a President, CEO, and Radio Host. A third tweet from "I Drive Online Backup" (@IDriveBackup) promotes online backup services.

AjaQwX1Z3
@AjaQwX1Z3

0 TWEETS 741 FOLLOWING 344 FOLLOWERS

AjaQwX1Z3
@AjaQwX1Z3

Joined June 2009

Tweet to AjaQwX1Z3

Who to follow · Refresh · View all

UPMC HEALTH PLAN Followed · UPMC Health Plan @UPMCHea... Follow Promoted

Microsoft Store Followed · Microsoft Followed

Li Zhaonian @paulzrn Followed by 欧明栋 and other. Follow

Popular accounts · Find friends

Trends · Change

Bestsell...

FOLLOWING 741 FOLLOWERS 344 More

nikhil_parekh Follow

Internet Marketer,Affiliate Marketer, World Class Public Speaker, Trainer, Coach & Mentor to Thousands

Anthony Gemma Follow

President, CEO, Candidate for US Congress, Radio and TV Talk Show Host, Social Media Guru, Author, Innovator, Creative Marketing...

I Drive Online Backup Follow

Online backup to protect your digital life! 5 GB free starter space. Sign up today at IDrive.com!

Danny Whitehouse Follow

@dannywhitehouse

Mia Rose Follow

@DrMiaRose

Psychologist // Prize-winning author of 'Awaken to Love' // Publisher of Soulwoman eMagazine // Founder of the Soulwoman Sanctuary.

Kevin Cottrell Follow

@kevincottrell

Austin REALTOR(R) with Prudential Texas Realty. Entrepreneur originally from Silicon Valley. Passionate about all things real estate and living...



Zombie Follower Detection

❖ Challenges

- Scalability: How to catch zombie followers from large graphs of millions of nodes and billions of edges? Can we explain the spikes on out-degree distributions?
- **Camouflage:** fake profile, no or little content, extra performance

Buy AB22 Propertwee
@Buy_AB22

0 TWEETS	20 FOLLOWING	2 FOLLOWERS
-------------	-----------------	----------------

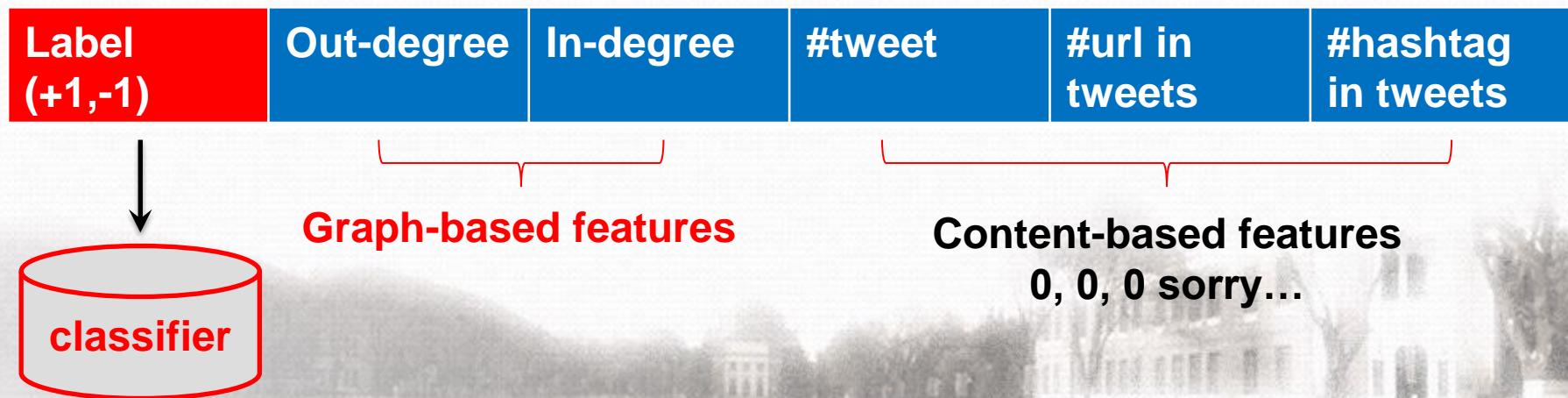
The image shows a grid of Twitter profiles. From top-left to bottom-right:
1. Buy AB22 Propertwee (@Buy_AB22): 0 tweets, 20 following, 2 followers.
2. B.J. Mendelson (@BJMendelson): Following 20, Followers 2. Bio: "I just post silly stuff that I found on the Internet. None of it mine."
3. someecards (@someecards): Following 20, Followers 2. Bio: "Welcome to the Twitter feed of somewhat acclaimed humor sites, Someecards.com and HappyPlace.com. You'll love not unfollowing us!"
4. Steven Johnson (@stevenjohnson): Following 20, Followers 2. Bio: "Author, (Latest: Future Perfect) TV host, (How We Got To Now, on PBS/BBC soon) Startup creator, (FEED, outside.in) Dad. (Three boys...)".
5. adventuregirl (@adventuregirl): Following 20, Followers 2. Bio: "Hi Everyone! I'm Stef Michaels- an avid lifestyles journalist, TV personality, adventurer. Co-founder of KEEN Digital Summit. Contributor to Yahoo! Travel."
6. People magazine (@PeopleMag): Following 20, Followers 2. Bio: "PEOPLE.com is the No. 1 site for celebrity news! Tweet your questions to our customer service team @Peoplemag_Help".
7. ashton kutcher (@aplusk): Following 20, Followers 2. Bio: "I make stuff, actually I make up stuff, stories mostly, collaborations of thoughts, dreams, and actions. Thats me."
8. Paul Pierce (@paulierce34): Following 20, Followers 2. Bio: "The one and only Truth. Founder of The @TruthonHealth".
9. Mashable (@mashable): Following 20, Followers 2. Bio: "News, resources, inspiration and fun for the connected generation. Tweets by".

Zombie Follower Detection

❖ Challenges

- Scalability: How to catch zombie followers from large graphs of millions of nodes and billions of edges? Can we explain the spikes on out-degree distributions?
- Camouflage: fake profile, no or little content, extra performance

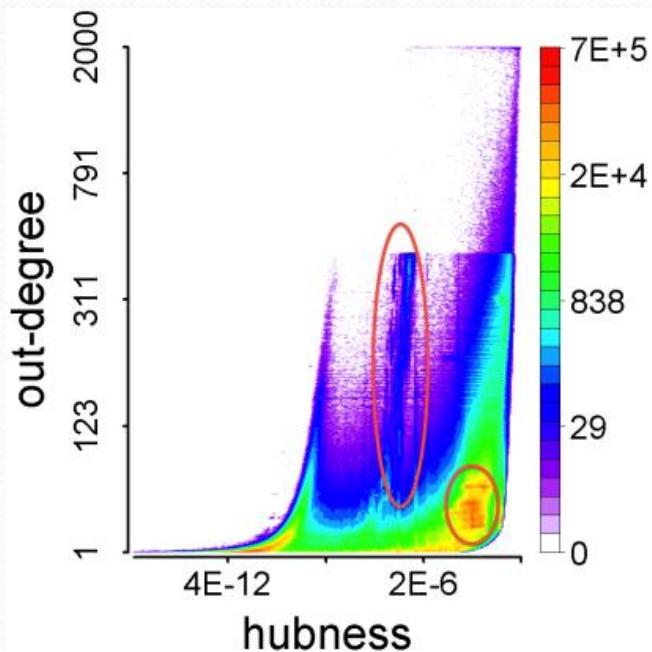
❖ Previous approaches



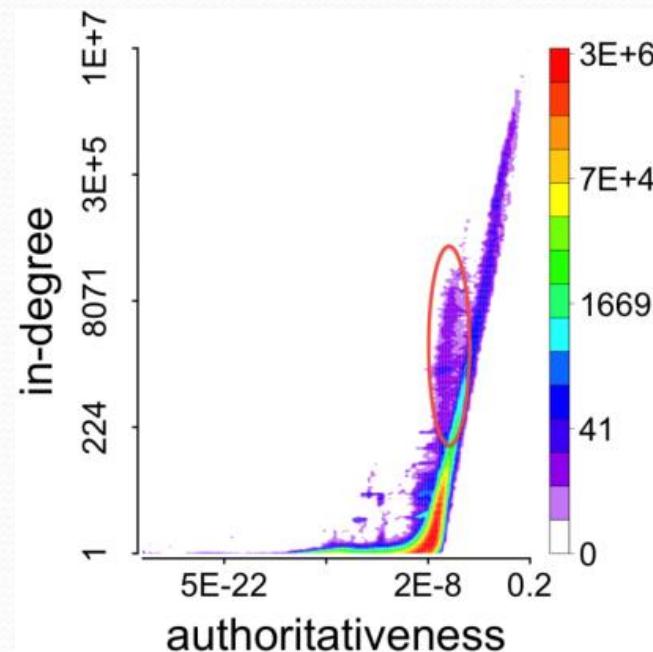


Graph-based Feature Space

Followers



Followees

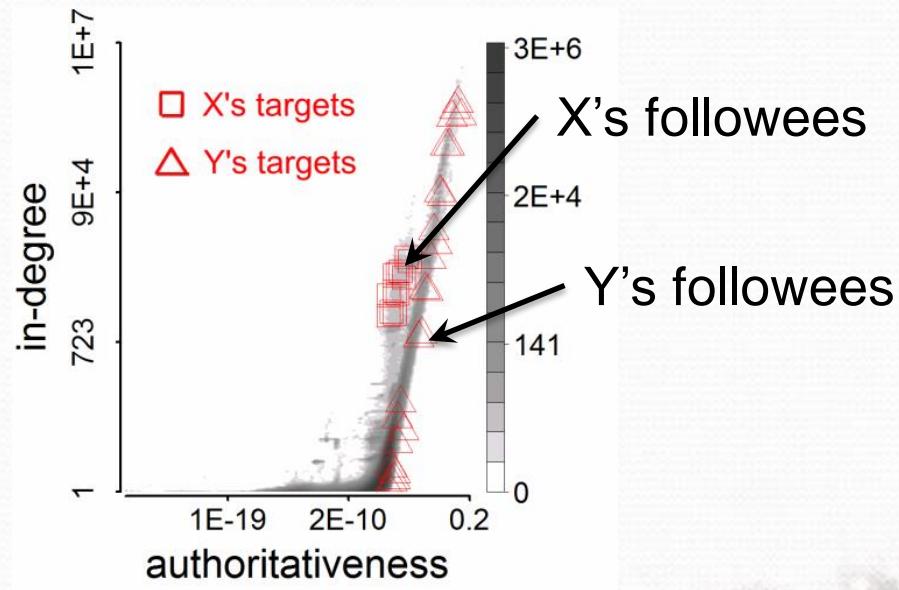
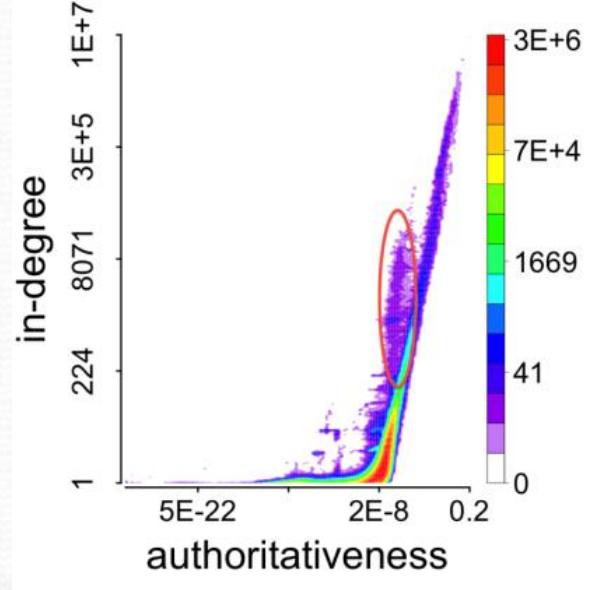


We spot some outliers, but are they suspicious?
Answer: **suspicious behavior patterns** create
suspicious graph structure



Compare Zombie Follower and Normal User

- ❖ X = @Buy_AB22: a zombie follower with 20 followees
- ❖ Y = a random user with 20 followees



- ❖ Suspicious behavior: similar with each other, different from normal.



Suspicious Behavior Pattern

- ❖ Behavior-based features
- ❖ Synchronicity: similarity between one's followees

$$sync(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{F}(u)} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times d(u)}$$

- ❖ Normality: similarity between one's followee and everyone

$$norm(u) = \frac{\sum_{(v,v') \in \mathcal{F}(u) \times \mathcal{U}} \mathbf{p}(v) \cdot \mathbf{p}(v')}{d(u) \times N}$$

Feature vector

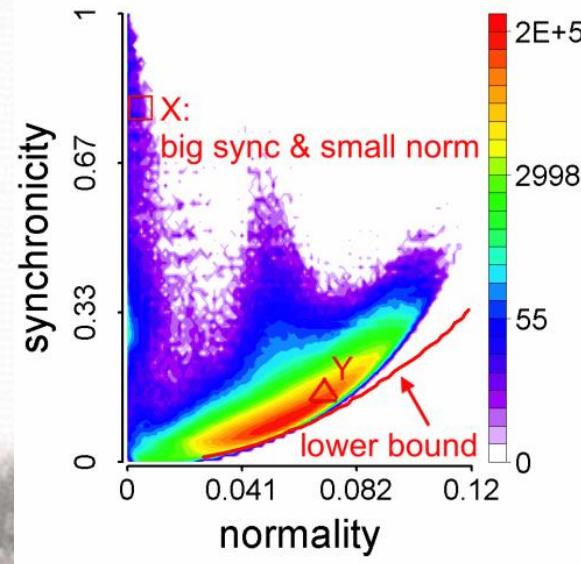
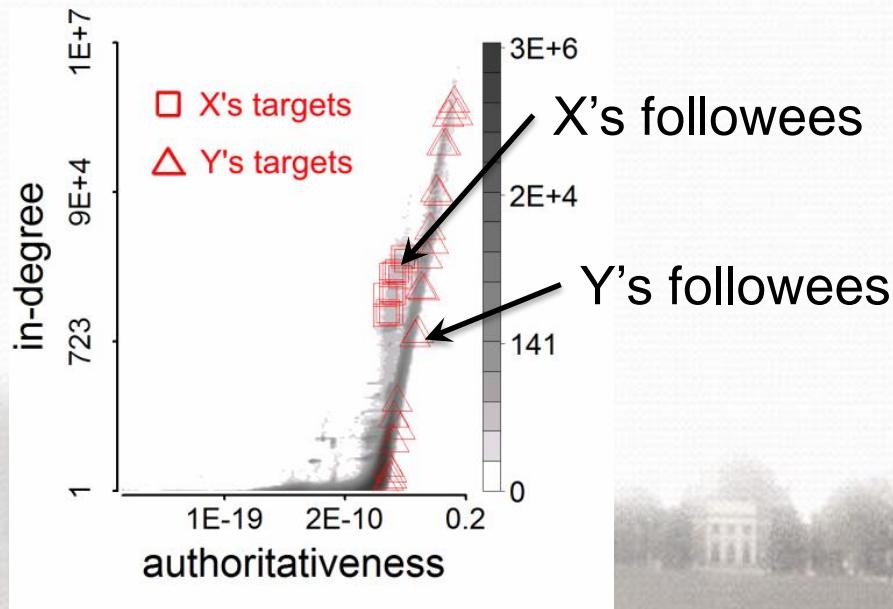
- ❖ **Theorem** *For any distribution, there is a parabolic lower limit in the synchronicity-normality plot.*

$$s_{min} = (-Mn^2 + 2n - s_b)/(1 - Ms_b)$$



Synchronicity-Normality Plot

- ❖ Use *CatchSync* to catch Zombie follower
 - **Synchronized** and **abnormal** behavior
 - **Big synchronicity** and **small normality**; away from the densest
 - Last step: develop a distance-based anomaly detection method!



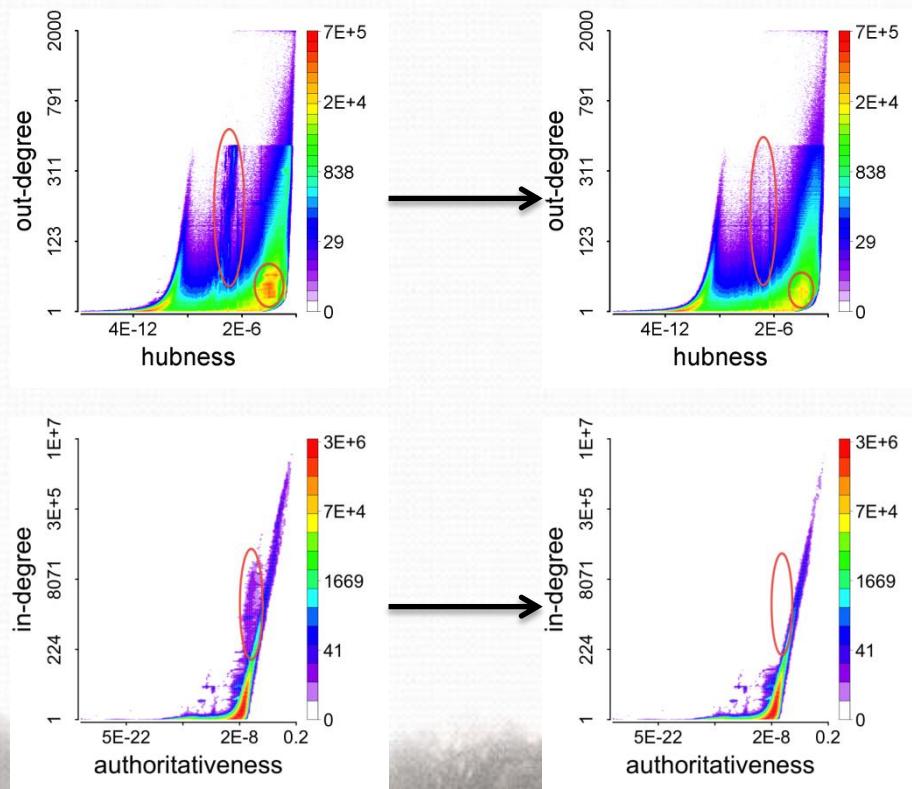


Goal 1: Do we catch the anomalies?

❖ Graph-based feature space

❖ Followers

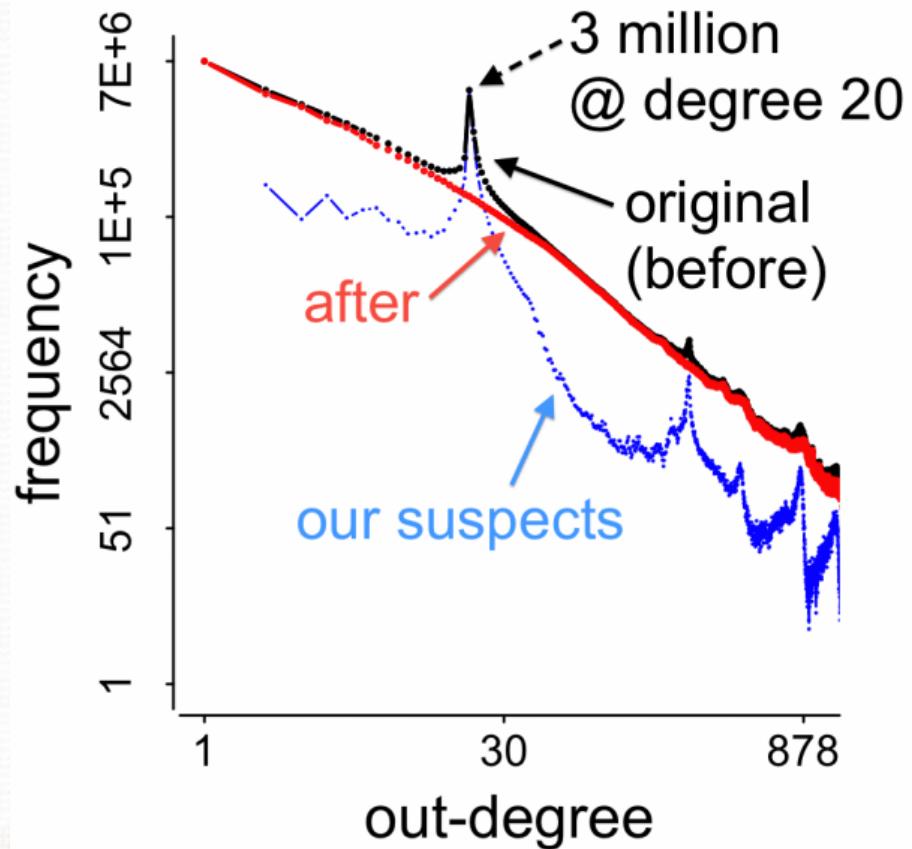
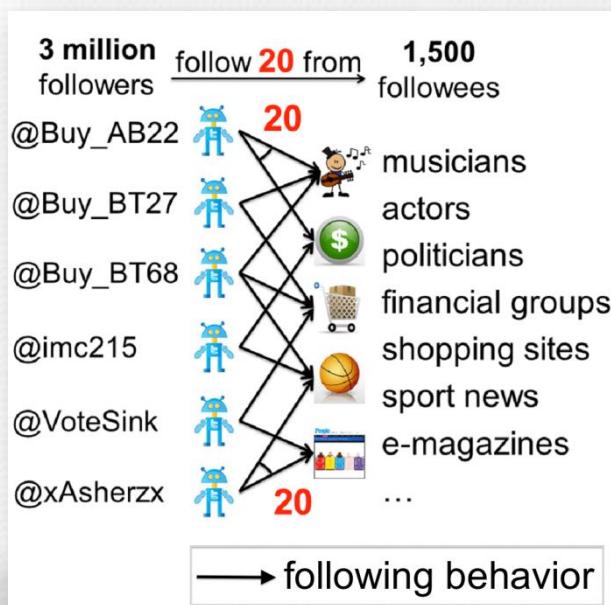
❖ Followees





Goal 1: Do we catch the anomalies?

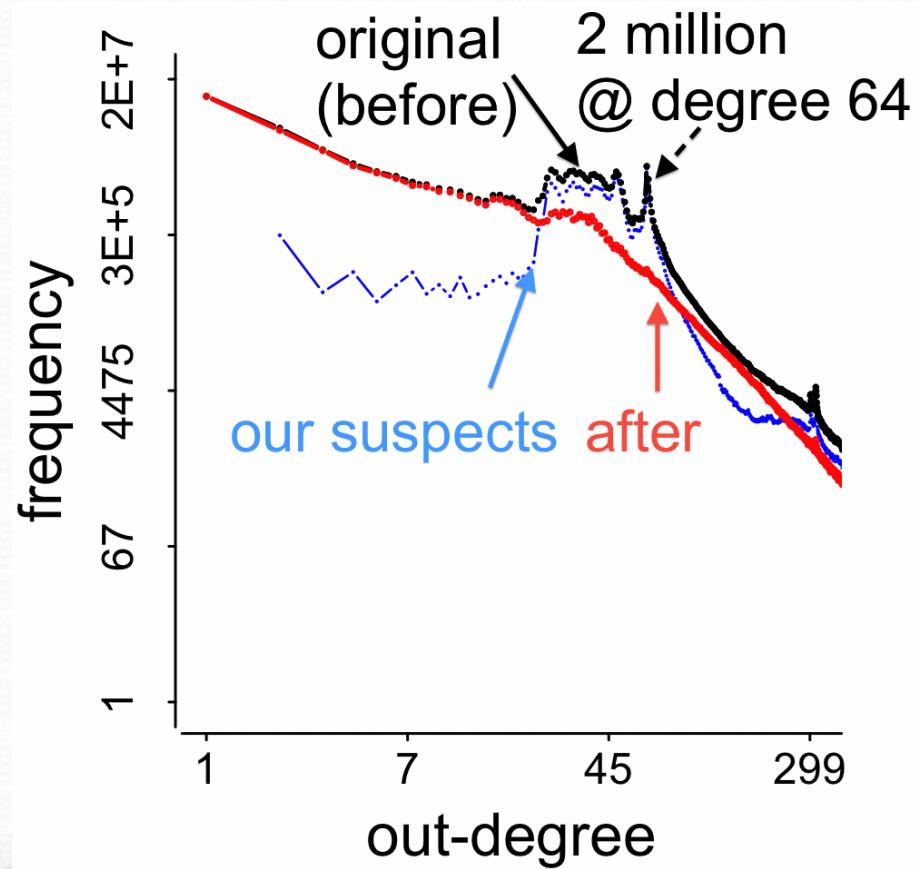
- ❖ Out-degree distribution
- ❖ Twitter





Goal 1: Do we catch the anomalies?

- ❖ Out-degree distribution
- ❖ Tencent Weibo



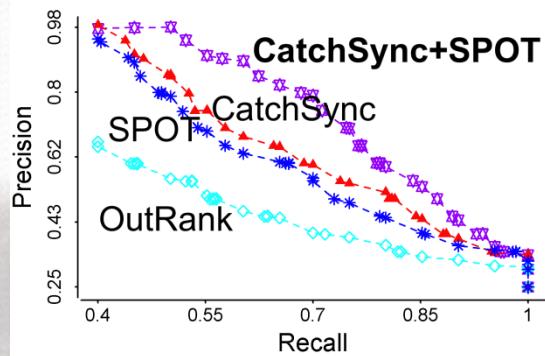


Goal 2: Do we catch suspicious users?

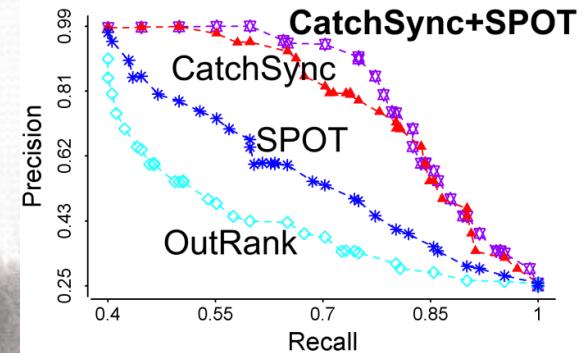
- ❖ Manually label 1,000 users as suspicious or not by reading their pages (profiles, tweets, followers, followings, etc.)
- ❖ Better than content-based method SPOT
- ❖ **Complementary** with SPOT

	TWITTERSG	WEIBOSG
CATCHSYNC	0.751	0.694
OUTRANK	0.412	0.377
SPOT	0.597	0.653
CATCHSYNC+SPOT	0.813	0.785

Twitter



Tencent Weibo

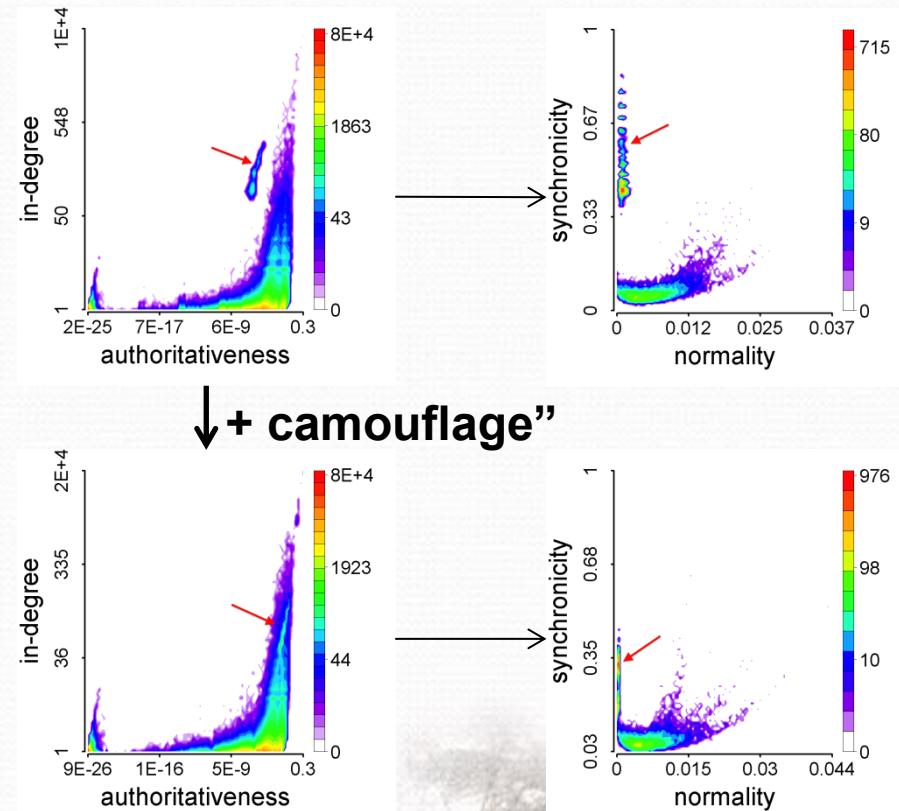




Goal 3: Is CatchSync anti-camouflage?

❖ Camouflage

- +10%, +50%
- +random, +popular



↓ + camouflage”

Hide in feature space

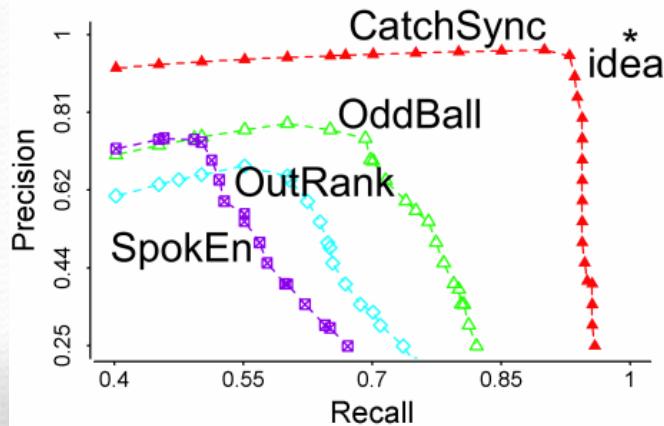
Show in SN-plot

Goal 3: Is CatchSync anti-camouflage?

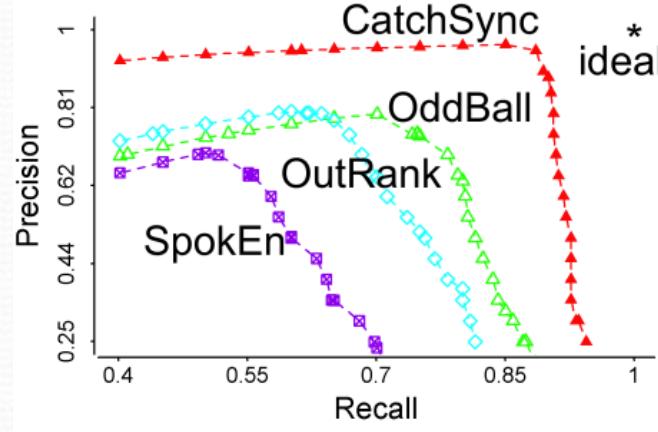
Accuracy:

Synthetic graph	SYNTH-3M	SYNTH-3M-RAND	SYNTH-3M-POP	
Camouflage (d_{camou})	None (0)	10%	50%	10% 50%
CATCHSYNC	0.956	0.910	0.764	0.885 0.792
ODDBALL	0.755	0.702	0.525	0.657 0.433
OUTRANK	0.725	0.678	0.516	0.694 0.392
SPOKEN	0.677	0.586	0.470	0.553 0.351

Random “Camouflage”:



Popular “Camouflage”:



Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos and Shiqiang Yang. CatchSync: Catching Synchronized Behavior in Large Directed Graphs. *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.

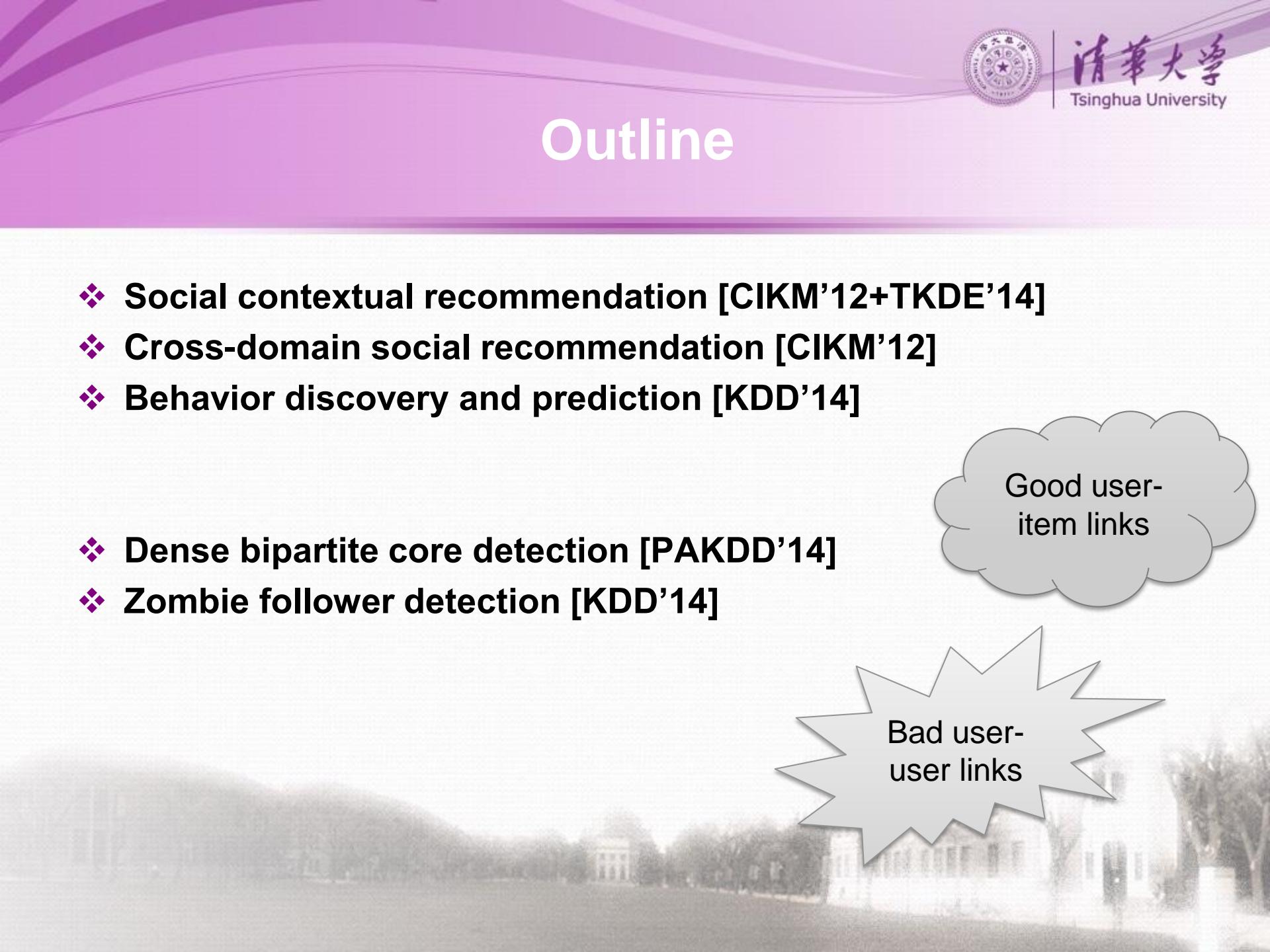
Meng Jiang, Peng Cui, Alex Beutel, Christos Faloutsos and Shiqiang Yang. Detecting Suspicious Following Behavior in Multimillion-Node Social Networks. *The 23rd international conference on World Wide Web companion (WWW)*, 2014. (Poster)



Outline

- ❖ Social contextual recommendation [CIKM'12+TKDE'14]
- ❖ Cross-domain social recommendation [CIKM'12]
- ❖ Behavior discovery and prediction [KDD'14]

- ❖ Dense bipartite core detection [PAKDD'14]
- ❖ Zombie follower detection [KDD'14]



Good user-item links

Bad user-user links



References

- ❖ **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos and Shiqiang Yang. CatchSync: Catching Synchronized Behavior in Large Directed Graphs. *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.
- ❖ **Meng Jiang**, Peng Cui, Fei Wang, Xinran Xu, Wenwu Zhu and Shiqiang Yang. FEMA: Flexible Evolutionary Multi-faceted Analysis for Dynamic Behavioral Pattern Discovery. *The 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2014.
- ❖ **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos and Shiqiang Yang. Inferring Strange Behavior from Connectivity Pattern in Social Networks. *The 18th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2014.
- ❖ **Meng Jiang**, Peng Cui, Alex Beutel, Christos Faloutsos and Shiqiang Yang. Detecting Suspicious Following Behavior in Multimillion-Node Social Networks. *The 23rd international conference on World Wide Web companion (WWW)*, 2014. (Poster)
- ❖ **Meng Jiang**, Peng Cui, Fei Wang, Wenwu Zhu and Shiqiang Yang. Scalable Recommendation with Social Contextual Information. *IEEE Transactions on Knowledge and Data Engineering (TKDE)*, 2014.
- ❖ **Meng Jiang**, Peng Cui, Rui Liu, Qiang Yang, Fei Wang, Wenwu Zhu and Shiqiang Yang. Social Contextual Recommendation. *The 21st ACM International Conference on Information and Knowledge Management (CIKM)*, 2012.
- ❖ **Meng Jiang**, Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu and Shiqiang Yang. Social Recommendation across Multiple Relational Domains. *The 21st ACM International Conference on Information and Knowledge Management (CIKM)*, 2012.
- ❖ Lu Liu, Feida Zhu, **Meng Jiang**, Jiawei Han, Lifeng Sun, Shiqiang Yang. Mining Diversity on Social Media Networks. *Multimedia Tools and Applications*, 2012.
- ❖ Lu Liu, Jie Tang, Jiawei Han, **Meng Jiang**, Shiqiang Yang. Mining Topic-Level Influence in Heterogeneous Networks. *The 19th ACM International Conference on Information and Knowledge Management (CIKM)*, 2010.



Acknowledgements

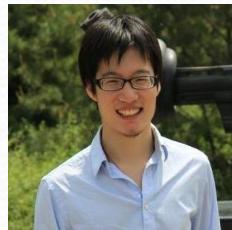
❖ Tsinghua University



Shiqiang Yang



Wenwu Zhu



Peng Cui



Lu Liu

❖ Carnegie Mellon University



Christos Faloutsos



Alex Beutel

❖ IBM T. J. Watson Research Center



Fei Wang



Thank you!

Welcome to visit my homepage:
<http://www.meng-jiang.com>

Top Work Publication & Demo Contact



Hi. I'm **Meng Jiang** (蒋朦).

I am a 4th-year Ph.D. candidate in Department of Computer Science and Technology, Tsinghua University. My advisor is Professor Shiqiang Yang (杨士强) and my research interests lie in **data mining**, **behavior analysis** and **social network analysis**. Here you can download my [CV](#) and [resume](#).

♥ New friends ♥ Discussions ♥ Collaborations