ICDM 2015
IEEE International Conference on Data Mining

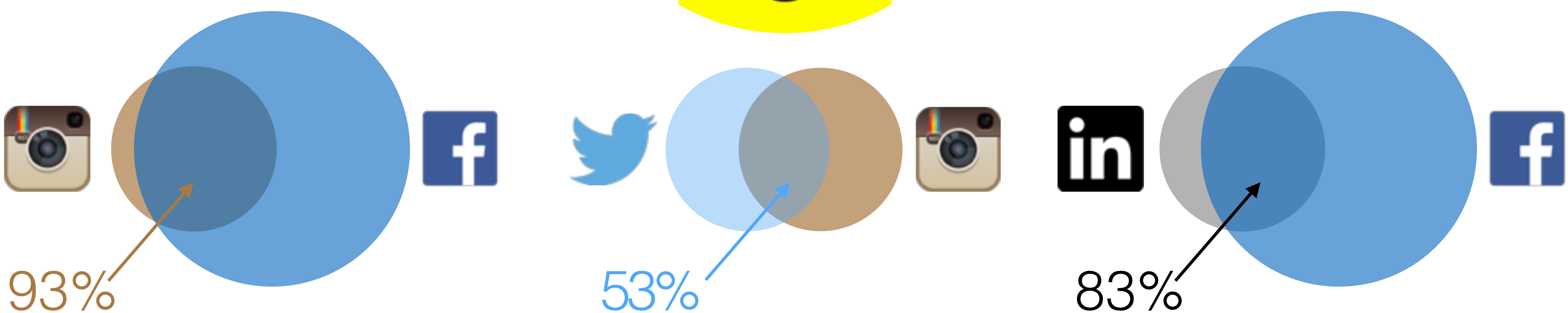
Multiple Anonymized Social Networks Alignment

Jiawei Zhang, Philip S. Yu

University of Illinois at Chicago, Chicago, IL

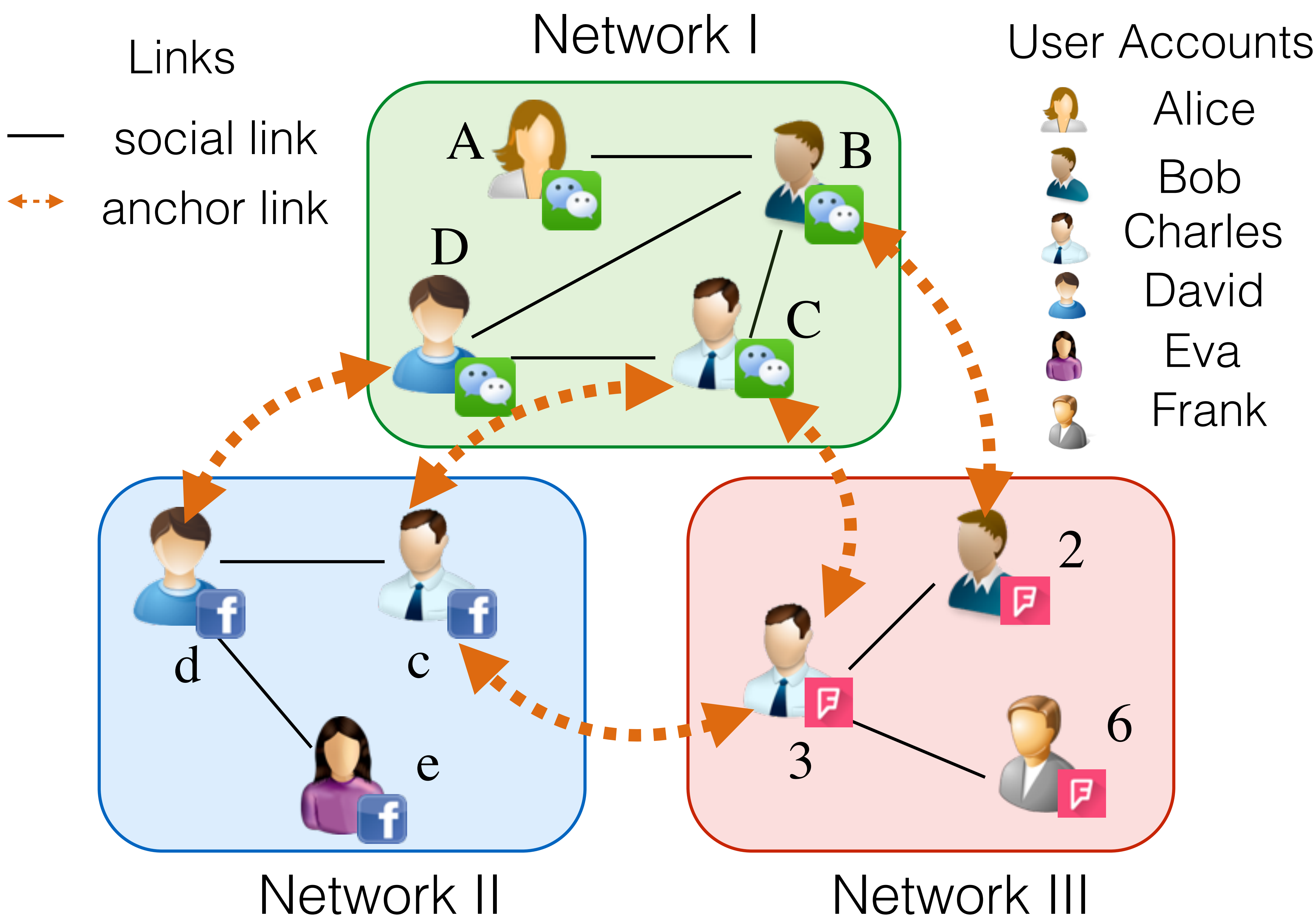
UIC UNIVERSITY OF ILLINOIS AT CHICAGO

People are using multiple social networks simultaneously



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

Align multiple anonymized social networks via the common users

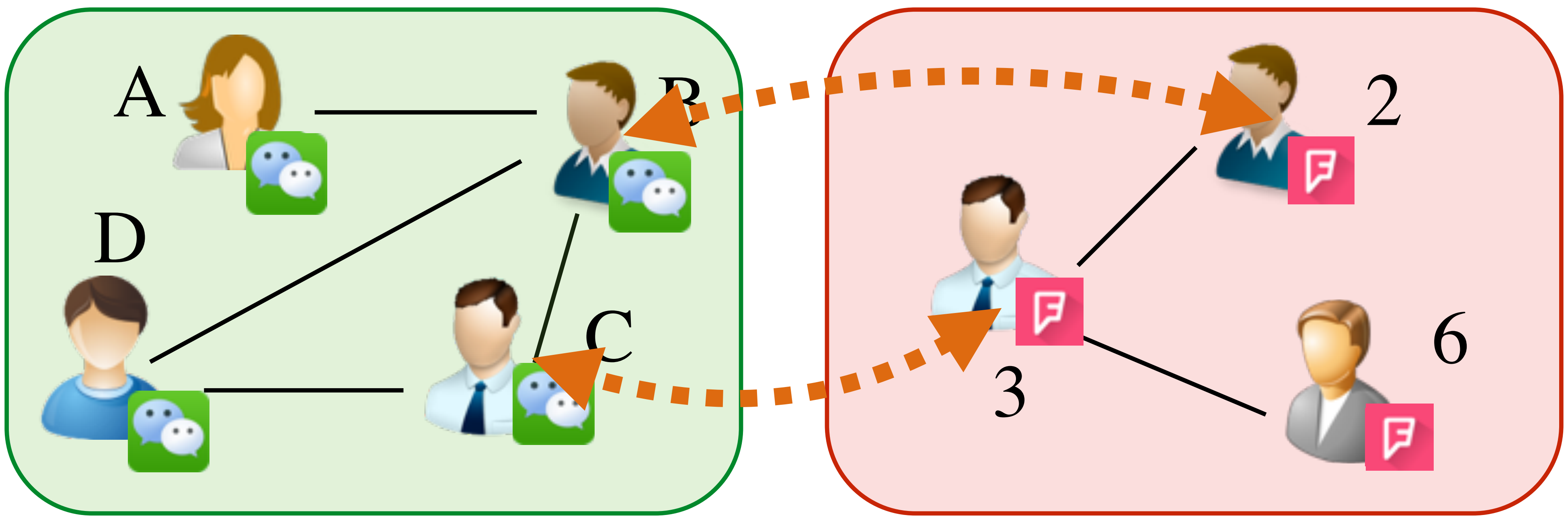


Motivation: fuse multiple information sources

- gain a more comprehensive understanding about users' social activities in all social networks
  - provide better recommendation services [Zhang ICDM 2013, Zhang WSDM 2014, Zhang CIKM 2015, Zhang IJCAI 2015]
  - detect more accurate community structures [Jin IEEE BigData 2014, Zhang IEEE IRI 2015, Zhang SDM 2015]
- cross-network information exchange and diffusion
  - resolve the information sparsity problem [Zhang ICDM 2013, Zhang WSDM 2014, Zhang IJCAI 2015, Zhang SDM 2015]
  - overcome the cold-start problem [Zhang ICDM 2013, Zhang SDM 2015]
  - broader influence in viral marketing [Zhan PAKDD 2015]

Challenge 1: networks are pre-anonymized

Solution: use the structure information only



	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 $\mathbf{S}^{(i)}$

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

Transition Matrix $\mathbf{T}^{(i,j)}$

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

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

Challenge 1: networks are pre-anonymized

Assumption: shared users have similar social structures in different networks

	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 $\mathbf{S}^{(i)}$

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

Transition Matrix $\mathbf{T}^{(i,j)}$

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

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

Via transition matrix $\mathbf{T}^{(i,j)}$(i.e., anchor links), we can map the social connections among shared users from network I to network II:

$$(\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)}$$

The optimal transition matrix $\mathbf{T}^{(i,j)}$(i.e., anchor links) should minimize the mapping cost

$$\min \left\| (\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)} - \mathbf{S}^{(j)} \right\|_F^2$$

Challenge 2: one-to-one constraint on anchor links

- Add one-to-one constraint on the transitional matrix $\mathbf{T}^{(i,j)}$
  - $\mathbf{T}^{(i,j)}$ Should be binary matrix
  - each row and each column of $\mathbf{T}^{(i,j)}$ should contain at most one entry being filled with value 1
- The complete objective function should be

$$
\begin{aligned}
\bar{\mathbf{T}}^{(i,j)} &= \arg\min_{\mathbf{T}^{(i,j)}} \left\| (\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)} - \mathbf{S}^{(j)} \right\|_F^2 \\
s.t. \quad & \mathbf{T}^{(i,j)} \in \{0,1\}^{|\mathcal{U}^{(i)}| \times |\mathcal{U}^{(j)}|}, \\
& \mathbf{T}^{(i,j)} \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1}, \\
& (\mathbf{T}^{(i,j)})^\top \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1},
\end{aligned}
$$

Challenge 3: multiple network alignment

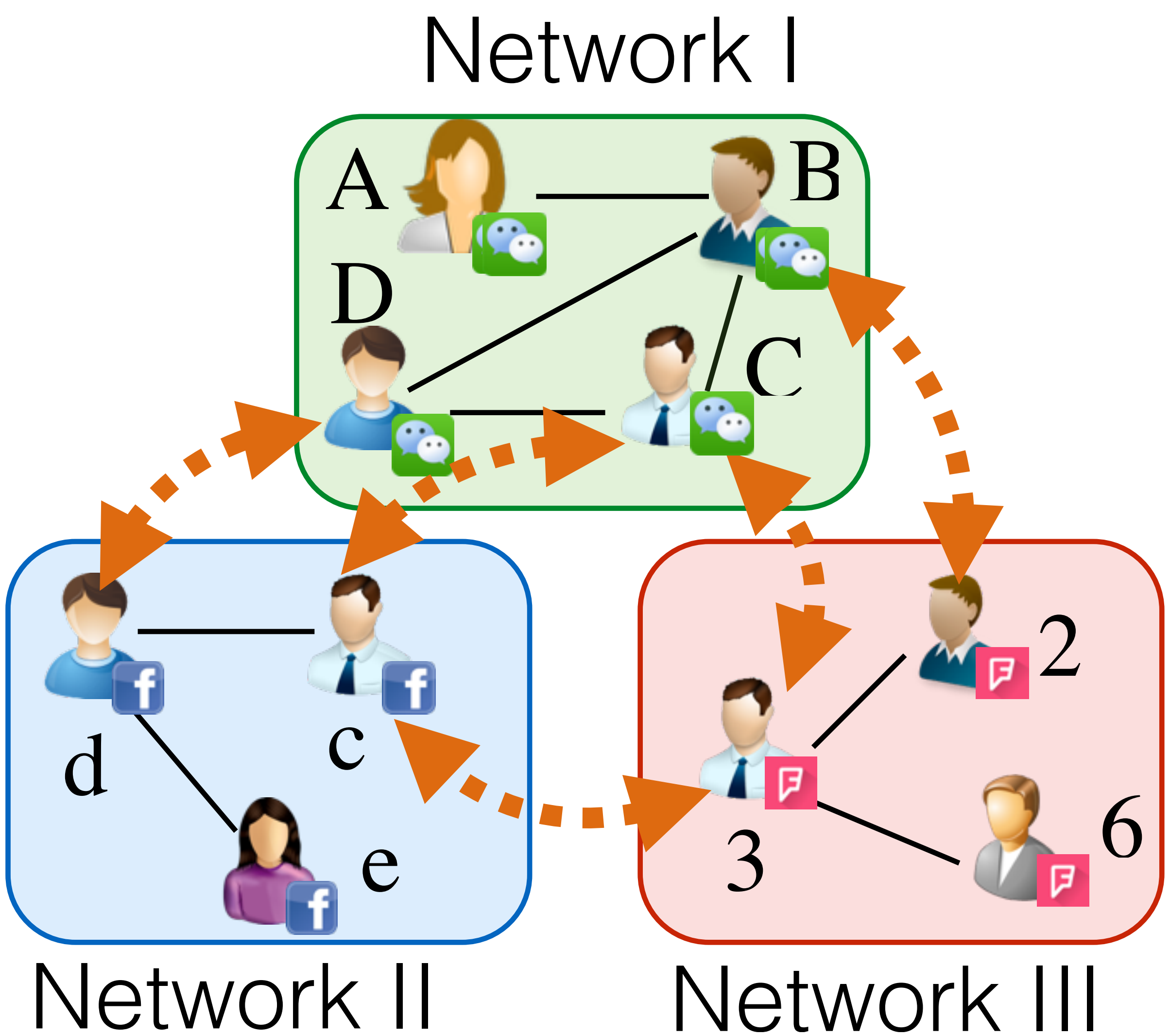
- Anchor links meet the "Transitivity Law"

$$\forall a, b, c \in \mathcal{X}, (a, b) \in \mathcal{R} \wedge (b, c) \in \mathcal{R} \rightarrow (a, c) \in \mathcal{R}.$$

- if (u, v) are connected by anchor links between network 1 and network 2;
- and (v, w) are connected by anchor links between network 2 and network 3;
- then (u, w) should be also connected by anchor links between network 1 and network 3

Challenge 3: multiple network alignment

- Anchor links meet the "Transitivity Law"



- The alignment results of any 3 networks should be consistent
- In other words, the mapping from network 1 to network 3 via paths
  - network 1 -> network 2 -> network 3
  - network 1 -> network 3

  should be similar
- we introduce the "Alignment Transitivity Penalty" cost function to be

$$C(\{G^{(i)}, G^{(j)}, G^{(k)}\})$$
$$= \left\| (\mathbf{T}^{(j,k)})^\top (\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)} \mathbf{T}^{(j,k)} - (\mathbf{T}^{(i,k)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,k)} \right\|_F^2 .$$

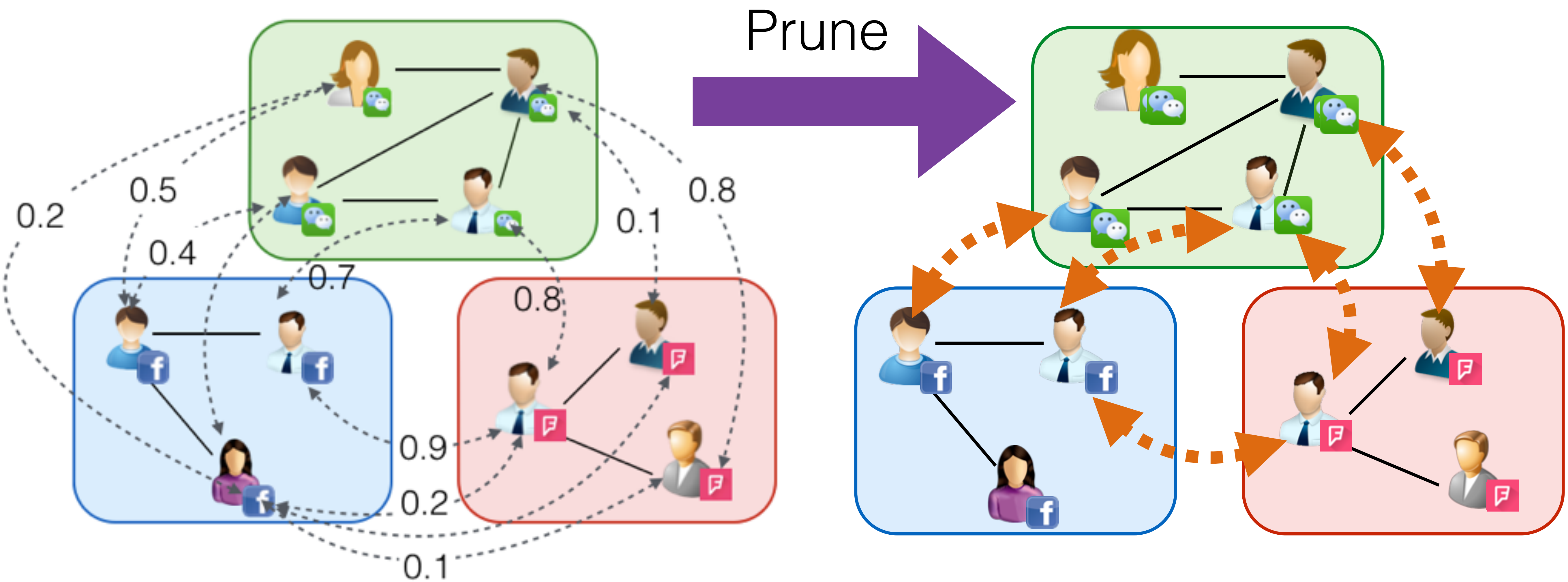
Joint Objective Function

$$
\begin{aligned}
&\bar{\mathbf{T}}^{(i,j)}, \bar{\mathbf{T}}^{(j,k)}, \bar{\mathbf{T}}^{(k,i)} \\
&= \arg\min_{\mathbf{T}^{(i,j)}, \mathbf{T}^{(j,k)}, \mathbf{T}^{(k,i)}} \left\| (\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)} - \mathbf{S}^{(j)} \right\|_F^2 \\
&+ \left\| (\mathbf{T}^{(j,k)})^\top \mathbf{S}^{(j)} \mathbf{T}^{(j,k)} - \mathbf{S}^{(k)} \right\|_F^2 + \left\| (\mathbf{T}^{(k,i)})^\top \mathbf{S}^{(k)} \mathbf{T}^{(k,i)} - \mathbf{S}^{(i)} \right\|_F^2 \\
&+ \alpha \cdot \left\| (\mathbf{T}^{(j,k)})^\top (\mathbf{T}^{(i,j)})^\top \mathbf{S}^{(i)} \mathbf{T}^{(i,j)} \mathbf{T}^{(j,k)} - \mathbf{T}^{(k,i)} \mathbf{S}^{(i)} (\mathbf{T}^{(k,i)})^\top \right\|_F^2 \\
&s.t.\ \mathbf{T}^{(i,j)} \in \{0,1\}^{|\mathcal{U}^{(i)}| \times |\mathcal{U}^{(j)}|}, \mathbf{T}^{(j,k)} \in \{0,1\}^{|\mathcal{U}^{(j)}| \times |\mathcal{U}^{(k)}|} \\
&\qquad \mathbf{T}^{(k,i)} \in \{0,1\}^{|\mathcal{U}^{(k)}| \times |\mathcal{U}^{(i)}|} \\
&\qquad \mathbf{T}^{(i,j)} \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1}, (\mathbf{T}^{(i,j)})^\top \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1}, \\
&\qquad \mathbf{T}^{(j,k)} \mathbf{1}^{|\mathcal{U}^{(k)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1}, (\mathbf{T}^{(j,k)})^\top \mathbf{1}^{|\mathcal{U}^{(j)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(k)}| \times 1}, \\
&\qquad \mathbf{T}^{(k,i)} \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(k)}| \times 1}, (\mathbf{T}^{(k,i)})^\top \mathbf{1}^{|\mathcal{U}^{(k)}| \times 1} \preccurlyeq \mathbf{1}^{|\mathcal{U}^{(i)}| \times 1},
\end{aligned}
$$

b to solve the objective function, ts

the constraints are relaxed

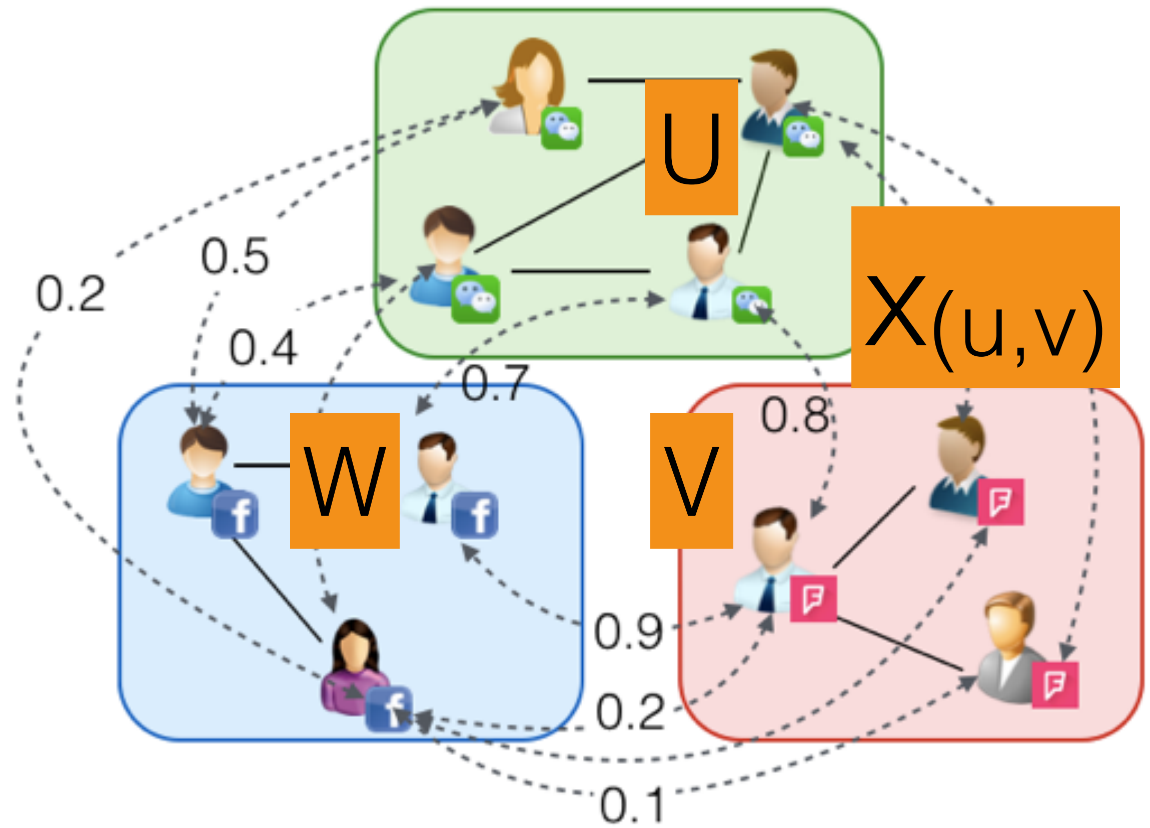
Challenge 4: Transitive Network Mapping



- Prune the non-existing anchor links introduced by constraint relaxation
- We introduce the **transitive multiple network matching** method, with considerations about the "one-to-one" constraint and "Transitivity Law" property of anchor links

Transitive multiple network matching

- Network matching:
  - pick the high confidence anchor links from the prediction results; prune the remaining ones of low confidence scores
- Constraints on network matching
  - one-to-one constraint, for each use
  - "Transitivity Law" constraint



- define $x_{(u,v)}$ to denote potential anchor link (u,v) is selected or not (1: selected, 0 otherwise).
- in the alignment of two networks, for each user u, at most one variable $x_{(u,v)}$ is assigned with value 1 (one-to-one constraint)
- in the alignment of any 3 networks, for any 3 users u, v, w
  - if $x_{(u,v)} = 1$, $x_{(v,w)} = 1$, then $x_{(u,w)} = 1$ ("Transitivity Law" constraint)
  - in other words $x_{(u,v)} + x_{(v,w)} + x_{(u,w)} \,!= 2$ for any user u, v, w in 3 networks

Transitive multiple network matching

$$\max_{\mathbf{x}^{(i,j)},\mathbf{x}^{(j,k)},\mathbf{x}^{(k,i)}} \sum_{l,m} x_{l,m}^{(i,j)} \mathbf{T}^{(i,j)}(l,m) + \sum_{l,m} x_{l,m}^{(i,j)} \mathbf{T}^{(i,j)}(l,m) + \sum_{l,m} x_{l,m}^{(i,j)} \mathbf{T}^{(i,j)}(l,m),$$

select anchor links of high confidences

$$s.t. \sum_{u_m^{(j)} \in \mathcal{U}^{(j)}} x_{l,m}^{(i,j)} \leq 1, \sum_{u_o^{(k)} \in \mathcal{U}^{(k)}} x_{l,o}^{(i,k)} \leq 1, \forall u_l^{(i)} \in \mathcal{U}^{(i)},$$

$$\sum_{u_l^{(i)} \in \mathcal{U}^{(i)}} x_{m,l}^{(j,i)} \leq 1, \sum_{u_o^{(k)} \in \mathcal{U}^{(k)}} x_{m,o}^{(j,k)} \leq 1, \forall u_m^{(j)} \in \mathcal{U}^{(j)},$$

$$\sum_{u_l^{(i)} \in \mathcal{U}^{(i)}} x_{o,l}^{(k,i)} \leq 1, \sum_{u_m^{(j)} \in \mathcal{U}^{(j)}} x_{o,m}^{(k,j)} \leq 1, \forall u_o^{(k)} \in \mathcal{U}^{(k)},$$

one-to-one constraint

$$x_{l,m}^{(i,j)} + x_{m,o}^{(j,k)} + x_{o,l}^{(k,i)} \neq 2, \forall l \in \{1, 2, \cdots, |\mathcal{U}^{(i)}|\},$$
$$\forall m \in \{1, 2, \cdots, |\mathcal{U}^{(j)}|\}, \forall o \in \{1, 2, \cdots, \mathcal{U}^{(k)}|,$$

transitivity law constraint

$$x_{l,m}^{(i,j)} \in \{0,1\}, \forall u_l^{(i)} \in \mathcal{U}^{(i)}, u_m^{(j)} \in \mathcal{U}^{(j)}.$$
$$x_{m,o}^{(j,k)} \in \{0,1\}, \forall u_m^{(j)} \in \mathcal{U}^{(j)}, u_o^{(k)} \in \mathcal{U}^{(k)}.$$
$$x_{o,l}^{(k,i)} \in \{0,1\}, \forall u_o^{(k)} \in \mathcal{U}^{(k)}, u_l^{(i)} \in \mathcal{U}^{(i)}.$$

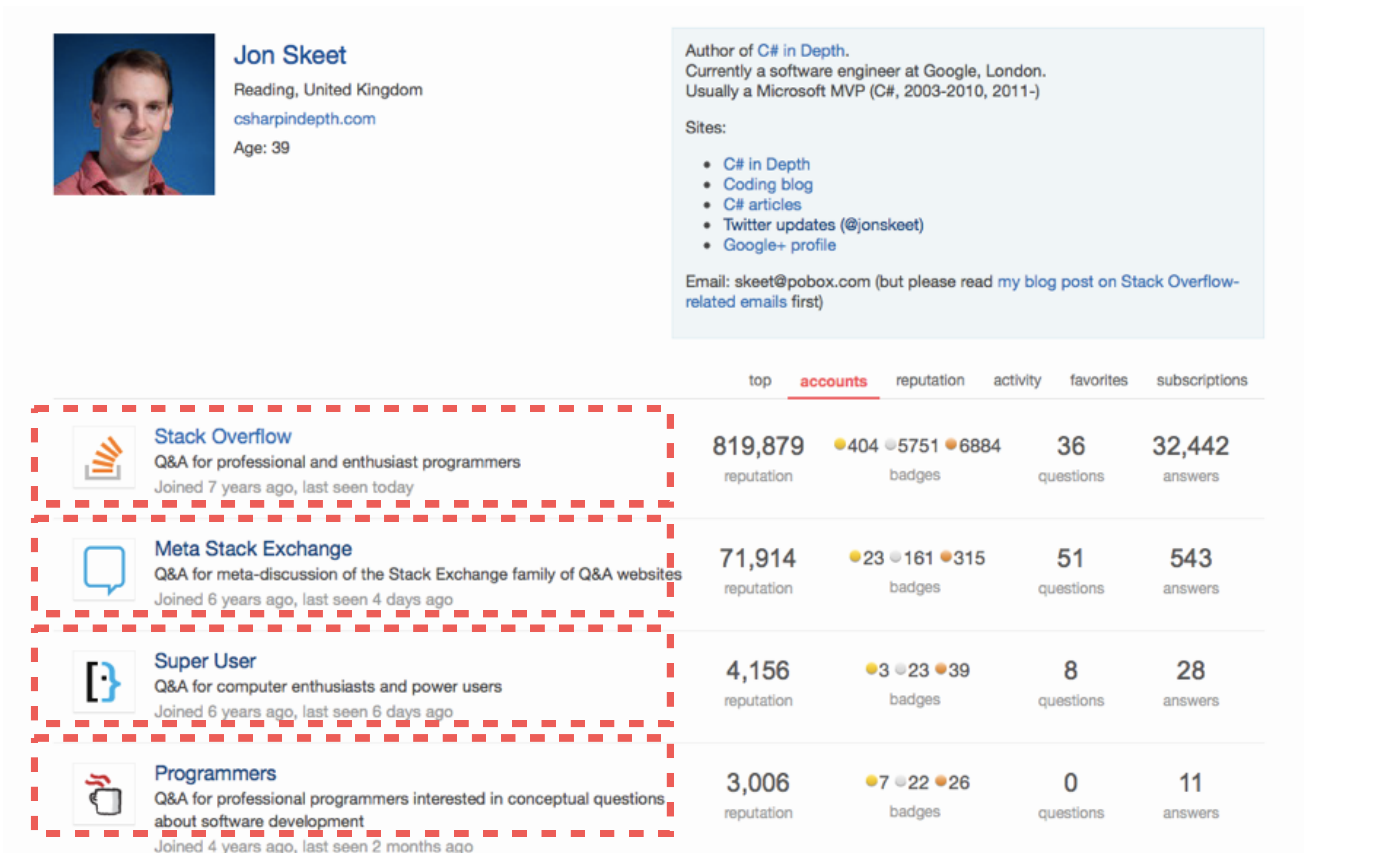
binary value of variables

Experiments

- Dataset - 3 Q&A sites
    - Stack Overflow (10,000 users)
    - Super User (10,000 users)
    - Programmers (10,000 users)
- Anchor links among these 3 networks
    - Stack Overflow - Super User: 3,677
    - Stack Overflow - Programmers: 2,626
    - Super User - Programmers: 1,953
- Experiment Setting: unsupervised learning settings
    - the anchor links are used as the ground truth for evaluating the prediction results
    - known anchor links are not involved in the model building

Experiments

- Ground Truth Anchor Links



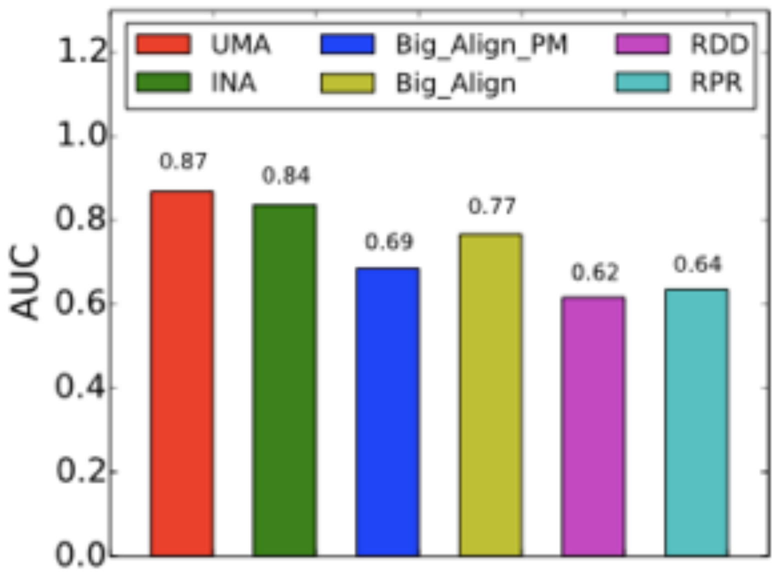
Experiments

	UMA	INA	Big-A	Big-A-PM	RDD	RPR
prediction	√	√	√ (pairwise)	√ (pairwise)		
matching	√			√ (pairwise)		
score					√	√

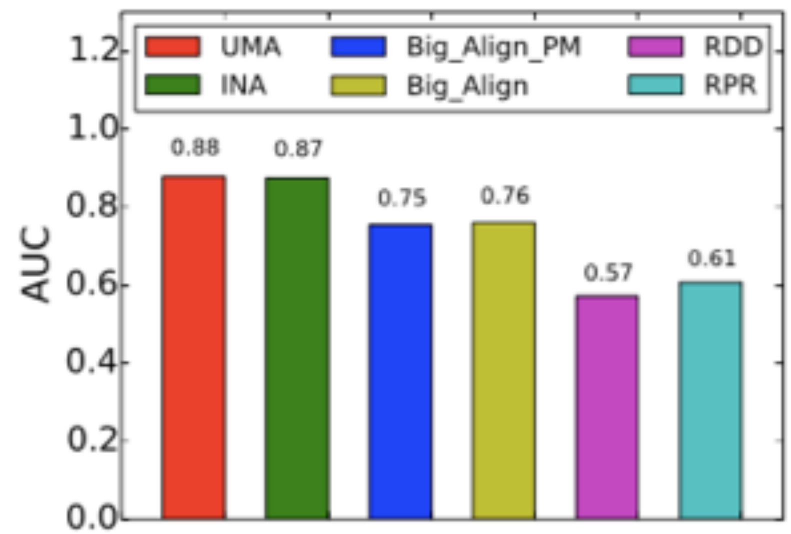
- Comparison Methods
  - UMA: multiple network alignment + matching
  - INA: multiple network alignment (the 1st step of UMA)
  - Big-Align: pairwise network alignment [13]
  - Big-Align-PM: Big-Align + pairwise network matching
  - RDD: calculate the Relative Degree Distance as the confidence scores of potential anchor links [13]
  - RPR: calculate the Relative PageRank score as the confidence scores of potential anchor links
- Evaluation Metrics
  - AUC and Precision@100
  - Accuracy, Precision, Recall, F1 (for methods UMA and Big-Align-PM with the matching step)

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

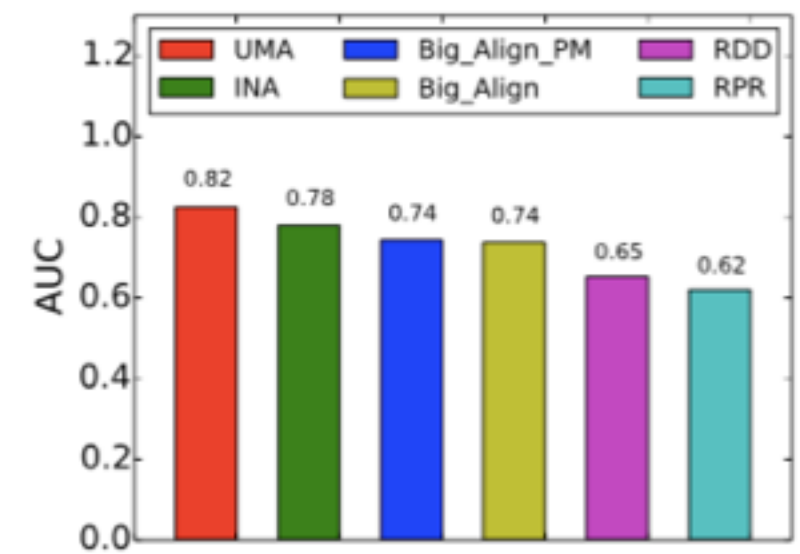
Experiment results



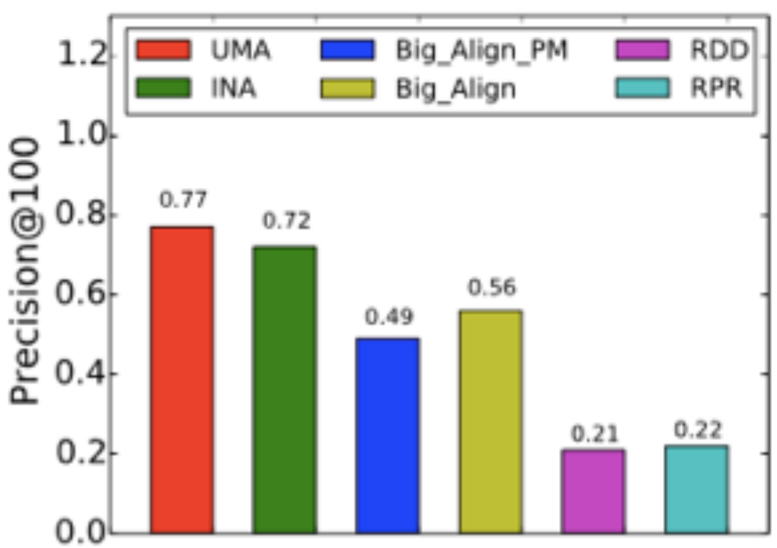
(a) AUC ($G^{(i)}$, $G^{(j)}$)



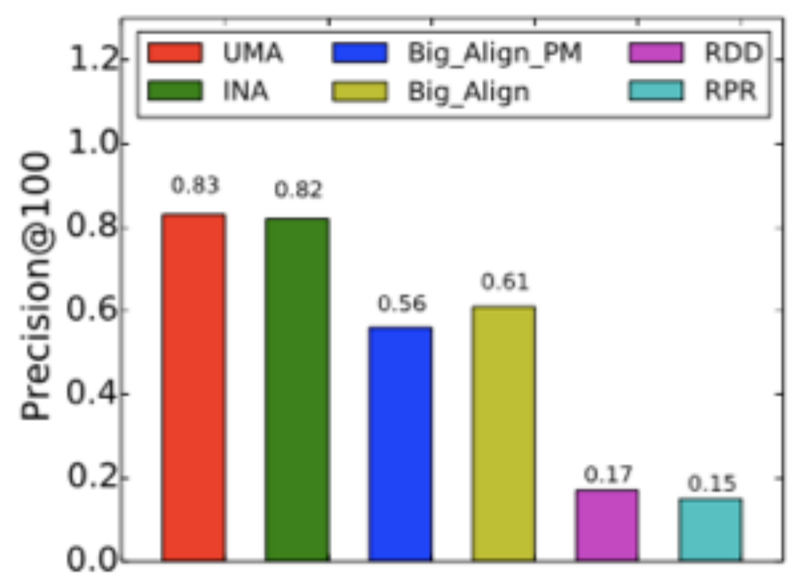
(b) AUC ($G^{(j)}$, $G^{(k)}$)



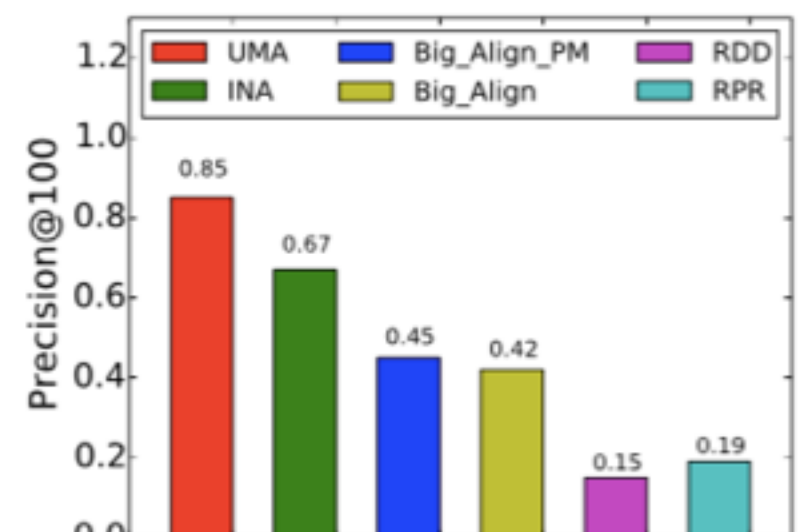
(c) AUC ($G^{(k)}$, $G^{(i)}$)



(d) Precision@100 ($G^{(i)}$, $G^{(j)}$)

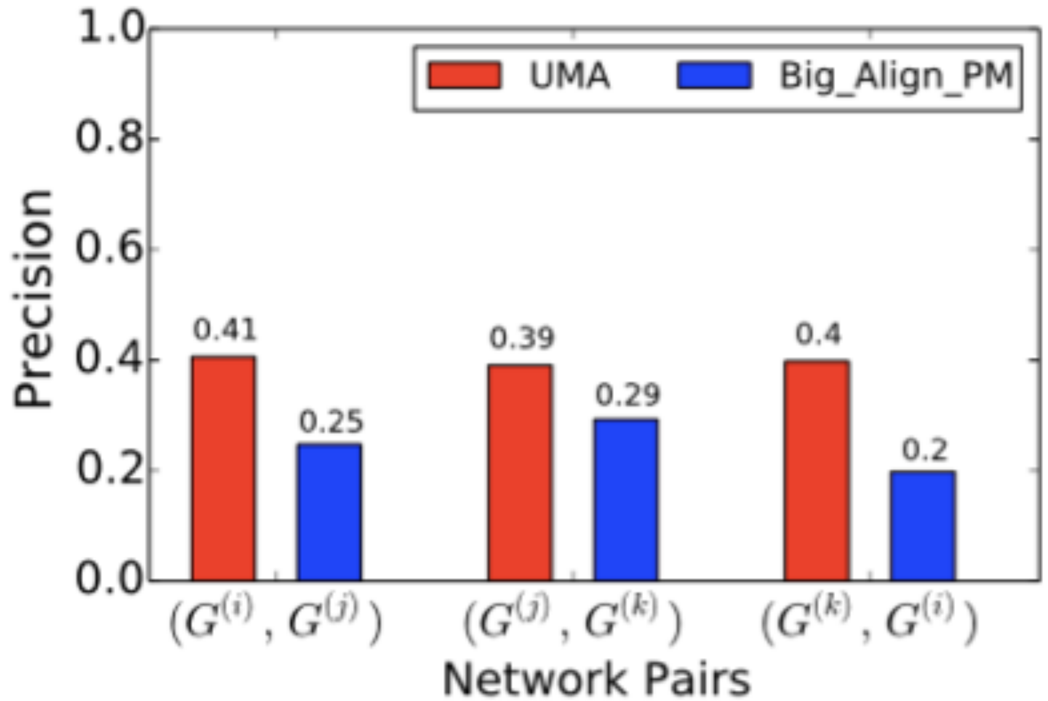


(e) Precision@100 ($G^{(j)}$, $G^{(k)}$)

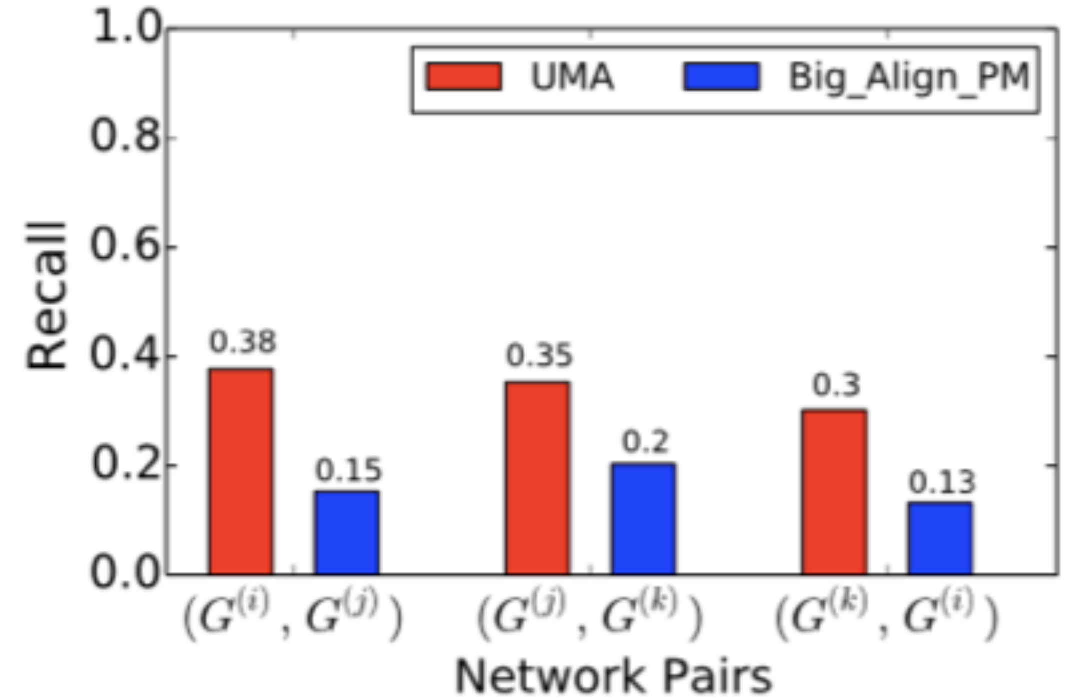


(f) Precision@100 ($G^{(k)}$, $G^{(i)}$)

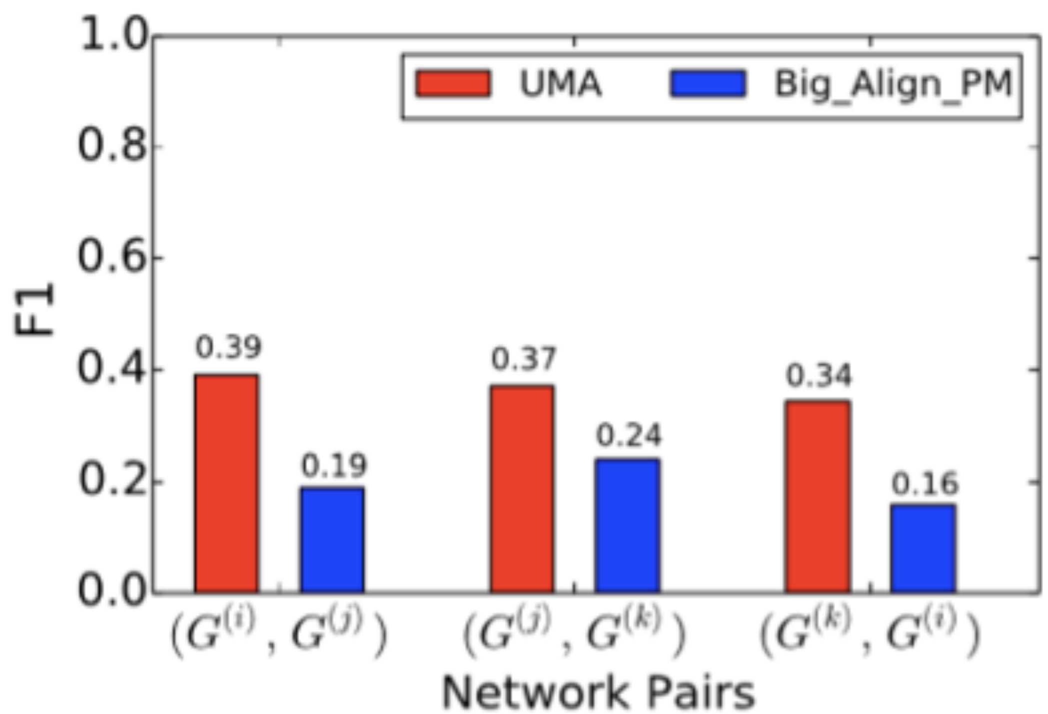
Experiment results



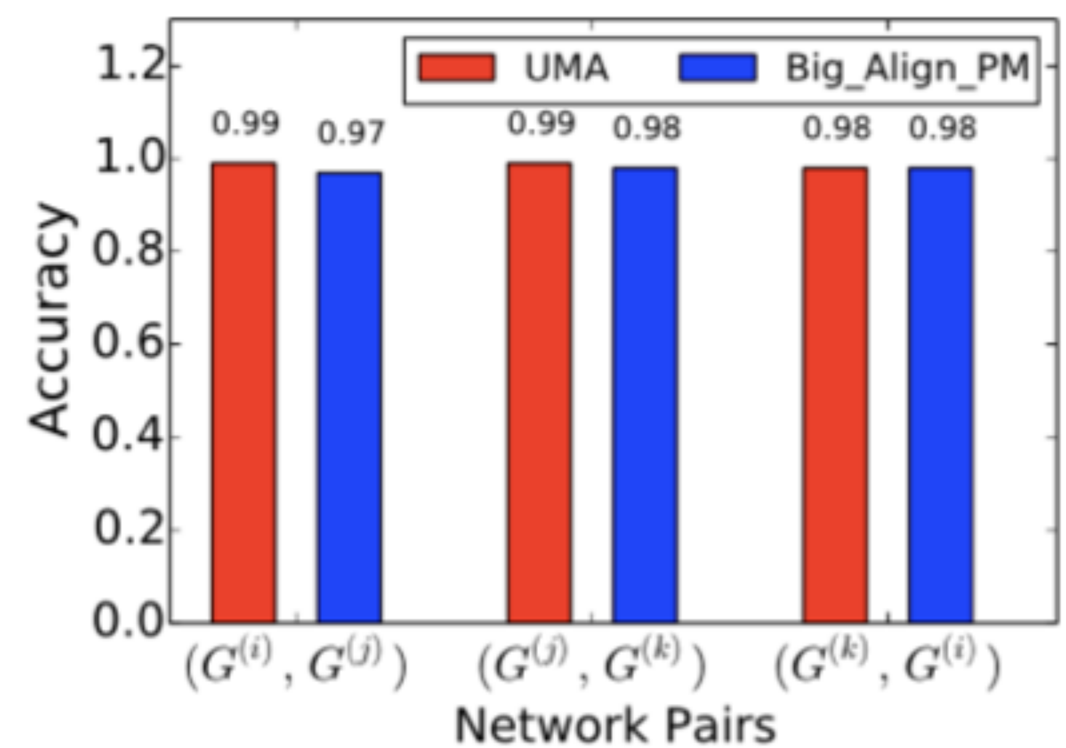
(a) Precision



(b) Recall



(c) F1



(d) Accuracy

Summary

- Problem studied
  - simultaneous alignment of multiple anonymized social networks
- Propose method:
  - an joint optimization function to minimize mapping cost and “transitivity penalty’’ cost with one-to-one constraint
  - a transitive multiple network matching algorithm to prune the non-existing anchor links with one-to-one constraint and “transitivity law” constraint

ICDM 2015
IEEE International Conference on Data Mining

Multiple Anonymized Social Networks Alignment

Q&A

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

UIC UNIVERSITY OF ILLINOIS AT CHICAGO