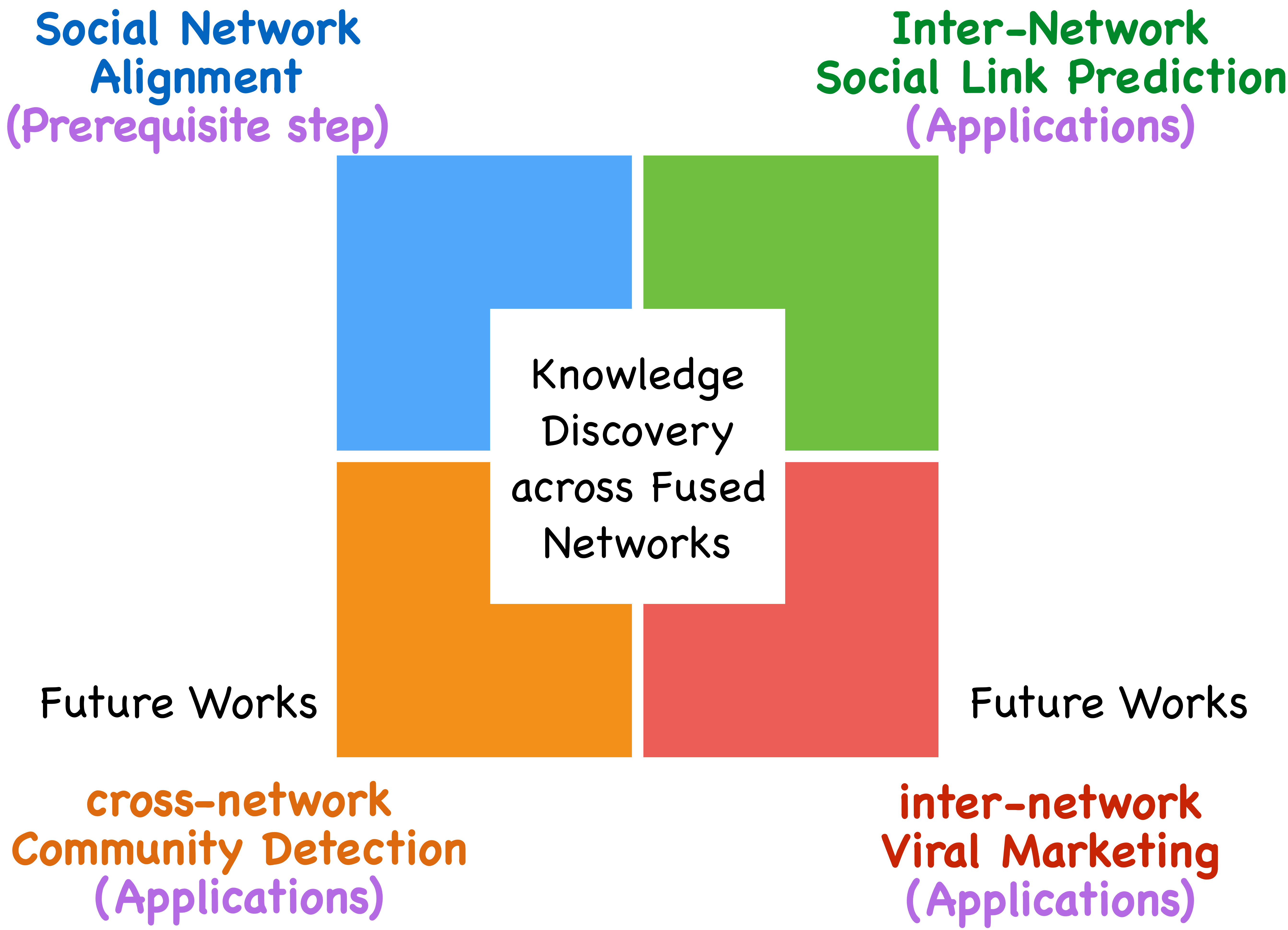
UIC Department of Computer Science

Fusion of Heterogeneous Social Networks for Synergistic Knowledge Discovery

Presenter: Jiawei Zhang
Date: March 14, 2016

Committee Members: Prof. Philip S. Yu (Chair),
Prof. Piotr J. Gmytrasiewicz,
Prof. Xinhua Zhang,
Prof. Yuheng Hu,
Prof. Xiangnan Kong

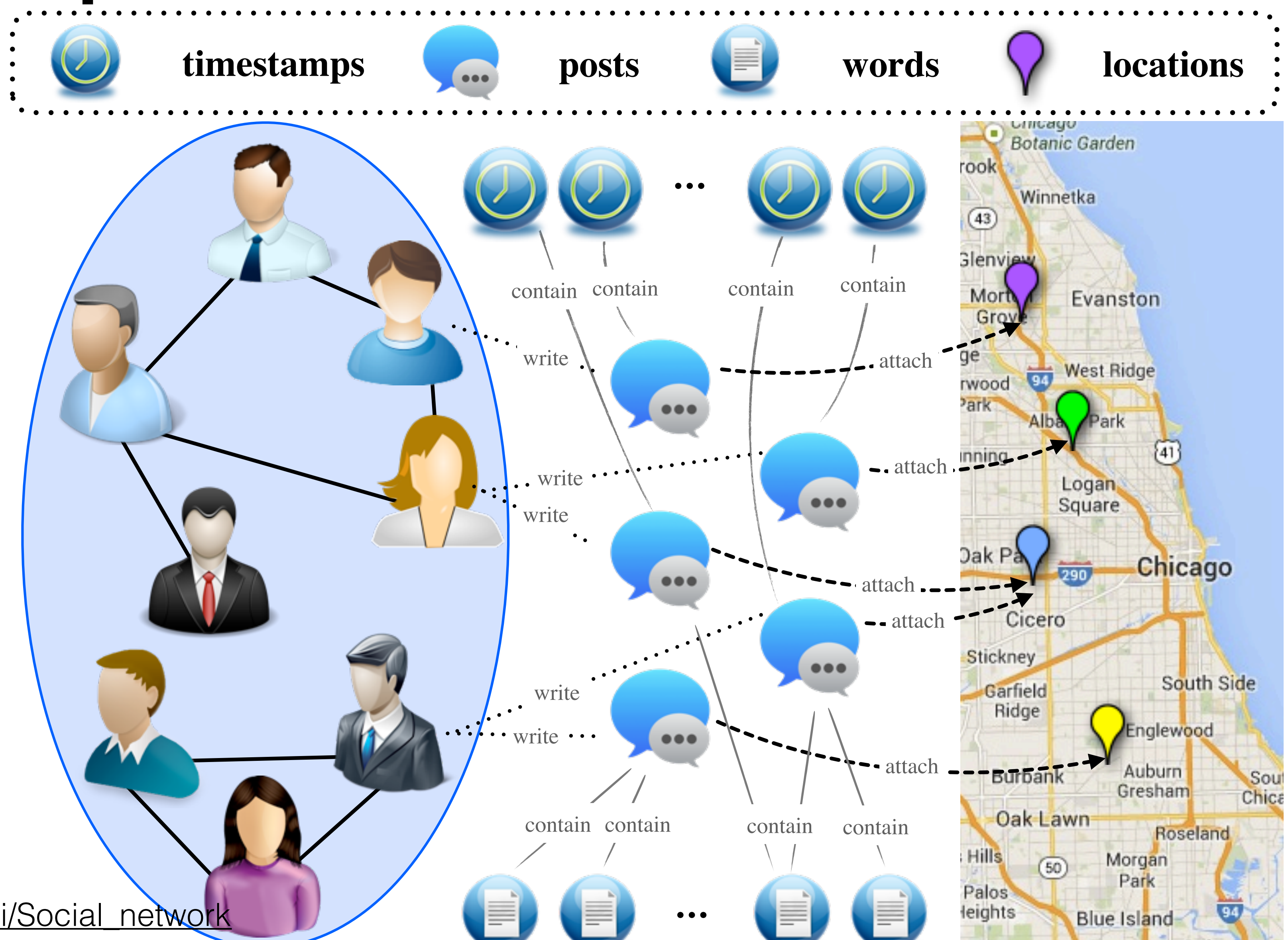

Social Network
Alignment
(Prerequisite step)
Inter-Network
Social Link Prediction
(Applications)
Knowledge
Discovery
across Fused
Networks
Future Works
Future Works
cross-network
Community Detection
(Applications)
inter-network
Viral Marketing
(Applications)

Outline

1. Background Knowledge and Basic Concepts
2. Problem 1: Network Alignment
3. Problem 2: Social Link Prediction across Networks
4. Future work 1: Community Detection across Networks
5. Future work 2: Viral Marketing across Networks
6. Summary and Selected References

Basic Concept Social Networks

- **Online Social Networks**
  - **Definition:** An online social network is a social structure made up of a set of social actors and a set of ties between these actors[1].
  - **Example:**



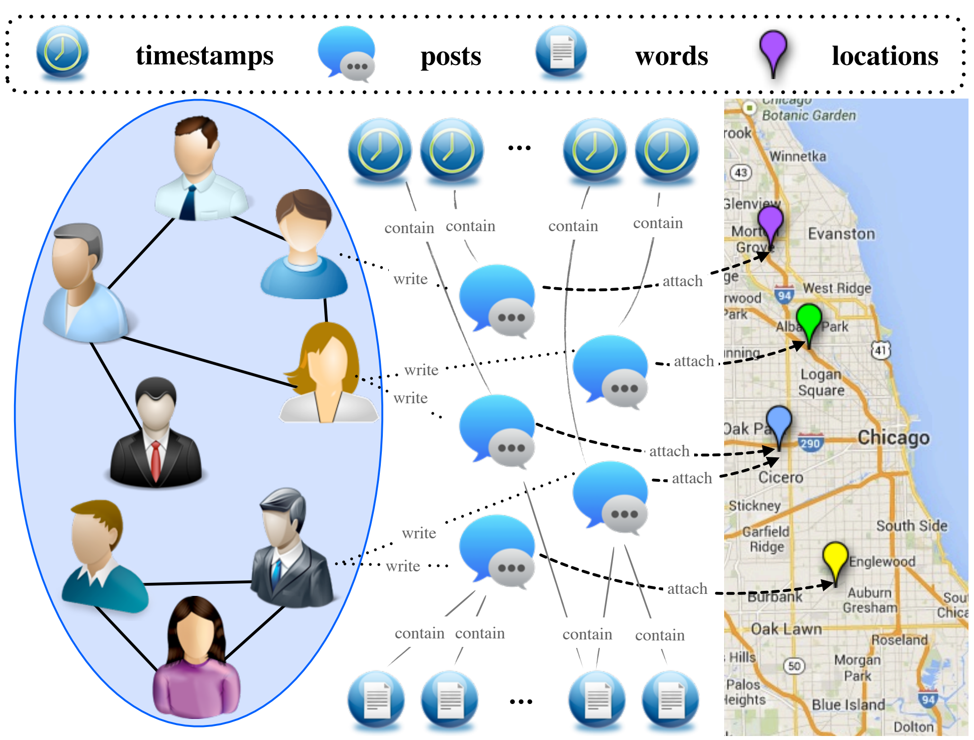
[1]: http://en.wikipedia.org/wiki/Social_network

Basic Concept Social Networks

- Representation: Heterogeneous Information Network

$$G = (V, E)$$

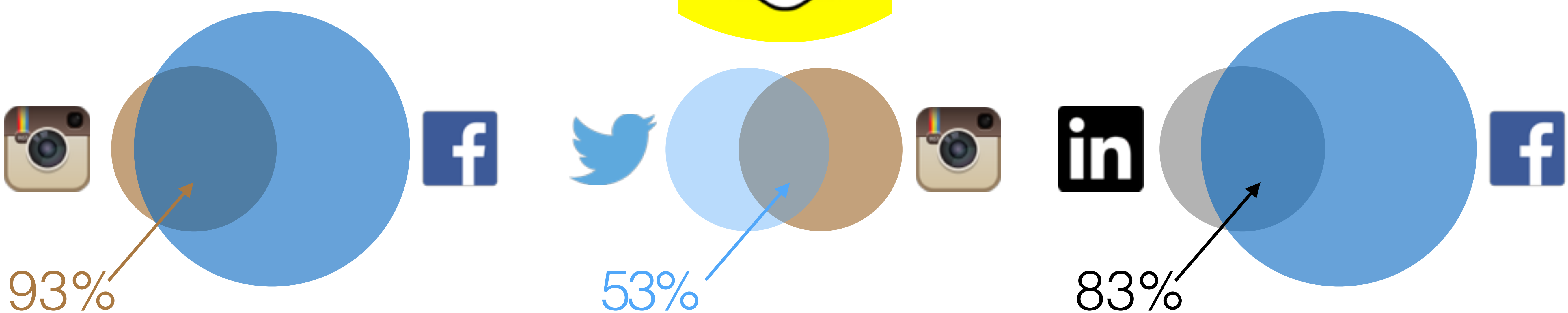
where $V = \bigcup_i V_i$ is the sets of various kinds of nodes in the network and V_i is the i_{th} kind of nodes in $\boldsymbol{G}$; $E = \bigcup_j E_j$ is the sets of various types of links in the network and E_j is the j_{th} kind of link in $\boldsymbol{G}$.



V = {user node set, post node set, word node set, time node set, location node set}

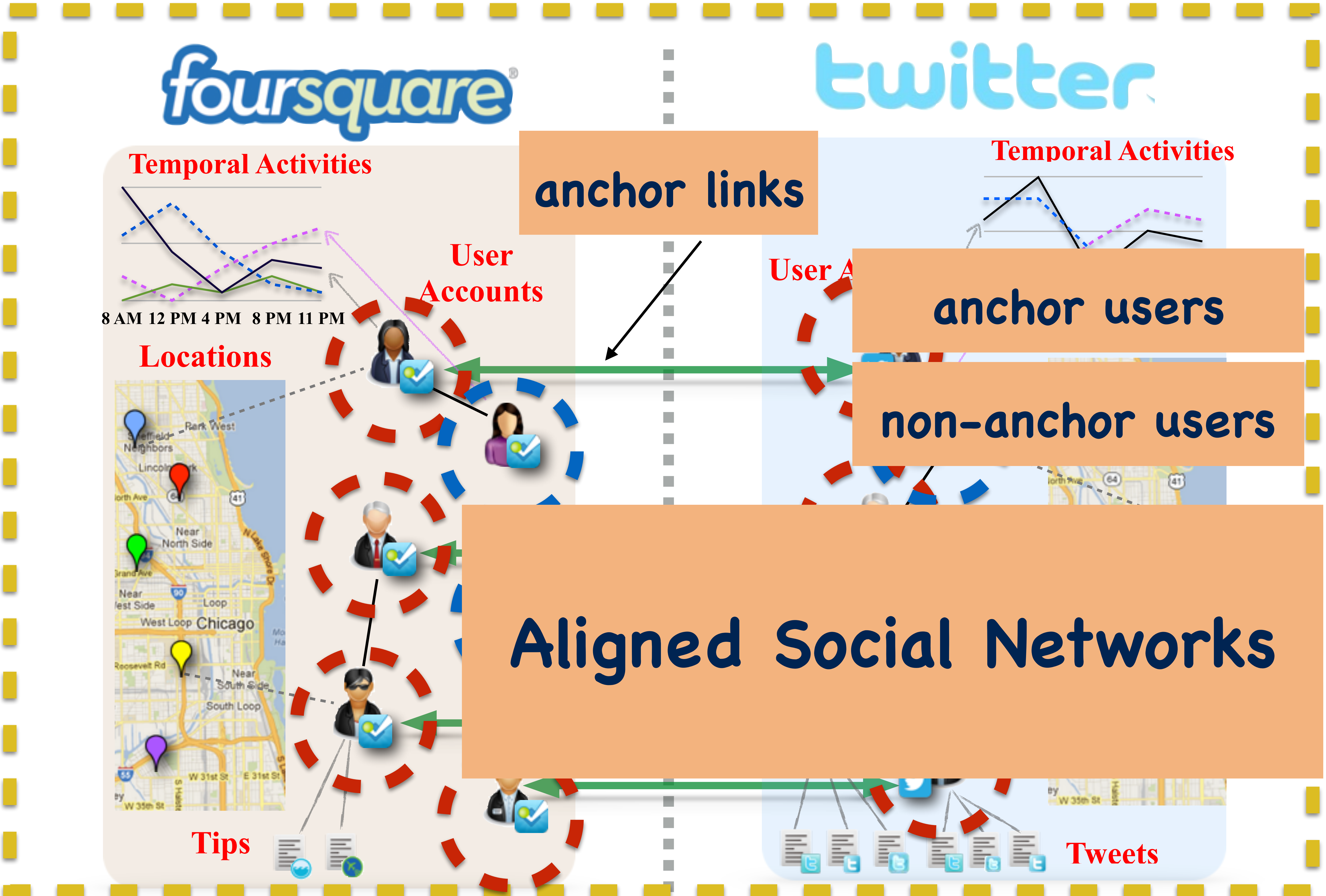
E = {user-user link set, user-post link set, post-word link set, post-time link set, post-location link set}

People are using multiple online 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.

Multiple Aligned Social Networks



Multiple Aligned Social Networks

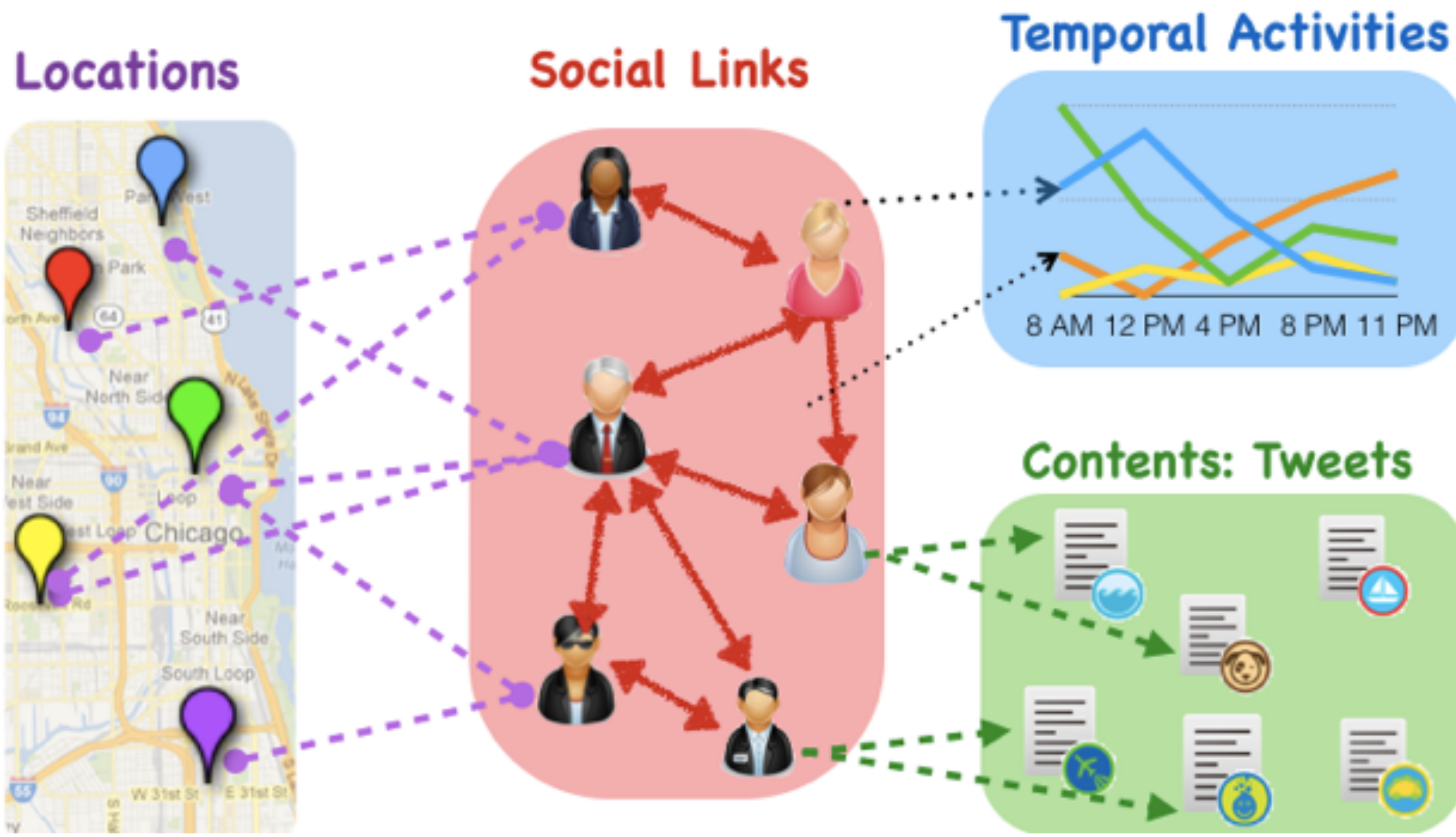
- **Multi Aligned Heterogeneous Social Networks**
  - **Definition:**

$$\mathcal{G} = (G_{set}, A_{set})$$

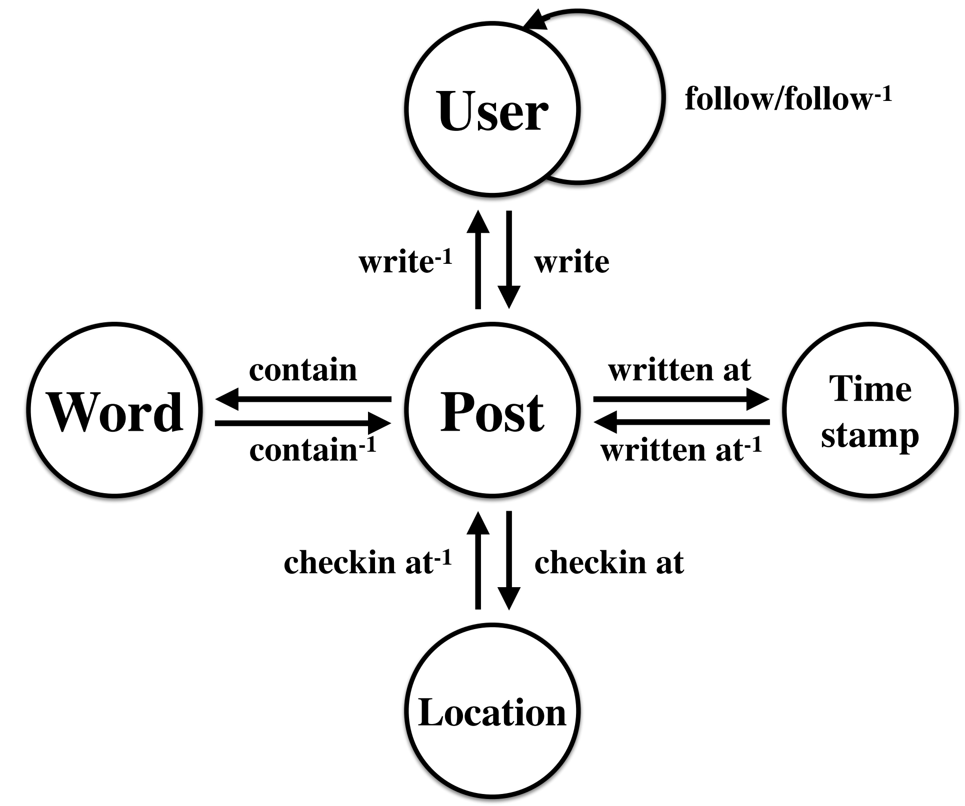
where $G_{set} = \{G^{(1)}, G^{(2)}, \cdots, G^{(|G_{set}|)}\}$ is the set of $|G_{set}|$ different ***heterogeneous information networks***;

$A_{set} = \{A^{(1,2)}, A^{(1,3)}, \cdots, A^{(|G_{set}|,|G_{set}|-1)}\}$ is the set of undirected ***anchor links*** among networks.

Network Schema and Intra-Network Social Meta Path

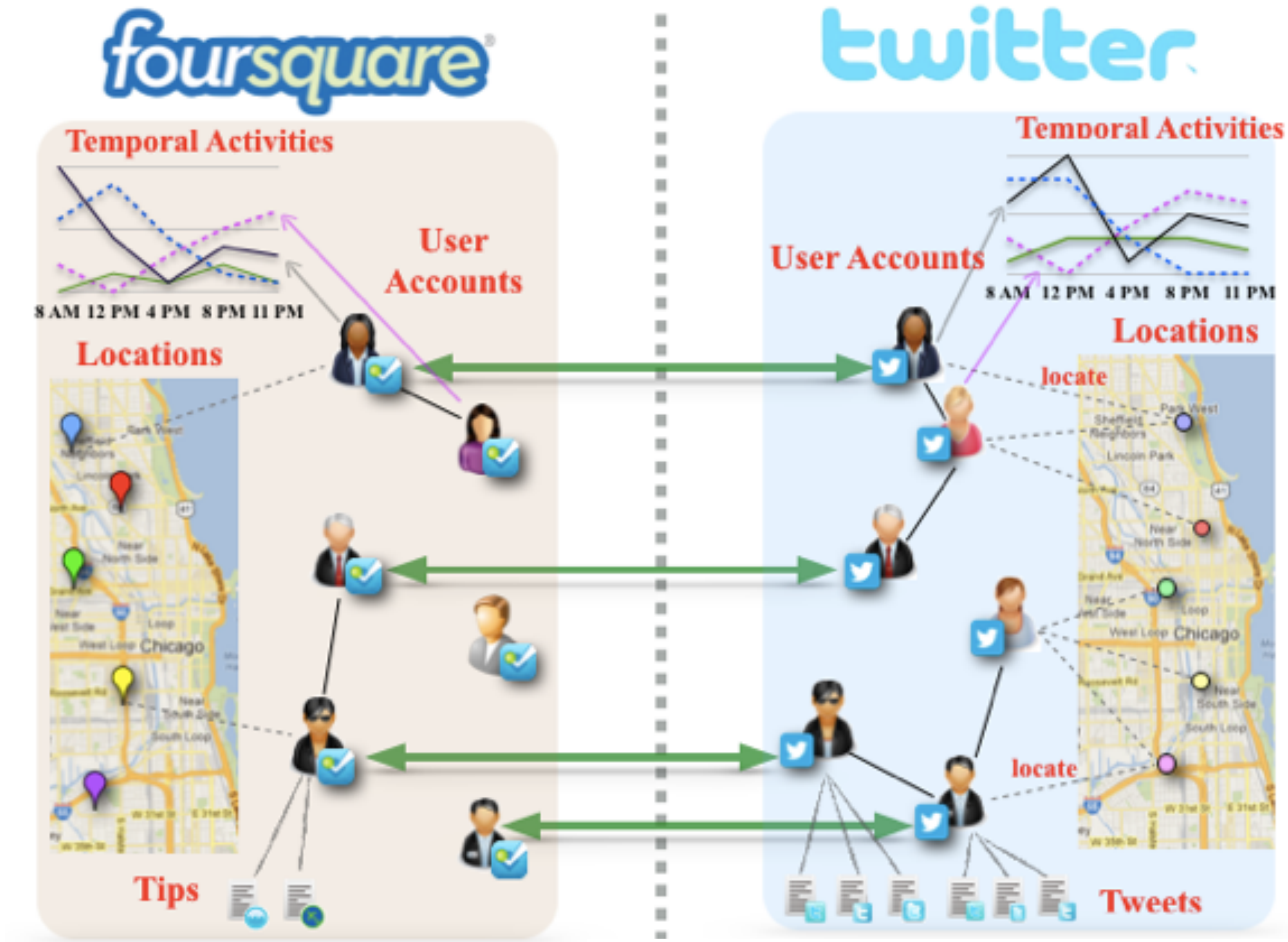


network schema



ID	Notation	Heterogeneous Network Meta Path	Semantics
1	$U \to U$	User $\xrightarrow{follow}$ User	Follow
2	$U \to U \to U$	User $\xrightarrow{follow}$ User $\xrightarrow{follow}$ User	Follower of Follower
3	$U \to U \leftarrow U$	User $\xrightarrow{follow}$ User $\xrightarrow{follow^{-1}}$ User	Common Out Neighbor
4	$U \leftarrow U \to U$	User $\xrightarrow{follow^{-1}}$ User $\xrightarrow{follow}$ User	Common In Neighbor
5	$U \to P \to W \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Word $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write^{-1}}$ User	Posts Containing Common Words
6	$U \to P \to T \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Time $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write^{-1}}$ User	Posts Containing Common Timestamps
7	$U \to P \to L \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{attach}$ Location $\xrightarrow{attach^{-1}}$ Post $\xrightarrow{write^{-1}}$ User	Posts Attaching Common Location Check-ins

Aligned Network Schema and Inter-Network Social Meta Path



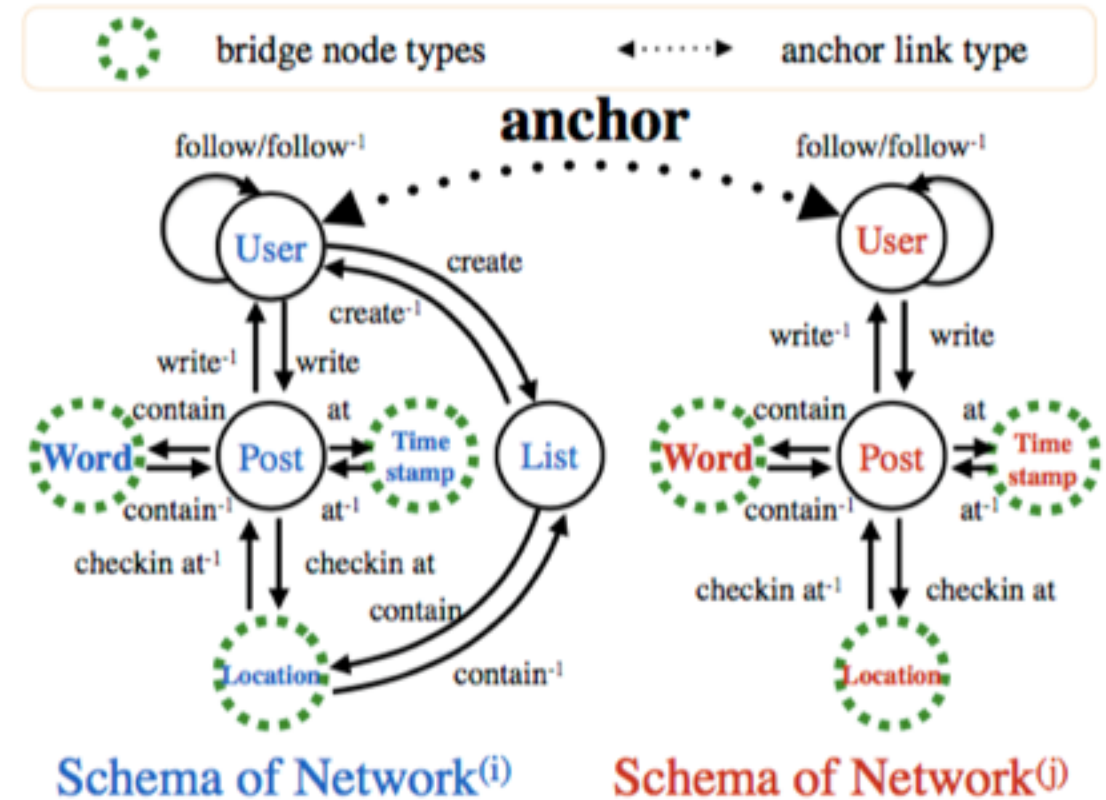


Fig. 1. Schema of aligned heterogeneous network.

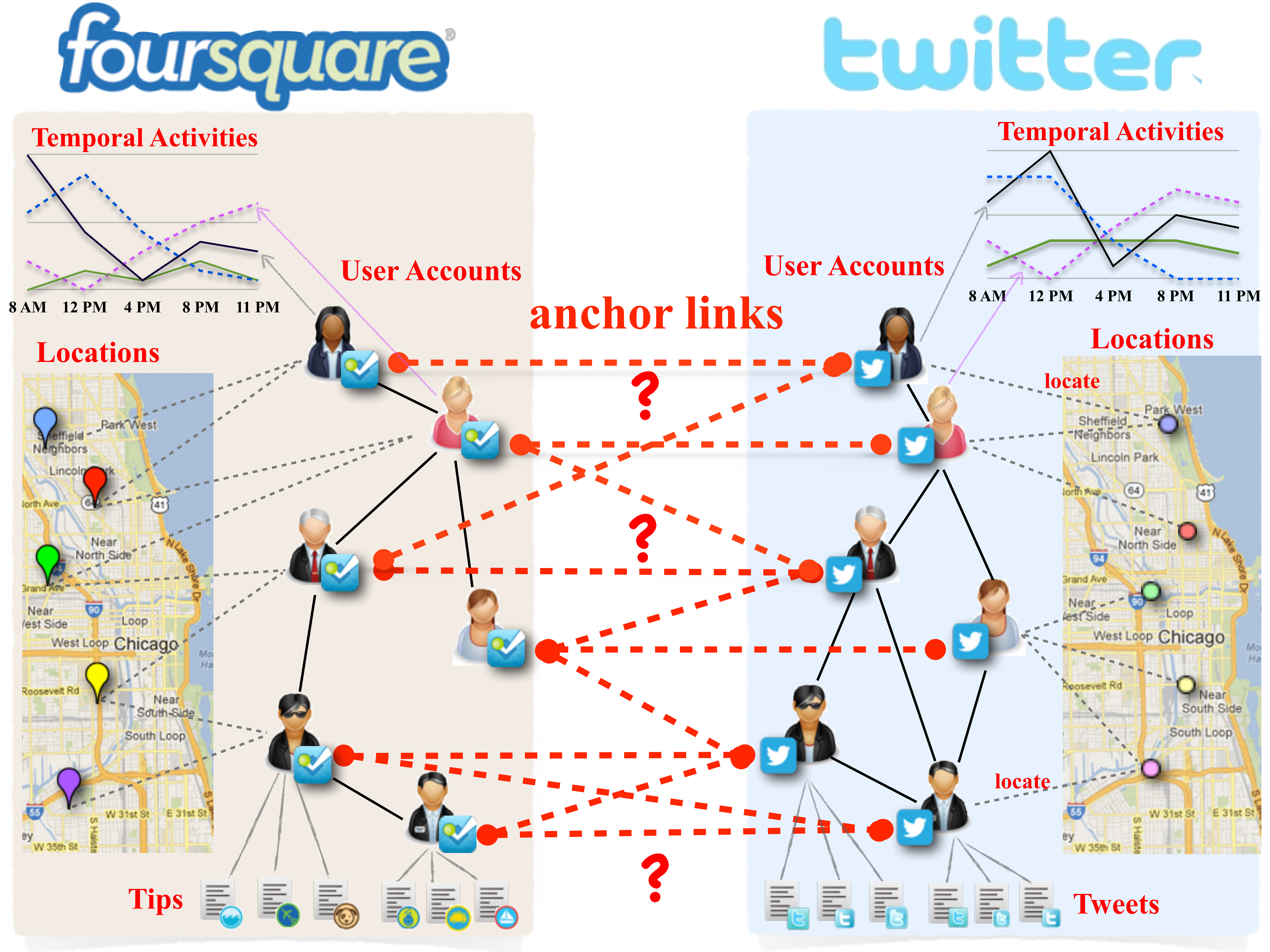
Definition 1: ... ...eta path $\Psi =$ $T_1 \xrightarrow{R_1} T_2 \xrightarrow{R_2}$ ... ...ork meta path ... $\xleftrightarrow{R_m} T_{m+1} =$ across G^i and ... $\Upsilon(U^i, U^j)$. ... is 1.

- *Common Out Neighbor Anchor Meta Path* (Ψ_1): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$" for short.
- *Common In Neighbor Anchor Meta Path* (Ψ_2): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common Out In Neighbor Anchor Meta Path* (Ψ_3): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common In Out Neighbor Anchor Meta Path* (Ψ_4): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$".

Outline

Jiawei Zhang and Philip S. Yu. PCT: Partial Co-Alignment of Social Networks. In: Proceedings of the 25th International World Wide Web Conference (WWW '16), Montreal, Canada, April 11-15, 2016.

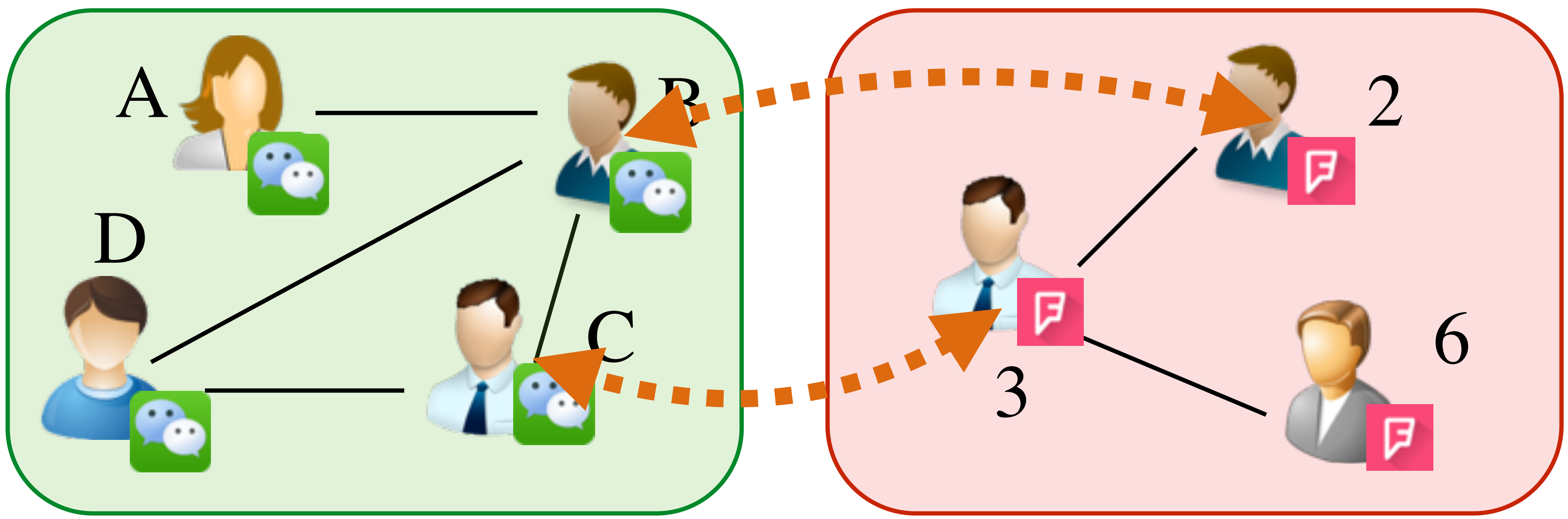
1. Background Knowledge and Basic Concepts
2. Problem 1: Network Alignment
3. Problem 2: Social Link Prediction across Networks
4. Future work 1: Community Detection across Networks
5. Future work 2: Viral Marketing across Networks
6. Summary and Selected References


foursquare
twitter
Temporal Activities
8 AM
12 PM
4 PM
8 PM
11 PM
Locations
User Accounts
anchor links
Temporal Activities
8 AM
12 PM
4 PM
8 PM
11 PM
User Accounts
Locations
locate
locate
Tips
Tweets

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 **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

$$\begin{array}{c|cccc} & A & B & C & D \\ \hline A & 0 & 1 & 0 & 0 \\ B & 1 & 0 & 1 & 1 \\ C & 0 & 1 & 0 & 1 \\ D & 0 & 1 & 1 & 0 \end{array}$$

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

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

Transition Matrix $\mathbf{P}$

$$\begin{array}{c|ccc} & 2 & 3 & 6 \\ \hline 2 & 0 & 1 & 0 \\ 3 & 1 & 0 & 1 \\ 6 & 0 & 1 & 0 \end{array}$$

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

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

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

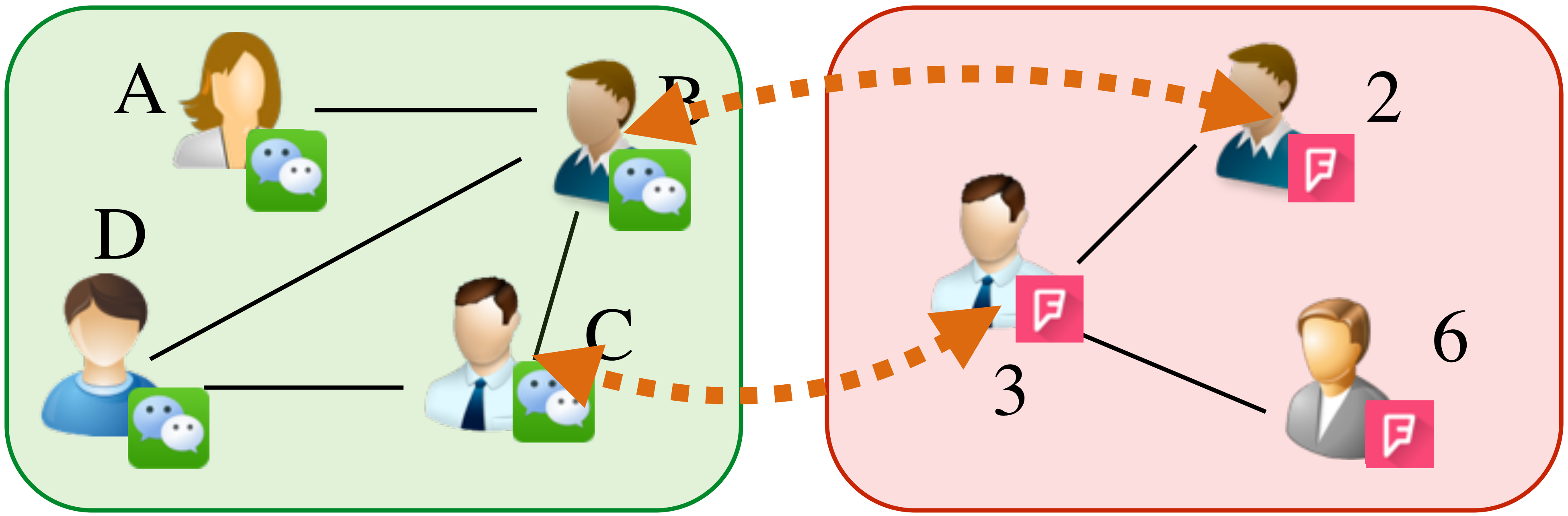
The optimal transition matrix **P** (i.e., anchor links) should minimize the mapping cost

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

Challenge 1: Heterogeneity of Social Networks

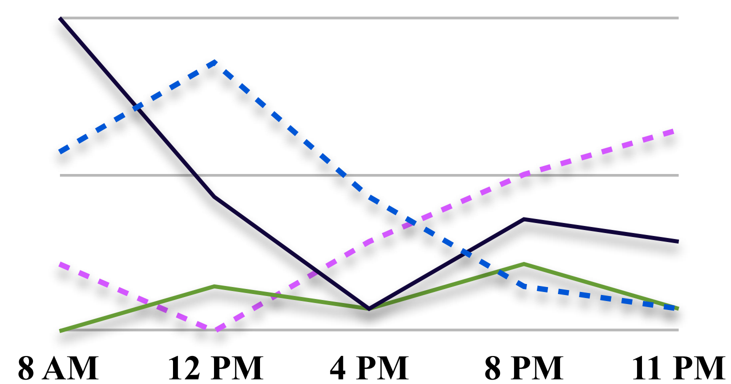
Heterogeneous **Link** and **Attribute** Information

- 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

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

Transition Matrix **P**

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

Similarity Matrix $\boldsymbol{\Lambda}$

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

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

Challenge 1: Heterogeneity of Social Networks

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

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

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

$$s.t. \quad \mathbf{P} \in \{0,1\}^{|\mathcal{U}^{(1)}| \times |\mathcal{U}^{(2)}|},$$

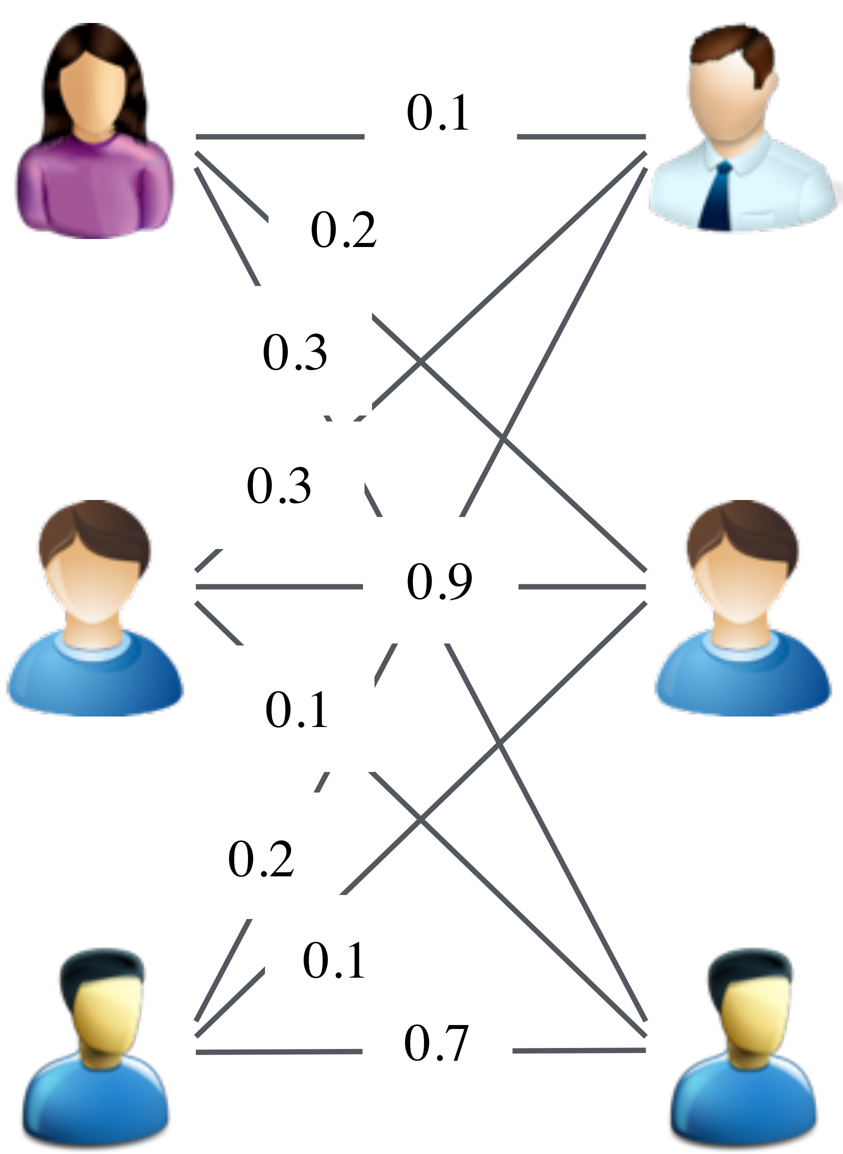
$$\mathbf{P}\mathbf{1}^{|\mathcal{U}^{(2)}| \times 1} \preceq \mathbf{1}^{|\mathcal{U}^{(1)}| \times 1}, \mathbf{P}^\top \mathbf{1}^{|\mathcal{U}^{(1)}| \times 1} \preceq \mathbf{1}^{|\mathcal{U}^{(2)}| \times 1}$$

One-to-One Constraint

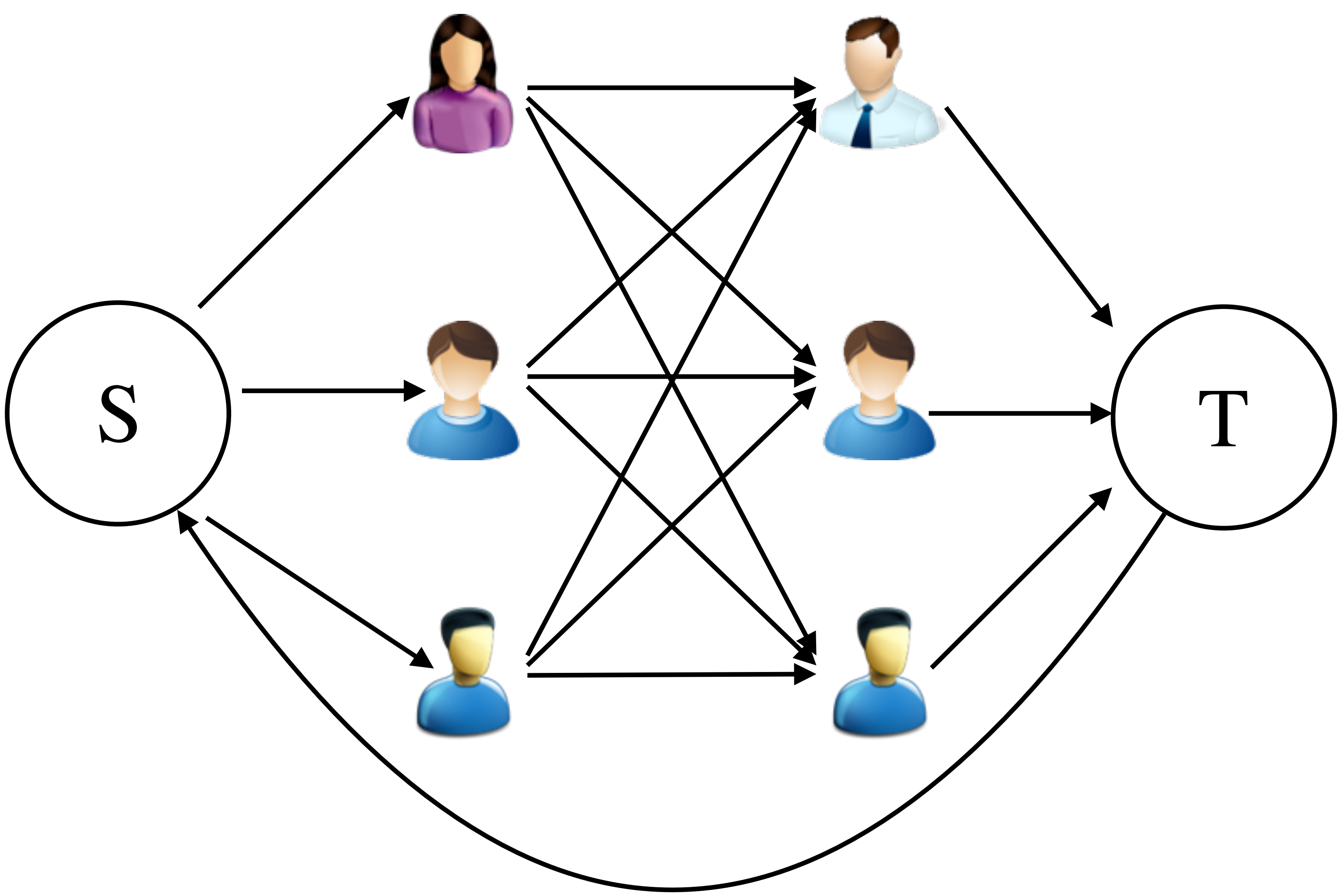
- Hard 0-1 non-linear integer programming problem, very challenging to address
- Relax the hard 0-1 constraint, **P** can take real values in range [0, 1]
- These introduced redundant anchor links will be pruned with a network matching post-processing step

Challenge 2: Redundant Link Pruning with Network Flow based Social Network Matching

User Preference Bipartite Graphs



Network Flow Graph



$$G_F = (\mathcal{N}_F, \mathcal{L}_F, \mathcal{W})$$

Challenge 2: Redundant Link Pruning with Network Flow based Social Network Matching

introduced network flow variable for user anchor links

user anchor link existence confidence scores (previous step)

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

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

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

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

mass balance constraints

variable bound constraints

Multiple Aligned Network Dataset

	property	network: Twitter	network: Foursquare
# node	user	5,223	5,392
	tweet/tip	9,490,707	48,756
	location	297,182	38,921
# link	friend/follow	164,920	31,312
	write	9,490,707	48,756
	locate	615,515	48,756

Anchor Links: 3,388

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

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

Experiment Results

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

θ = 1: full alignment setting

θ = 5: 20% alignment setting

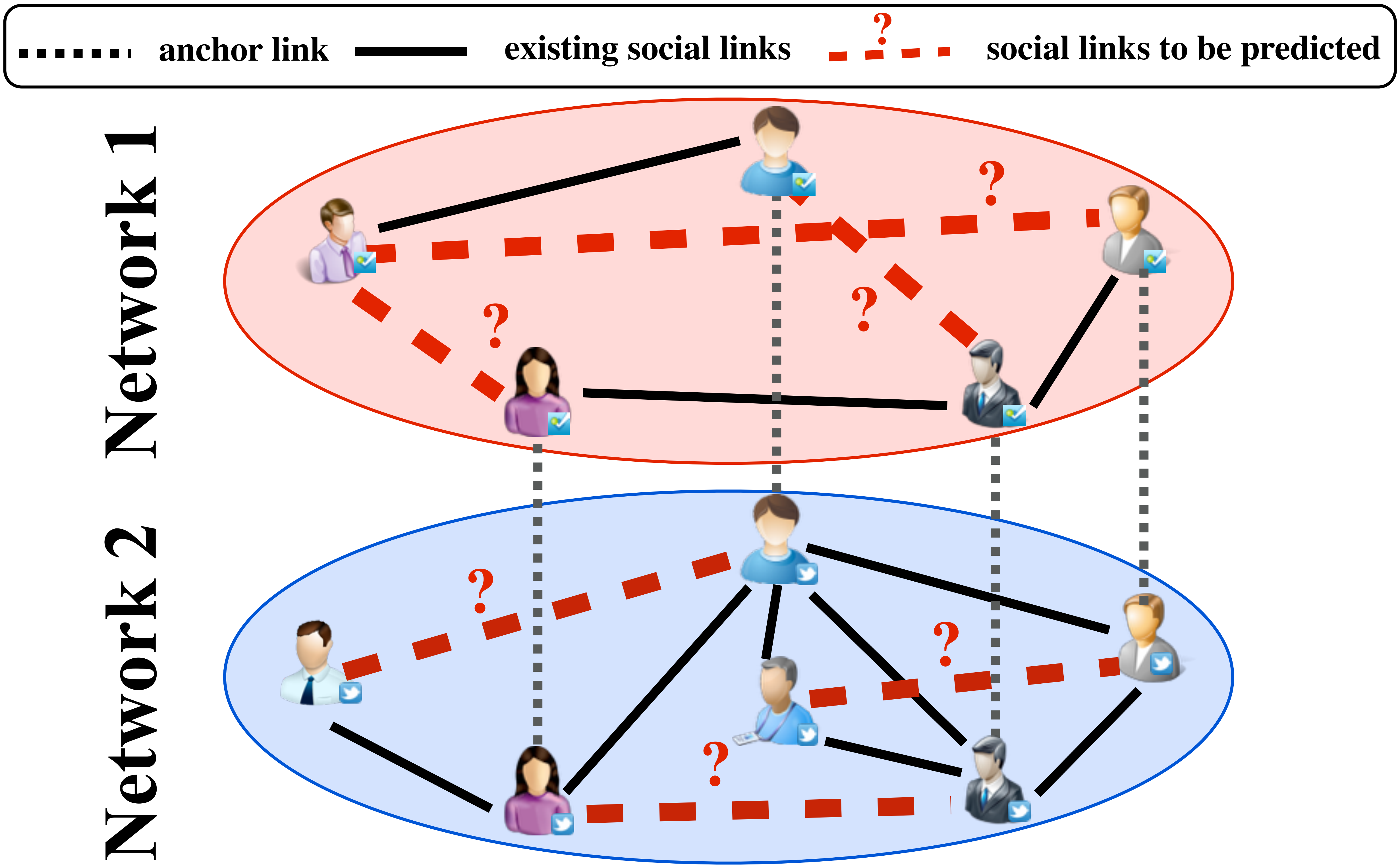
measure	methods	θ = 1	2	3	4	5
AUC	UNICOAT	**0.868**	**0.831**	**0.814**	**0.804**	**0.799**
	BIGALIGNEXT	0.813	0.779	0.759	0.752	0.749
	BIGALIGN	0.568	0.557	0.555	0.552	0.550
	ISOEXT	0.818	0.782	0.762	0.754	0.61
	ISO	0.547	0.529	0.52	0.518	0.516
	RDD	0.531	0.530	0.523	0.514	0.508
Prec@100	UNICOAT	**0.705**	**0.688**	**0.657**	**0.640**	**0.556**
	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

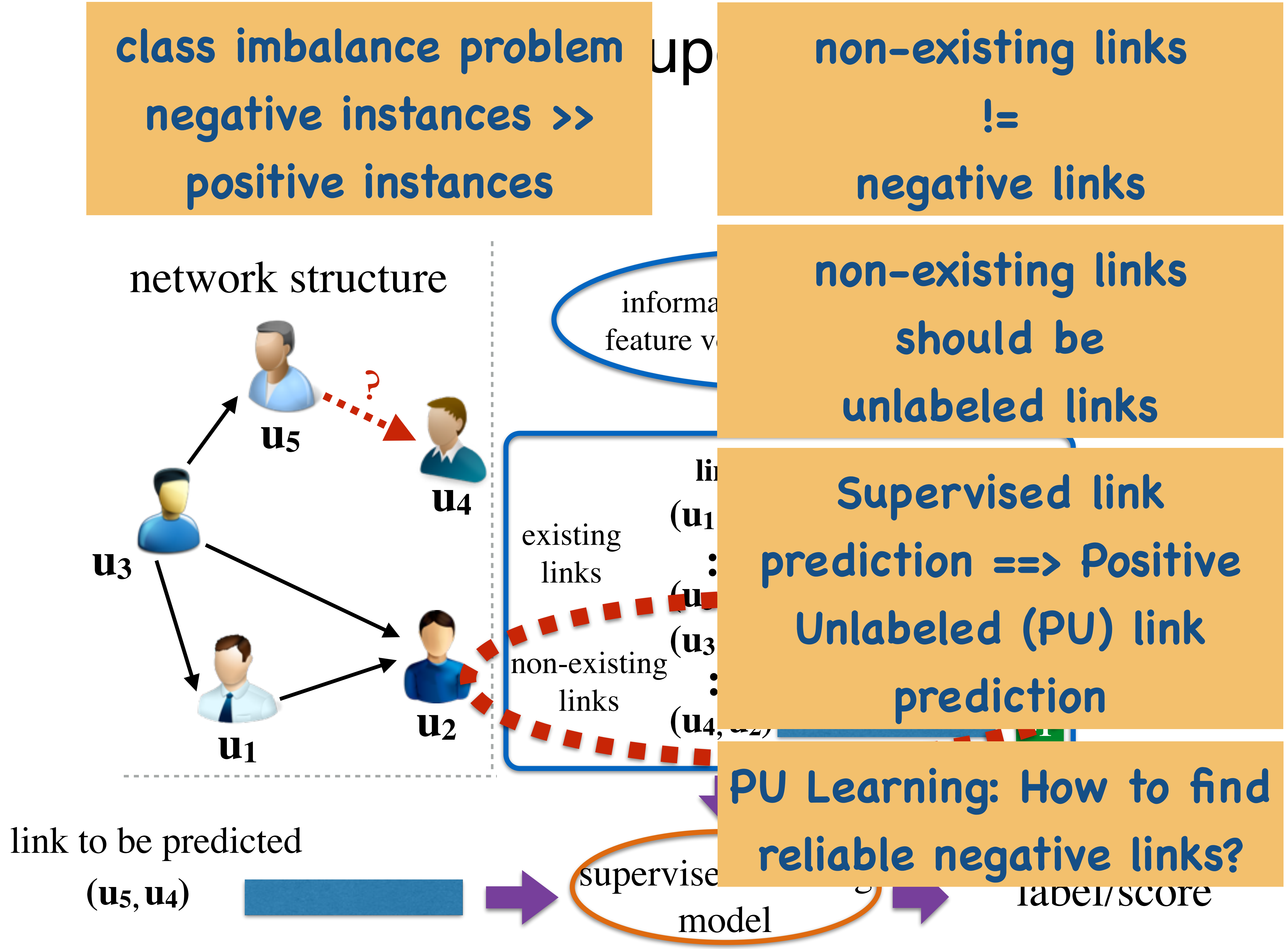
Outline

Jiawei Zhang, Philip S. Yu and Zhi-Hua Zhou. Meta-path based Multi-network Collective Link Prediction. In: Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '14), New York City, NY, August 24-27, 2014.

1. **Background Knowledge and Basic Concepts**
2. **Problem 1: Network Alignment**
3. **Problem 2: Social Link Prediction across Networks**
4. **Future work 1: Community Detection across Networks**
5. **Future work 2: Viral Marketing across Networks**
6. **Summary and Selected References**

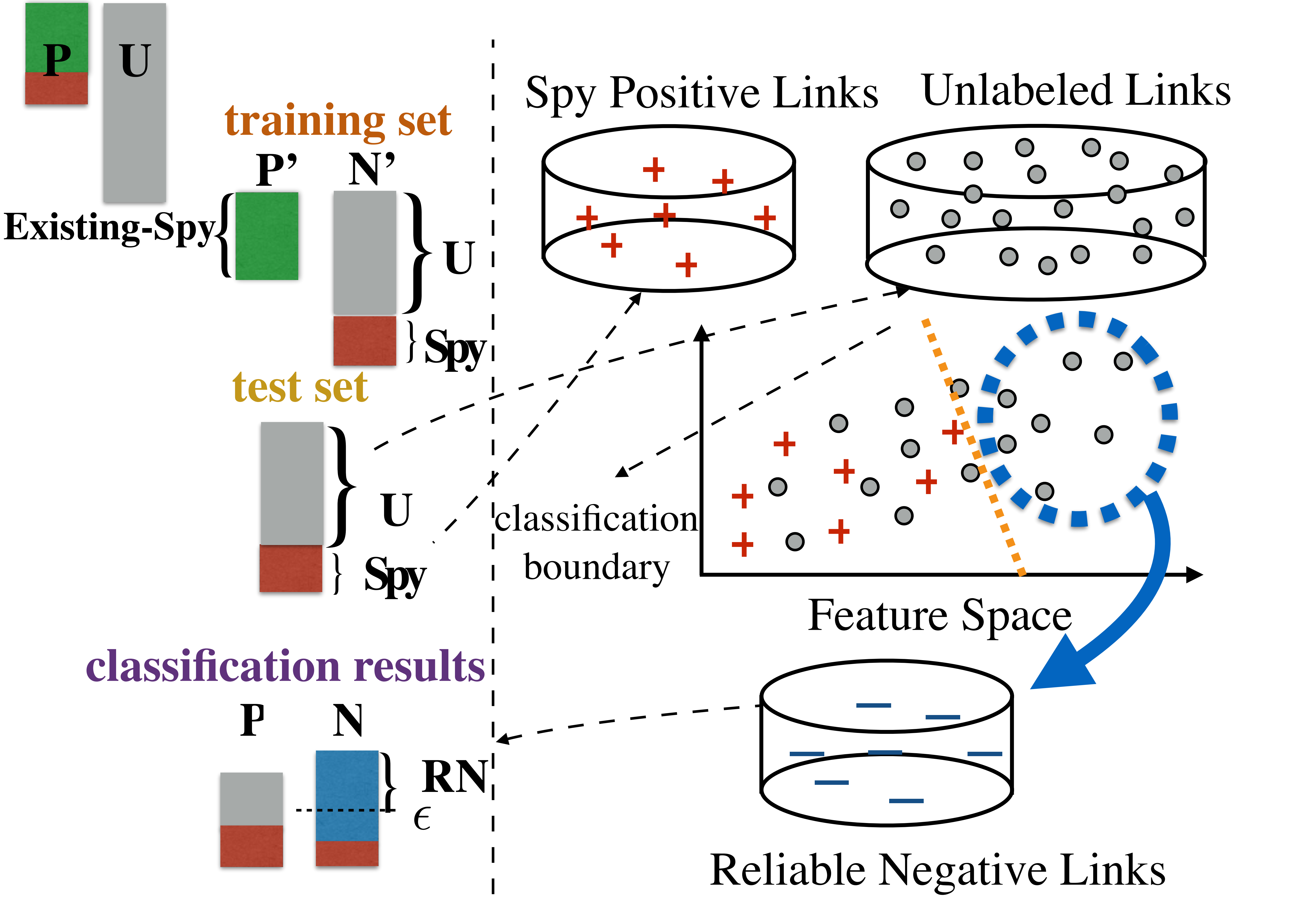
Predicting social links across multiple aligned networks simultaneously



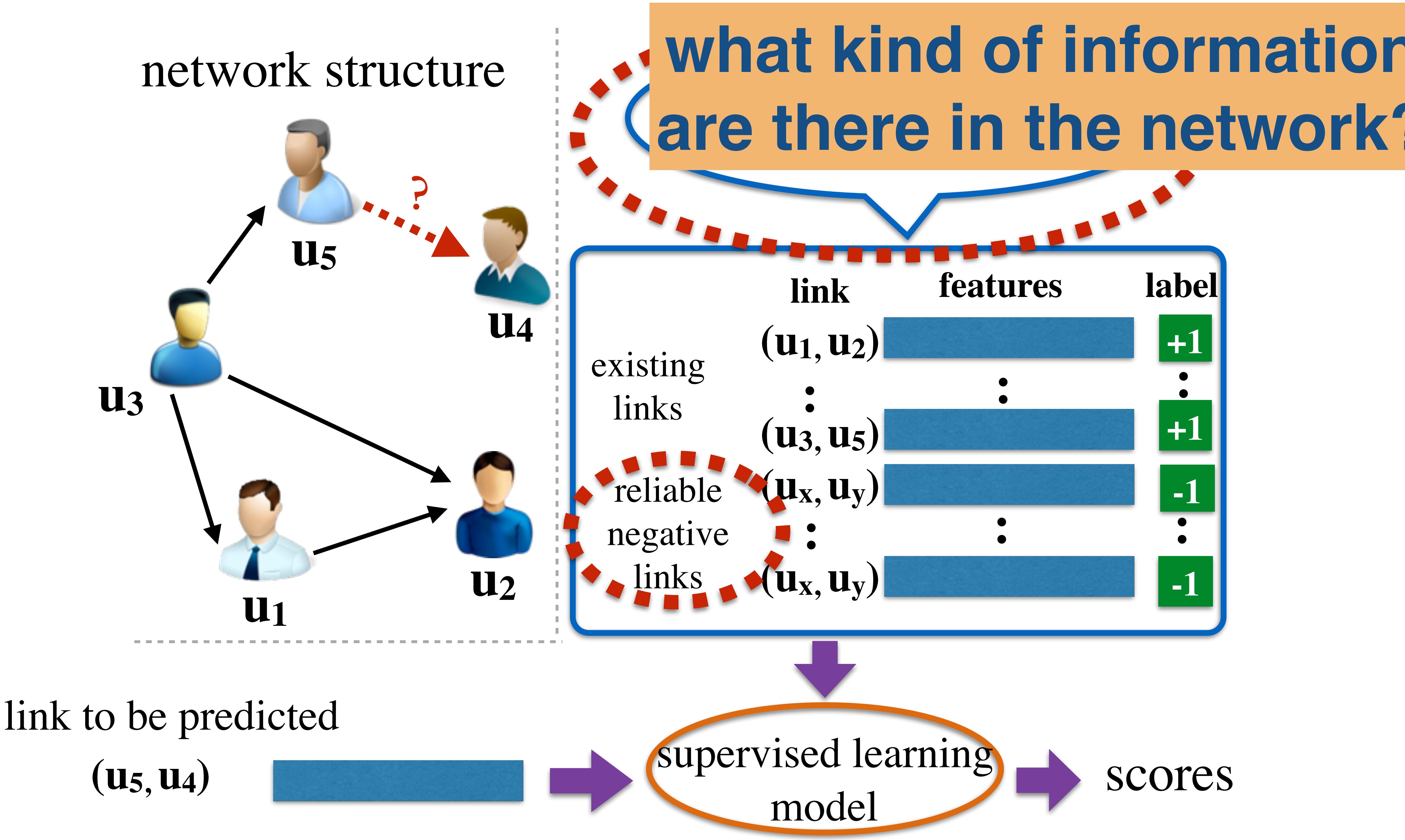


class imbalance problem
negative instances >>
positive instances
non-existing links
!=
negative links
non-existing links
should be
unlabeled links
Supervised link
prediction ==> Positive
Unlabeled (PU) link
prediction
PU Learning: How to find
reliable negative links?
network structure
?
u_5
u_4
u_3
u_2
u_1
existing
links
non-existing
links
link to be predicted
(u_5, u_4)
model

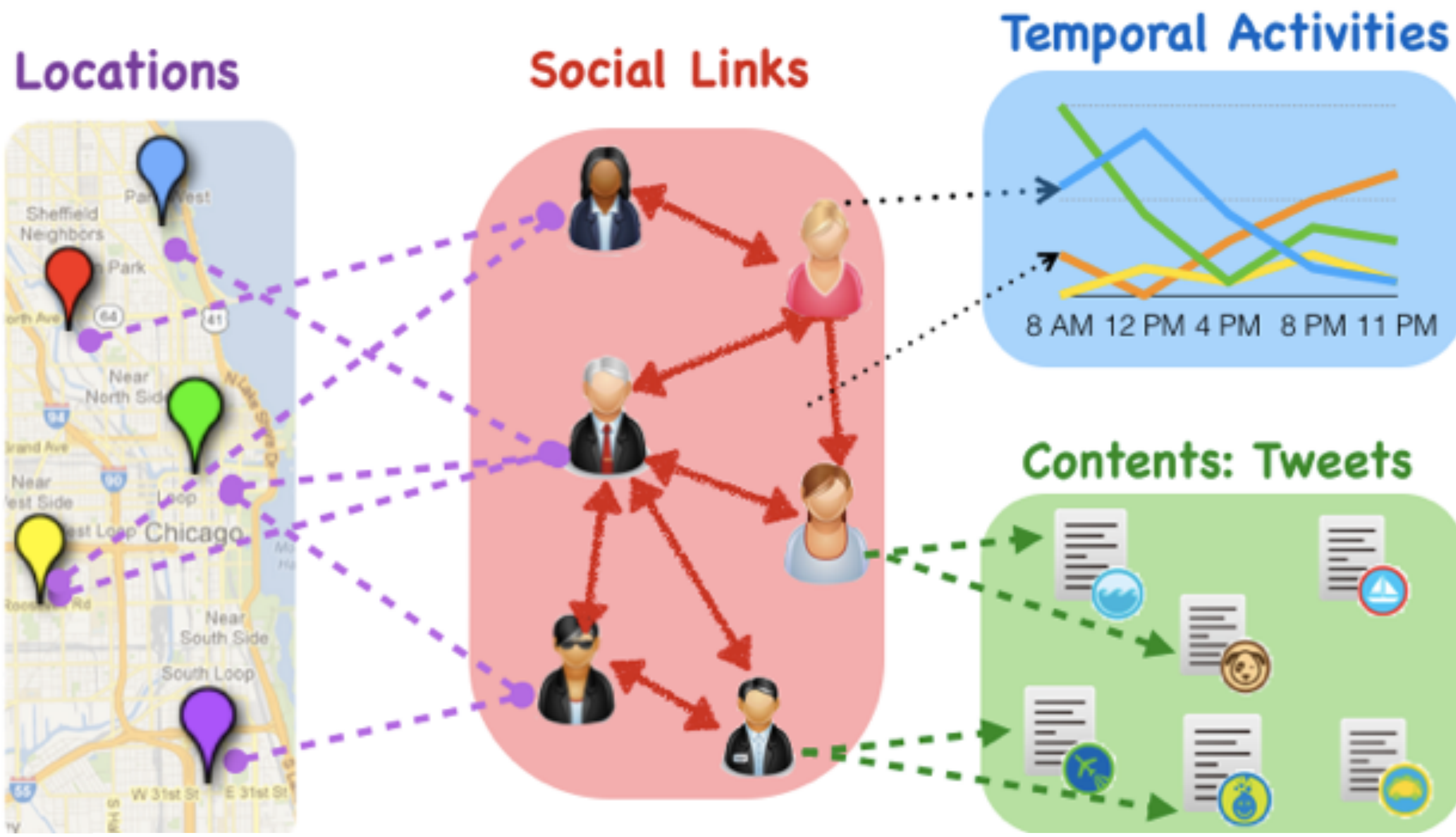
Reliable Negative Links Extraction



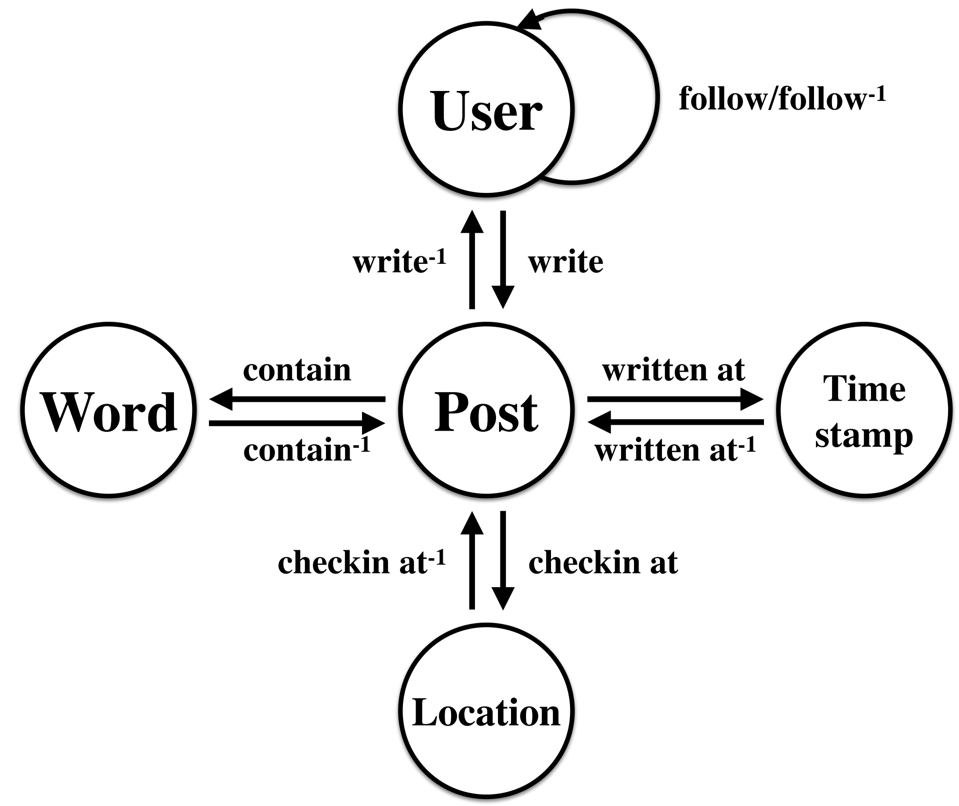
PU Link Prediction Setting



Network Schema and Intra-Network Social Meta Path



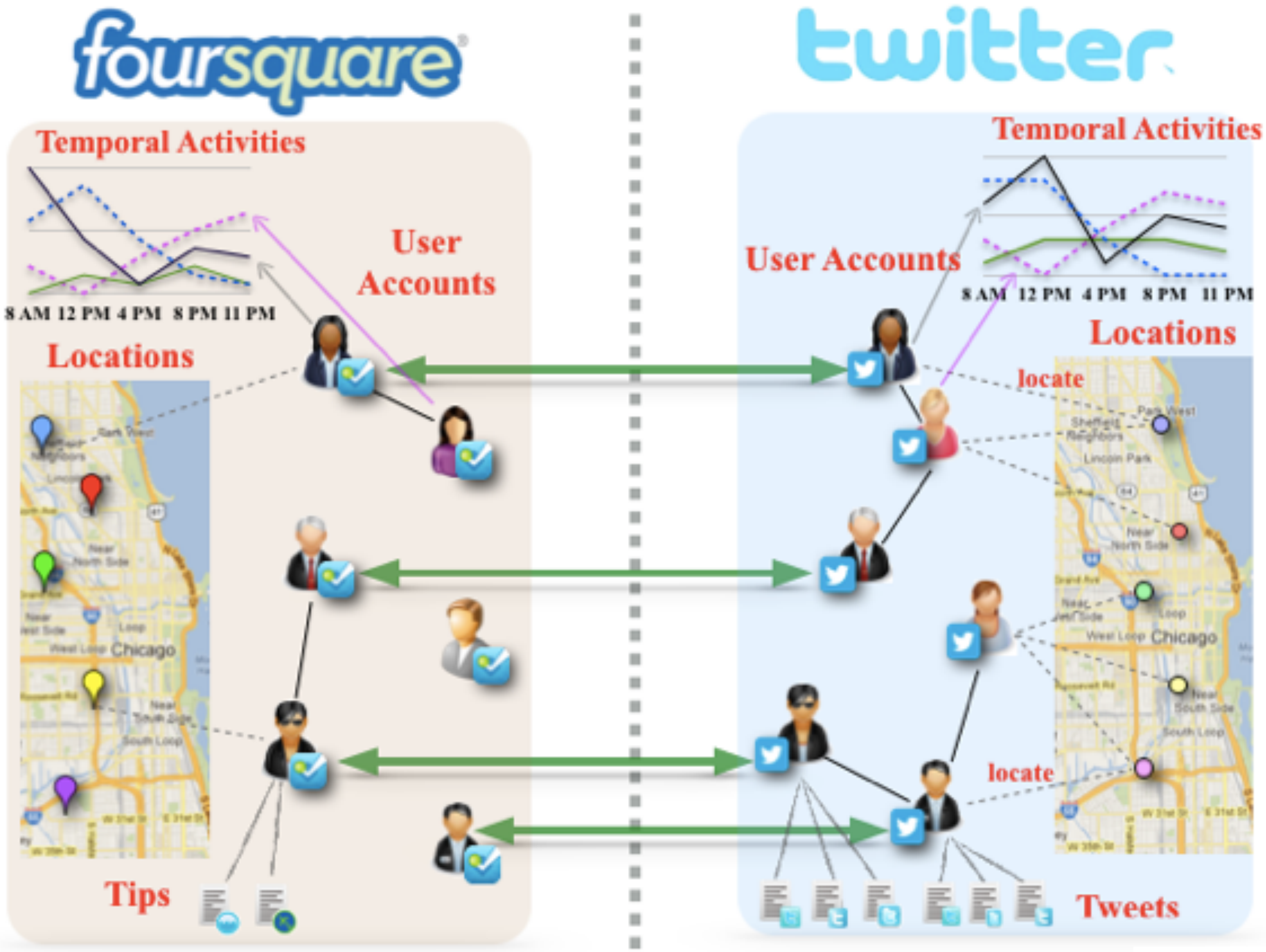
network schema



ID	Notation	Heterogeneous Network Meta Path	Semantics
1	$U \rightarrow U$	User $\xrightarrow{follow}$ User	Follow
2	$U \rightarrow U \rightarrow U$	User $\xrightarrow{follow}$ User $\xrightarrow{follow}$ User	
3	$U \rightarrow U \leftarrow U$	User $\xrightarrow{follow}$ User $\xrightarrow{follow^{-1}}$ User	
4	$U \leftarrow U \rightarrow U$	User $\xrightarrow{follow^{-1}}$ User $\xrightarrow{follow}$ User	
5	$U \rightarrow P \rightarrow W \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Word $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write}$	
6	$U \rightarrow P \rightarrow T \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Time $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write}$	
7	$U \rightarrow P \rightarrow L \leftarrow P \leftarrow U$	User $\xrightarrow{write}$ Post $\xrightarrow{attach}$ Location $\xrightarrow{attach^{-1}}$ Post $\xrightarrow{write^{-1}}$ User	Posts Attaching Common Location Check-ins

count the # meta path instances in the network connecting certain user pairs (u, v)

Aligned Network Schema and Inter-Network Social Meta Path



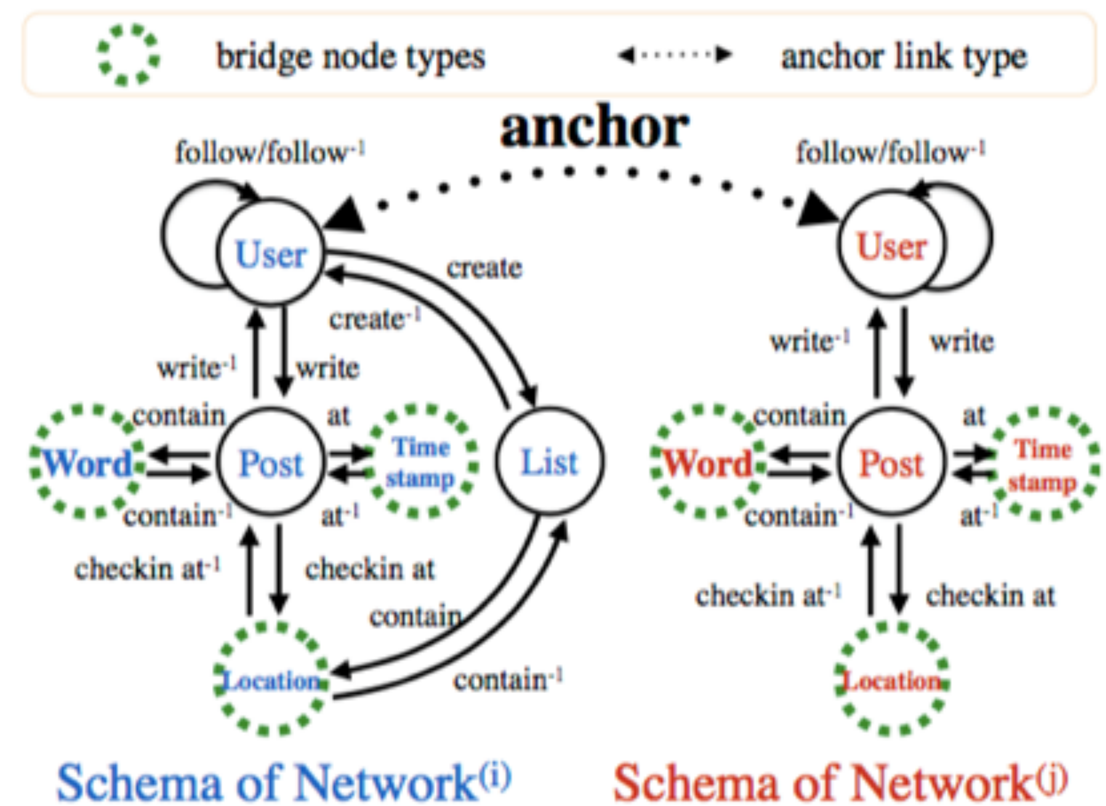
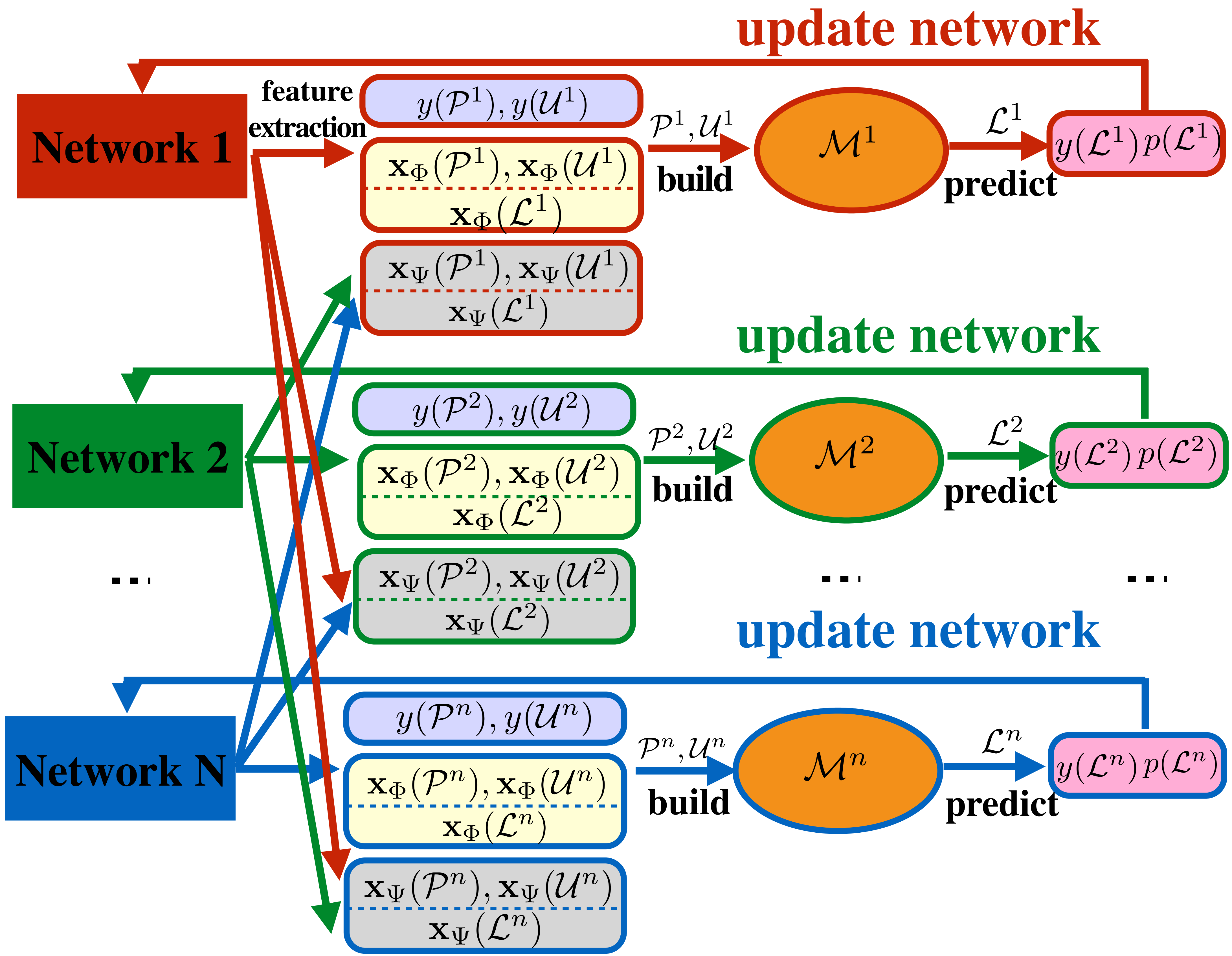


Fig. 1. Schema of aligned heterogeneous network.

- *Common Out Neighbor Anchor Meta Path* (Ψ_1): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$" for short.
- *Common In Neighbor Anchor Meta Path* (Ψ_2): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common Out In Neighbor Anchor Meta Path* (Ψ_3): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common In Out Neighbor Anchor Meta Path* (Ψ_4): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Location Checkin Anchor Meta Path 1* (Ψ_5): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{checkin\ at} Location \xleftarrow{checkin\ at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{L} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Location Checkin Anchor Meta Path 2* (Ψ_6): $User^{(i)} \xrightarrow{create} List^{(i)} \xrightarrow{contain} Location \xleftarrow{checkin\ at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{I}^{(i)} \rightarrow \mathcal{L} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Timestamps Anchor Meta Path* (Ψ_7): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{at} Time \xleftarrow{at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{T} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Word Usage Anchor Meta Path* (Ψ_8): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{contain} Word \xleftarrow{contain} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{W} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".

Multi-Network Collective Link Prediction Framework



Experiment Settings

- Ground truth: existing social link among users
    - hide part of the existing links in the test set
    - build model to discover these links

- Comparison Methods
    - MLI (Multi-network Link Identifier)
    - LI (Link Identifier): predict links in each network independently
    - SCAN(Supervised Cross-Aligned-Network link prediction): supervised link prediction
    - SCAN_s (SCAN with source network): features are extracted based on inter-network meta paths
    - SCAN_t (SCAN with target network): features are extracted based on intra-network meta paths

- Evaluation Metrics
    - AUC, Accuracy, F1

simultaneous link prediction is better than independent link prediction

Experiment Results

PU link prediction setting can improve the results

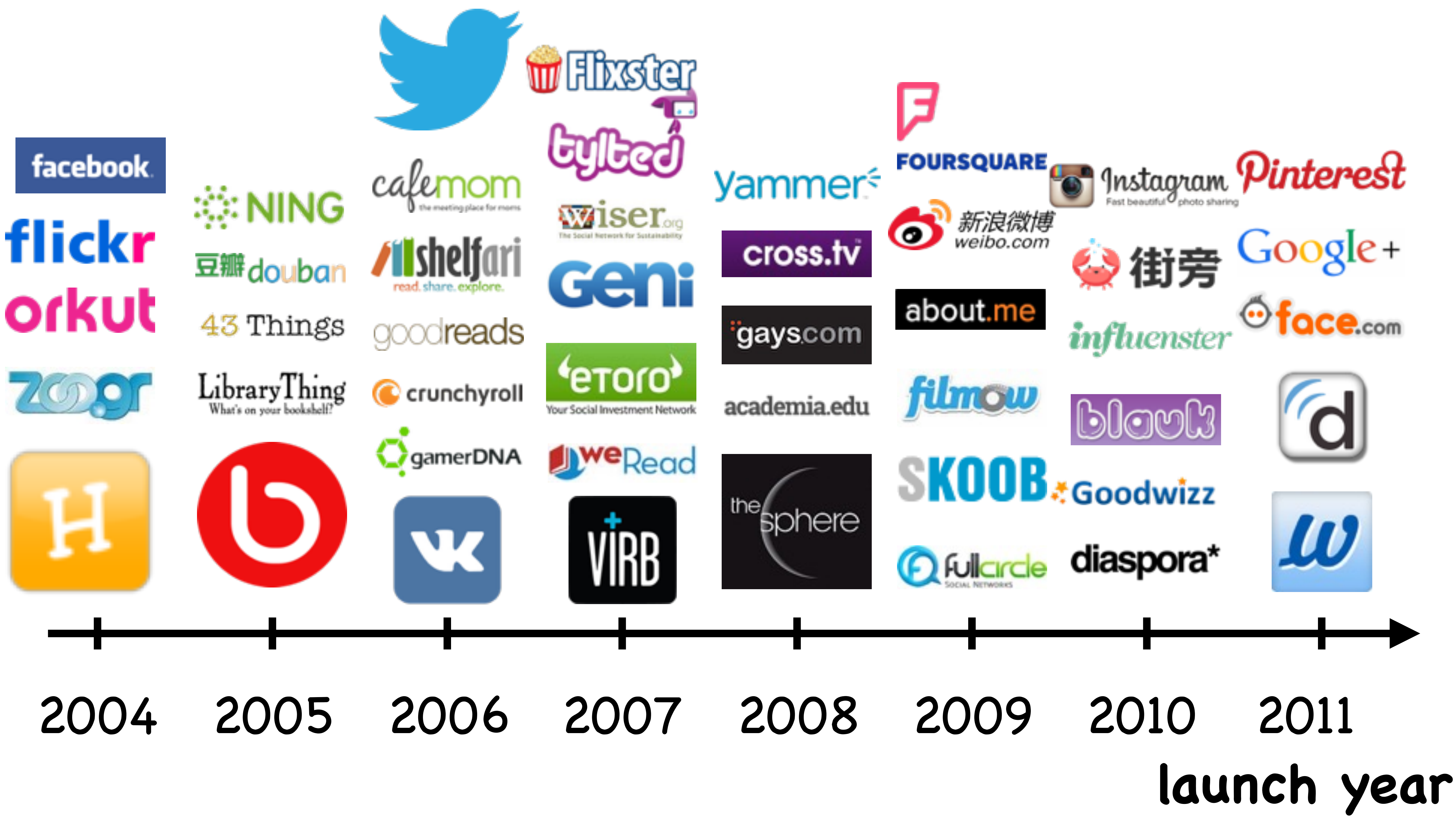
using features based on intra-network meta paths and inter-network meta paths simultaneously can achieve better results

network	measure	methods	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Foursquare	AUC	M_{LI}	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		LI	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		SCAN	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		$SCAN_T$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		$SCAN_S$	0.524±0.013	0.524±0.017	0.524±0.012	0.524±0.005	0.524±0.002
	Accuracy	M_{LI}	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		LI	0.568±0.013	0.624±0.053	0.699±0.004	0.722±0.006	0.761±0.01
		SCAN	0.558±0.007	0.6±0.006	0.683±0.071	0.714±0.009	0.721±0.007
		$SCAN_T$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		$SCAN_S$	0.548±0.011	0.548±0.055	0.548±0.007	0.548±0.008	0.548±0.007
	F1	M_{LI}	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		LI	0.63±0.017	0.635±0.015	0.66±0.007	0.684±0.01	0.715±0.016
		SCAN	0.6±0.02	0.609±0.006	0.614±0.031	0.632±0.018	0.645±0.018
		$SCAN_T$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
		$SCAN_S$	0.56±0.016	0.56±0.041	0.56±0.015	0.56±0.015	0.56±0.013

Outline

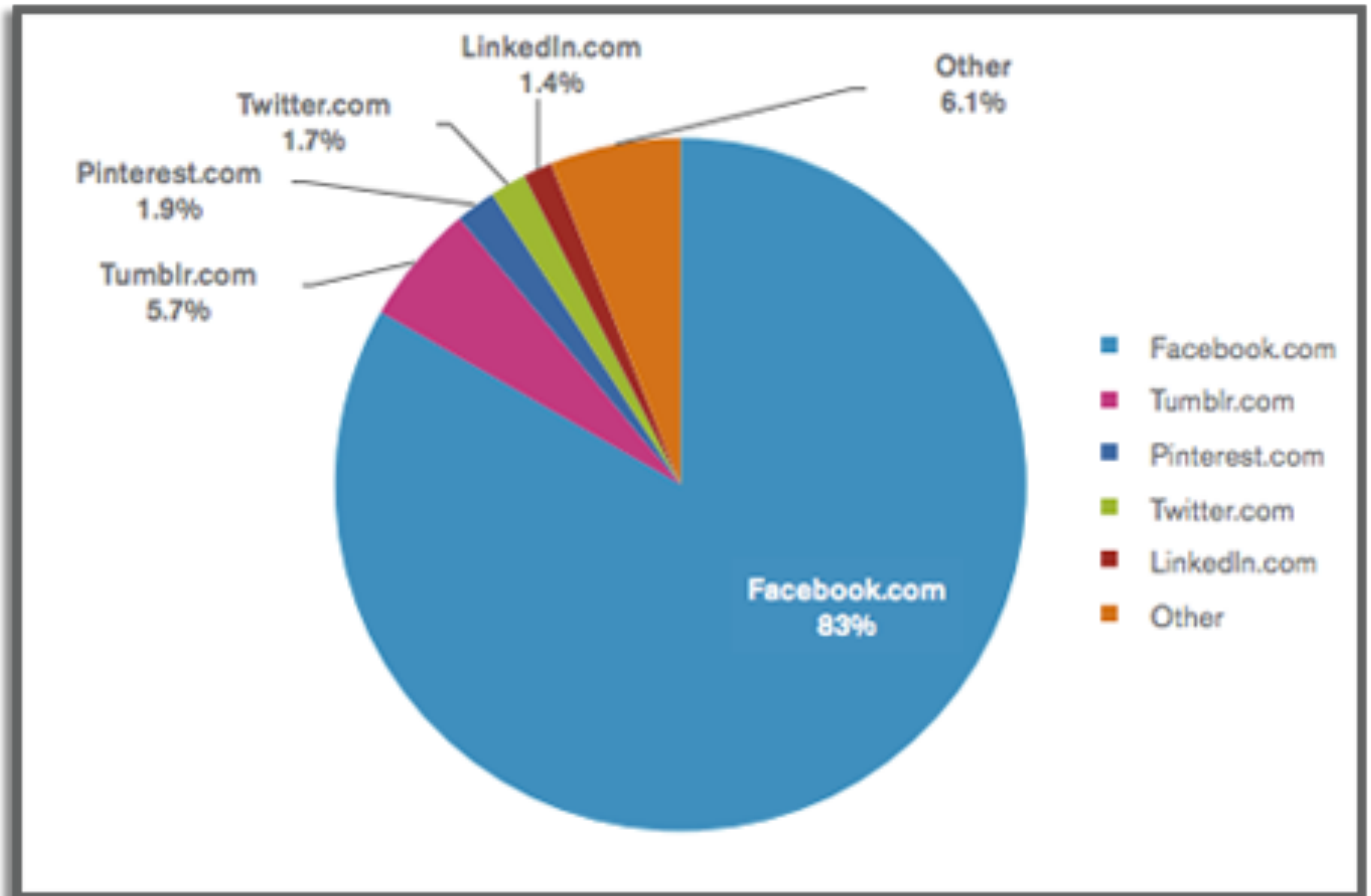
1. Background Knowledge and Basic Concepts
2. Problem 1: Network Alignment
3. Problem 2: Social Link Prediction across Networks
4. Future work 1: Community Detection across Networks
5. Future work 2: Viral Marketing across Networks
6. Summary and Selected References

New Social Networks Emerge Every Year



http://en.wikipedia.org/wiki/List_of_social_networking_websites

Emerging Networks Attract Limited Usages



Share of time spent

Emerging Networks Contains Sparse Information

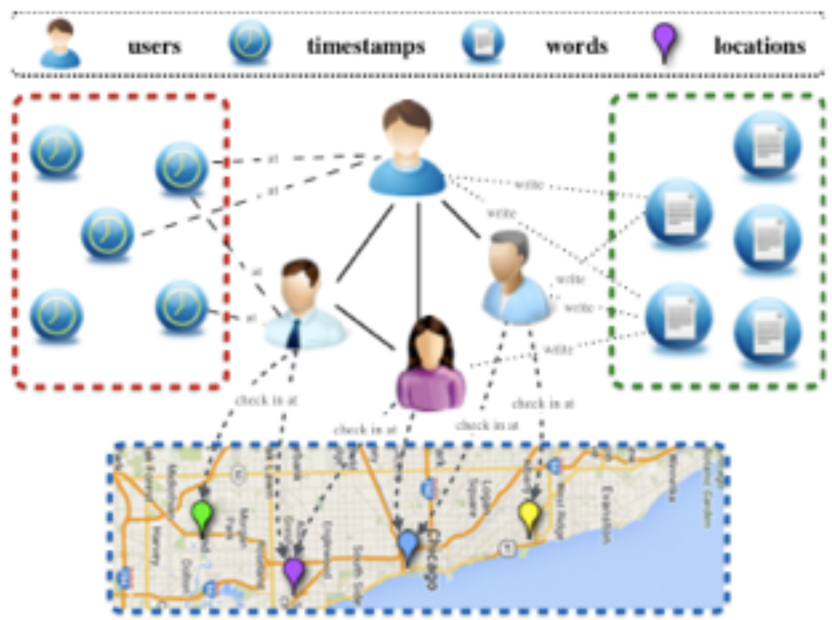
Emerging Network
Community Detection

Hard to calculate effective closeness measures among users due to the sparse information

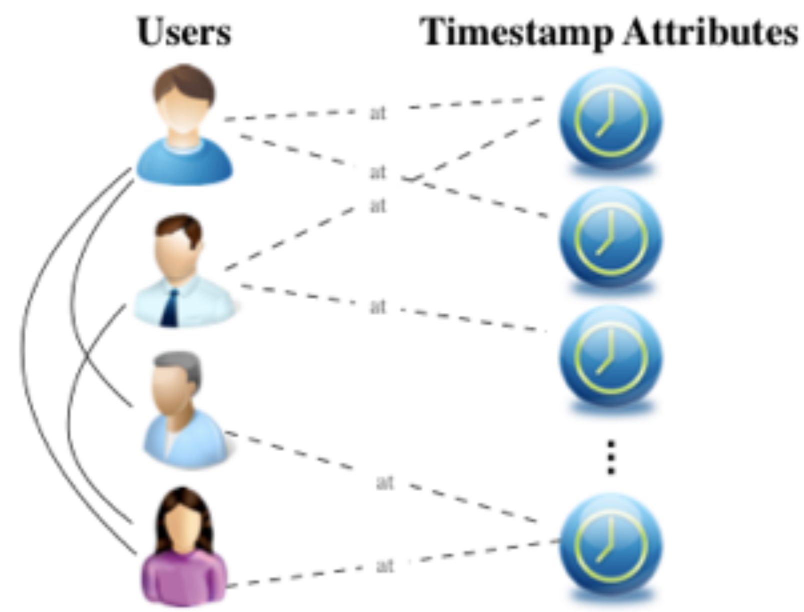
closeness measures among users:
Intimacy

Challenge: Information Sparsity Problem

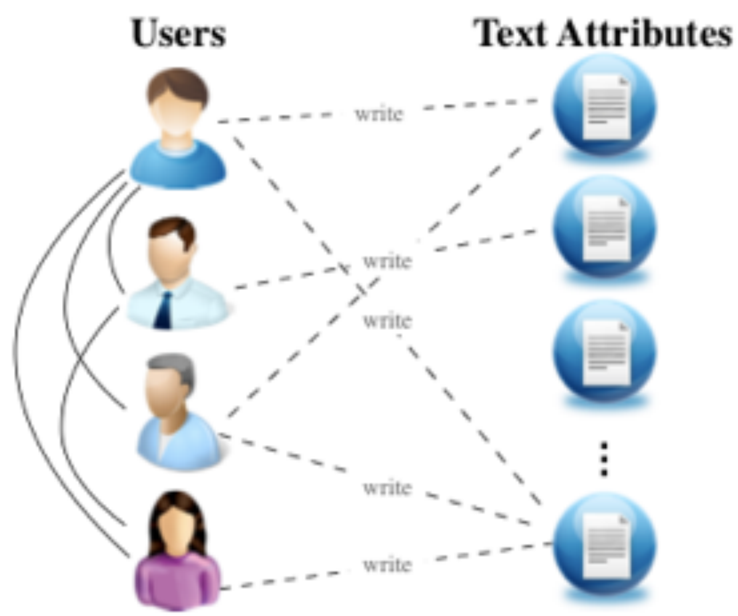
- Solution: use both Link and Attribute information



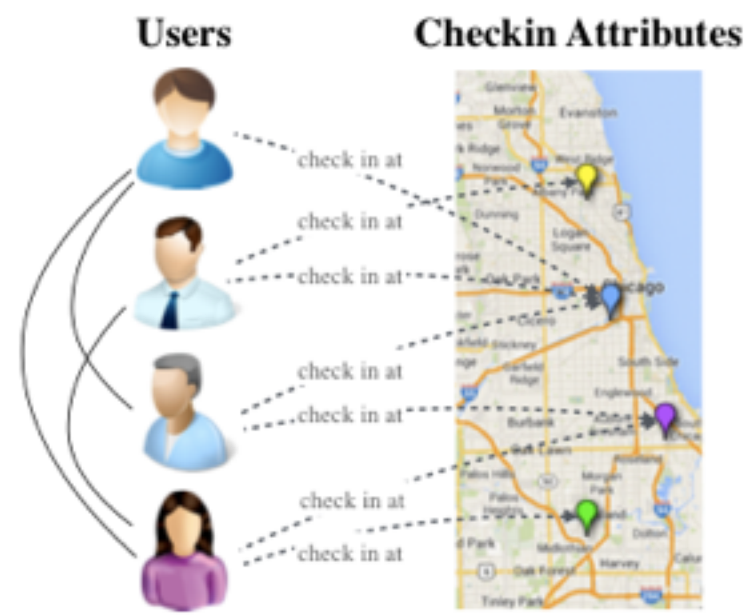
(a) augmented network



(b) timestamp attribute

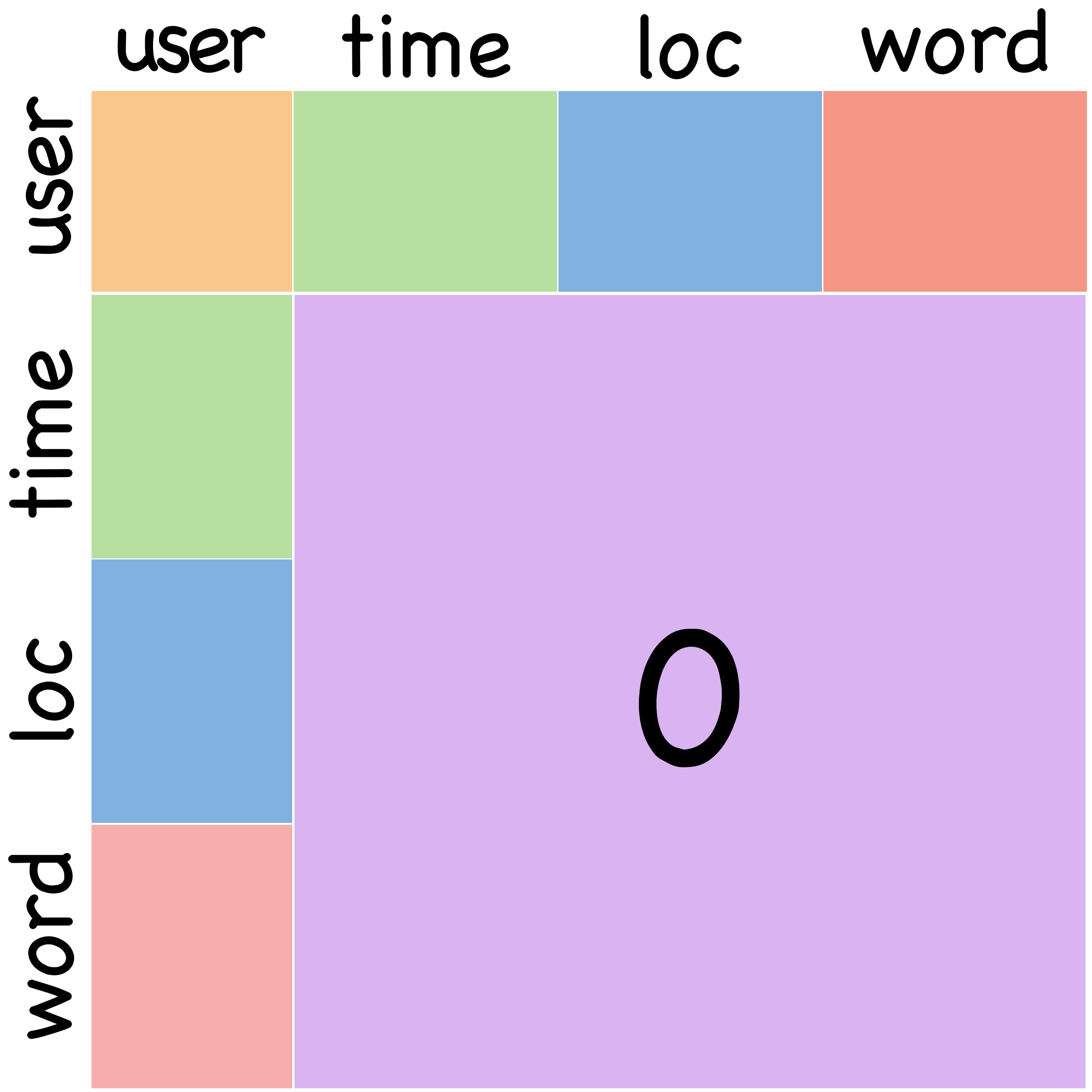


(c) text attribute



(d) checkin attribute

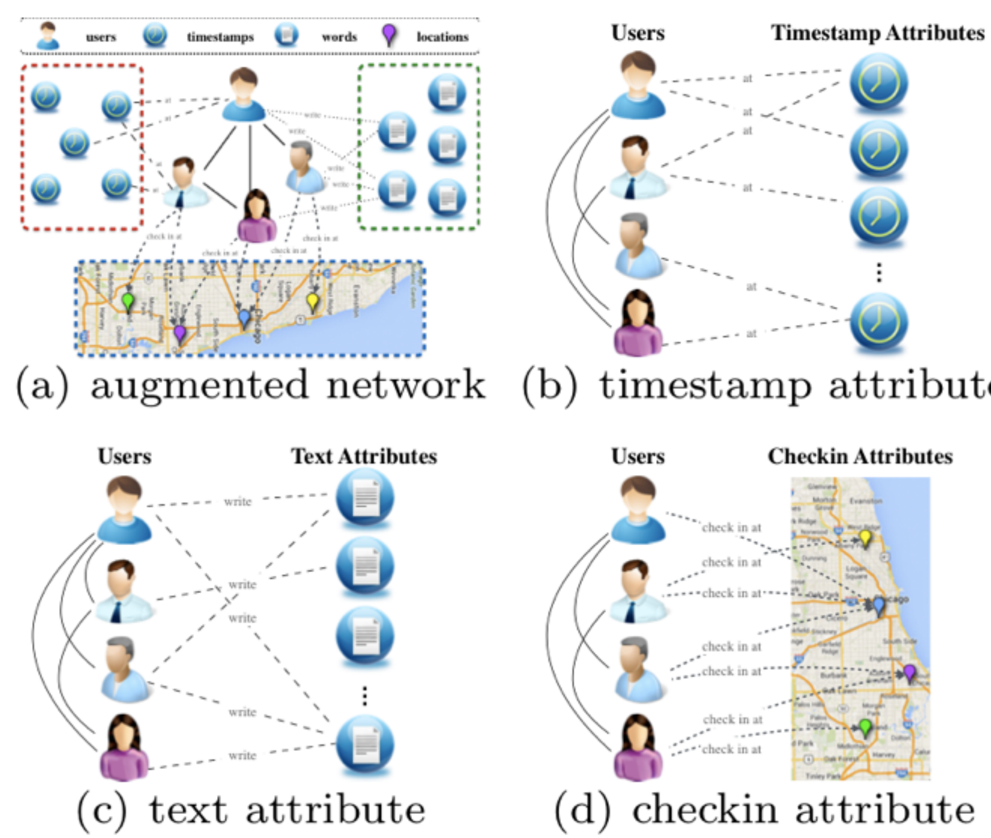
Intimacy Calculation with both Connection and Attribute Information



$$\tilde{\mathbf{Q}}_{aug} = \begin{bmatrix} \tilde{\mathbf{Q}} & \tilde{\mathbf{R}} \\ \tilde{\mathbf{S}} & \mathbf{0} \end{bmatrix}$$

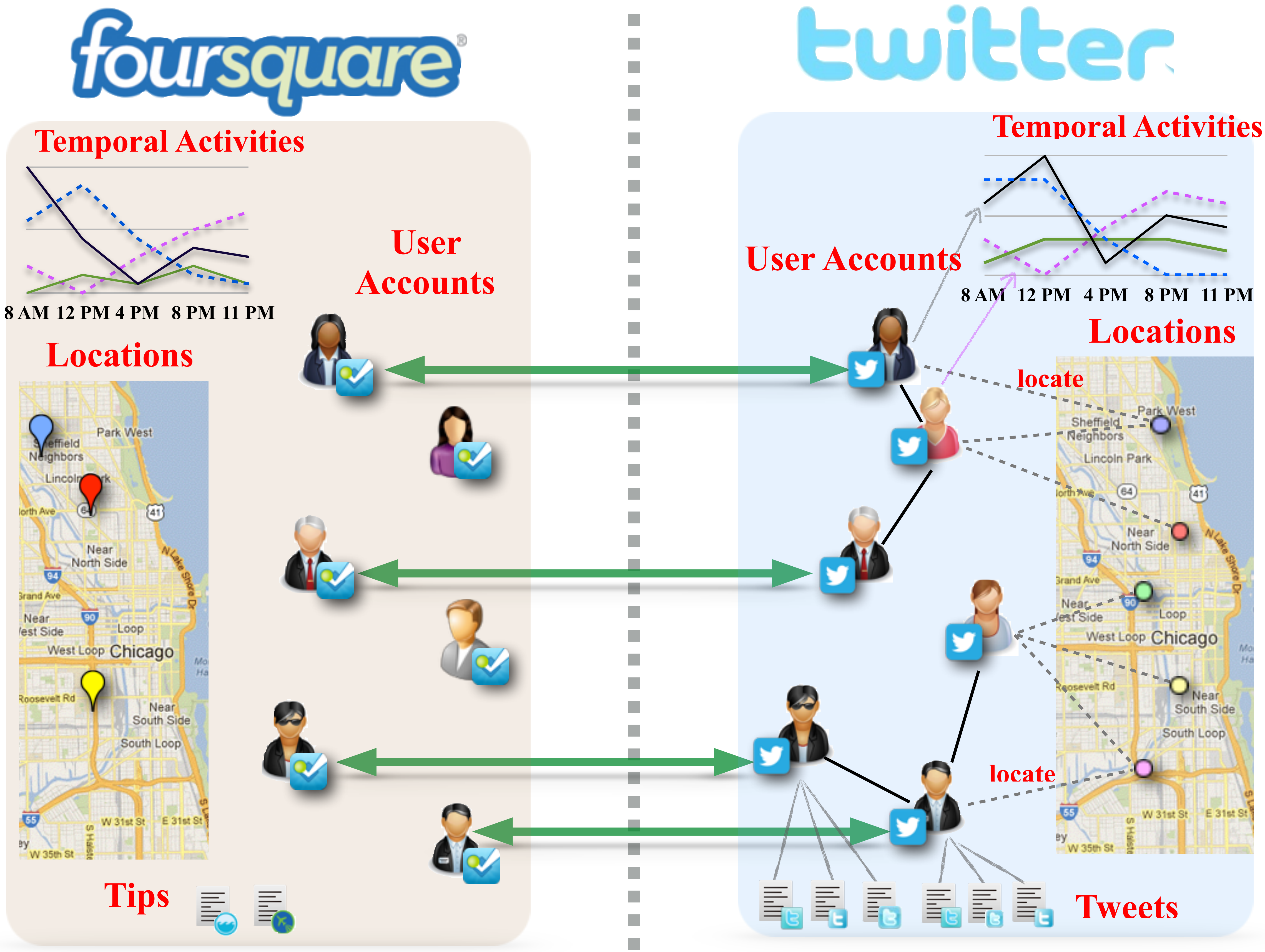
network transitional matrix

weighted normalized adjacency matrices
(1) among users
(2) between users and attributes

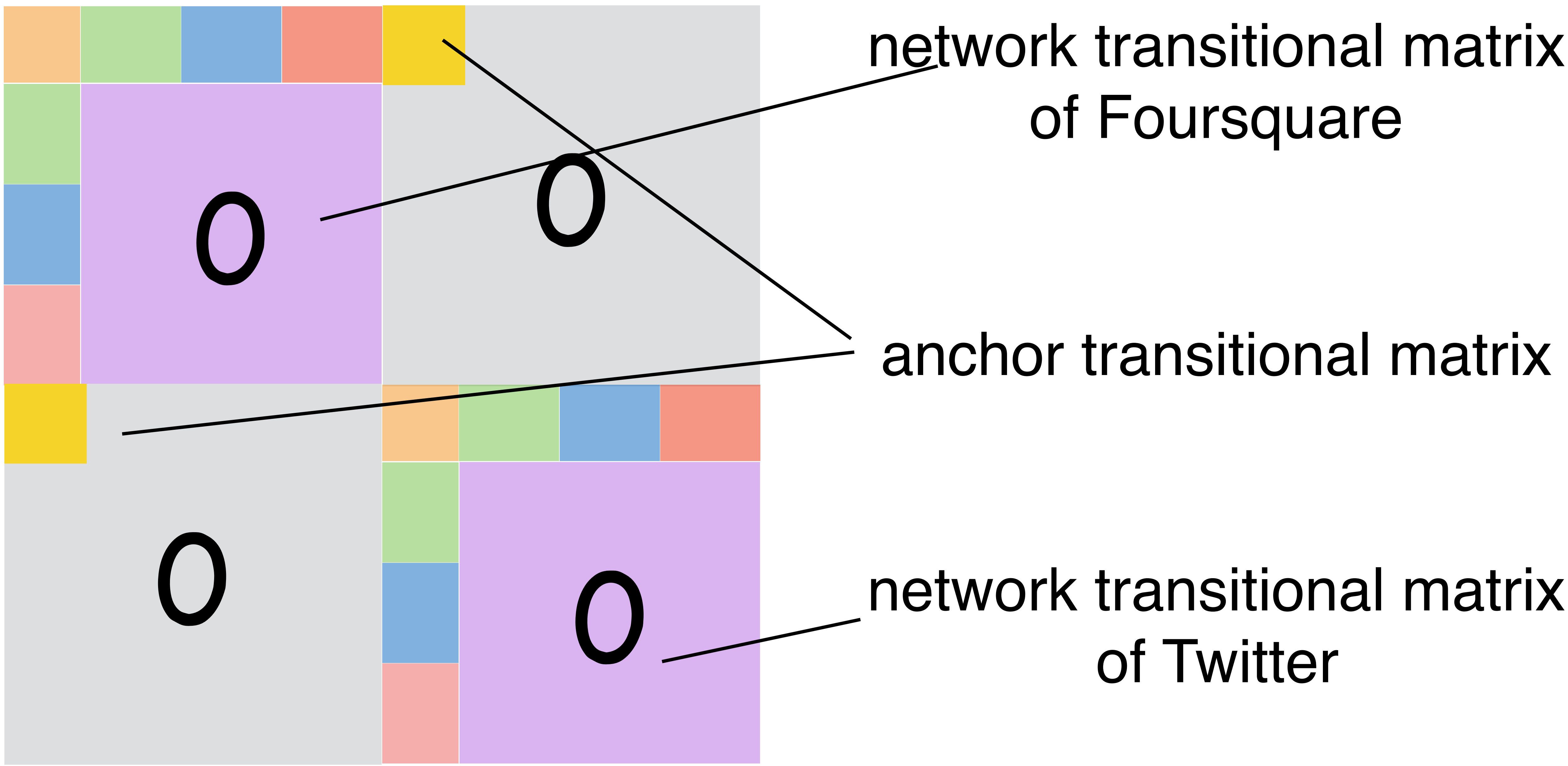


(a) augmented network (b) timestamp attribute
(c) text attribute (d) checkin attribute

Users use multiple social networks simultaneously



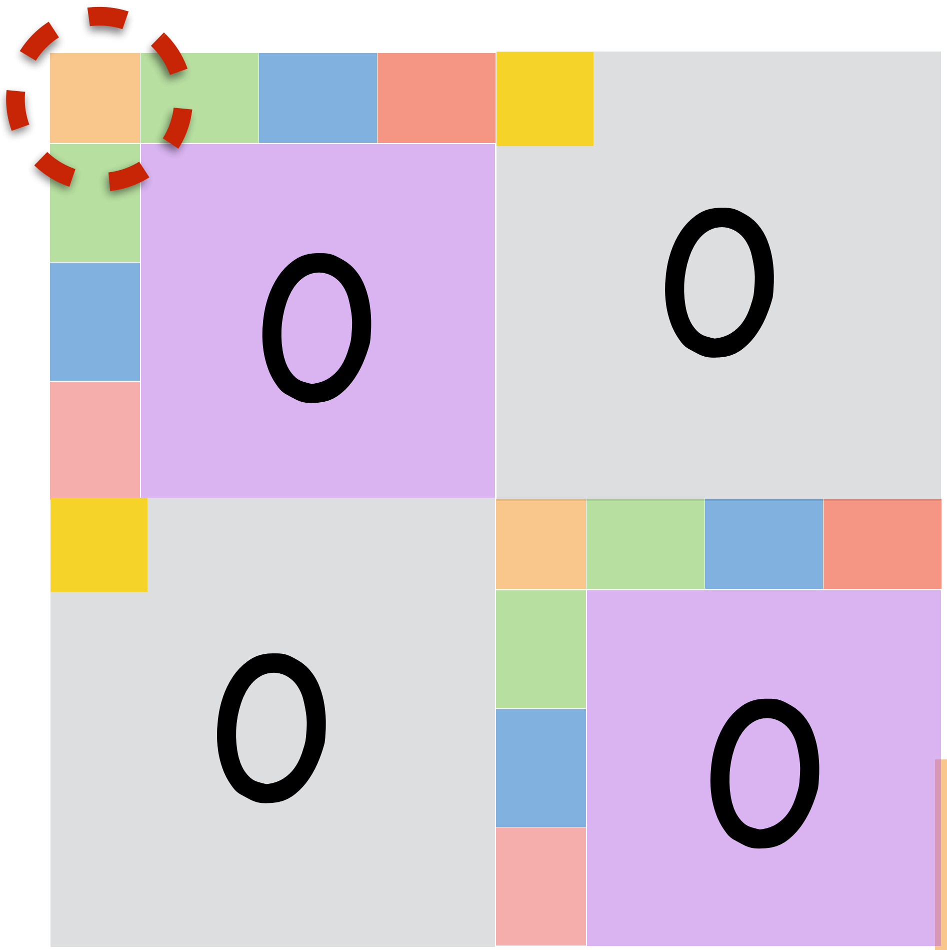
Intimacy Calculation with Information across Aligned Networks



$$\bar{\mathbf{Q}}_{align} = \begin{bmatrix} \bar{\mathbf{Q}}^{t}_{aug} & \bar{\mathbf{T}}^{t,s} \\ \bar{\mathbf{T}}^{s,t} & \bar{\mathbf{Q}}^{s}_{aug} \end{bmatrix}$$

weighted aligned network transitional matrix

Intimacy Calculation with Information across Aligned Networks



$$\left(\mathbf{I} + \alpha \bar{\mathbf{Q}}_{align}\right)^{\tau}$$

high-dimensional
stationary aligned
network transitional matrix

we only care about the intimacy matrix among users (lower dimension)

$$\underbrace{\bar{\mathbf{H}}_{align}}_{\text{intimacy matrix among users in Foursquare}} = \left(\mathbf{I} + \alpha \bar{\mathbf{Q}}_{align}\right)^{\tau} \underbrace{(1 : |\mathcal{V}^t|, 1 : |\mathcal{V}^t|)}_{\text{sub-matrix at the upper left corner}}$$

Outline

1. Background Knowledge and Basic Concepts
2. Problem 1: Network Alignment
3. Problem 2: Social Link Prediction across Networks
4. Future work 1: Community Detection across Networks
5. Future work 2: Viral Marketing across Networks
6. Summary and Selected References

Future Work 1: Viral Marketing across Aligned Networks

- Viral Marketing Background Knowledge

Linear Threshold (LT) Model

- A node v has threshold θ_v ~ $U[0,1]$
- A node v is influenced by each neighbor w according to a *weight* b_{vw} such that

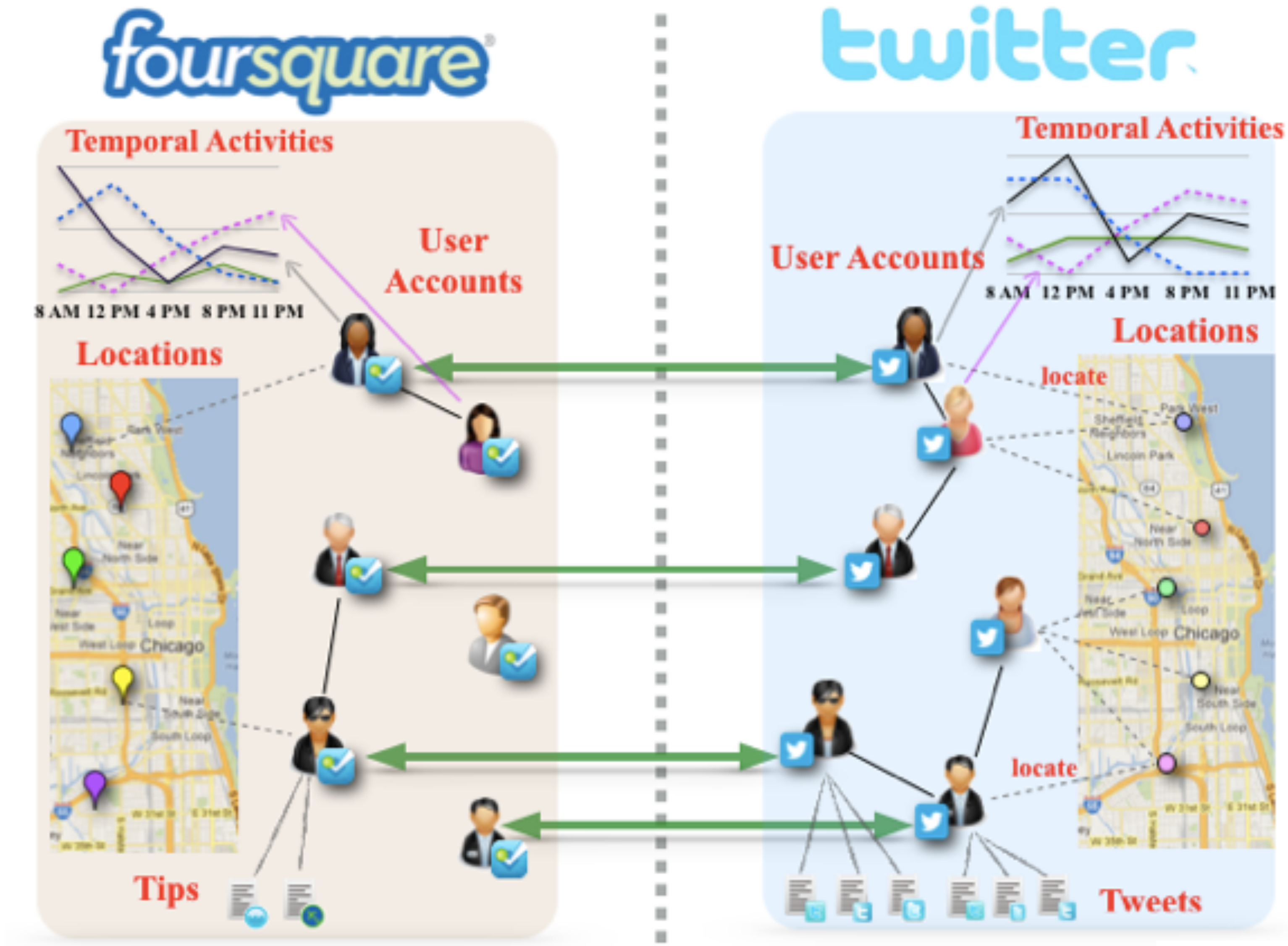
$$\sum_{w \text{ neighbor of } v} b_{v,w} \leq 1$$

- A node v becomes active when at least

$$\sum_{w \text{ active neighbor of } v} b_{v,w} \geq \theta_v$$

- It will stop if no other nodes can be activated

Aligned Network Schema and Inter-Network Social Meta Path



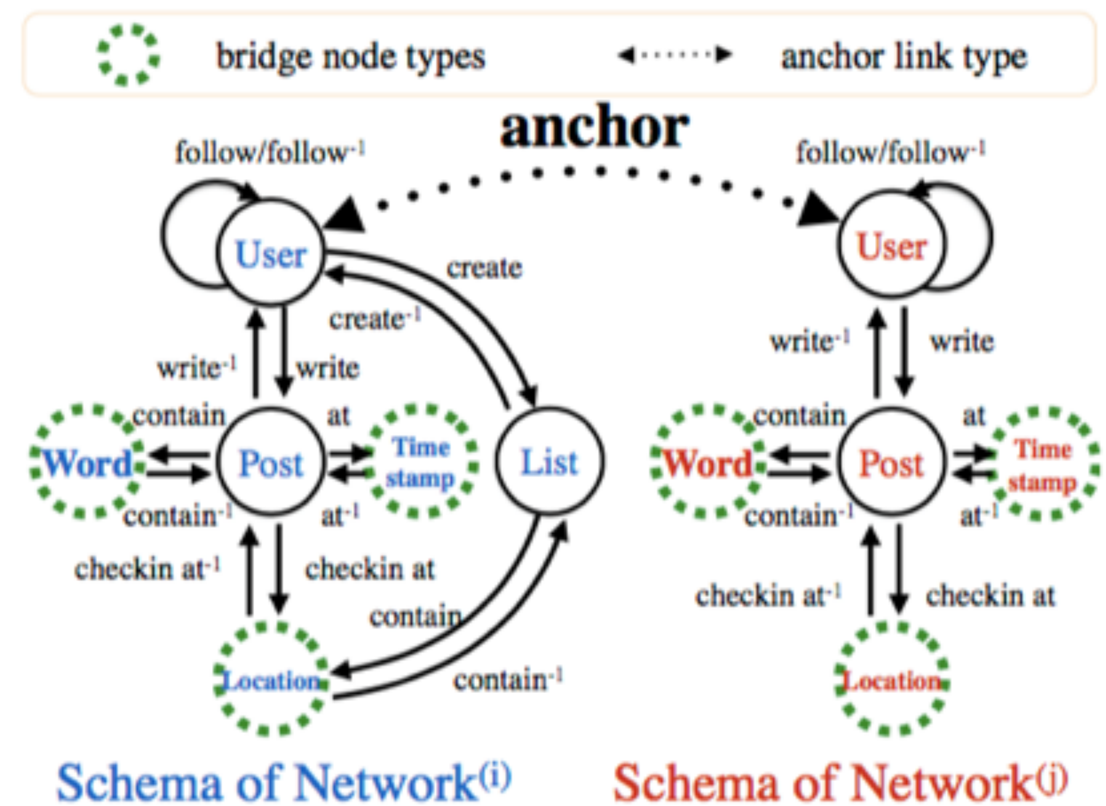


Fig. 1. Schema of aligned heterogeneous network.

- *Common Out Neighbor Anchor Meta Path* (Ψ_1): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$" for short.
- *Common In Neighbor Anchor Meta Path* (Ψ_2): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common Out In Neighbor Anchor Meta Path* (Ψ_3): $User^{(i)} \xrightarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xrightarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \rightarrow \mathcal{U}^{(j)}$".
- *Common In Out Neighbor Anchor Meta Path* (Ψ_4): $User^{(i)} \xleftarrow{follow} User^{(i)} \xleftrightarrow{Anchor} User^{(j)} \xleftarrow{follow} User^{(j)}$ or "$\mathcal{U}^{(i)} \leftarrow \mathcal{U}^{(i)} \leftrightarrow \mathcal{U}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Location Checkin Anchor Meta Path 1* (Ψ_5): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{checkin\ at} Location \xleftarrow{checkin\ at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{L} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Location Checkin Anchor Meta Path 2* (Ψ_6): $User^{(i)} \xrightarrow{create} List^{(i)} \xrightarrow{contain} Location \xleftarrow{checkin\ at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{I}^{(i)} \rightarrow \mathcal{L} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Timestamps Anchor Meta Path* (Ψ_7): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{at} Time \xleftarrow{at} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{T} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".
- *Common Word Usage Anchor Meta Path* (Ψ_8): $User^{(i)} \xrightarrow{write} Post^{(i)} \xrightarrow{contain} Word \xleftarrow{contain} Post^{(j)} \xleftarrow{write} User^{(j)}$ or "$\mathcal{U}^{(i)} \rightarrow \mathcal{P}^{(i)} \rightarrow \mathcal{W} \leftarrow \mathcal{P}^{(j)} \leftarrow \mathcal{U}^{(j)}$".

Details in Propagation

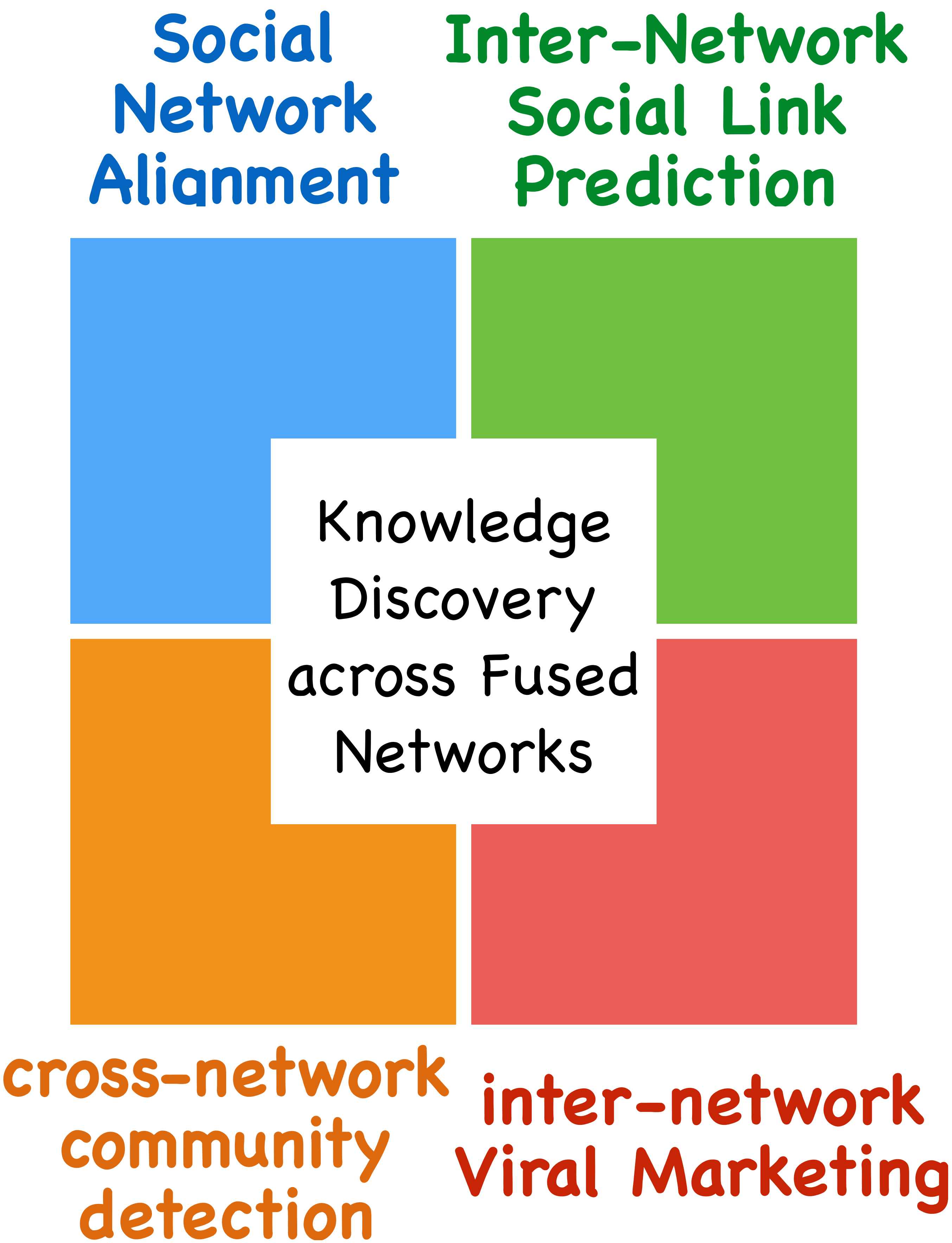
- Diffusion weight：
$$\phi^i_{(u,v)} = \frac{2|P^i_{(u,v)}|}{|P^i_{(u,)}| + |P^i_{(,v)}|}$$
- Activation Probability:
$$g^{(1)}_{v,i}(t+1) = \frac{\sum_{u \in \Gamma_{in}(v,i)} \phi^i_{(u,v)} \varphi(u,t)}{\sum_{u \in \Gamma_{in}(v,i)} \phi^i_{(u,v)}}$$
- Aggregation: logistic function
$$p^{(1)}_v(t+1) = \frac{e^{\sum_{(i)} \rho^{(1)}_i g^{(1)}_{v,i}(t+1) + \sum_{(j)} \omega^{(1)}_j h^{(1)}_{v,j}(t+1)}}{1 + e^{\sum_{(i)} \rho^{(1)}_i g^{(1)}_{v,i}(t+1) + \sum_{(j)} \omega^{(1)}_j h^{(1)}_{v,j}(t+1)}}$$

Aligned Heterogeneous Network Influence Maximization

1. The problem is NP-hard.
2. Based on our diffusion model, influence function is monotone.
3. Based on our diffusion model, influence function is submodular.
4. Algorithm: Step-wise greedy algorithm achieves a (1-1/e)-approximation.

Outline

1. Background Knowledge and Basic Concepts
2. Problem 1: Network Alignment
3. Problem 2: Social Link Prediction across Networks
4. Future work 1: Community Detection across Networks
5. Future work 2: Viral Marketing across Networks
6. **Summary and Selected References**



Summary of problems across aligned networks

- Social Network Alignment: to identify shared users across aligned networks;
- Social Link Prediction: to infer the friendship among users across aligned networks;
- Community Detection: to detect the community structure formed by users across aligned networks;
- Viral Marketing: viral marketing across aligned social networks to achieve broader influence.

References

Network Alignment and Anchor Link Inference

[1] **Jiawei Zhang** and Philip S. Yu. PCT: Partial Co-Alignment of Social Networks. In: Proceedings of the 25th International World Wide Web Conference (WWW '16), Montreal, Canada, April 11-15, 2016.

[2] **Jiawei Zhang** and Philip S. Yu. Multiple Anonymized Social Networks Alignment. In: Proceedings of the IEEE International Conference on Data Mining (ICDM '15), Atlantic City, NJ, USA, November 14-17, 2015.

[3] **Jiawei Zhang** and Philip S. Yu. Integrated Anchor and Social Link Predictions across Partially Aligned Social Networks. In: Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI '15), Buenos Aires, Argentinean, July 25-31, 2015.

[4] **Jiawei Zhang**, Weixiang Shao, Senzhang Wang, Xiangnan Kong and Philip S. Yu. PNA: Partial Network Alignment with Generic Stable Matching. In: Proceedings of the 16th IEEE International Conference on Information Reuse and Integration (IEEE IRI '15), San Francisco, CA, USA, August 13-15, 2015.

[5] Xiangnan Kong, **Jiawei Zhang** and Philip S. Yu. Inferring Anchor Links across Heterogeneous Social Networks. In: Proceedings of the 22nd ACM International Conference on Information and Knowledge Management (CIKM '13), San Francisco, CA, October 27-November 1, 2013.

Link Prediction and Recommendation

[6] **Jiawei Zhang**, Philip S. Yu and Zhi-Hua Zhou. Meta-path based Multi-network Collective Link Prediction. In: Proceedings of the 20th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '14), New York City, NY, August 24-27, 2014.

[7] **Jiawei Zhang**, Xiangnan Kong and Philip S. Yu. Transferring Heterogeneous Links across Location-Based Social Networks. In: Proceedings of the 7th International ACM Conference on Web Search and Data Mining (WSDM '14), New York City, NY, February 24-28, 2014.

[8] **Jiawei Zhang**, Xiangnan Kong and Philip S. Yu. Predicting Social Links for New Users across Aligned Heterogeneous Social Networks. In: Proceedings of the 13th IEEE International Conference on Data Mining (ICDM '13), Dallas, Texas, December 7-10, 2013.

References

Clustering and Community Detection

[9] **Jiawei Zhang** and Philip S. Yu. Community Detection for Emerging Networks. In: Proceedings of the 15th SIAM International Conference on Data Mining (SDM '15), Vancouver, British Columbia, Canada, Apr 30-May 2, 2015.

[10] **Jiawei Zhang** and Philip S. Yu. MCD: Mutual Clustering across Multiple Heterogeneous Networks. In: Proceedings of IEEE BigData Congress (IEEE BigData Congress '15), New York City, NY, June 27-July 2, 2015.

[11] Weixiang Shao, **Jiawei Zhang**, Lifang He and Philip S. Yu. Multi-Source Multi-View Clustering via Discrepancy Penalty. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN '16), Vancouver, Canada, July 24-29, 2016.

[12] Songchang Jin, **Jiawei Zhang**, Philip S. Yu, Shuqiang Yang and Aiping Li. Synergistic Partitioning in Multiple Large Scale Social Networks. In: Proceedings of the 2014 IEEE International Conference on Big Data (IEEE BigData '14), Washington, DC, October 27-30, 2014

Information Diffusion and Viral Marketing

[13] Qianyi Zhan, **Jiawei Zhang**, Senzhang Wang, Philip S. Yu and Junyuan Xie. Influence Maximization across Partially Aligned Heterogenous Social Networks. In: Proceedings of the 19th Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD '15), Ho Chi Minh City, Vietnam, May 19-22, 2015.

[14] **Jiawei Zhang**, Philip S. Yu, Yuanhua Lv, and Qianyi Zhan. Information Propagation at Workplace via Diffusion Channel Aggregation. Submitted to the Proceedings of the 22nd ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD).

References

Enterprise and Professional Social Networks Knowledge Discovery

[15] **Jiawei Zhang**, Philip S. Yu and Yuanhua Lv. Organizational Chart Inference. In: Proceedings of the 21st ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD ’15), Sydney, Australia, August 10-13, 2015.

[16] **Jiawei Zhang**, Yuanhua Lv and Philip S. Yu. Enterprise Social Link Prediction. In: Proceedings of the 24th ACM International Conference on Information and Knowledge Management (CIKM ’15), Melbourne, Australia, October 19-23, 2015.

[17] Qingbo Hu, Sihong Xie, **Jiawei Zhang**, Qiang Zhu and Philip S. Yu. HeteroSales: Utilizing Heterogeneous Social Networks to Identify the Next Enterprise Customer. In: Proceedings of the 25th International World Wide Web Conference (WWW ’16), Montreal, Canada, April 11-15, 2016.

Geo-spatial Information Fusion and Mining

[18] **Jiawei Zhang**, Xiao Pan, Moyin Li and Philip S. Yu. Bicycle-Sharing System Analysis and Trip Prediction. In: Proceedings of the 17th International Conference on Mobile Data Management (MDM '16), Porto, Portugal, June 13-16, 2016.

Other Correlated Papers

[19] Senzhang Wang, Honghui Zhang, **Jiawei Zhang**, Philip S. Yu and Zhoujun Li. Inferring Diffusion Networks with Structure Transfer. In: Proceedings of the 20th International Conference on Database Systems for Advanced Applications (DASFAA ’15), Hanoi, Vietnam, April 20-23, 2015.

[20] **Jiawei Zhang**, Qianyi Zhan, and Philip S. Yu. Concurrent Alignment of Multiple Anonymized Social Networks with Generic Stable Matching. Submitted to the Springer Advances in Intelligent Systems and Computing (AISC) Journal.

UIC Department of Computer Science

Fusion of Heterogeneous Social Networks for Synergistic Knowledge Discovery

Q & A

Jiawei Zhang
jzhan9@uic.edu