Identifying Your Customers in Social Networks

Chun-Ta Lu

Hong-Han Shuai

Philip S. Yu

University of Illinois at Chicago

Presenter: Chun-Ta Lu

Identifying Your Customers in Social Networks

1. Motivation
2. Problem & Challenge
3. Methods
4. Experiments
5. Conclusion

hulu

NETFLIX

Google play

Spotify®

豆瓣 douban

amazon

淘宝网
Taobao.com

Knowing your customers

Knowing your customers

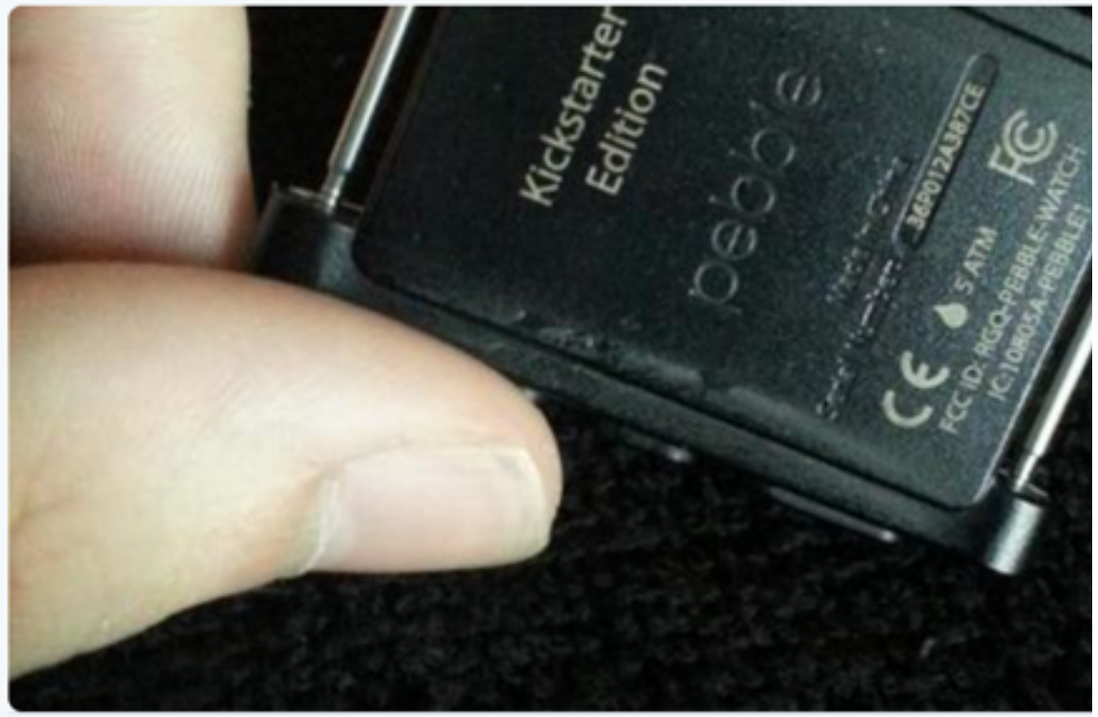
Challenges:

* **Lack of users' info**
* **Lack of feedback**


Tyler T. @TuckertCTD · Aug 6
@Pebble I've had this white Kickstarter Edition since day one... still the best thing I've ever bought. #FreshHotFly
Kickstarter Edition
pebble
5 ATM
FCC ID: RGQ-PEBBLE-WATCH
IC:10805A-PEBBLE1
7
23

Tyler T. @TuckertCTD · Aug 6

@**Pebble** I've had this white Kickstarter Edition since day one. thing I've ever **bought**. #FreshHotFly



Marcus Wright @marcuswtech · Jul 3

Friend just **bought** one of these! **pebble** e-paper cherry red watch p-cr001 amzn.to/1qJMyPP #**pebble** #smartwatch

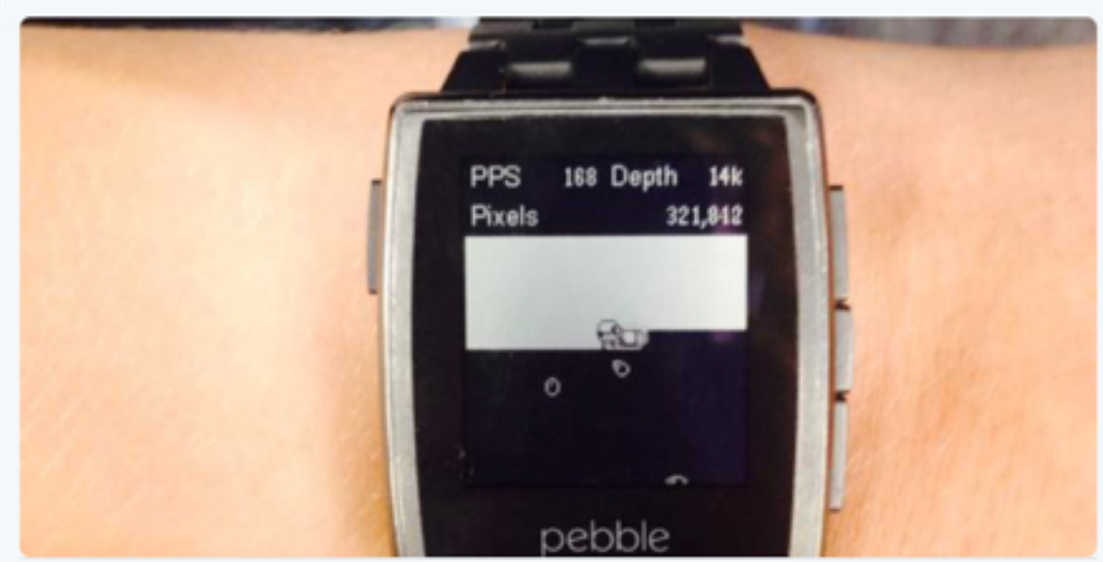
Jeremy Yancey @jeremy_yancey · May 1

Finally caved...curiosity got the better of me and I **bought** a #**Pebble** #smartwatch Love it! ift.tt/1fDrVkm

3 1

Joah Gerstenberg @therealjoahg · Aug 21

Just **bought** a drill gun on #pebbleminer :) @**pebble** #pebblesteel



3 14

Identifying Your Customers in Social Networks

Potential Applications:

1. Analyze your customers' opinions


Wednesday
12:11
April 2
13°
pebble


Wednesday
12:11
April 2
13°
pebble

Identifying Your Customers in Social Networks

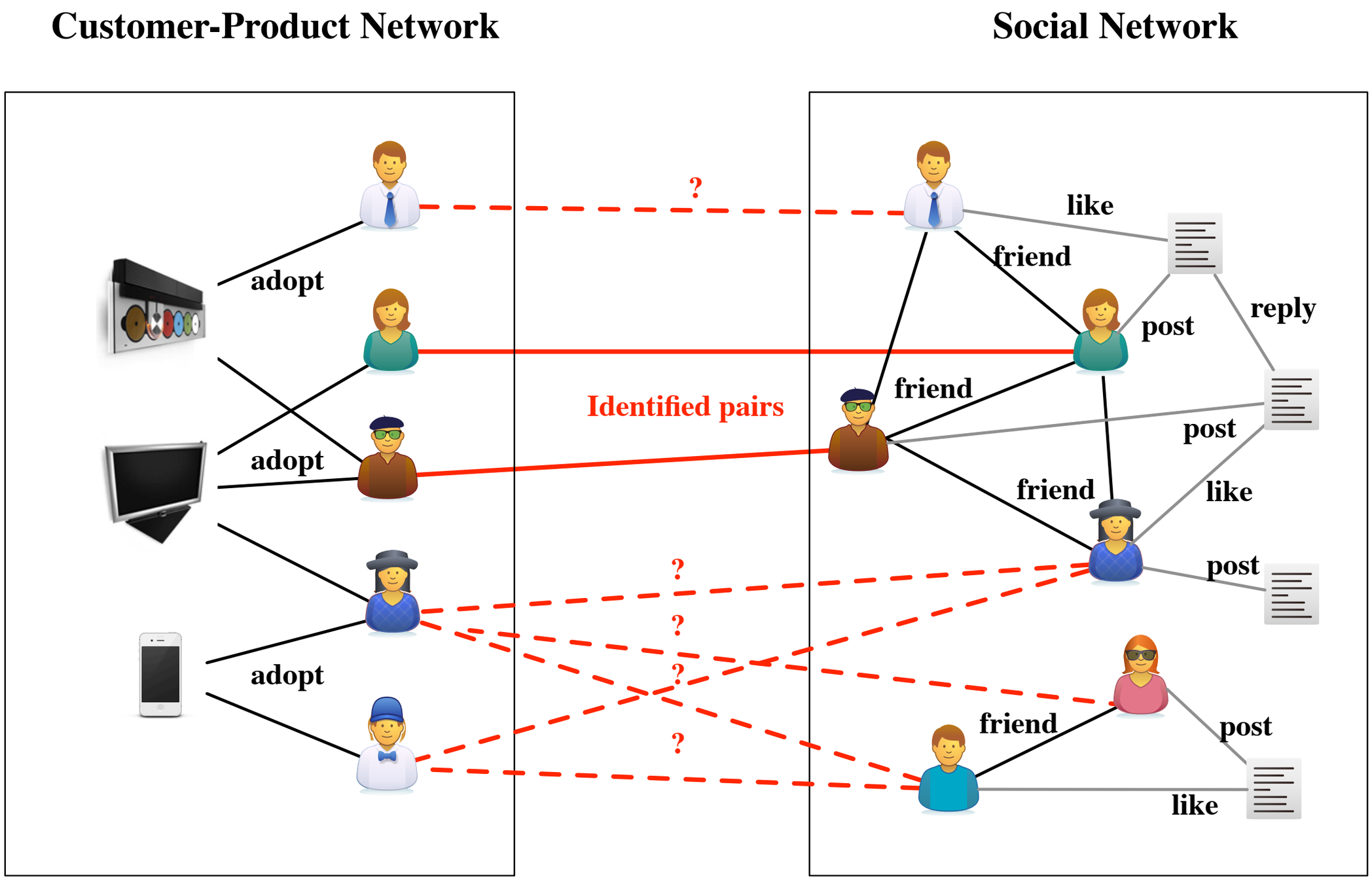
Potential Applications:

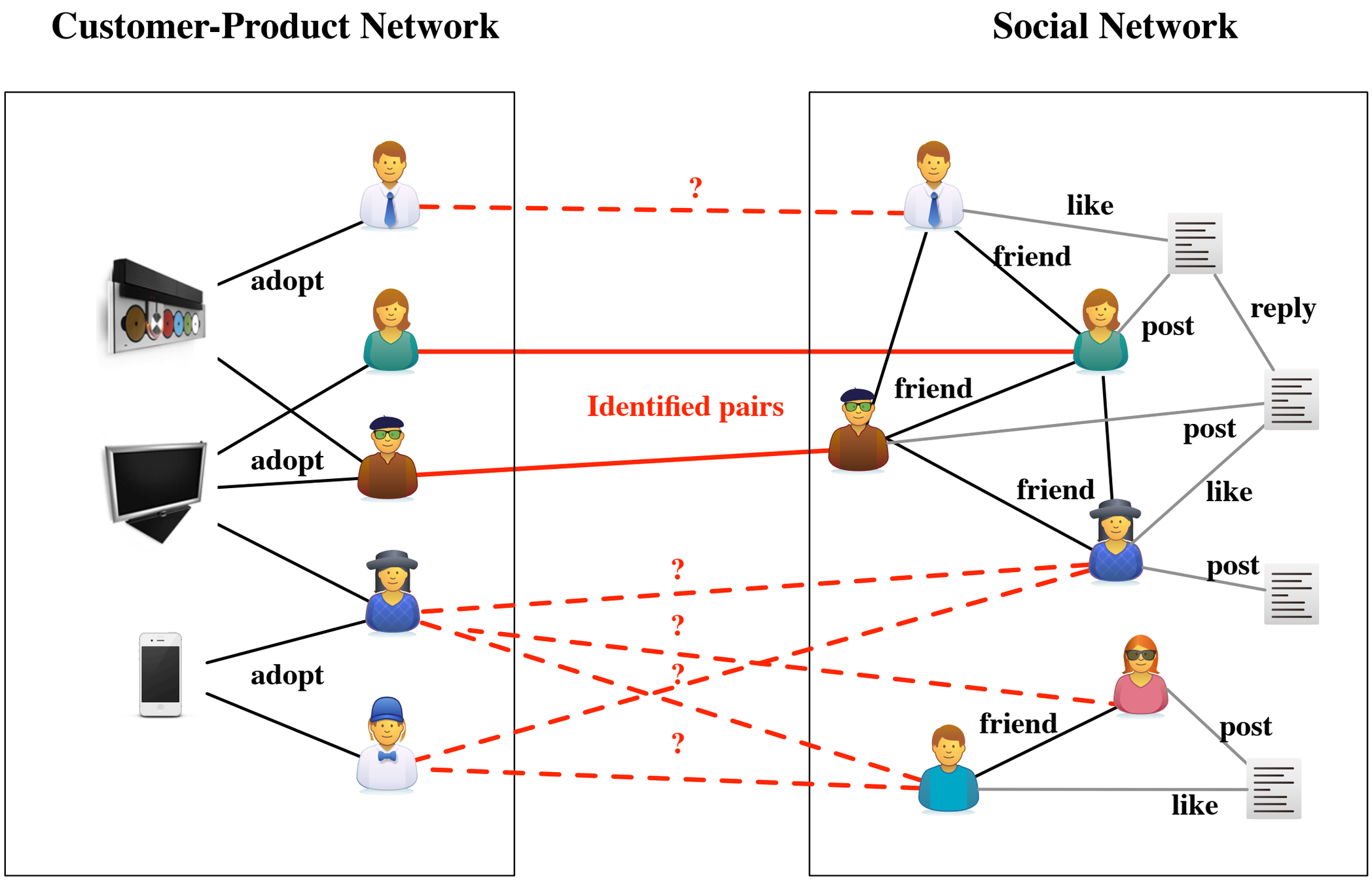
1. Analyze your customers' opinions

Identifying Your Customers in Social Networks

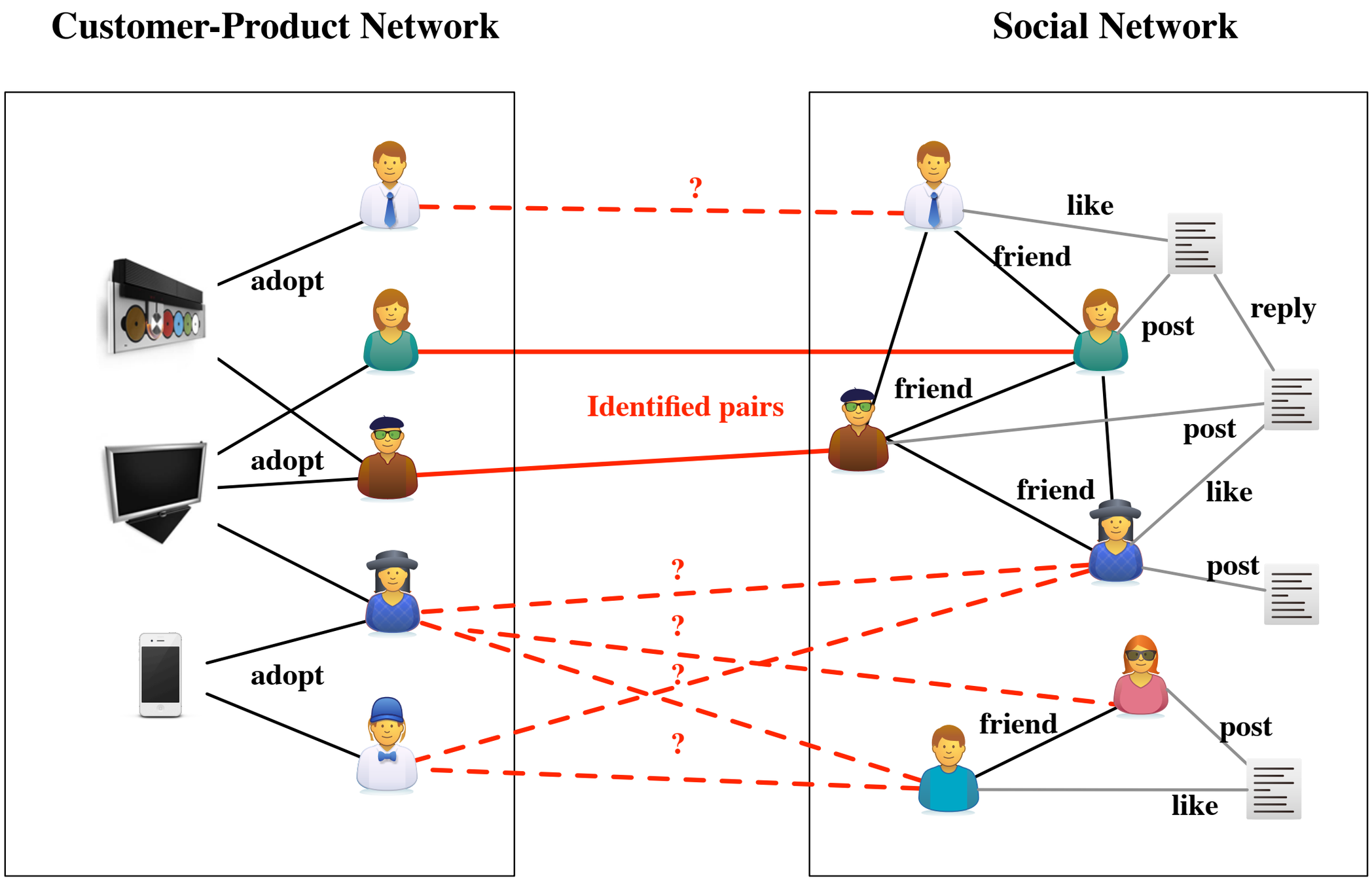
Potential Applications:

1. Analyze your customers' opinions
2. Personalized Product Recommendation
3. Discover the communities of your customers
4. Maximize product adoption


Customer-Product Network
Social Network
adopt
adopt
adopt
?
Identified pairs
?
?
?
?
like
friend
reply
post
friend
post
friend
like
post
friend
post
like

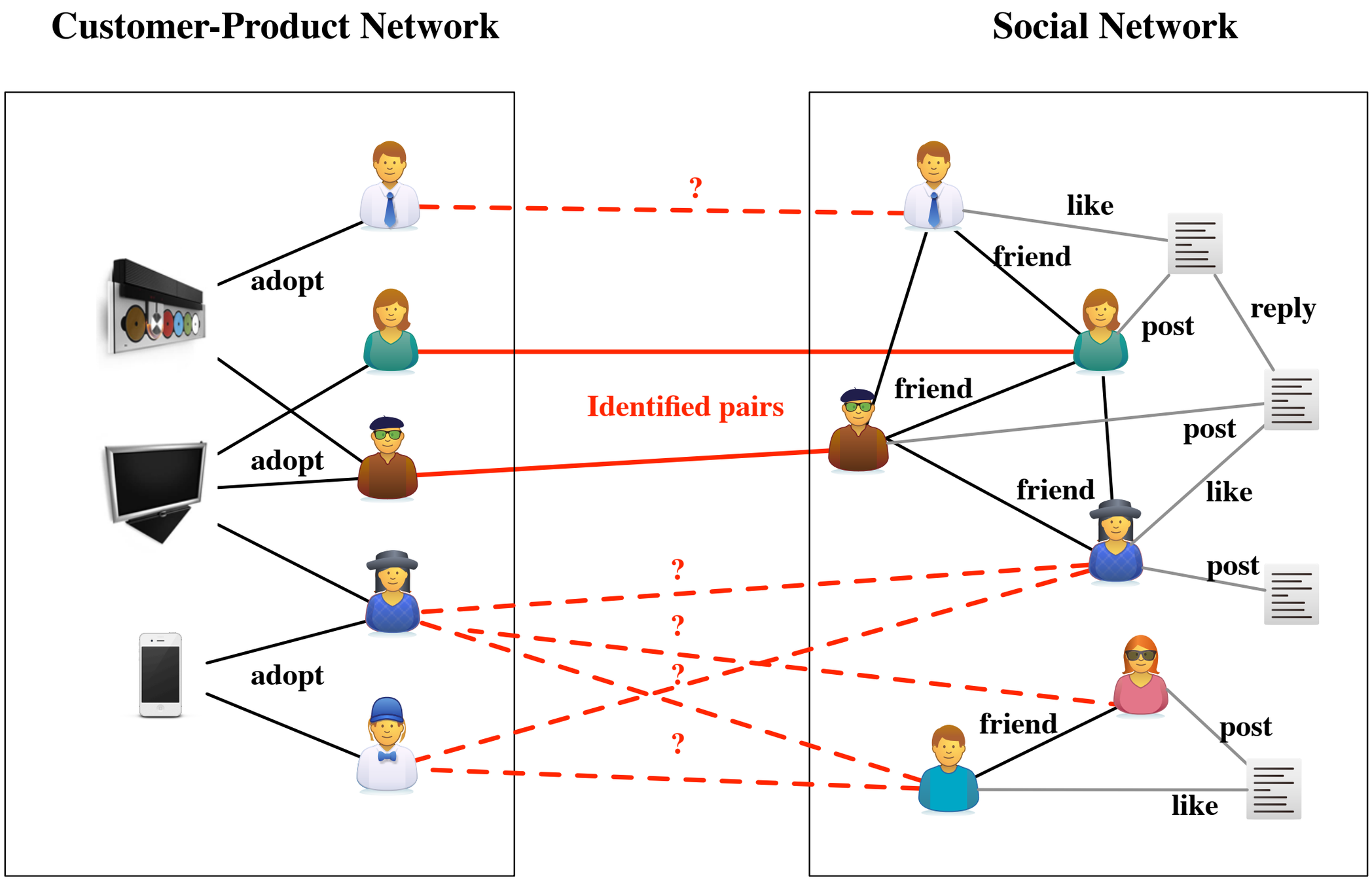


Challenges: 1. Difference in network schema



Challenges: 1. Difference in network schema

2. One-to-one matching



Challenges: **1. Difference in network schema**

2. One-to-one matching

3. Partially aligned networks

Method: Customer-Social Identification (CSI)

Method: Customer-Social Identification (CSI)

1. Modeling user similarity across networks with different schema

Method: Customer-Social Identification (CSI)

1. Modeling user similarity across networks with different schema
   - User profile similarity
   - User interest similarity

Method: Customer-Social Identification (CSI)

1. Modeling user similarity across networks with different schema
   - User profile similarity
   - User interest similarity
2. Identifying customers in partially aligned networks (w.r.t one to one constraint)

Method: Customer-Social Identification (CSI)

1. Modeling user similarity across networks with different schema
   - User profile similarity
   - User interest similarity
2. Identifying customers in partially aligned networks (w.r.t one to one constraint)
   - Find the top-K pairs

User profile similarity

Username*:

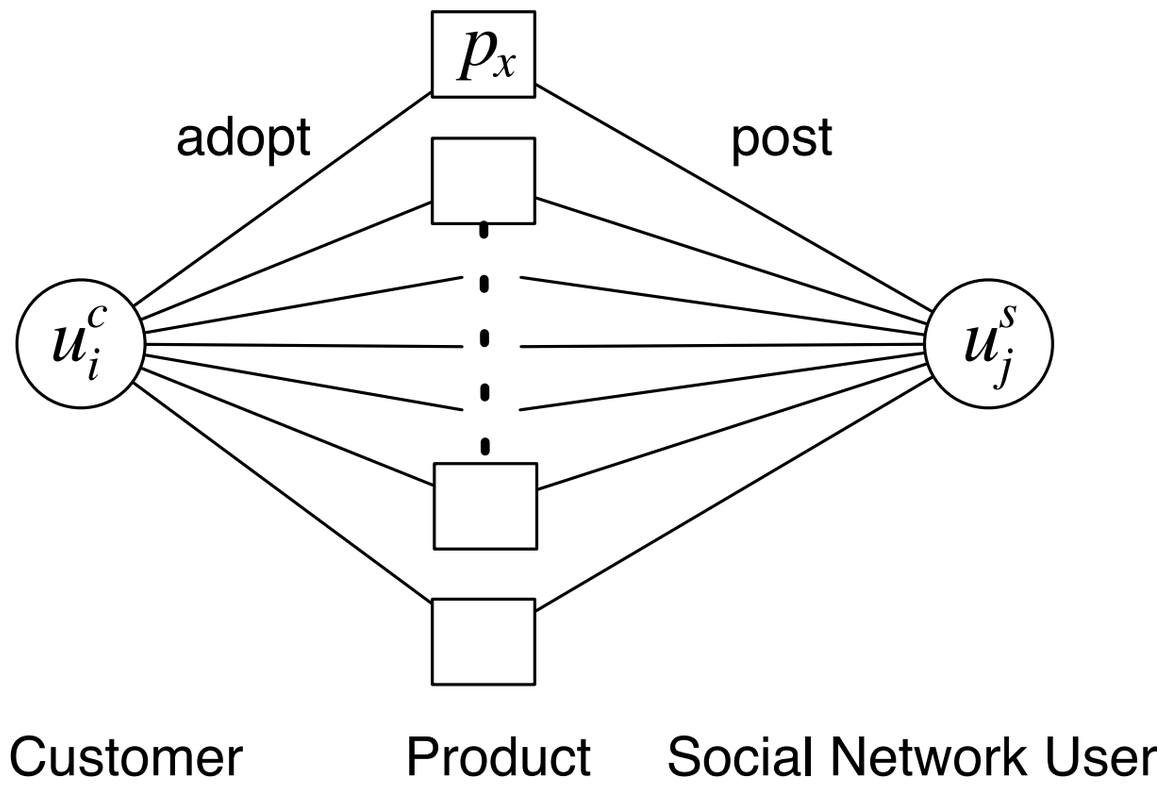
- Exact username match,
- Jaccard similarity
- Distance travled when typing usernames
- Longest common subsequence
- Levenshtein edit distance

Email address:

Matching identities are used for training

* R. Zafarani and H. Liu. Connecting users across social media sites: a behavioral-modeling approach. In KDD'13.

User interest similarity



1) Common Interests (CI):

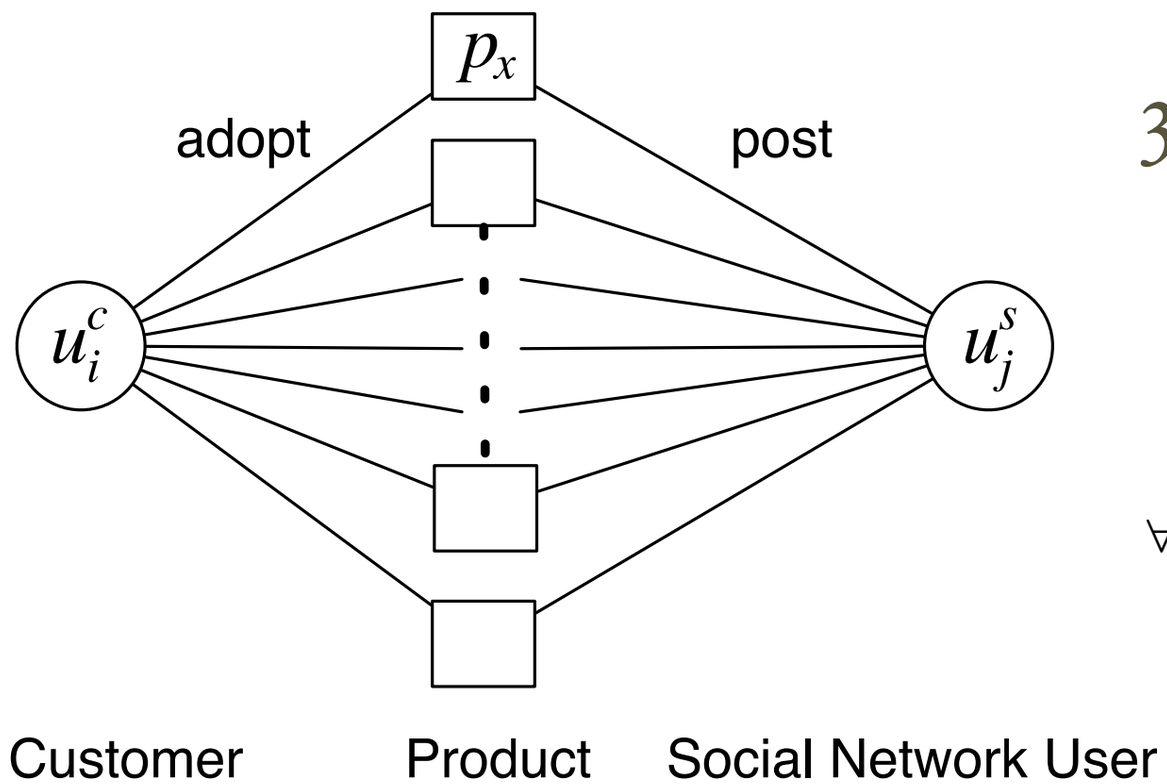
$$|\Gamma_p(u_i^c) \cap \Gamma_p(u_j^s)|$$

2) Jaccard's Coefficient (JC):

$$\frac{|\Gamma_p(u_i^c) \cap \Gamma_p(u_j^s)|}{|\Gamma_p(u_i^c) \cup \Gamma_p(u_j^s)|}$$

$\Gamma_p(u_i^c)$: neighbors of u_i^c

User interest similarity



3) Admic/Adar Index (AA)

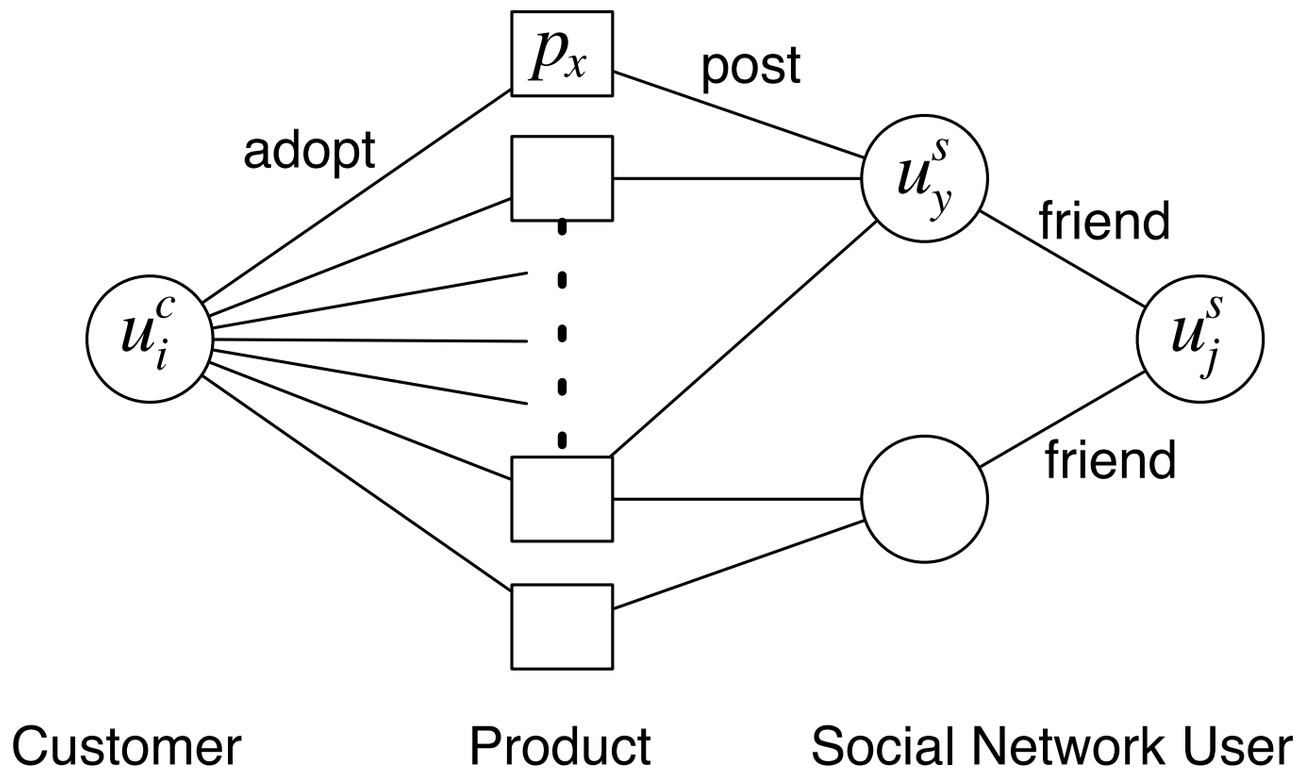
$$\sum_{\forall p_x \in \Gamma_p(u_i^c) \cap \Gamma_p(u_j^s)} log^{-1}\left(\frac{|\Gamma_u^c(p_x)| + |\Gamma_u^s(p_x)|}{2}\right)$$

4) Resource Allocation Index (RA)

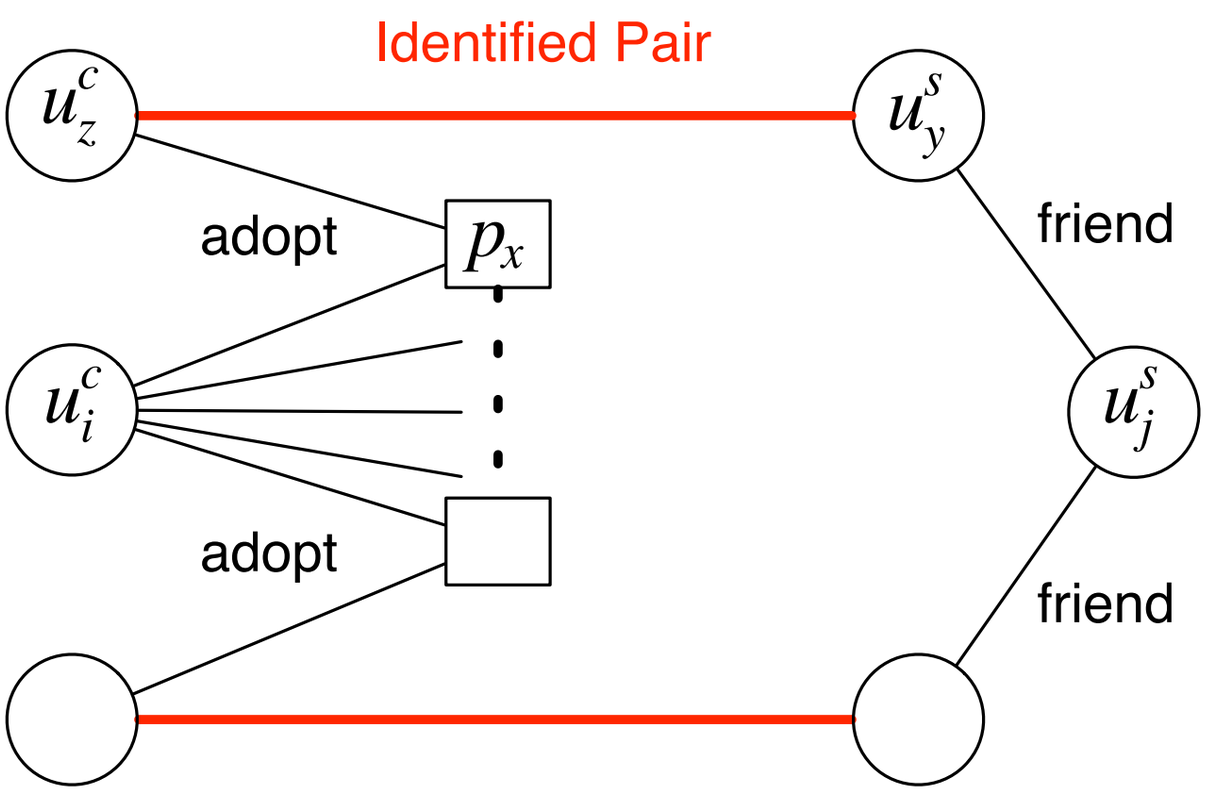
$$\sum_{\forall p_x \in \Gamma_p(u_i^c) \cap \Gamma_p(u_j^s)} \left(\frac{|\Gamma_u^c(p_x)| + |\Gamma_u^s(p_x)|}{2}\right)^{-1}$$

Weighted by the uniqueness of products

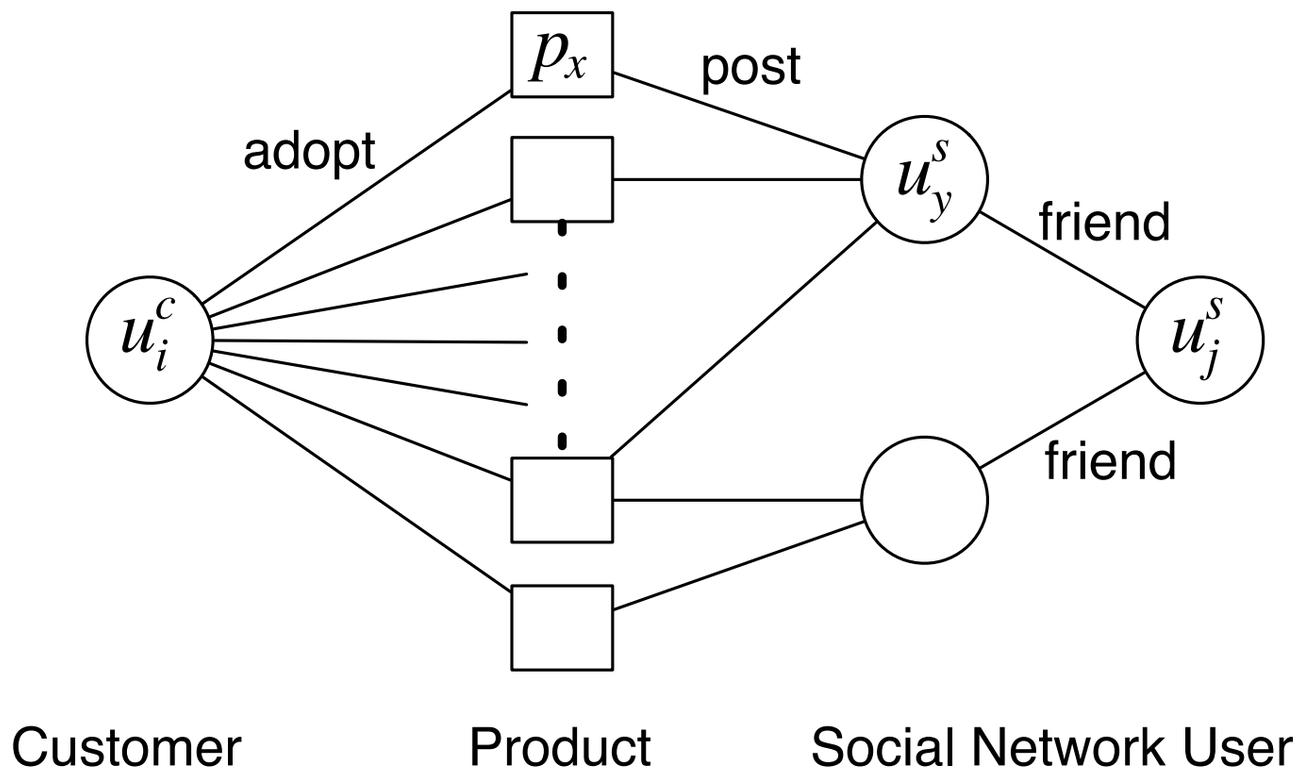
User interest similarity

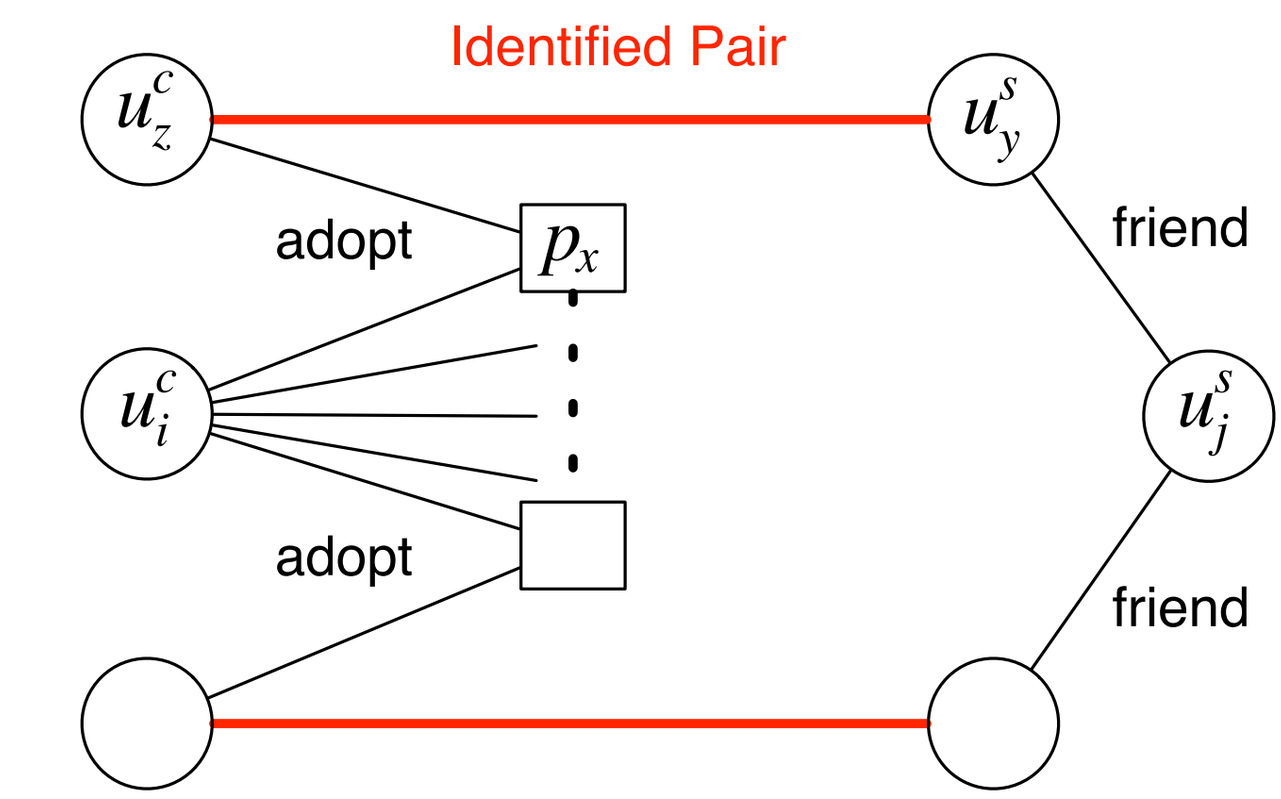


Inactive users' interests can be inferred by friends' interests



User interest similarity





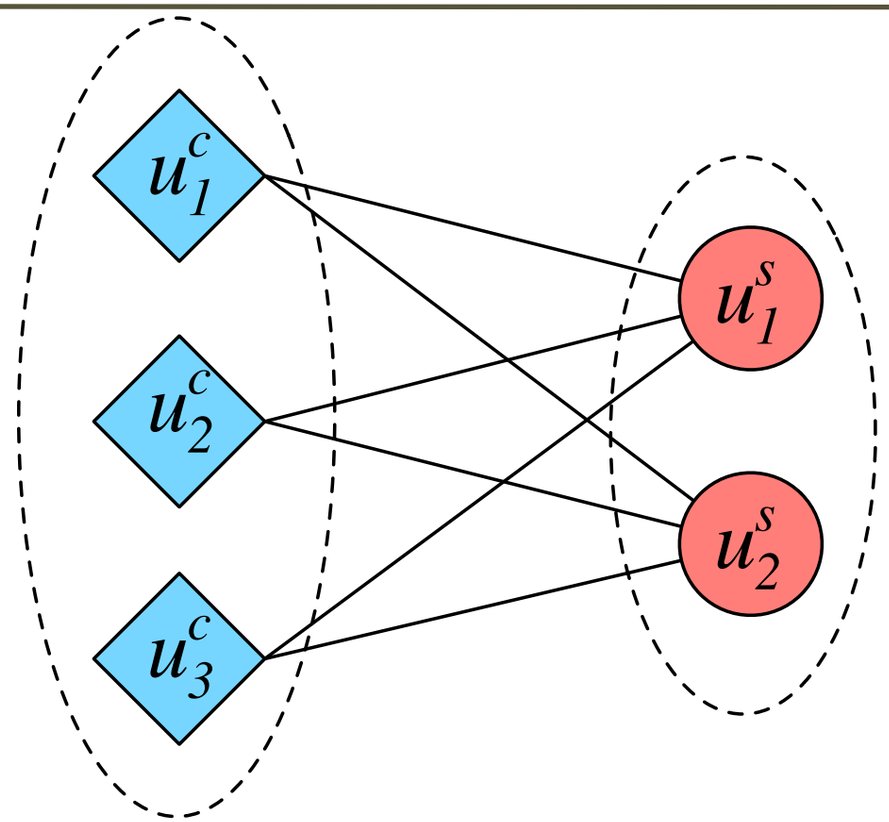
Inactive users' interests can be inferred by friends' interests

5) Katz's Index

$$\sum_{l=1}^{l_{max}} \beta^l \cdot |paths^{\langle l \rangle}_{u_i^c, u_j^s}|$$

Sum of the number of paths. Weighted by the length of the paths.

Identifying customers in partially aligned networks



	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

Similarity score of (u_i^c, u_j^s)

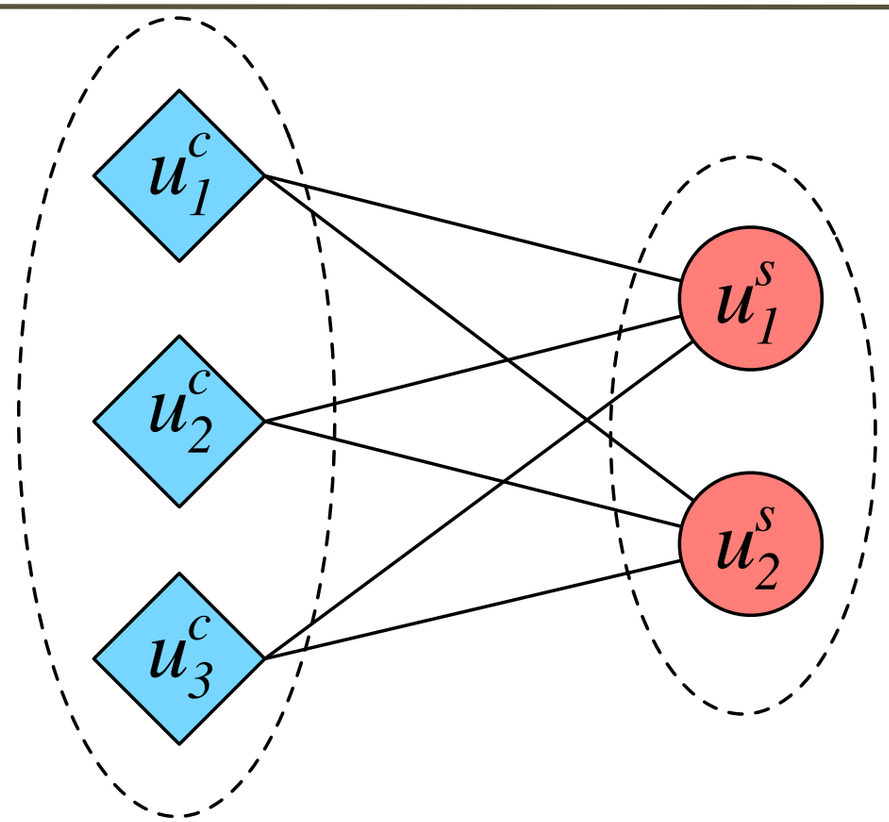
Not one-to-one

Customer

Social Network User

Not the best for

Identifying customers in partially aligned networks



	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

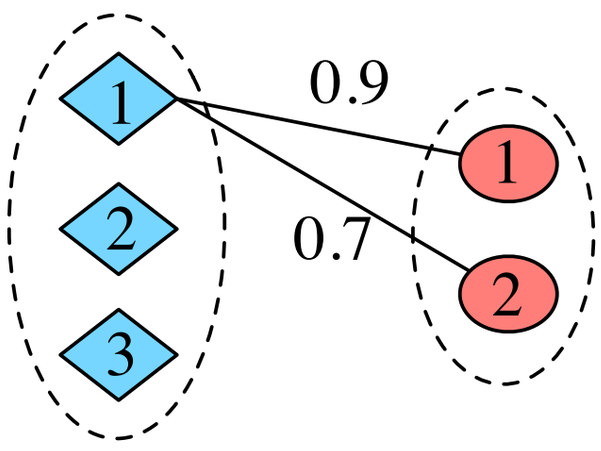
Similarity score of (u_i^c, u_j^s)

Not one-to-one

Customer

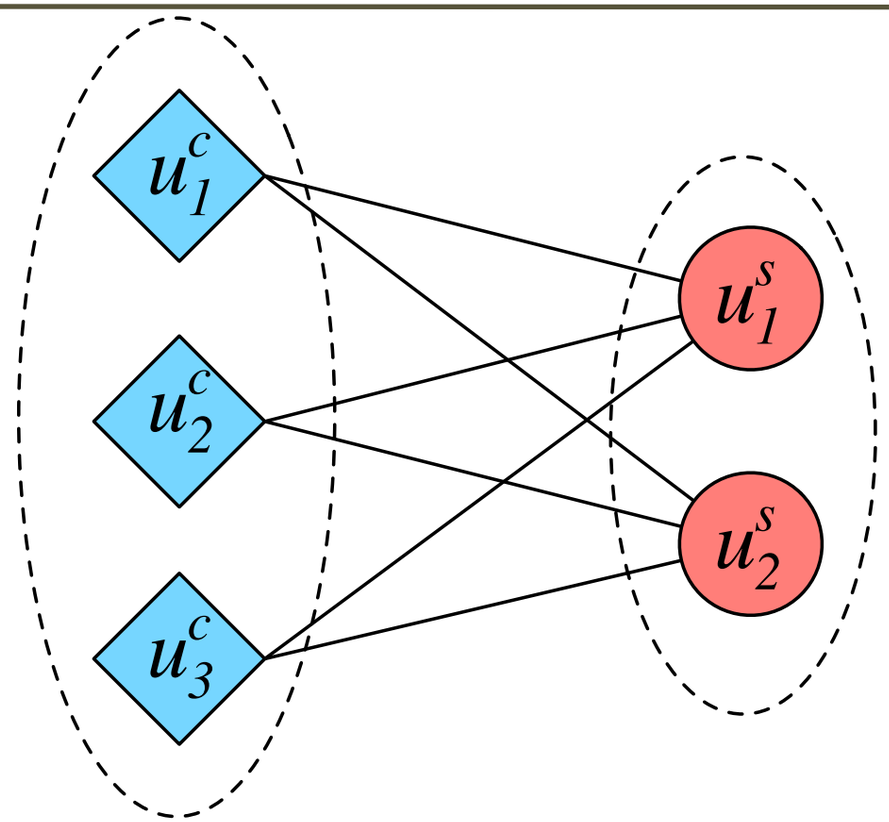
Social Network User

Not the best for



1) link prediction
(threshold = 0.5)

Identifying customers in partially aligned networks



	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

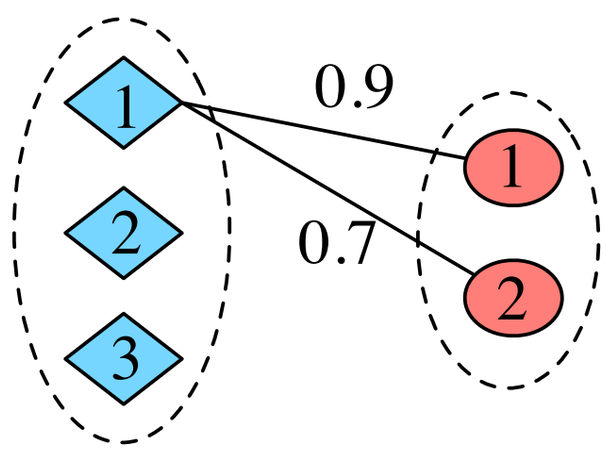
Similarity score of (u_i^c, u_j^s)

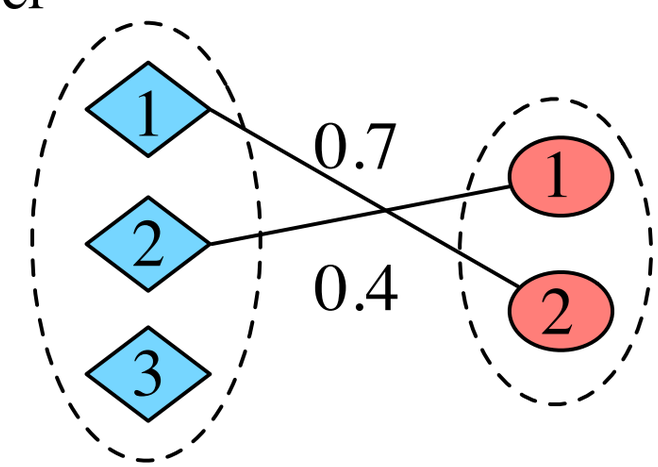
Not one-to-one

Customer

Social Network User

Not the best for

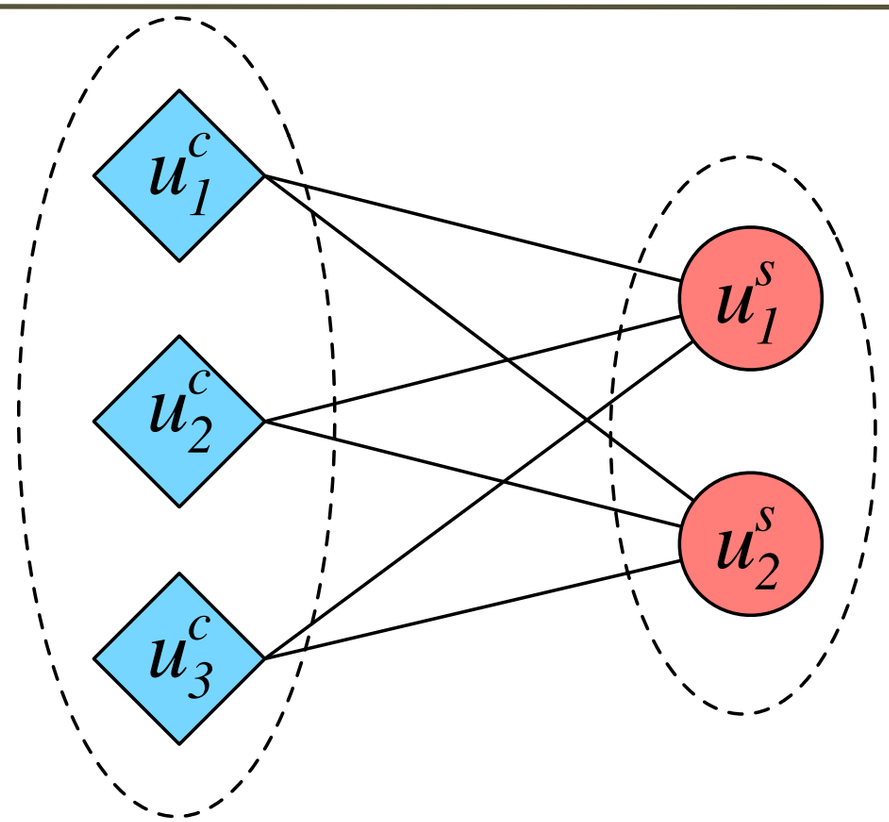




1) link prediction (threshold = 0.5)

2) maximize sum of weights (1:1 constrained)

Identifying customers in partially aligned networks



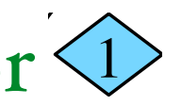
	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

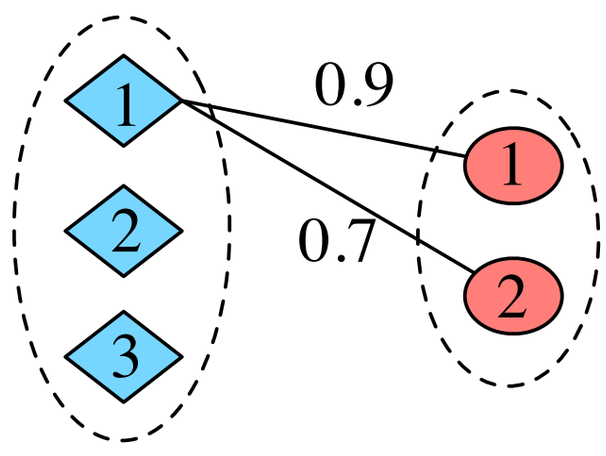
Similarity score of (u_i^c, u_j^s)

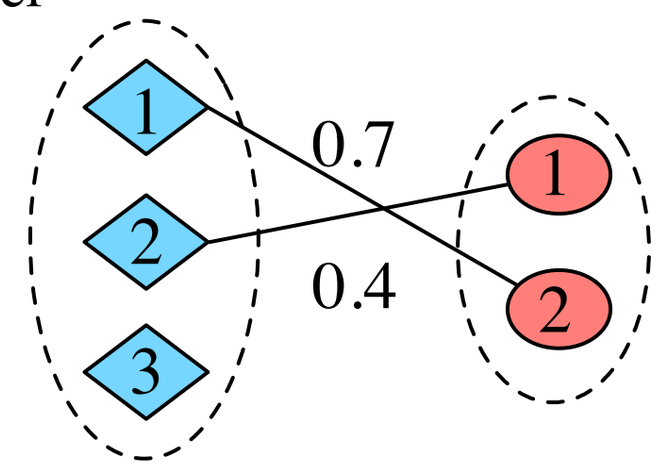
Not one-to-one

Customer

Social Network User

Not the best for 

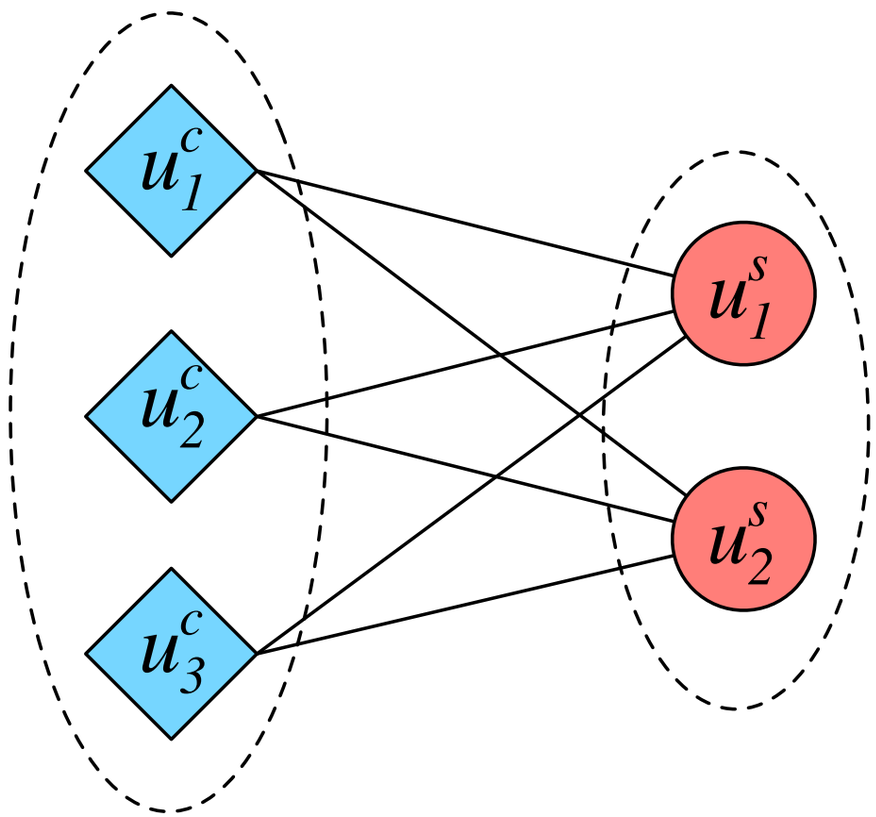




1) link prediction (threshold = 0.5)

2) maximize sum of weights (1:1 constrained)

Identifying customers in partially aligned networks



	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

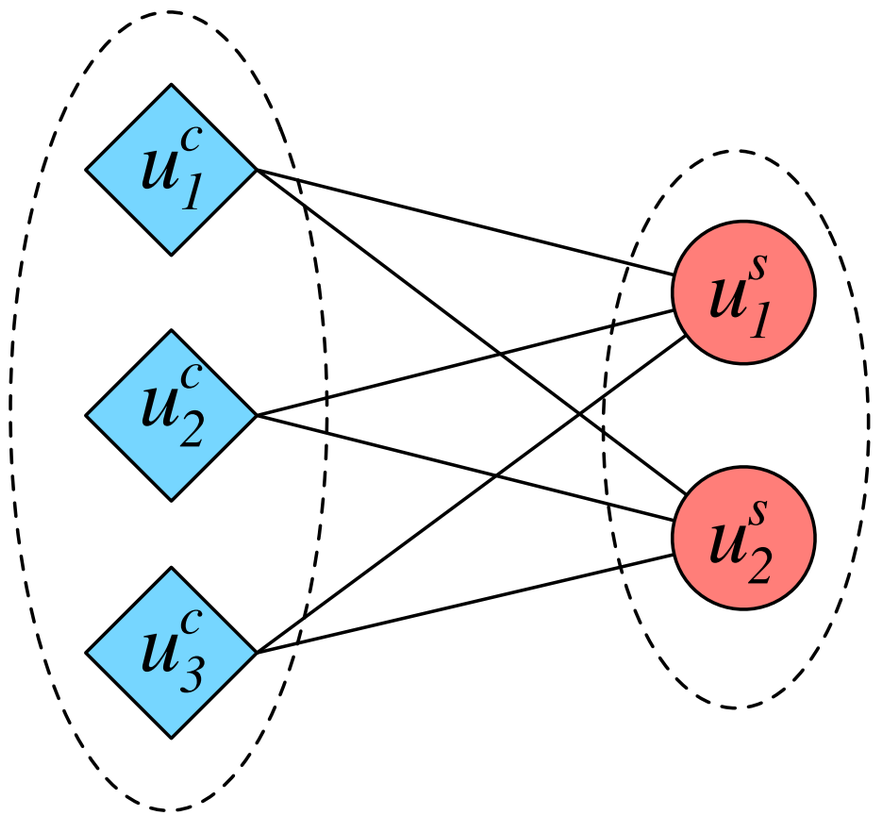
Similarity score of (u_i^c, u_j^s)

CSI:
Find top-K pairs with the highest similarities

Customer

Social Network User

Identifying customers in partially aligned networks

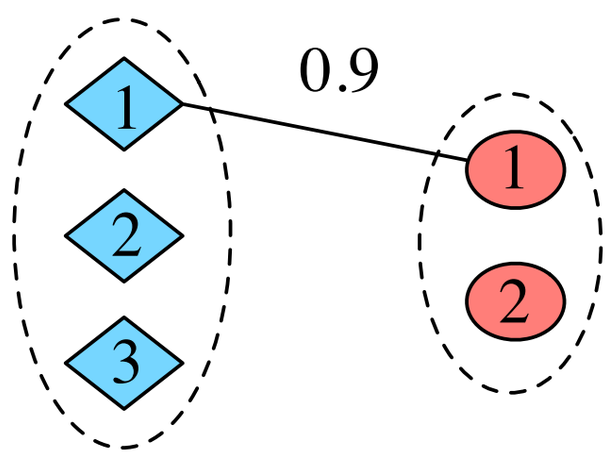


	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

Similarity score of (u_i^c, u_j^s)

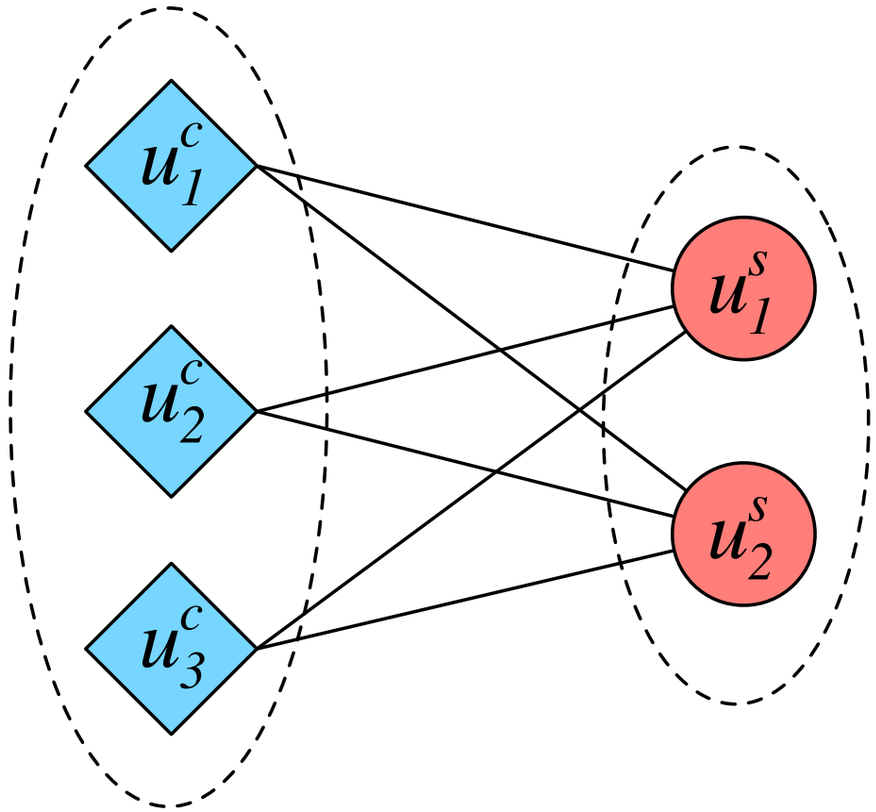
CSI:
Find top-K pairs with the highest similarities

Customer Social Network User



3) CSI method (K=1)

Identifying customers in partially aligned networks

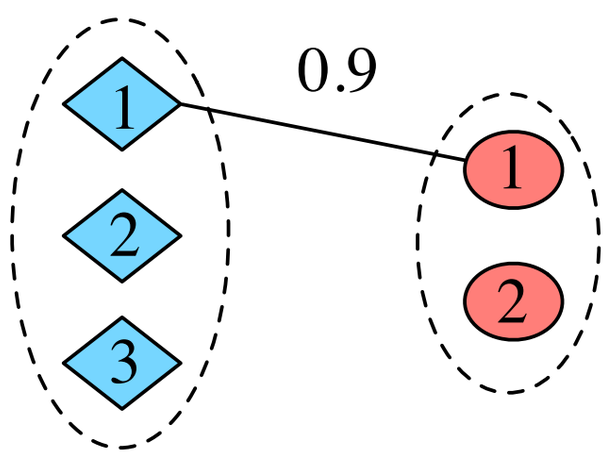


	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

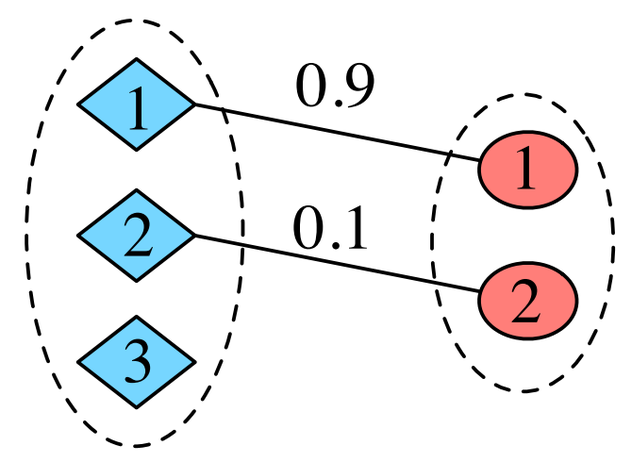
Similarity score of (u_i^c, u_j^s)

CSI:
Find top-K pairs with the highest similarities

Customer Social Network User

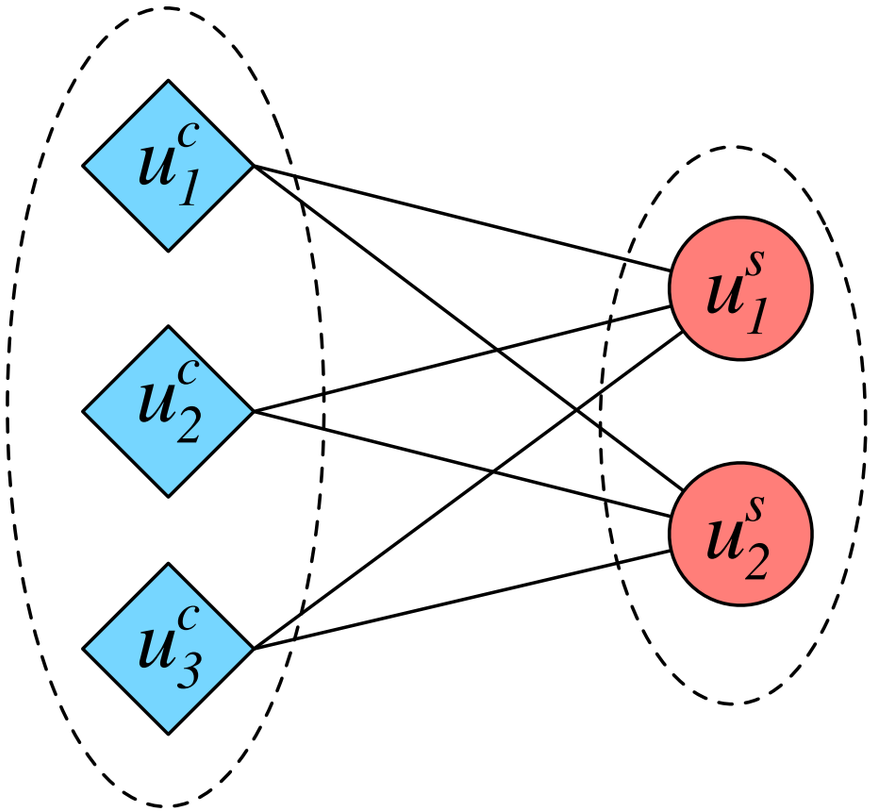


3) CSI method (K=1)



4) CSI method (K=2)

Identifying customers in partially aligned networks

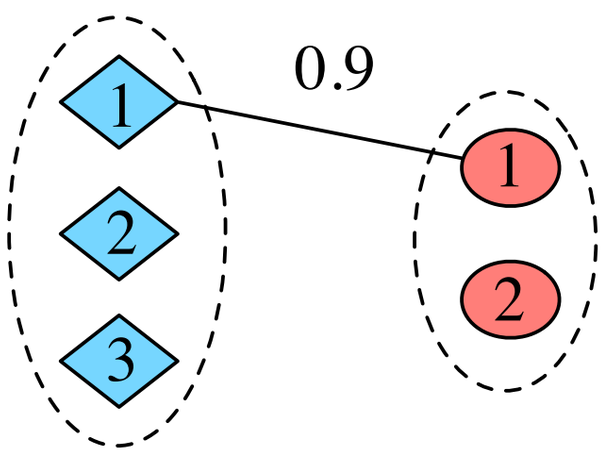


	u_1^s	u_2^s
u_1^c	0.9	0.7
u_2^c	0.4	0.1
u_3^c	0.2	0

Similarity score of (u_i^c, u_j^s)

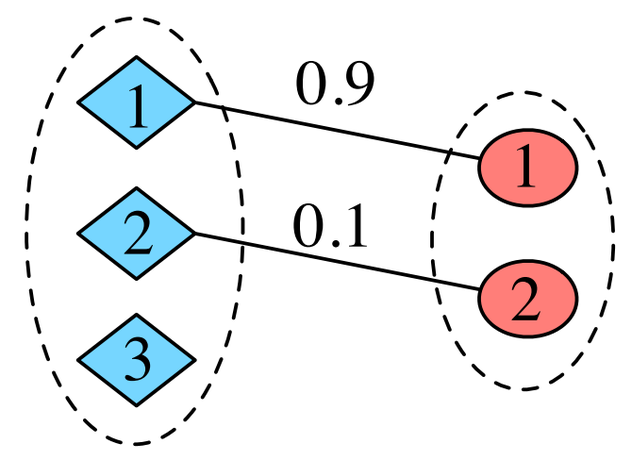
CSI:
Find top-K pairs with the highest similarities

Customer Social Network User



3) CSI method (K=1)

Fact: only one pair



4) CSI method (K=2)

Dataset: Kickstarter and Twitter

Kickstarter-Twitter			
	property	**original networks**	**sampled networks**
# node	projects	3,725	3,725
	customers	545,638	20,514
	social network users	178,792	43,675
# link	adoption	868,050	39,480
	post	234,550	58,988
	social links	5,467,565	513,651
	identified pairs	1,819	1,819
	negative pairs	1.3 billion	93,436

Sample positive pairs and negative pairs

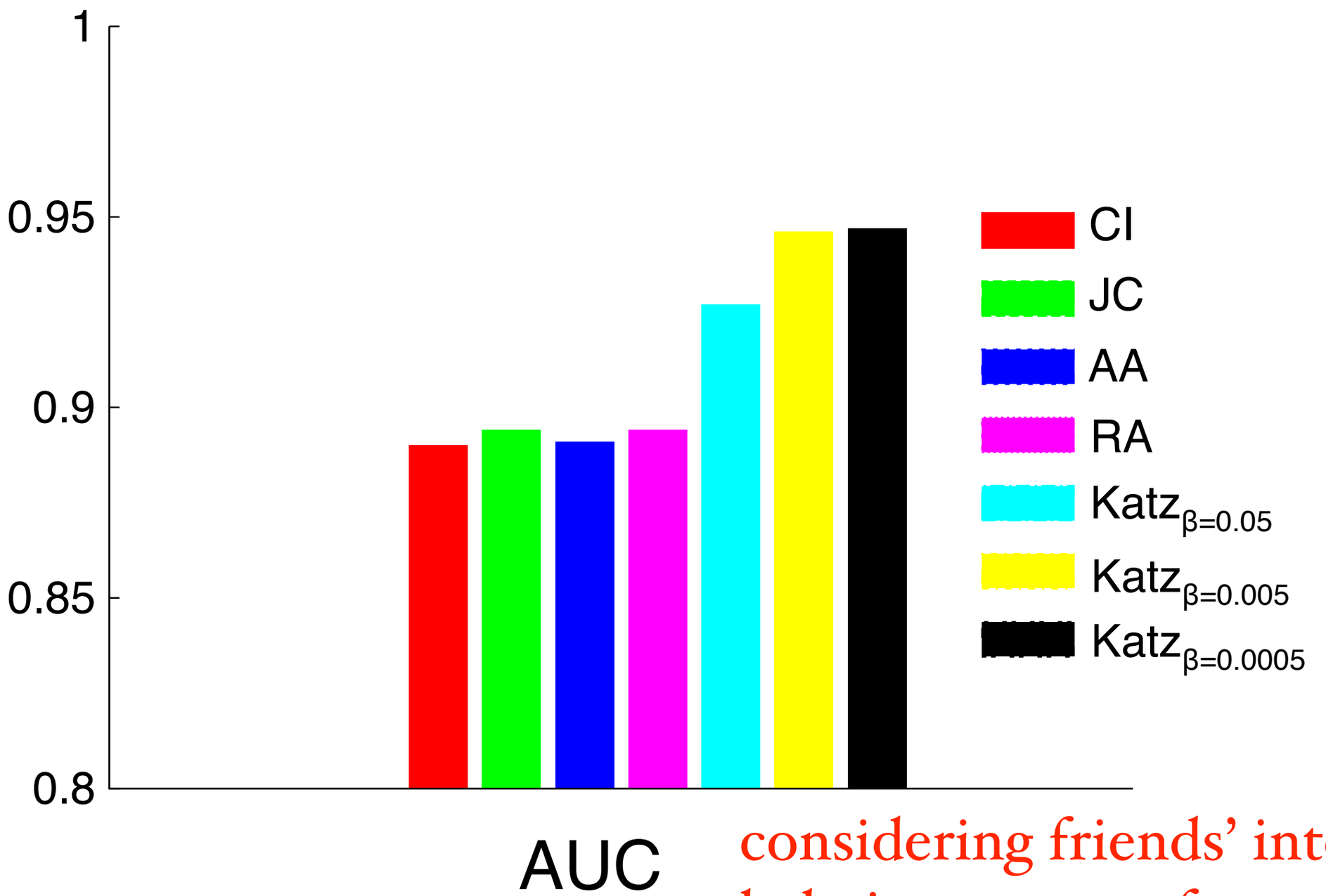
Experiments

Comparative methods

- Unsupervised Link Prediction
  - Common Interests (CI)
  - Jaccard Coefficient (JC)
  - Adamic/ Adar (AA)
  - Resource Allocation (RA)
  - Katz
- Supervised Link Prediction
  - Profile
  - Interest
  - Profile+Interest
- Customer-Social Identification (CSI):

Experiments

- Unsupervised Link Prediction



considering friends' interests help improve performance

Experiments

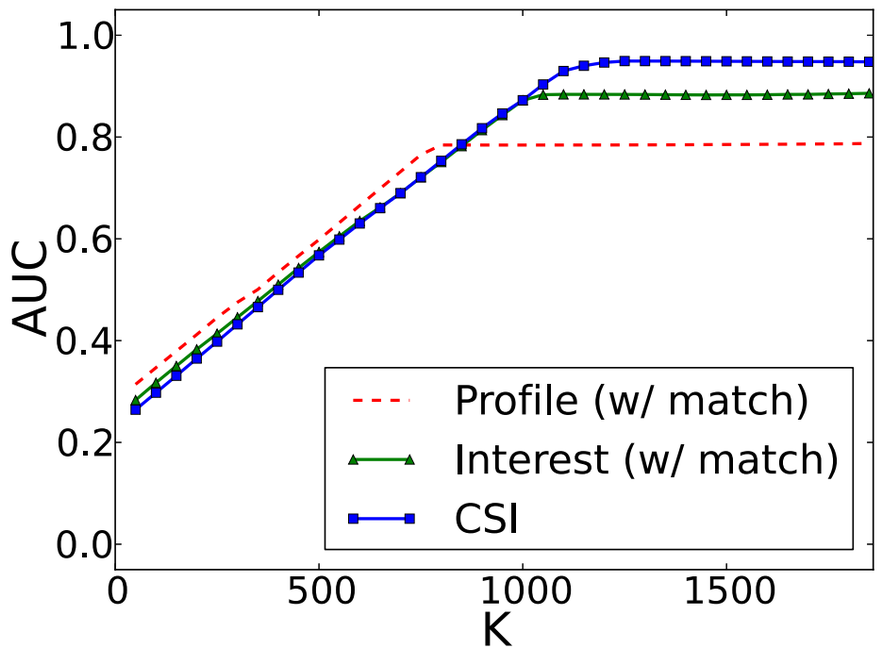
Measure	Methods	Imbalance Ratio 20	Imbalance Ratio 50
F1-score	Profile	0.671 ± 0.003	0.661 ± 0.004
	Interest	0.815 ± 0.011	0.747 ± 0.006
	Profile+Interest	0.875 ± 0.014	0.803 ± 0.017
	CSI	$\mathbf{0.915 \pm 0.003}$	$\mathbf{0.878 \pm 0.004}$
Precision	Profile	$\mathbf{0.977 \pm 0.001}$	$\mathbf{0.935 \pm 0.004}$
	Interest	0.932 ± 0.005	0.868 ± 0.017
	Profile+Interest	0.941 ± 0.005	0.896 ± 0.008
	CSI	0.92 ± 0.003	0.881 ± 0.003
Recall	Profile	0.511 ± 0.004	0.511 ± 0.004
	Interest	0.725 ± 0.019	0.656 ± 0.012
	Profile+Interest	0.818 ± 0.027	0.729 ± 0.033
	CSI	$\mathbf{0.91 \pm 0.003}$	$\mathbf{0.875 \pm 0.004}$
AUC	Profile	0.791 ± 0.001	0.792 ± 0.001
	Interest	0.933 ± 0.019	0.903 ± 0.018
	CSI	$\mathbf{0.958 \pm 0.004}$	$\mathbf{0.958 \pm 0.003}$

$$\text{Imbalance Ratio} = \frac{\#\text{negative instances}}{\#\text{positive instances}}$$

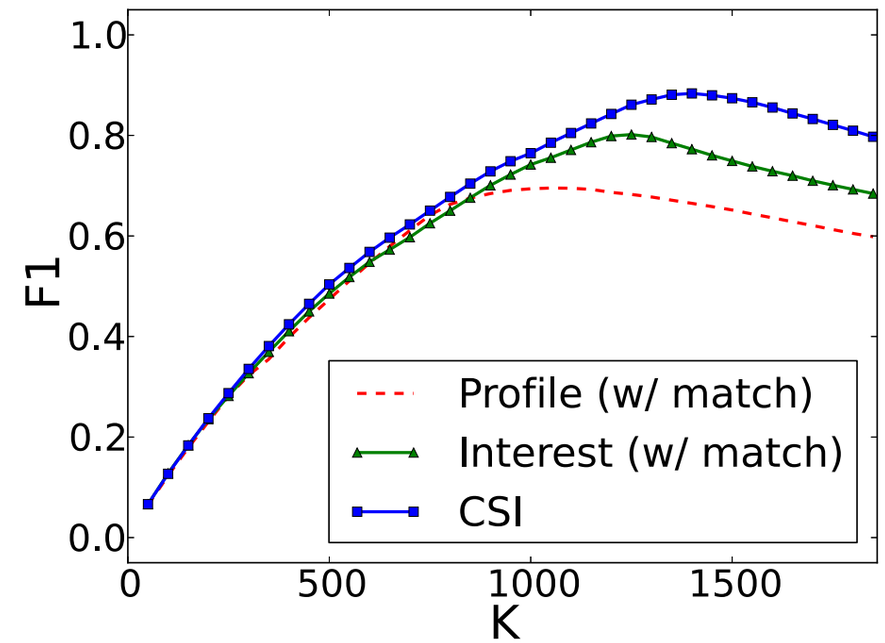
Experiments

Measure	Methods	Training Pairs (%)	
		20%	60%
F1-score	Profile	0.342 ± 0.007	0.661 ± 0.004
	Interest	0.486 ± 0.044	0.695 ± 0.012
	Profile+Interest	0.620 ± 0.029	0.771 ± 0.014
	CSI	$\mathbf{0.875 \pm 0.005}$	$\mathbf{0.877 \pm 0.006}$
Precision	Profile	$\mathbf{0.911 \pm 0.006}$	$\mathbf{0.936 \pm 0.002}$
	Interest	0.900 ± 0.030	0.892 ± 0.003
	Profile+Interest	0.938 ± 0.016	0.900 ± 0.006
	CSI	0.875 ± 0.005	0.876 ± 0.006
Recall	Profile	0.211 ± 0.006	0.511 ± 0.004
	Interest	0.335 ± 0.041	0.569 ± 0.016
	Profile+Interest	0.464 ± 0.033	0.674 ± 0.022
	CSI	$\mathbf{0.875 \pm 0.005}$	$\mathbf{0.877 \pm 0.006}$
AUC	Profile	0.773 ± 0.025	0.794 ± 0.002
	Interest	0.903 ± 0.018	0.894 ± 0.002
	CSI	$\mathbf{0.958 \pm 0.003}$	$\mathbf{0.958 \pm 0.004}$

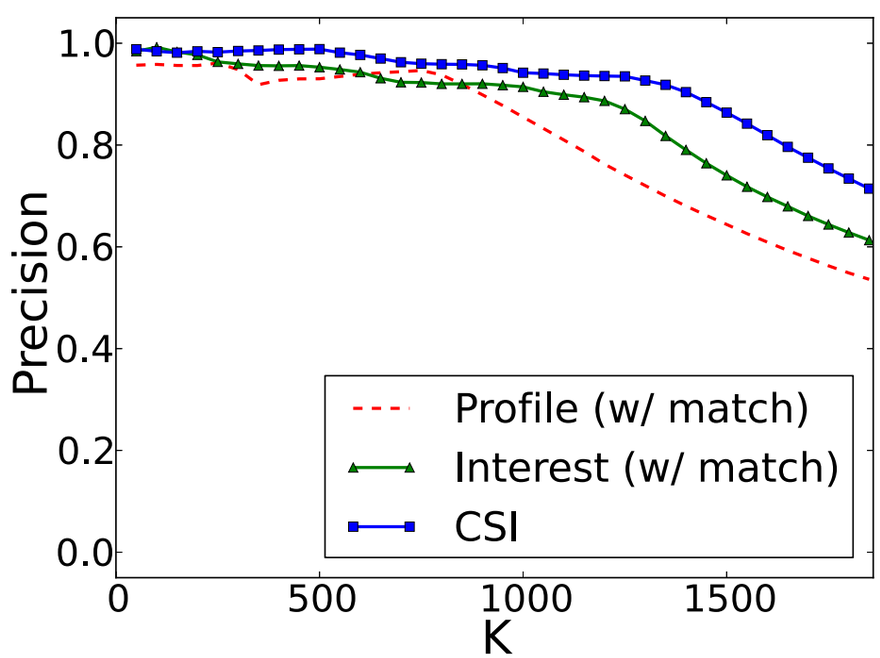
Experiments



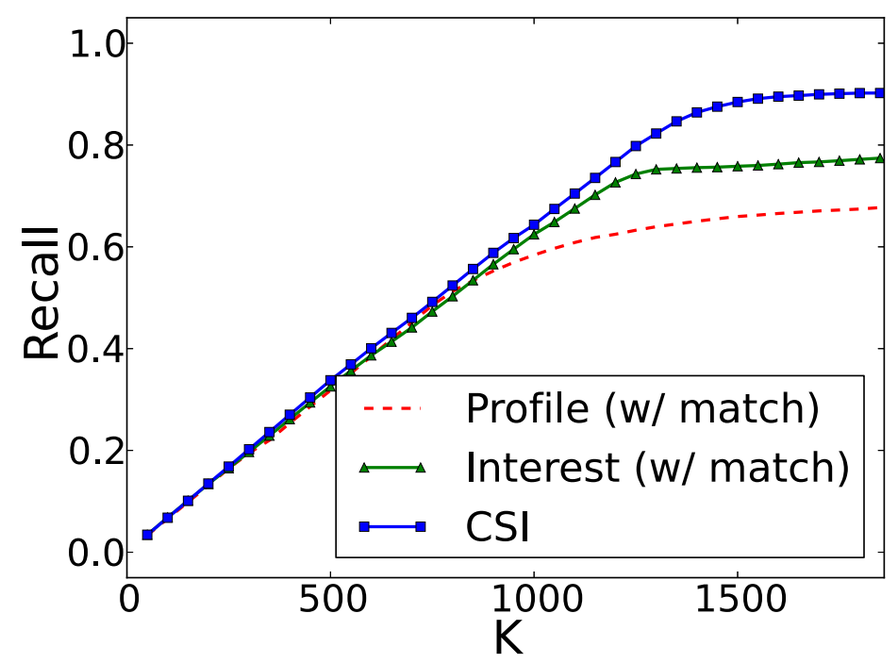
(a) AUC



(b) F1

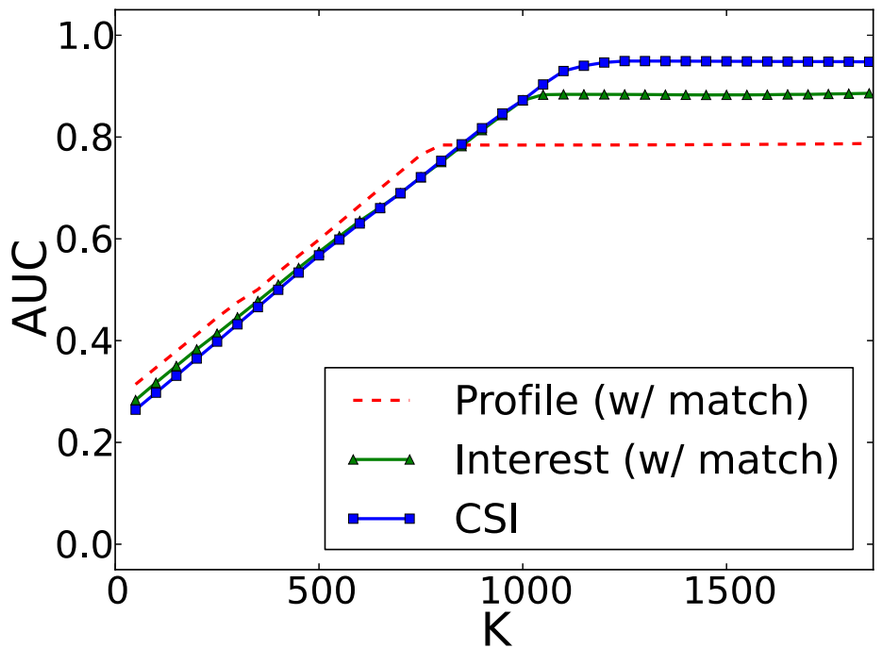


(c) Precision

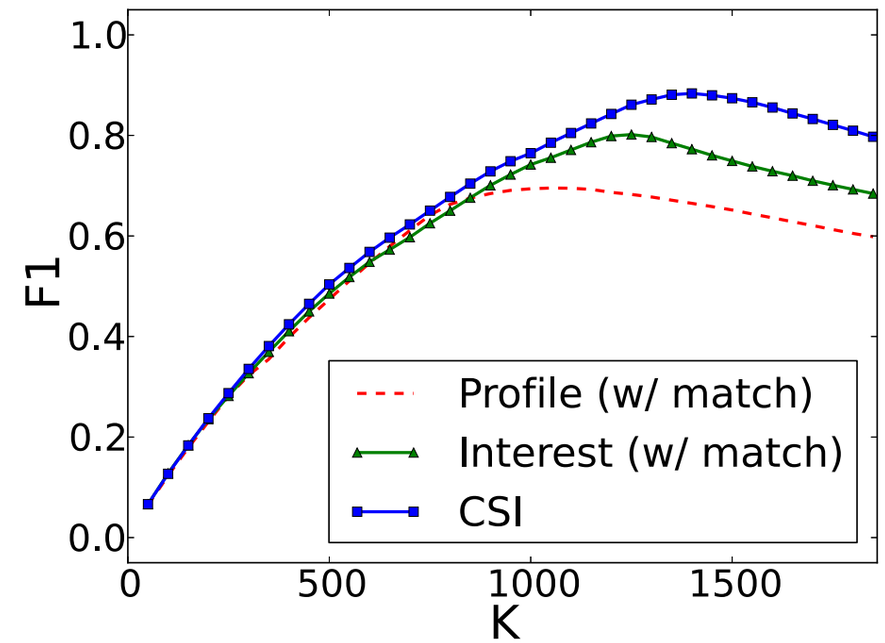


(d) Recall

Experiments

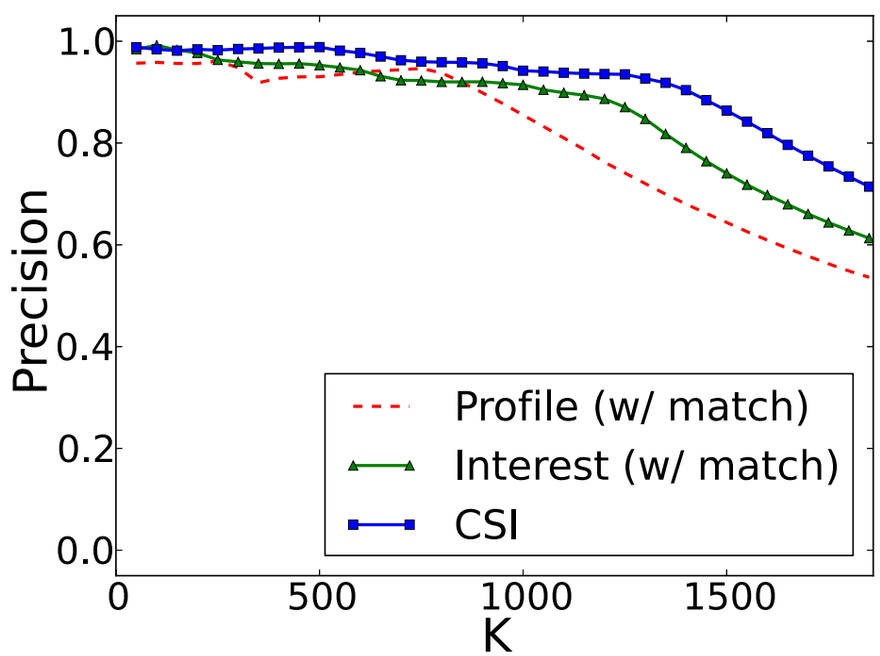


(a) AUC

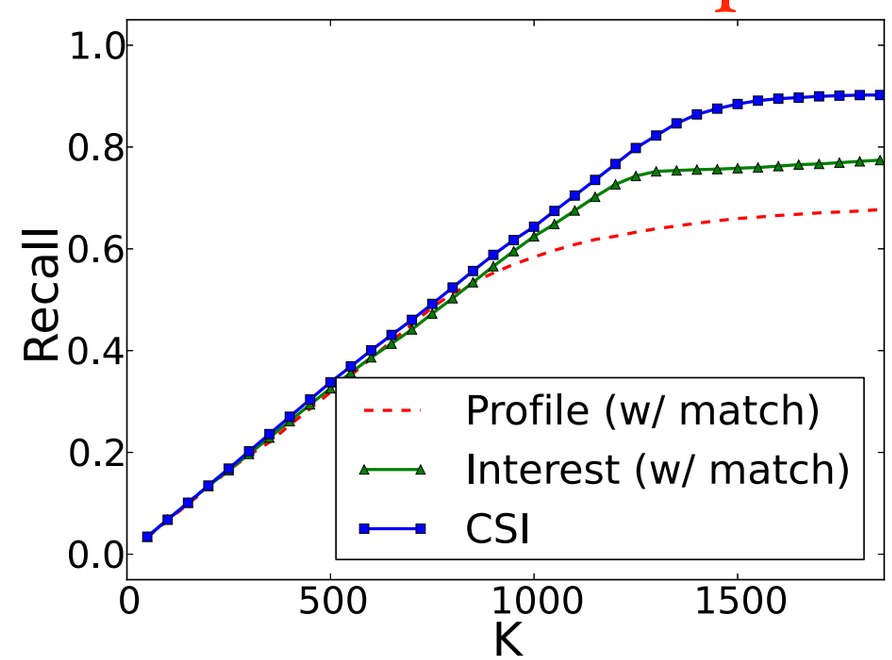


(b) F1

achieve the best when K = # true pairs



(c) Precision



(d) Recall

Conclusion

1. Profile similarities and interest similarities are both useful for identifying customers

2. Friends' interests help improve similarity measurements

3. Proposed CSI method is stable even using high imbalanced training sets or using very few training instances.

Thank you!