WWW2016
MONTRÉAL • CANADA

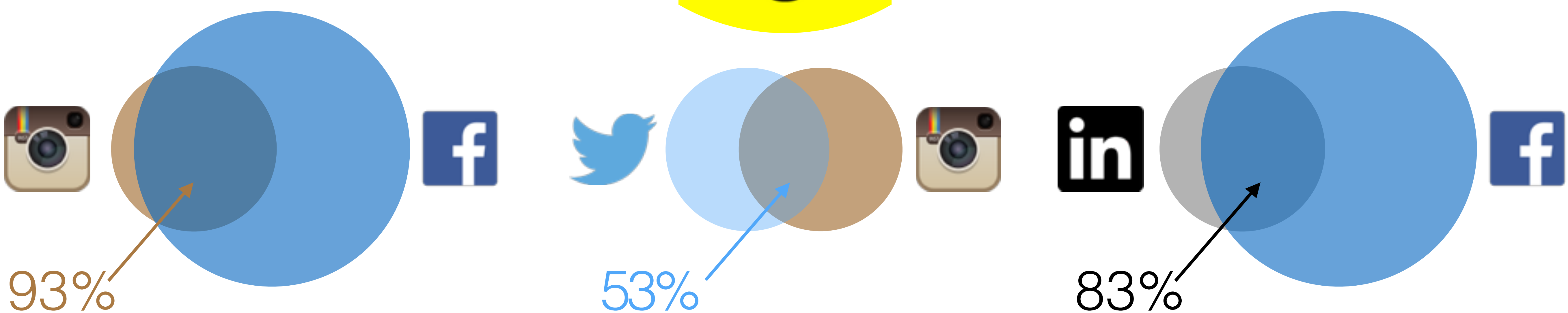
PCT: Partial Co-Alignment of Social Networks

Jiawei Zhang[1] and Philip S. Yu[1,2]

[1] University of Illinois at Chicago, Chicago, IL, USA
[2] Tsinghua University, Beijing, China

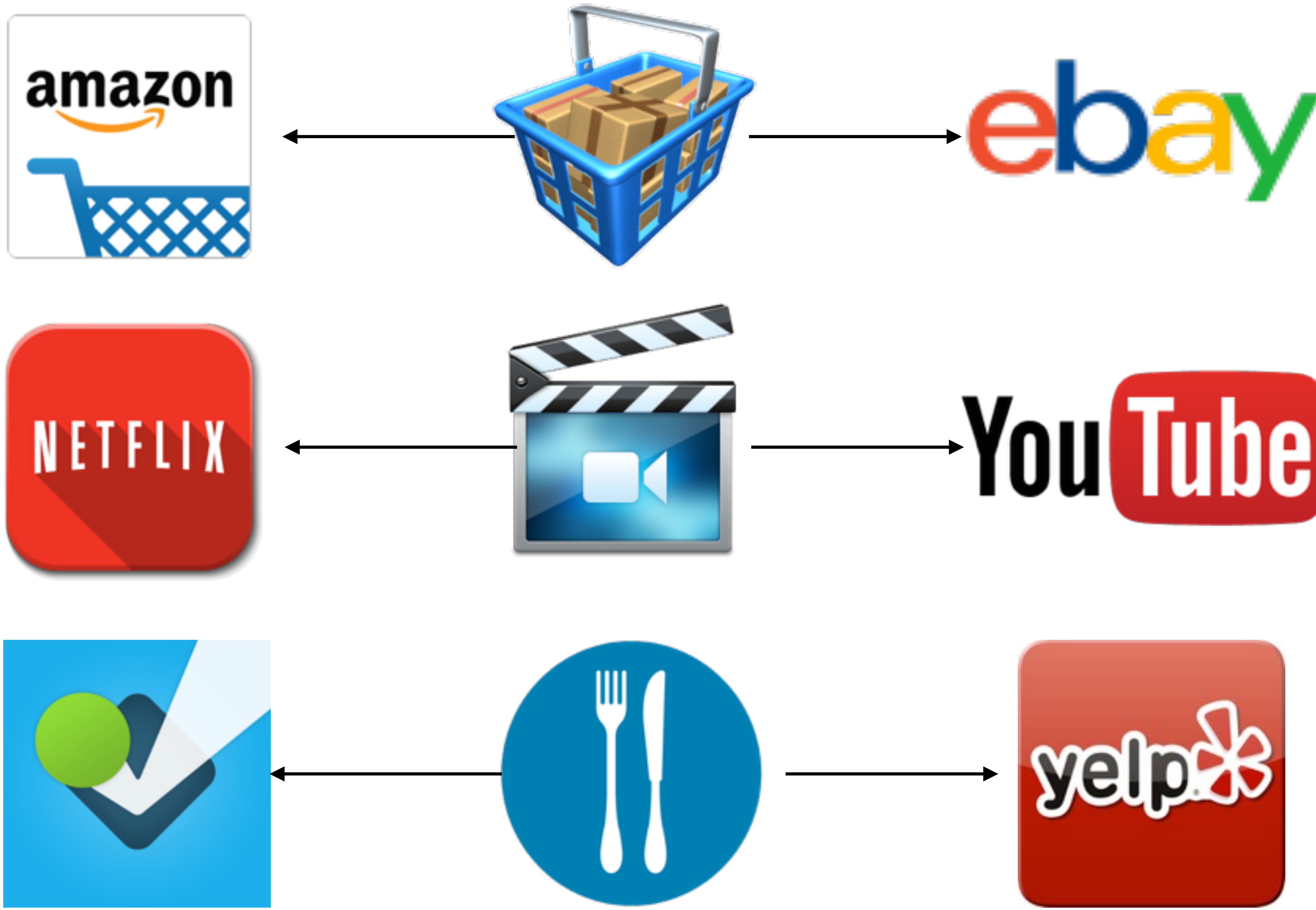
UIC UNIVERSITY OF ILLINOIS AT CHICAGO

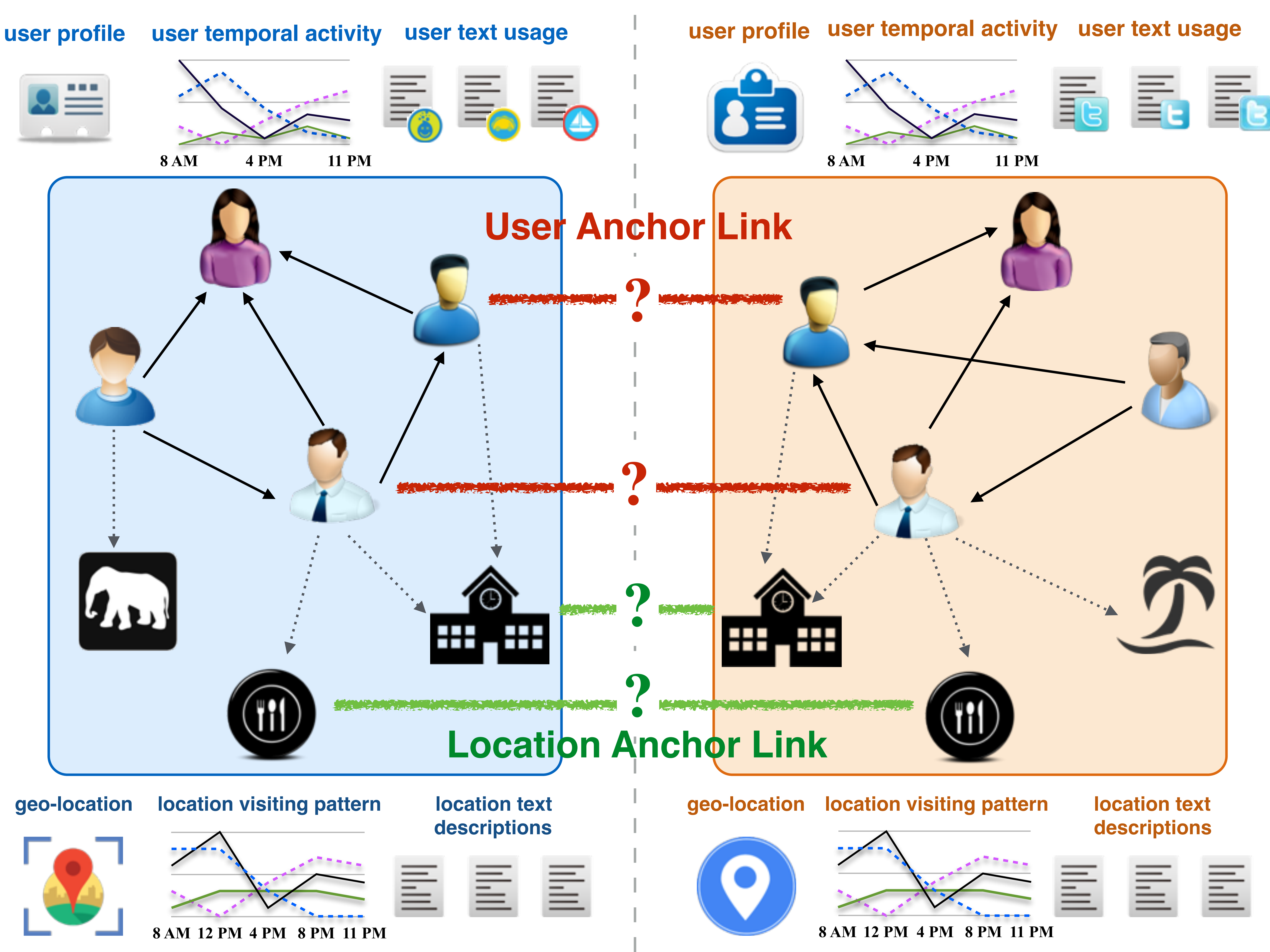
People are using multiple social networks simultaneously nowadays



[1] Zhang et al. PNA: Partial Network Alignment with Generic Stable Matching, 2015 IEEE IRI.
[2] Duggan et al. Social media update 2013.

Other information entities appear in multiple sites concurrently



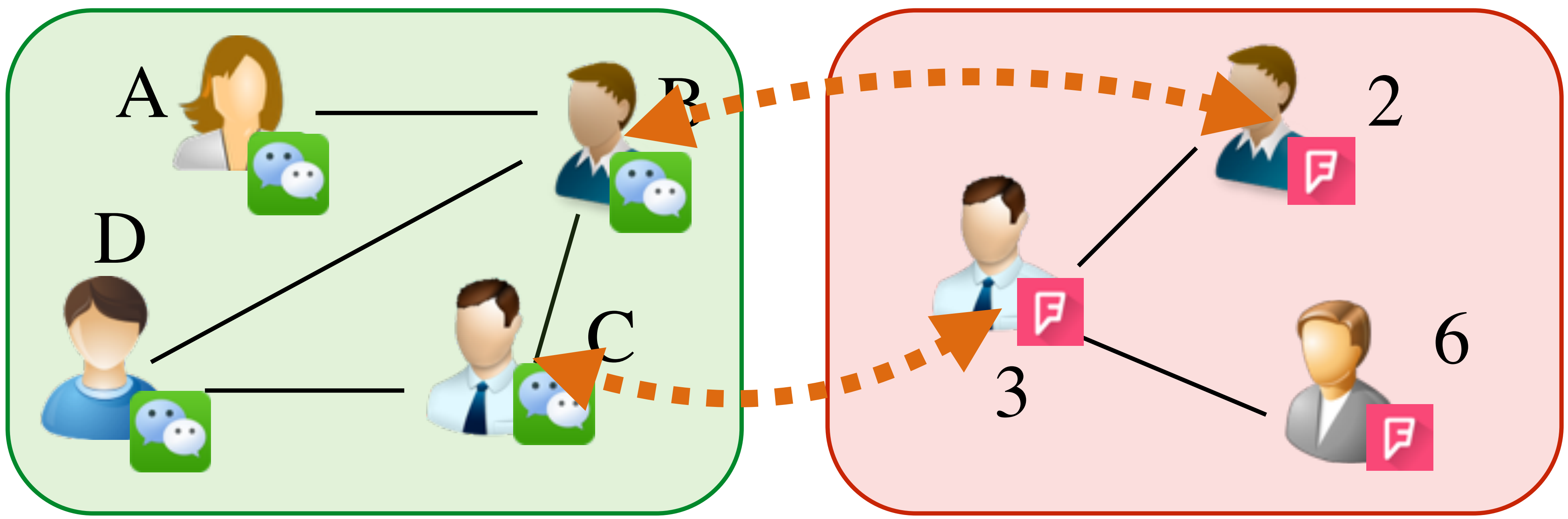


user profile
user temporal activity
user text usage
8 AM
4 PM
11 PM
user profile
user temporal activity
user text usage
8 AM
4 PM
11 PM
User Anchor Link
?
?
?
?
Location Anchor Link
geo-location
location visiting pattern
location text descriptions
8 AM 12 PM 4 PM 8 PM 11 PM
geo-location
location visiting pattern
location text descriptions
8 AM 12 PM 4 PM 8 PM 11 PM

Challenge 1: Heterogeneity of Social Networks

Heterogeneous **Link** and **Attribute** Information

- User Anchor Link Inference with **Link** Information



	A	B	C	D
A	**0**	**1**	**0**	**0**
B	**1**	**0**	**1**	**1**
C	**0**	**1**	**0**	**1**
D	**0**	**1**	**1**	**0**

Adjacency Matrix $\mathbf{S}^{(1)}$

	2	3	6
A	**0**	**0**	**0**
B	**1**	**0**	**0**
C	**0**	**1**	**0**
D	**0**	**0**	**0**

Transition Matrix $\mathbf{P}$

	2	3	6
2	**0**	**1**	**0**
3	**1**	**0**	**1**
6	**0**	**1**	**0**

Adjacency Matrix $\mathbf{S}^{(2)}$

User Anchor Link Inference with **Link** Information

Assumption: shared users have similar social structures in different networks

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1 & 0 & 0 \\ B & 1 & 0 & 1 & 1 \\ C & 0 & 1 & 0 & 1 \\ D & 0 & 1 & 1 & 0 \end{array} \qquad \begin{array}{c|ccc} & 2 & 3 & 6 \\ \hline A & 0 & 0 & 0 \\ B & 1 & 0 & 0 \\ C & 0 & 1 & 0 \\ D & 0 & 0 & 0 \end{array} \qquad \begin{array}{c|ccc} & 2 & 3 & 6 \\ \hline 2 & 0 & 1 & 0 \\ 3 & 1 & 0 & 1 \\ 6 & 0 & 1 & 0 \end{array}$$

Adjacency Matrix $\mathbf{S}^{(1)}$ Transition Matrix $\mathbf{P}$ Adjacency Matrix $\mathbf{S}^{(2)}$

Via transition matrix **P** (i.e., anchor links), we can map the social connections among shared users from network I to network II:

$$\mathbf{P}^{\top} \mathbf{S}^{(1)} \mathbf{P}$$

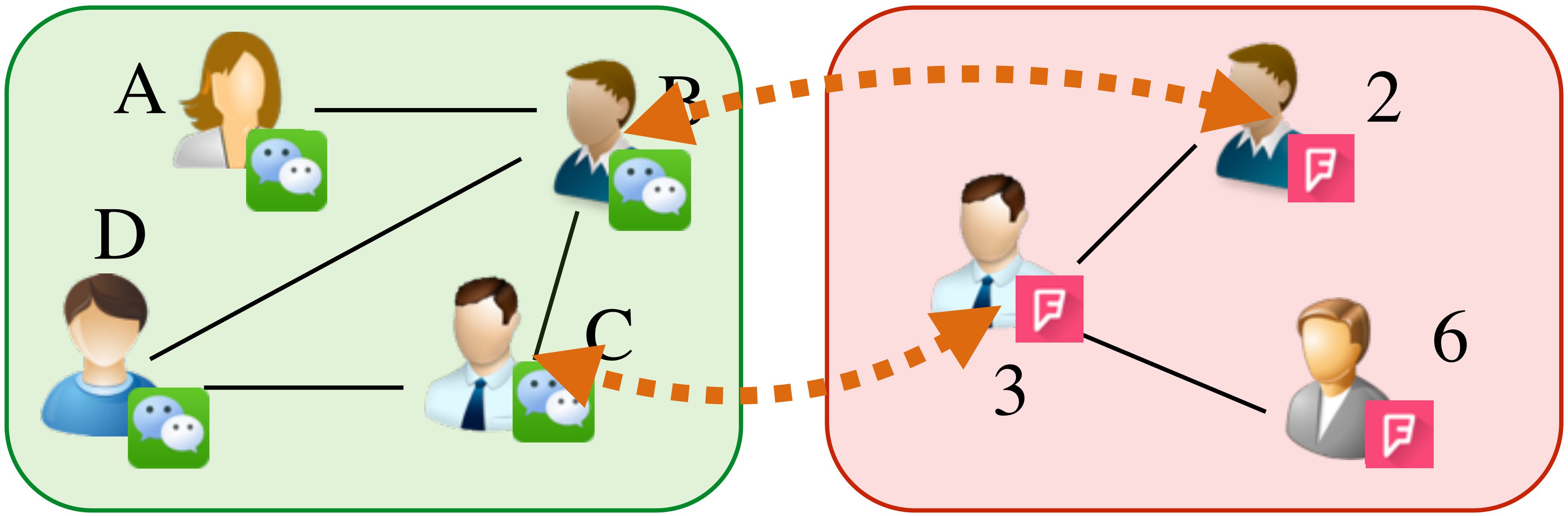
The optimal transition matrix **P** (i.e., anchor links) should minimize the mapping cost

$$\min \left\| \mathbf{P}^{\top} \mathbf{S}^{(1)} \mathbf{P} - \mathbf{S}^{(2)} \right\|_F^2$$

Challenge 1: Heterogeneity of Social Networks

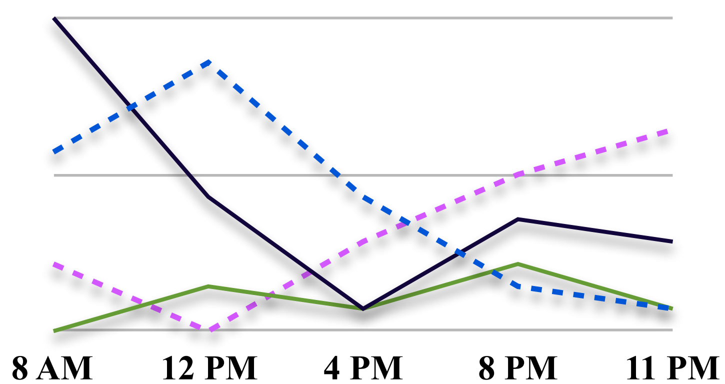
Heterogeneous **Link** and **Attribute** Information

- User Anchor Link Inference with **Attribute** Information



user profile

user temporal activity



user text usage

cross-network user similarity measures:

Name: $\frac{|n(u_i^{(1)}) \cap n(u_l^{(2)})|}{|n(u_i^{(1)}) \cup n(u_l^{(2)})|}$ Time: $\mathbf{t}(u_i^{(1)})^\top \mathbf{t}(u_l^{(2)})$ Word: $\frac{\mathbf{w}(u_i^{(1)})^\top \cdot \mathbf{w}(u_l^{(2)})}{\left\|\mathbf{w}(u_i^{(1)})\right\| \cdot \left\|\mathbf{w}(u_l^{(2)})\right\|}$

User Anchor Link Inference with **Attribute** Information

Assumption: shared users have similar attribute information in different networks

user similarity = (name_sim + time_sim + text_sim)/3

	2	3	6
A	0	0	0
B	1	0	0
C	0	1	0
D	0	0	0

Transition Matrix **P**

	2	3	6
A	1/3	2/3	1
B	1	1/3	0
C	2/3	1	0
D	0	1/3	2/3

Similarity Matrix $\mathbf{\Lambda}$

The optimal transition matrix **P** (i.e., anchor links) should maximize the mapped user similarities

$$\max \|\mathbf{P} \circ \mathbf{\Lambda}\|_1$$

Challenge 1: Heterogeneity of Social Networks

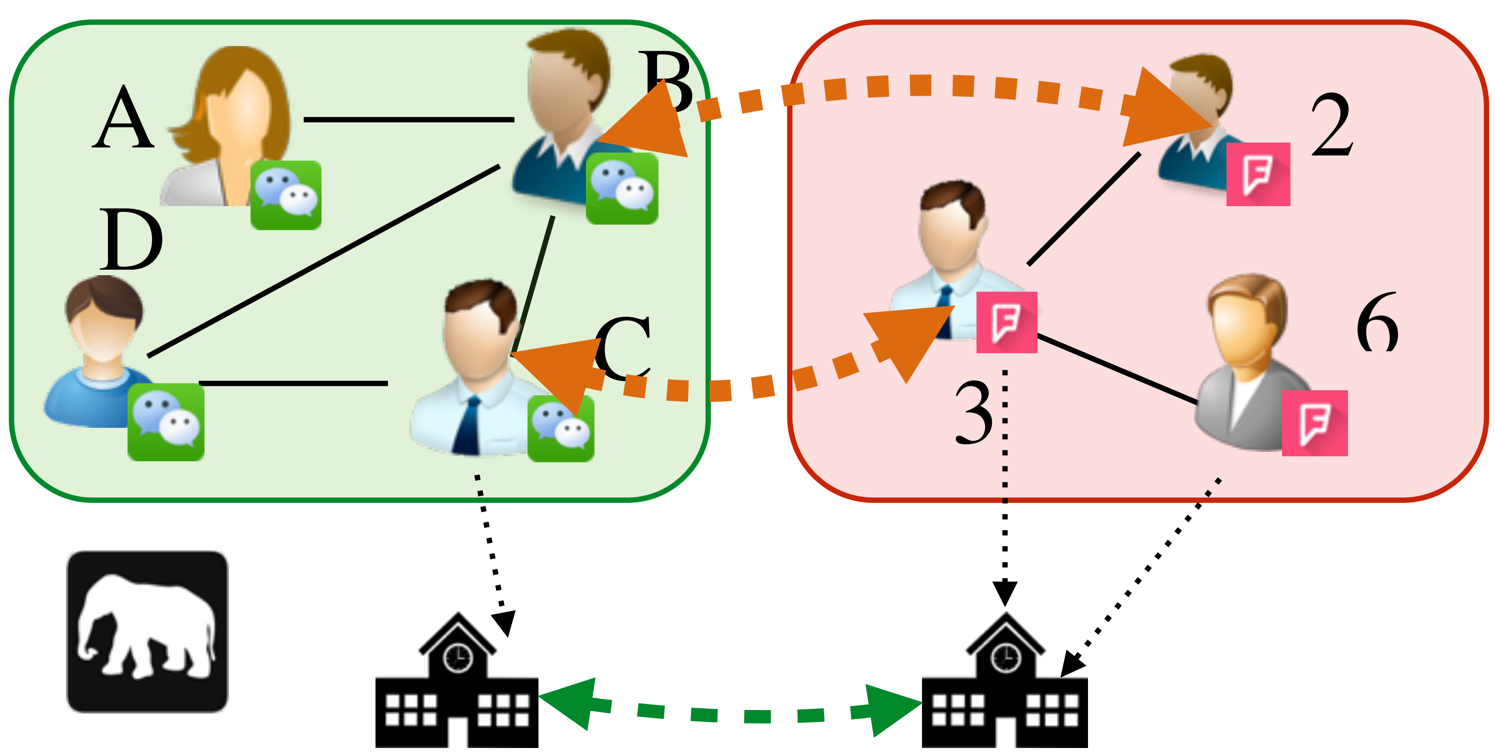
Heterogeneous **Link** and **Attribute** Information

- User Anchor Link Inference with **Link** and **Attribute** information

$$\arg\min_{\mathbf{P}} \left\| \mathbf{P}^\top \mathbf{S}^{(1)} \mathbf{P} - \mathbf{S}^{(2)} \right\|_F^2 - \alpha \cdot \left\| \mathbf{P} \circ \mathbf{\Lambda} \right\|_1$$

- Similarly, Location Anchor Link Inference with **Link** and **Attribute** information:

$$\arg\min_{\mathbf{P},\mathbf{Q}} \left\| \mathbf{P}^\top \mathbf{L}^{(1)} \mathbf{Q} - \mathbf{L}^{(2)} \right\|_F^2 - \alpha \cdot \left\| \mathbf{Q} \circ \mathbf{\Theta} \right\|_1$$



Challenge 1: Heterogeneity of Social Networks

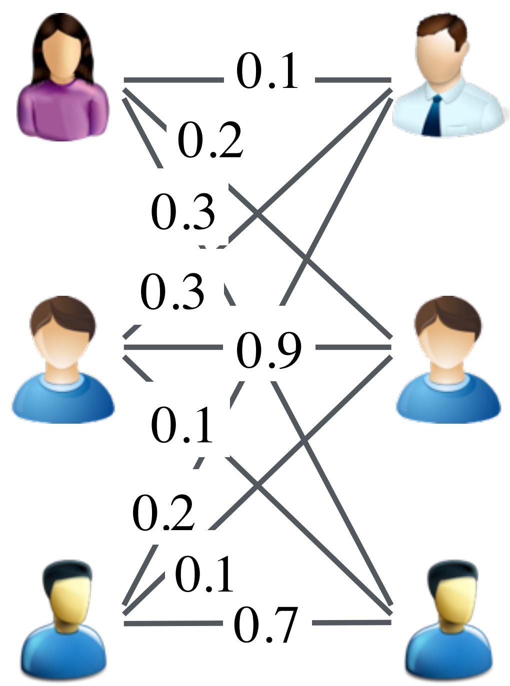
Heterogeneous Link and Attribute Information

$$
\begin{aligned}
\mathbf{P}^*, \mathbf{Q}^* = \arg\min_{\mathbf{P},\mathbf{Q}} & \left\| \mathbf{P}^\top \mathbf{S}^{(1)} \mathbf{P} - \mathbf{S}^{(2)} \right\|_F^2 + \left\| \mathbf{P}^\top \mathbf{L}^{(1)} \mathbf{Q} - \mathbf{L}^{(2)} \right\|_F^2 \\
& - \alpha \cdot \left\| \mathbf{P} \circ \mathbf{\Lambda} \right\|_1 - \alpha \cdot \left\| \mathbf{Q} \circ \mathbf{\Theta} \right\|_1 , \\
s.t. \quad & \mathbf{P} \in \{0,1\}^{|\mathcal{U}^{(1)}| \times |\mathcal{U}^{(2)}|}, \mathbf{Q} \in \{0,1\}^{|\mathcal{L}^{(1)}| \times |\mathcal{L}^{(2)}|}, \\
& \mathbf{P}\mathbf{1}^{|\mathcal{U}^{(2)}| \times 1} \leq \mathbf{1}^{|\mathcal{U}^{(1)}| \times 1}, \mathbf{P}^\top \mathbf{1}^{|\mathcal{U}^{(1)}| \times 1} \leq \mathbf{1}^{|\mathcal{U}^{(2)}| \times 1}, \\
& \mathbf{Q}\mathbf{1}^{|\mathcal{L}^{(2)}| \times 1} \leq \mathbf{1}^{|\mathcal{L}^{(1)}| \times 1}, \mathbf{Q}^\top \mathbf{1}^{|\mathcal{L}^{(1)}| \times 1} \leq \mathbf{1}^{|\mathcal{L}^{(2)}| \times 1}.
\end{aligned}
$$

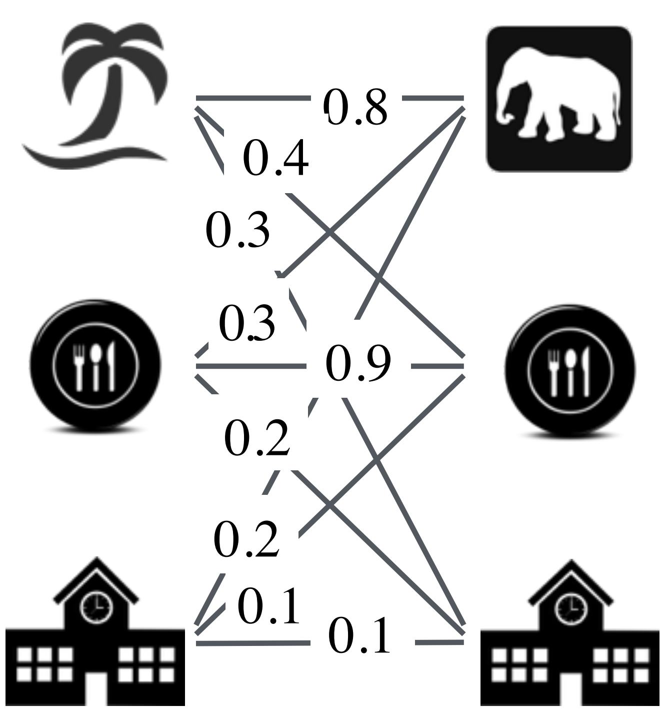
- Hard 0-1 programming problem, very challenging to address
- Relax the hard 0-1 constraint, **P** and **Q** can take real values in range [0, 1]
- These introduced redundant user/location anchor links will be pruned with a network matching post-processing step

Challenge 2: Redundant Link Pruning with Network Flow based Co-Matching

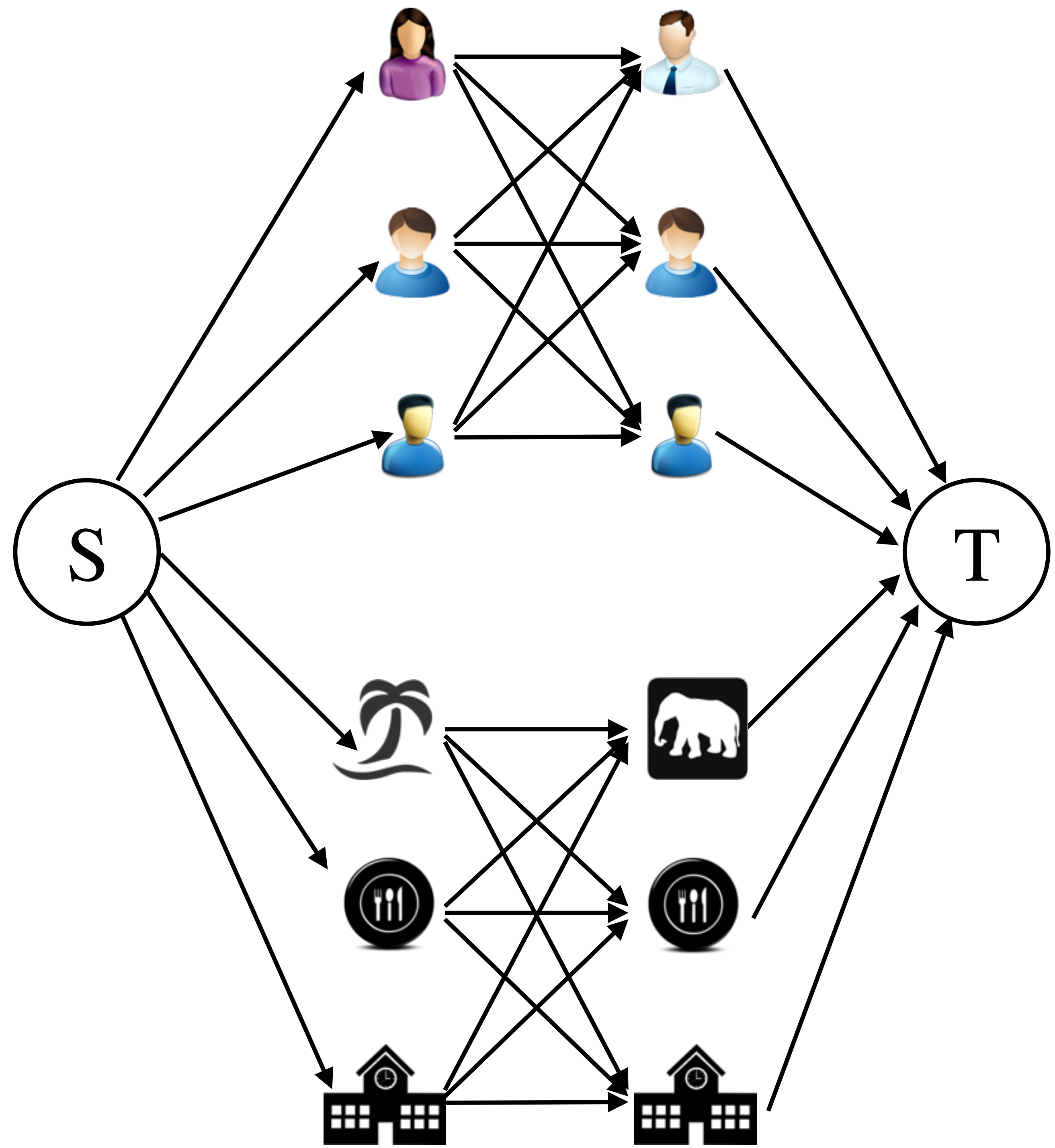
User Preference Bipartite Graphs



Location Preference Bipartite Graphs



Co-Matching Network Flow Graph



Challenge 2: Redundant Link Pruning with Network Flow based Co-Matching

introduced network flow variable for user anchor links

user anchor link existence confidence scores (previous step)

$$\max \sum_{(u,v)\in(\mathcal{U}^{(1)}\times\mathcal{U}^{(2)})} \boxed{F(u,v)} \cdot \boxed{\mathcal{W}_{\mathcal{U}}(u,v)} +$$

$$\sum_{(m,n)\in(\mathcal{L}^{(1)}\times\mathcal{L}^{(2)})} \boxed{F(m,n)\cdot \mathcal{W}_{\mathcal{L}}(m,n),}$$

location anchor link variables and confidence scores

$$s.t.\ 0 \leq F(u,v) \leq 1, \forall (u,v) \in \{S\} \times (\mathcal{U}^{(1)} \cup \mathcal{L}^{(1)}) \cup (\mathcal{U}^{(2)} \cup \mathcal{L}^{(2)}) \times \{T\},$$

$$F(u,v) \in \{0,1\}, \forall (u,v) \in \mathcal{U}^{(1)} \times \mathcal{U}^{(2)} \cup \mathcal{L}^{(1)} \times \mathcal{L}^{(2)},$$

$$\sum_{w\in\mathcal{N}_F,(w,u)\in\mathcal{L}_F} F(w,u) = \sum_{v\in\mathcal{N}_F,(u,v)\in\mathcal{L}_F} F(u,v).$$

mass balance constraints

variable bound constraints

Dataset for Experiments

- Dataset Statistical Information

Table 2: Properties of the Heterogeneous Networks

		network	
	property	**Twitter**	**Foursquare**
# node	user	5,223	5,392
	tweet/tip	9,490,707	48,756
	location	297,182	38,921
# link	friend/follow	164,920	76,972
	write	9,490,707	48,756
	locate	615,515	48,756

Detailed Experiment Settings

- Comparison Methods
  - UNICOAT: Model proposed in this paper, involves link inference and post-pruning steps.
  - BigAlign: Bipartite Network Alignment with Link Information [12]
  - BigAlignExt: Bipartite Network Alignment + Matching
  - ISO: User Anchor Link Inference with Link Information [12]
  - ISOExt: User Anchor Link Inference + Matching
  - RDD: a unsupervised anchor link inference method

	UNICOAT	Big-A	Big-A-E	ISO	ISO-E
prediction	√	√ (Bipartite)	√ (Bipartite)	√ (user anchor link)	√ (user anchor link)
matching	√		√		√
Link Info.	√	√	√	√	√
Attribute Info.	√				

- Evaluation Metrics
  - AUC, Precision@100
  - Precision, Recall, F1, Accuracy (Methods with Matching Step Only)

[12] D. Koutra, H. Tong, and D. Lubensky. Big-align: Fast bipartite graph alignment. In *ICDM*, 2013

Experiment Results

User Anchor Link Inference

Table 3: Performance comparison of different methods for inferring user anchor links (UNICOAT here denotes the first step of UNICOAT only).

measure	methods	θ = 1	2	3	4	5
AUC	UNICOAT	**0.868**	**0.831**	**0.814**	**0.804**	**0.799**
	BIGALIGNEXT	0.813	0.779	0.759	[illegible]	0.749
	BIGALIGN	0.568	0.557	0.555	[illegible]	0.550
	ISOEXT	0.818	0.782	0.762	0.754	0.61
	ISO	0.547	0.529	0.52	0.518	0.516
	RDD	0.531	0.530	0.523	0.514	[illegible]
Prec@100	UNICOAT	**0.705**	**0.688**	**0.657**	**0.640**	**[illegible]**
	BIGALIGNEXT	0.587	0.507	0.472	0.434	0.327
	BIGALIGN	0.347	0.284	0.265	0.228	0.220
	ISOEXT	0.427	0.391	0.373	0.352	0.301
	ISO	0.301	0.253	0.225	0.216	0.208
	RDD	0.234	0.228	0.207	0.172	0.127

Location Anchor Link Inference

Table 4: Performance comparison of different methods for inferring location anchor links (UNICOAT here denotes the first step of UNICOAT only).

measure	methods	θ = 1	2	3	4	5
AUC	UNICOAT	**0.822**	**0.815**	**0.796**	**0.794**	**0.753**
	BIGALIGNEXT	0.698	0.695	0.672	0.667	0.662
	BIGALIGN	0.592	0.586	0.576	0.572	0.56
	RDD	0.54	0.526	0.52	0.506	0.504
Prec@100	UNICOAT	**0.695**	**0.658**	**0.636**	**0.610**	**0.535**
	BIGALIGNEXT	0.507	0.434	0.372	0.328	0.327
	BIGALIGN	0.407	0.325	0.293	0.284	0.275
	RDD	0.216	0.204	0.183	0.182	0.157

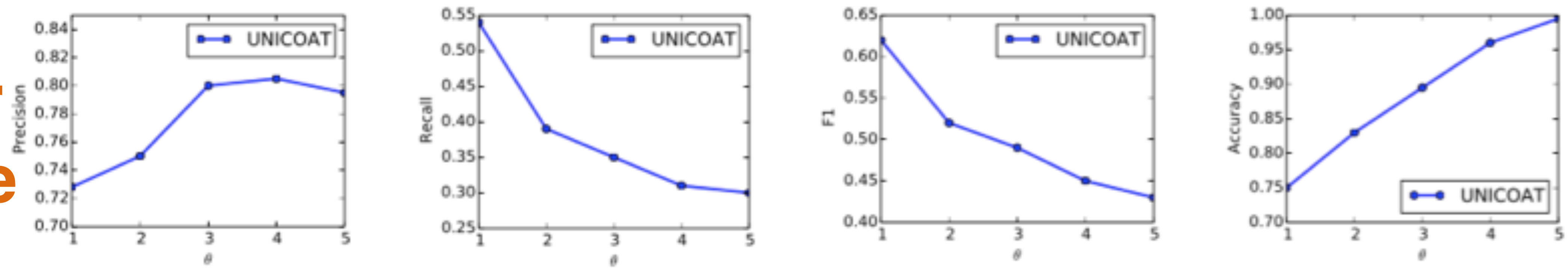
Alignment Ratio: $\theta = \frac{\#\text{total item}}{\#\text{anchor item}}$

θ = 1: full alignment setting

θ = 5: 20% alignment setting

Parameter Sensitivity Analysis and Alternative Updating Convergence Analysis

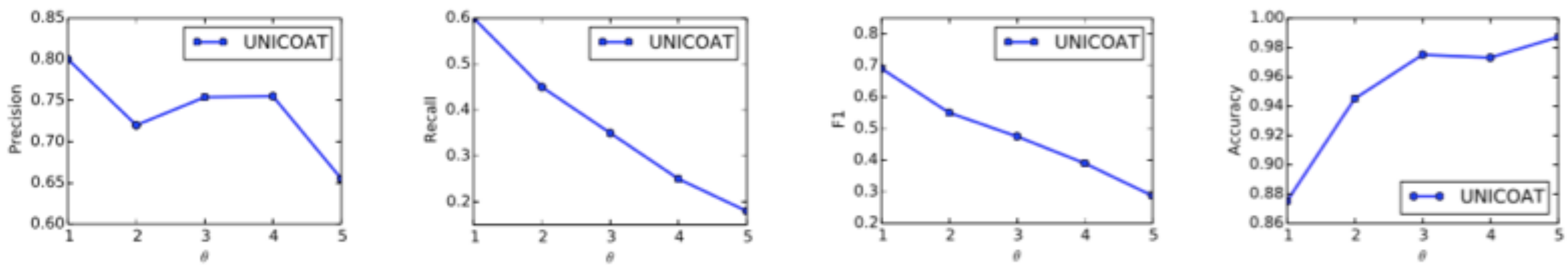
User Anchor Link Inference



(a) Precision (b) Recall (c) F1 (d) Accuracy

Figure 6: Performance of methods with matching in inferring user anchor links (UNICOAT here includes both two steps of UNICOAT).

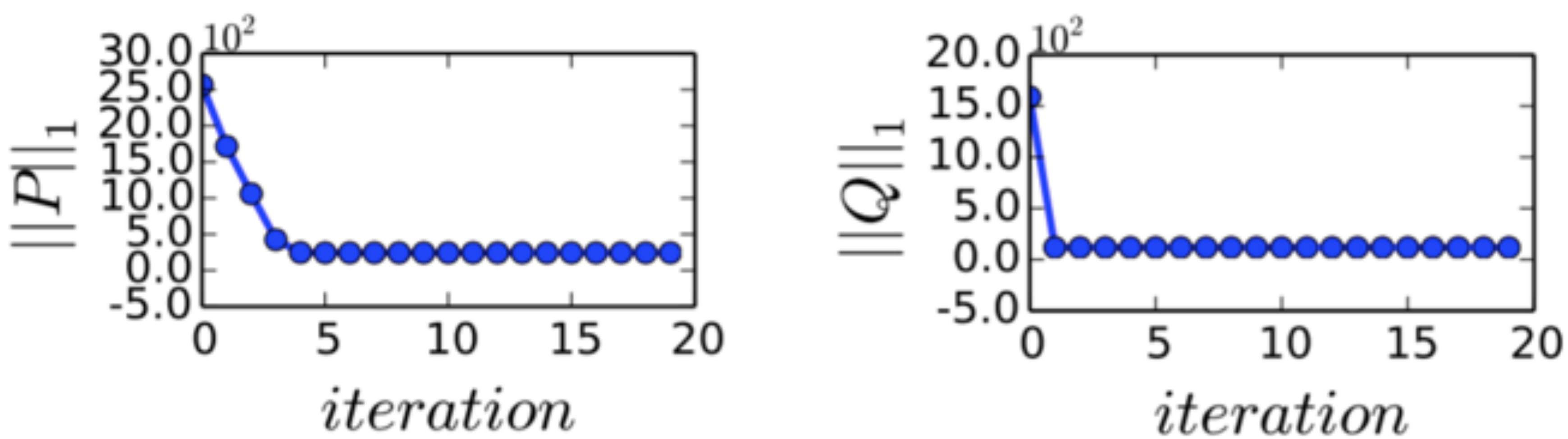
Location Anchor Link Inference



(a) Precision (b) Recall (c) F1 (d) Accuracy

Figure 7: Performance of methods with matching in inferring location anchor links (UNICOAT here includes both two steps of UNICOAT).

Convergence Analysis in Inferring P and Q



Summary

- Problem Studied:
  - Partial Co-Alignment of Social Networks, i.e., Simultaneous Inference of User Anchor Links and Location
- Proposed Model:
  - A joint optimization function to minimize the mapping cost and maximize the mapped item similarity concurrently for user and location anchor links with one-to-one constraint
  - A post-processing step to simultaneously prune the redundant user/location anchor links introduced due to the constraint relaxation.

PCT: Partial Co-Alignment of Social Networks

Q&A

Jiawei Zhang, Philip S. Yu
jzhan9@uic.edu, psyu@cs.uic.edu