Link Prediction in (Partially) Aligned Heterogeneous Social Networks

By: Sina Sajadmanesh
Advisor: Dr. Hamid Reza Rabiee

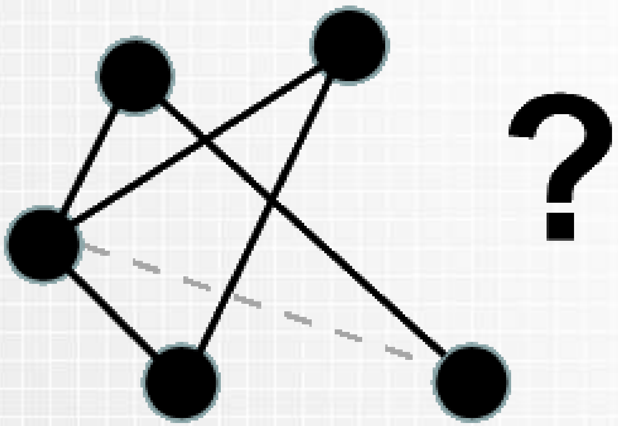
Outline

- Introduction
  - Problem Formulation
  - Applications
- Link Prediction in Homogeneous Networks
- Link Prediction in Heterogeneous Networks
- Link Prediction in Aligned Networks
- Future Works

Introduction

❖Problem

- Based on a snapshot of network, predicting the set of potential links to be formed in the future is formally defined as **Link Prediction Problem**.
- First proposed by Liben-Novel and Klienberg in CIKM 2003



Applications

- Social networks and E-commerce
  - Recommender systems
  - Friend recommendation
- Bioinformatics
  - Prioritization of candidate disease genes
  - Drug discovery
- Security
  - Identify missing links between criminals
  - Controlling computer viruses

Outline

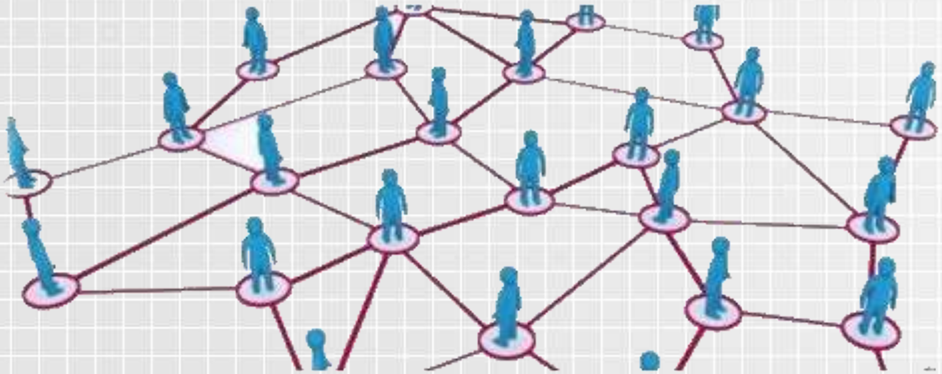
- Introduction
- Link Prediction in Homogeneous Networks
  - Unsupervised methods
  - Supervised methods
- Link Prediction in Heterogeneous Networks
- Link Prediction in Aligned Networks
- Future Works

Link Prediction in Homogeneous Networks

- Homogeneous Network

$$G = (V, E)$$

  - If V contains one single type nodes and E contains one single type of links, then G is a **homogeneous network**



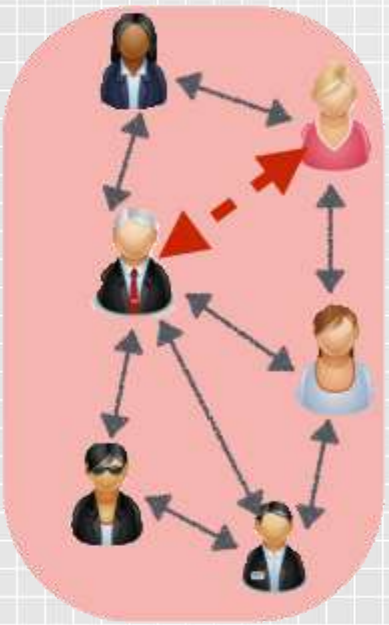
Link Prediction in Homogeneous Networks

- Unsupervised methods
  - Measuring the closeness among nodes
  - Assuming that close nodes are more likely to be connected
- Unsupervised Link Predicators
  - Local neighbor based predicators
  - Path based predicators
  - Random-Walk based predicators

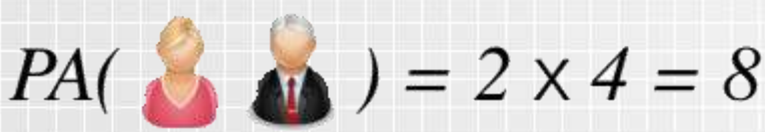
Unsupervised Methods

- Local Neighbor based Link Predicators
  - Preferential Attachment

$$PA(u, v) = |\Gamma(u)||\Gamma(v)|$$



user u neighbor $\Gamma(u)$

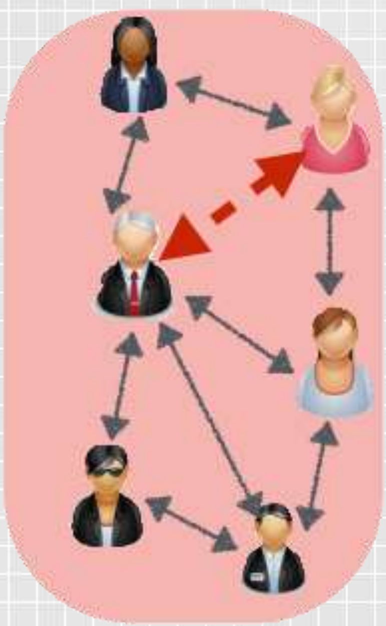


Unsupervised Methods

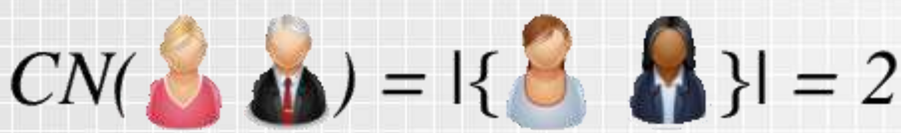
❖ Local Neighbor based Link Predicators

- Common Neighbor

$$CN(u, v) = |\Gamma(u) \cap \Gamma(v)|$$



user u neighbor $\Gamma(u)$



Unsupervised Methods

- Local Neighbor based Link Predicators
  - Jaccard Coefficient

$$JC(u,v) = \frac{|\Gamma(u) \cap \Gamma(v)|}{|\Gamma(u) \cup \Gamma(v)|}$$

  - Adamic/Adar

$$AA(u,v) = \sum_{w \in (\Gamma(u) \cap \Gamma(v))} \frac{1}{\log(|\Gamma(w)|)}$$

  - Resource Allocation

$$RA(u,v) = \sum_{w \in (\Gamma(u) \cap \Gamma(v))} \frac{1}{|\Gamma(w)|}$$

Unsupervised Methods

❖ Path based Link Predicators

- Shortest Path

$$SP(u, v) = \min\{|P_{u \rightsquigarrow v}|\}$$

- Katz Score

$$Katz(u, v) = \sum_{l=1}^{\infty} \beta^l |P^l_{u \rightsquigarrow v}|$$

$$= (I - \beta A)^{-1} - I$$

Unsupervised Methods

- Random-Walk based Link Predicators
  - Hitting Time

$$HT(u,v) = 1 + \sum_{w \in \Gamma(u)} P_{u,w} HT(w,v)$$

  - Commute Time

$$CT(u,v) = HT(u,v) + HT(v,u)$$

Link Prediction in Homogeneous Networks

❖ Other Unsupervised Methods

- Matrix Factorization methods
- Maximum Likelihood Methods
- Probabilistic Methods

Link Prediction in Homogeneous Networks

❖ Supervised Link Prediction

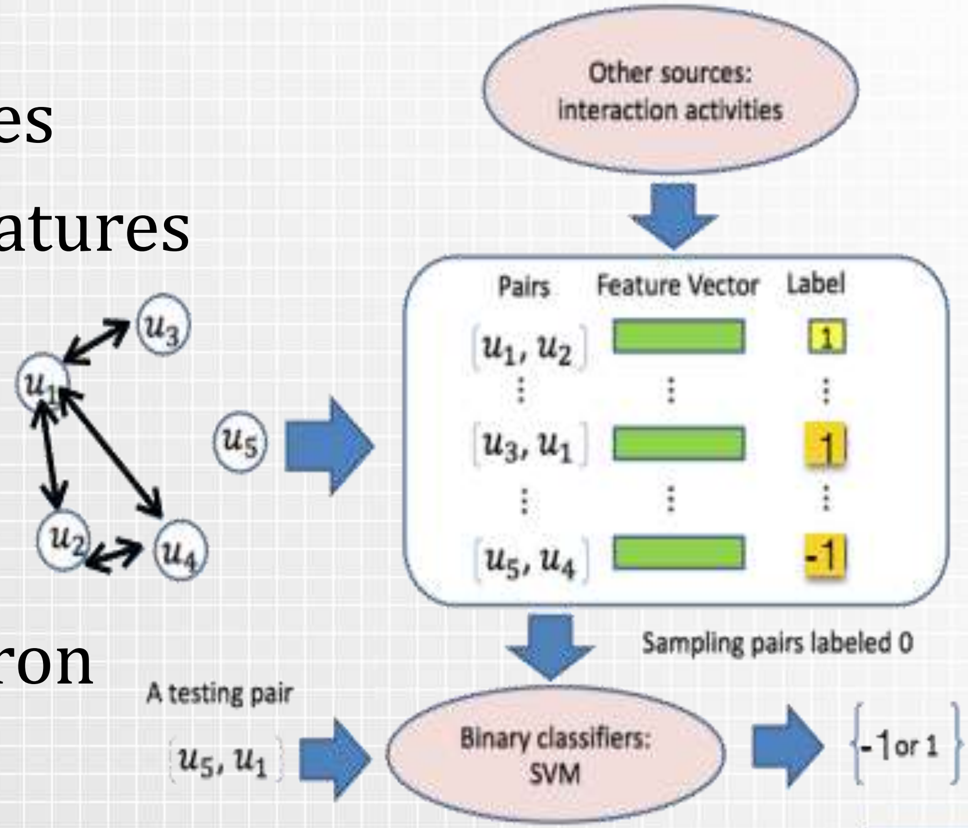
- Learning a binary classifier that will predict whether a link exists between a given pair of nodes or not.
- First proposed by Hassan et. al in SDM 2006

Supervised Link Prediction

- Dataset
  - We need two snapshot of the network for training
- Positive Samples
  - Links that are missing in former snapshot, but are formed in latter snapshot
- Negative Samples
  - Links that are missing in both snapshots

Supervised Link Prediction

- Features
  - Topological Features
  - Attribute-based Features
- Classifiers
  - SVM
  - Decision Trees
  - Multilayer Perceptron
  - KNN
  - Naïve Bayes



Outline

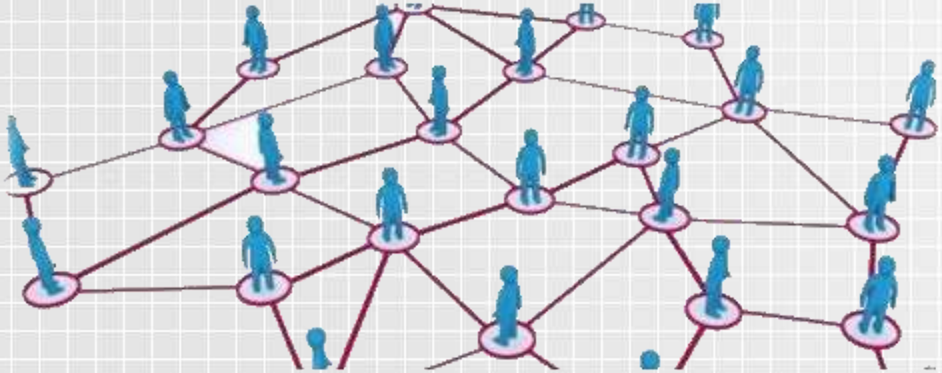
- Introduction
- Link Prediction in Homogeneous Networks
- Link Prediction in Heterogeneous Networks
  - Relationship Prediction
  - Collective Link Prediction
- Link Prediction in Aligned Networks
- Future Works

Link Prediction in Heterogeneous Networks

❖Heterogeneous Network

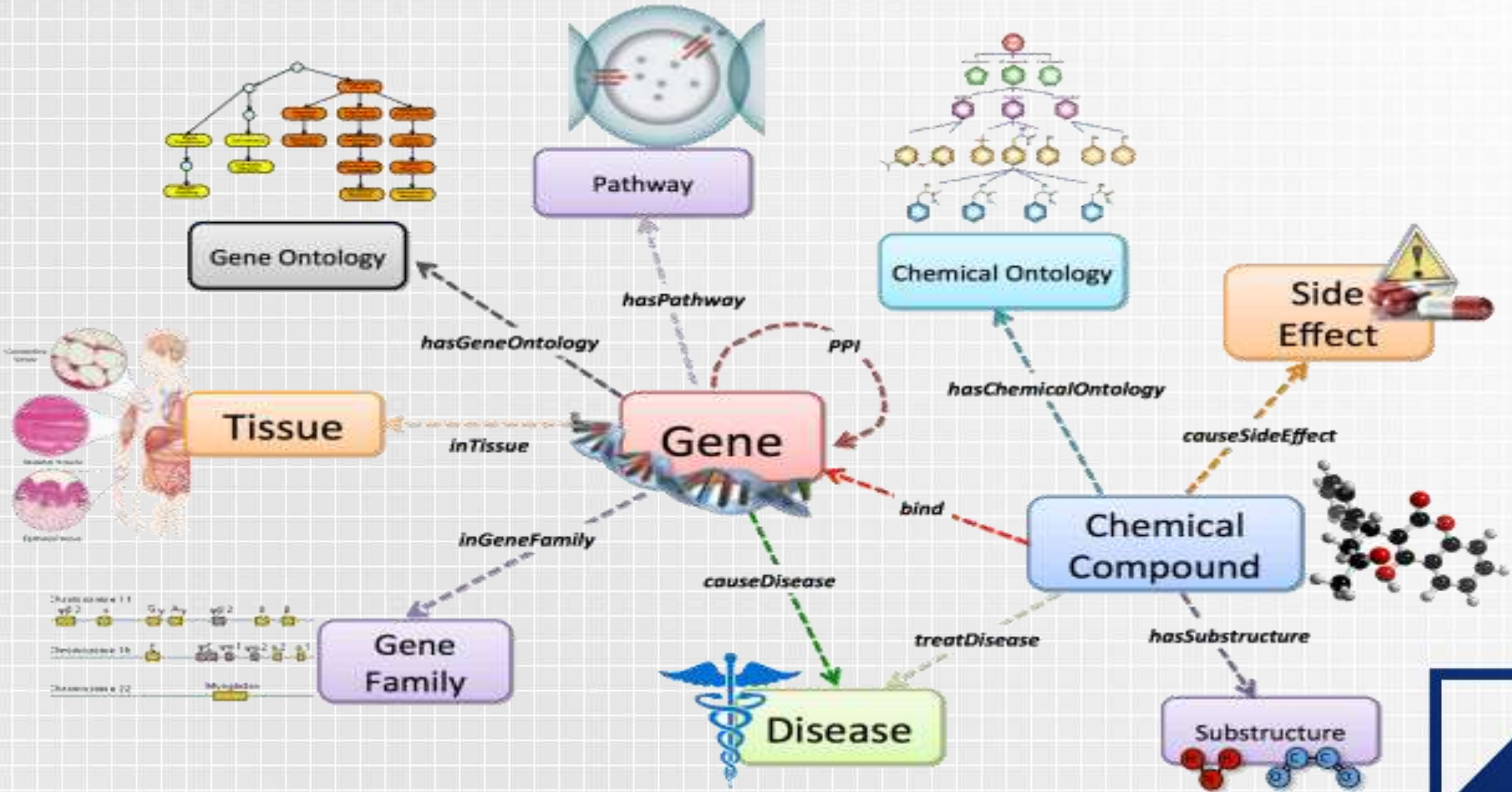
$$G = (V, E)$$

- $V = \bigcup_i V_i$ is the sets of various kinds of nodes and V_i is the i_{th} kind of nodes in G
- $E = \bigcup_j E_j$ is the sets of various types of links and E_j is the j_{th} kind of links in G



Link Prediction in Heterogeneous Networks

❖ Heterogeneous Network Schema



Link Prediction in Heterogeneous Networks

❖ From Link Prediction to Relationship Prediction

- A relationship between two objects could be a composition of two or more links
  - E.g. two authors have a co-author relationship if they have co-written a paper
- Need to redesign topological features in heterogeneous networks

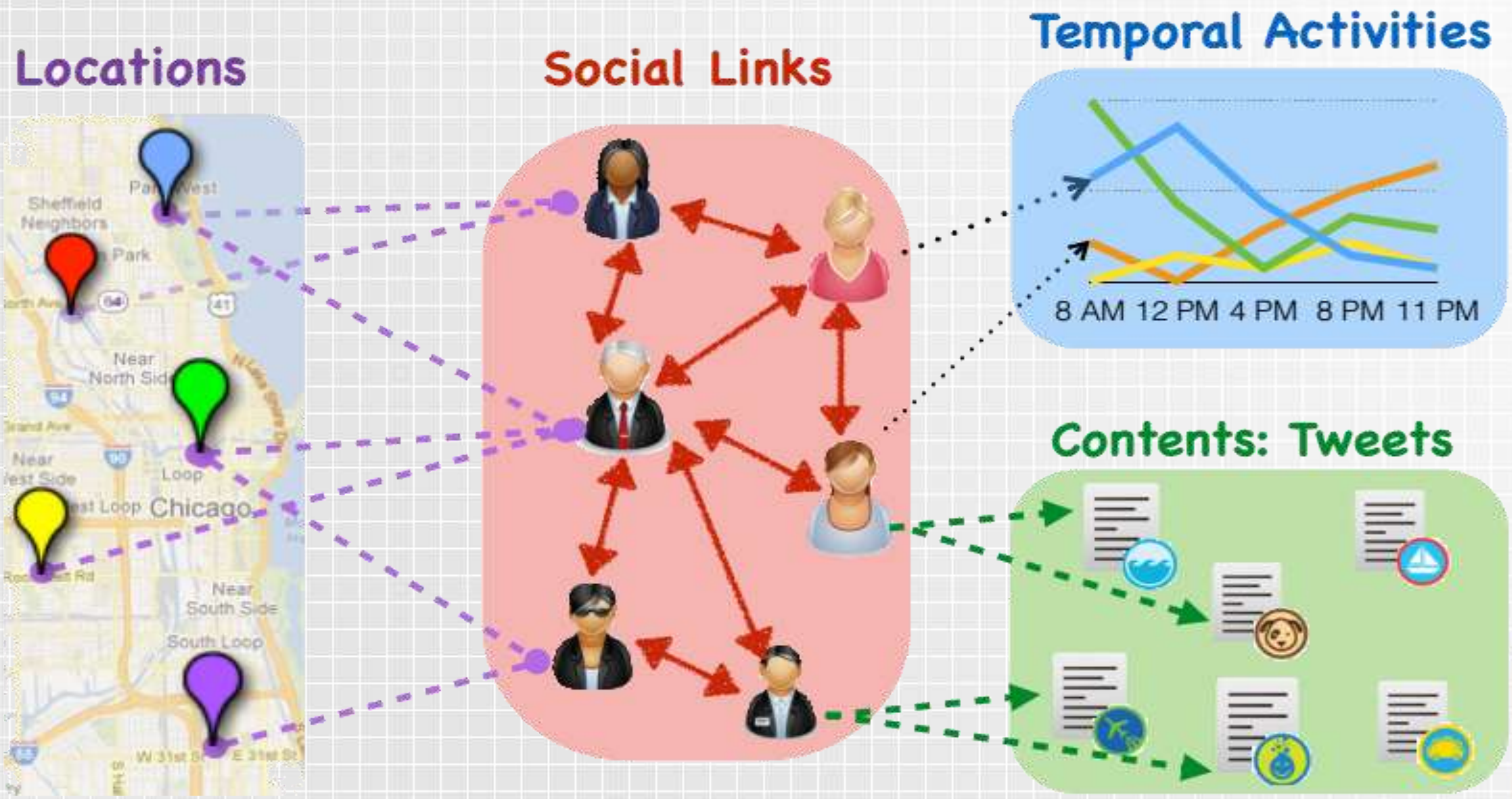
Supervised Relationship Prediction

❖ Feature Extraction

- Heterogeneous Features
  - Based on heterogeneous structure of network
- Meta-Path based Features
  - Uses the concepts of meta-paths
  - Different meta paths represent different semantic meanings
  - Number of path instanced of a meta-path Φ

Supervised Relationship Prediction

- Case Study
  - Social Link Prediction



Supervised Relationship Prediction

- Heterogeneous Features
  - Social Features
    - Common Neighbor
    - Jaccard Coefficient
    - Adamic/Adar
  - Spatial Features
    - Common Locations
    - Jaccard Coefficient of Common Locations
    - Average Geographic distance of locations
    - ...

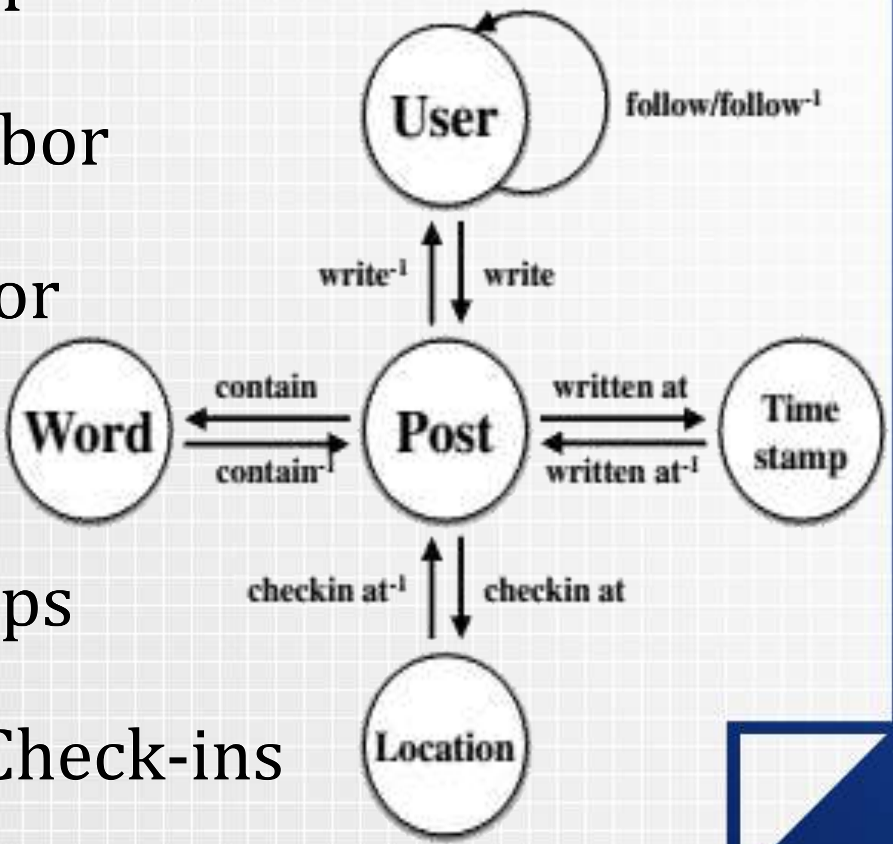
Supervised Relationship Prediction

- Heterogeneous Features
  - Temporal Features
    - Let T(u) be 24 hour activity vector of user u
    - Inner product of T(u) and T(v)
    - Cosine similarity of T(u) and T(v)
  - Text Content Features
    - Let w(u) be the bag-of-words vector of user u weighted by TF-IDF
    - Inner product of w(u) and w(v)
    - Cosine similarity of w(u) and w(v)

Supervised Relationship Prediction

❖ Meta-Path based Features

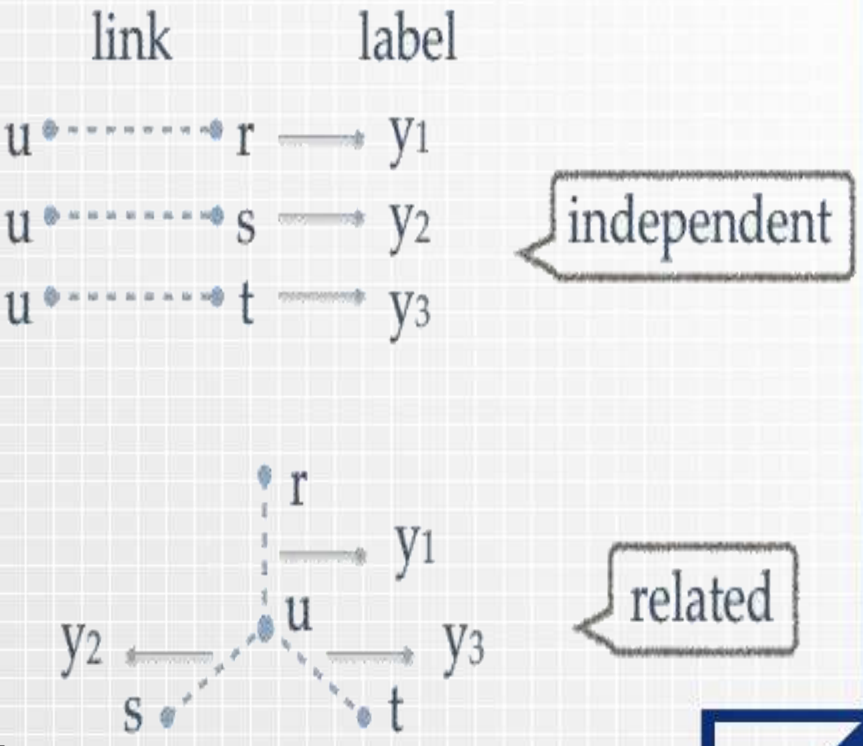
- Φ_1: Follower of Follower
  - U→U→U
- Φ_2: Common Out Neighbor
  - U→U←U
- Φ_3: Common In Neighbor
  - U←U→U
- Φ_4: Common Words
  - U→P→W←P←U
- Φ_5: Common Timestamps
  - U→P→T←P←U
- Φ_6: Common Location Check-ins
  - U→P→L←P←U



Link Prediction in Heterogeneous Networks

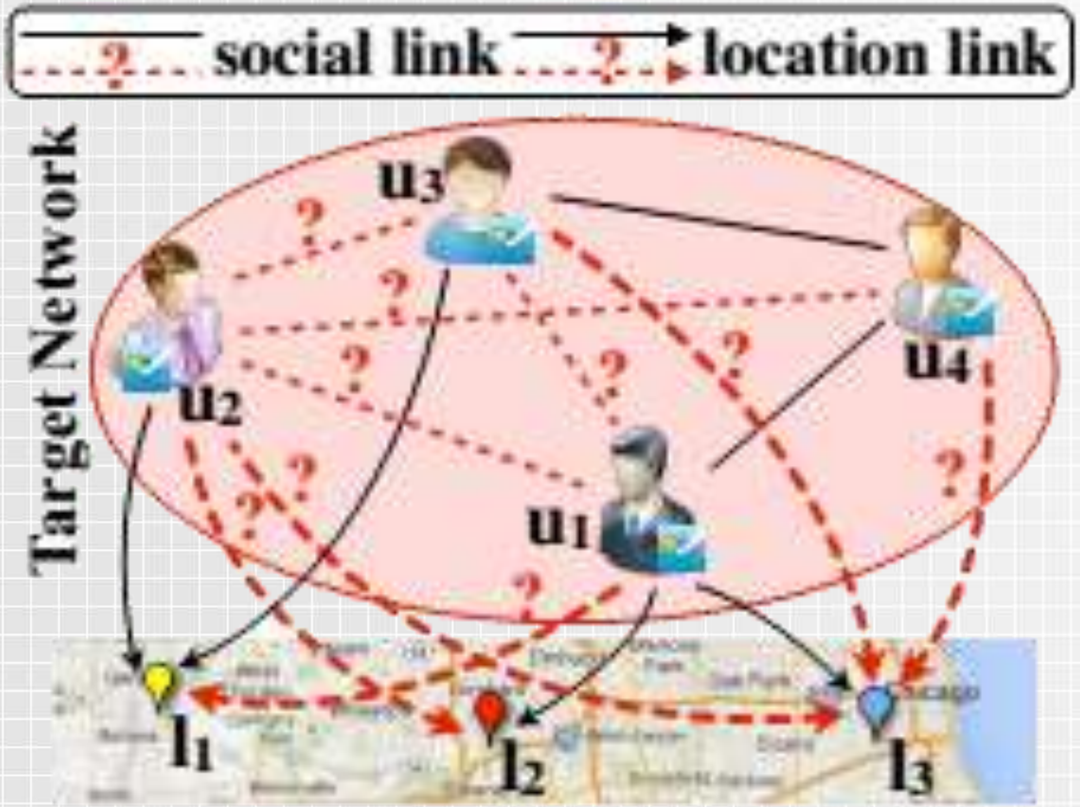
❖ Collective Link Prediction

- Conventional link prediction approaches assume that links are independent identically distributed (i.i.d).
- But in heterogeneous networks, different type of links are correlated and mutually influential.



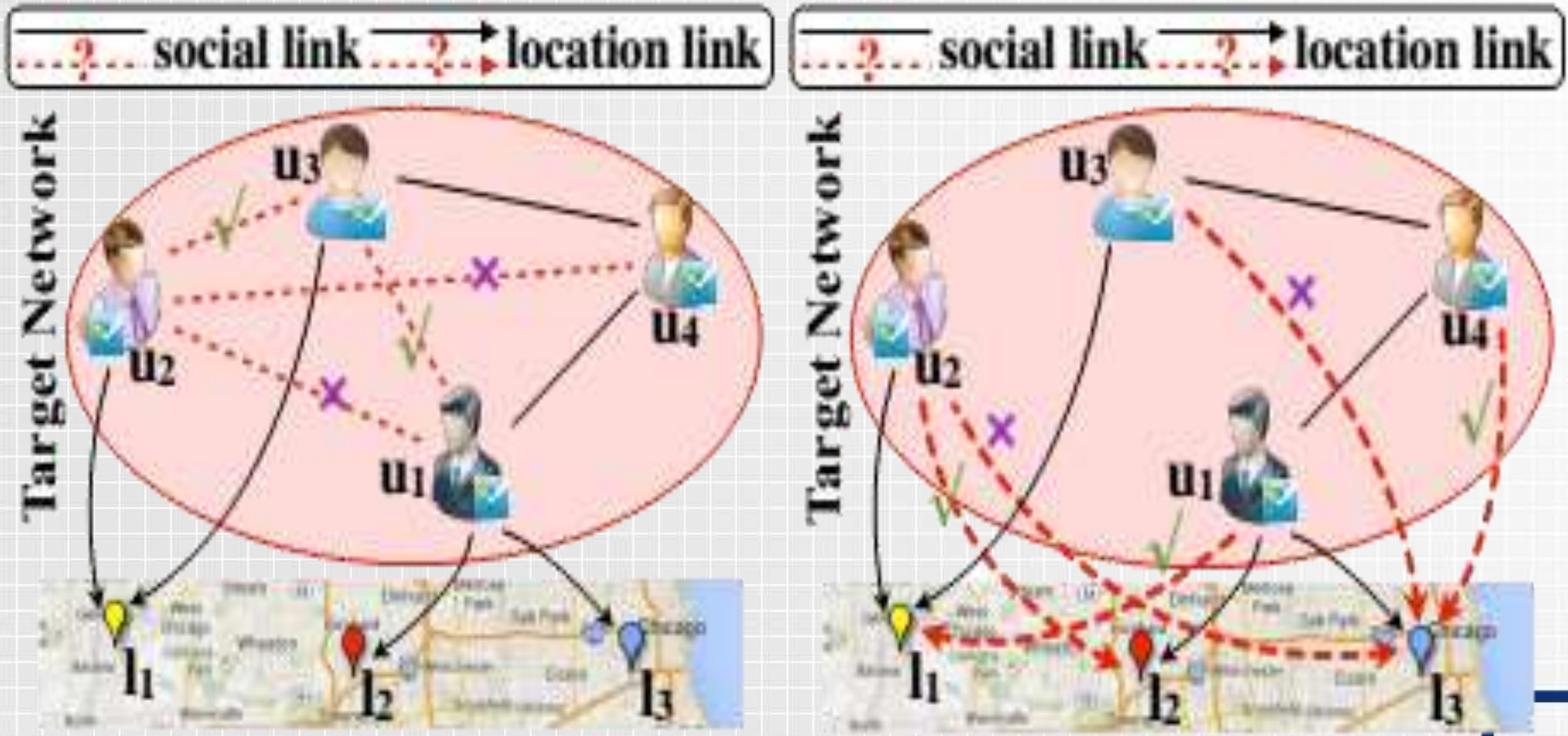
Link Prediction in Heterogeneous Networks

❖Input Social Network



Link Prediction in Heterogeneous Networks

❖ Independent Social and Location Link Prediction



Link Prediction in Heterogeneous Networks

- **Traditional Link Prediction**

$$\hat{Y}_s = arg\max_{Y_s} P(y(L_s) = Y_s)$$

$$\hat{Y}_l = arg\max_{Y_l} P(y(L_l) = Y_l)$$

  - L_s and L_l are the sets of potential social and location links
  - $P(y(L_s) = Y_s)$ is the probability scores achieved when links in L_s are assigned with Labels Y_s
  - $\hat{Y}_s$ and $\hat{Y}_l$ are the sets of optimal labels

Link Prediction in Heterogeneous Networks

- Collective Link Prediction

$$\hat{Y}_s, \hat{Y}_l = arg \max_{Y_s, Y_l} \begin{bmatrix} P(y(L_s) = Y_s | y(L_l) = Y_l) \\ \times P(y(L_l) = Y_l | y(L_s) = Y_s) \end{bmatrix}$$

Iterative Solution

$$\hat{Y}_s^{(t)} = arg \max_{Y_s} \begin{matrix} P(y(L_s) = Y_s \mid \\ y(L_s) = \hat{Y}_s^{(t-1)}, y(L_l) = \hat{Y}_l^{(t-1)}) \end{matrix}$$

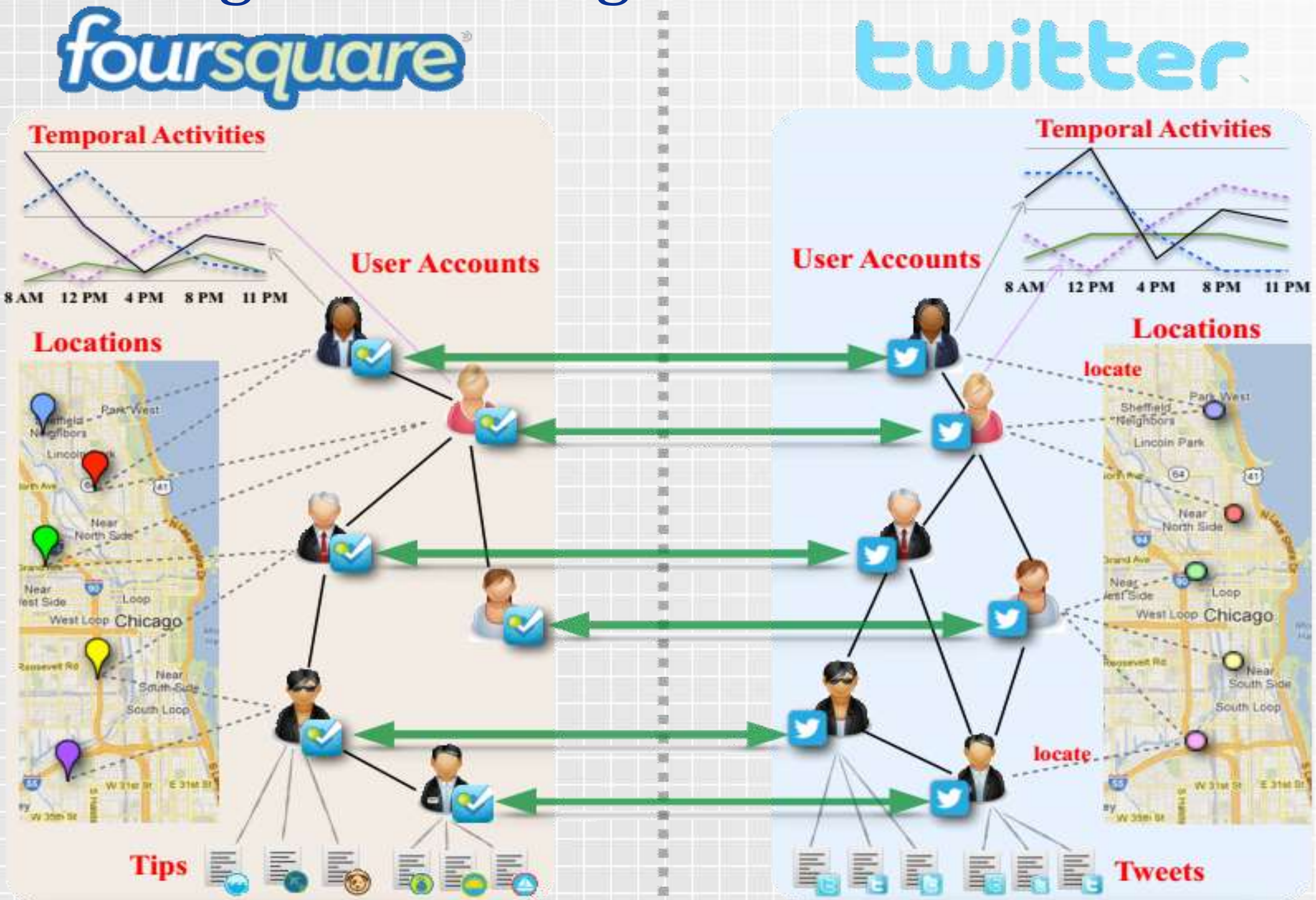
$$\hat{Y}_l^{(t)} = arg \max_{Y_l} \begin{matrix} P(y(L_l) = Y_l \mid \\ y(L_s) = \hat{Y}_s^{(t)}, y(L_l) = \hat{Y}_l^{(t-1)}) \end{matrix}$$

Outline

- Introduction
- Link Prediction in Homogeneous Networks
- Link Prediction in Heterogeneous Networks
- Link Prediction in Aligned Networks
  - Link Transfer
  - Anchor Link Inference
- Future Works

Link Prediction in Aligned Networks

❖ Multi Aligned Heterogeneous Social Networks



Link Prediction in Aligned Networks

- Link Transfer
  - Information Sparsity
    - New Network Problem
    - Cold Start Problem
  - Need to transfer knowledge from another domain
    - Transfer Learning
    - Use aligned network as a source
  - Context
    - Fully aligned networks
    - Partially aligned networks

Link Prediction in Aligned Networks

❖ Link Transfer across Fully Aligned Networks

- Prediction using only the target network:

$$P(y(u^t, v^t) = 1 | G^t) = P\big(y(u^t, v^t) = 1 | x(u^t, v^t)\big)$$

- Prediction using source and target network:

$$P(y(u^t, v^t) = 1 | G^t, G^s)$$

$$= P(y(u^t, v^t) = 1 | [x(u^t, v^t)^T, x(u^s, v^s)^T, y(u^s, v^s)]^T)$$

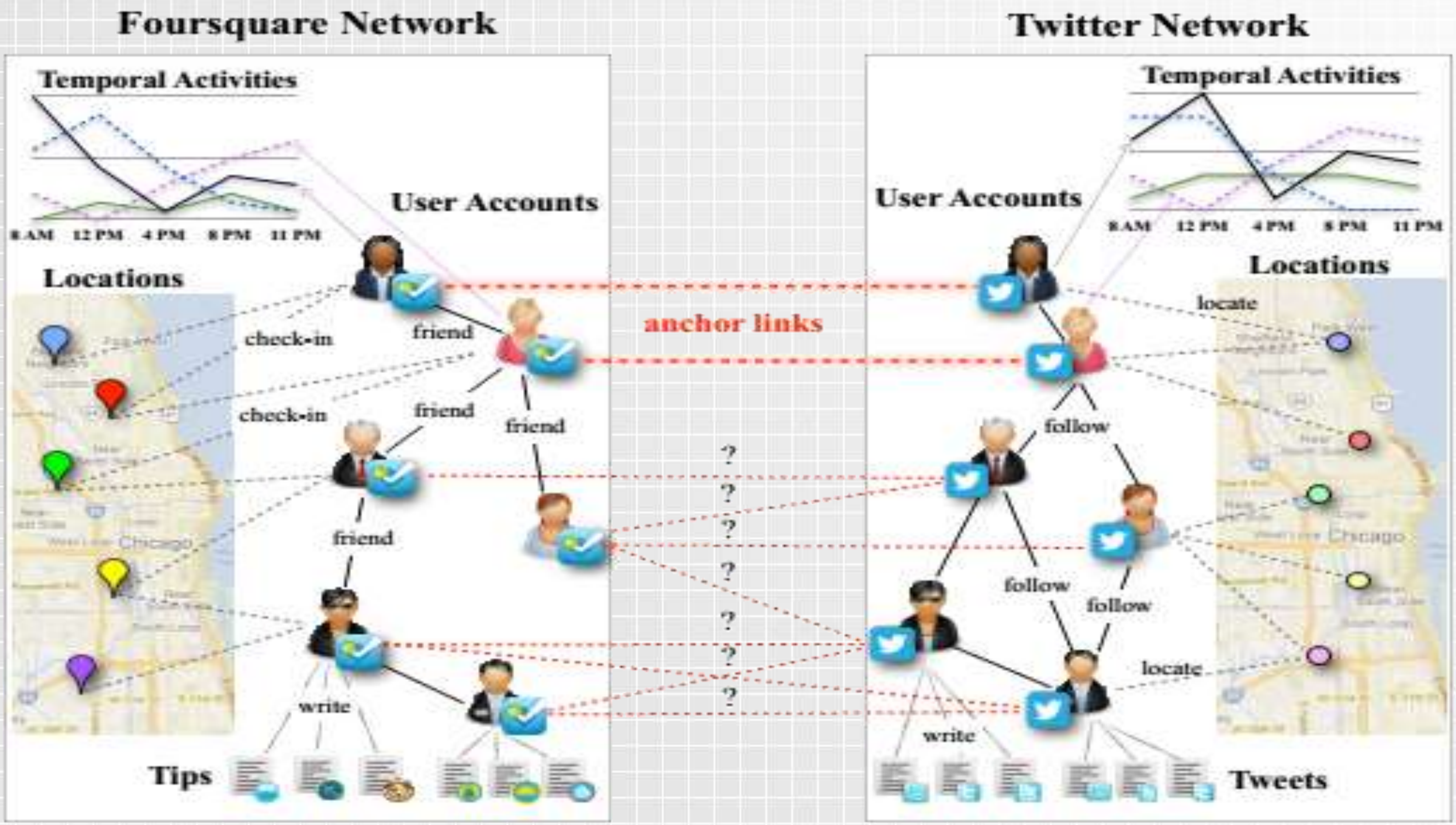
Link Prediction in Aligned Networks

❖ Link Transfer across Partially Aligned Networks

- Solution: Inter-Network Meta-Paths
- Let $\gamma(U^t, U^s)$ be the anchor meta-path
- $\Psi_1 : \gamma(U^t, U^s) - \Phi(U^s, U^s) - \gamma(U^s, U^t)$
- $\Psi_2 : \Phi(U^t, U^t) - \gamma(U^t, U^s) - \Phi(U^s, U^s) - \gamma(U^s, U^t)$
- $\Psi_3 : \gamma(U^t, U^s) - \Phi(U^s, U^s) - \gamma(U^s, U^t) - \Phi(U^t, U^t)$
- $\Psi_3 : \Phi(U^t, U^t) - \gamma(U^t, U^s) - \Phi(U^s, U^s) - \gamma(U^s, U^t) - \Phi(U^t, U^t)$

Link Prediction in Aligned Networks

❖Anchor Link Inference



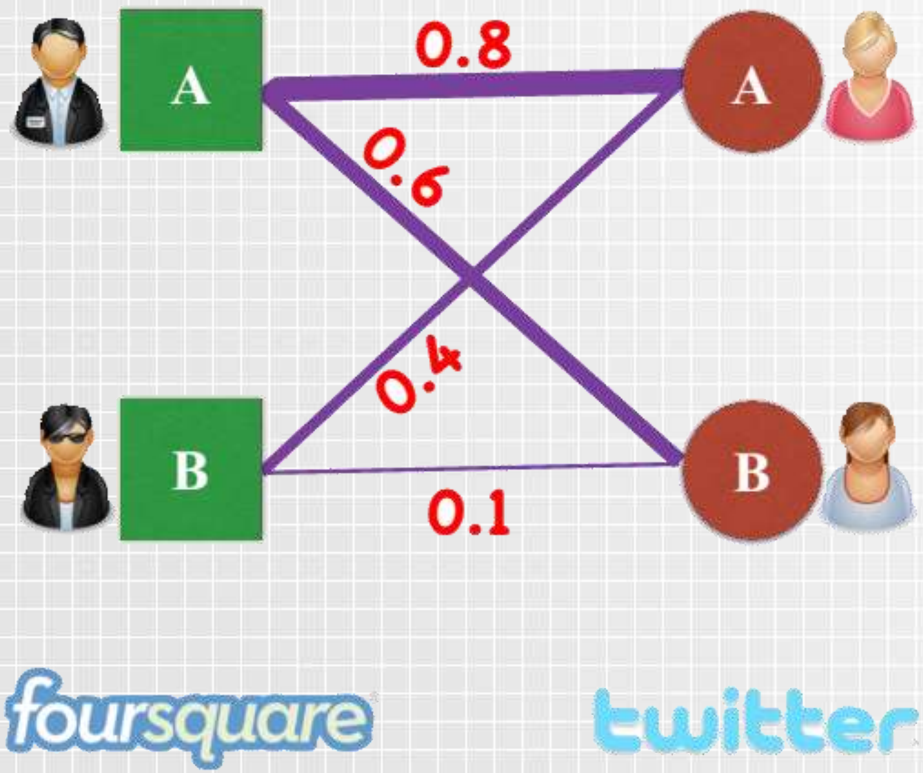
Anchor Link Inference

❖Supervised Method

- Social Features
  - Extended Common Neighbors
  - Extended Jaccard Coefficient
  - Extended Adamic/Adar
- Spatial Features
- Temporal Features
- Text Content Features

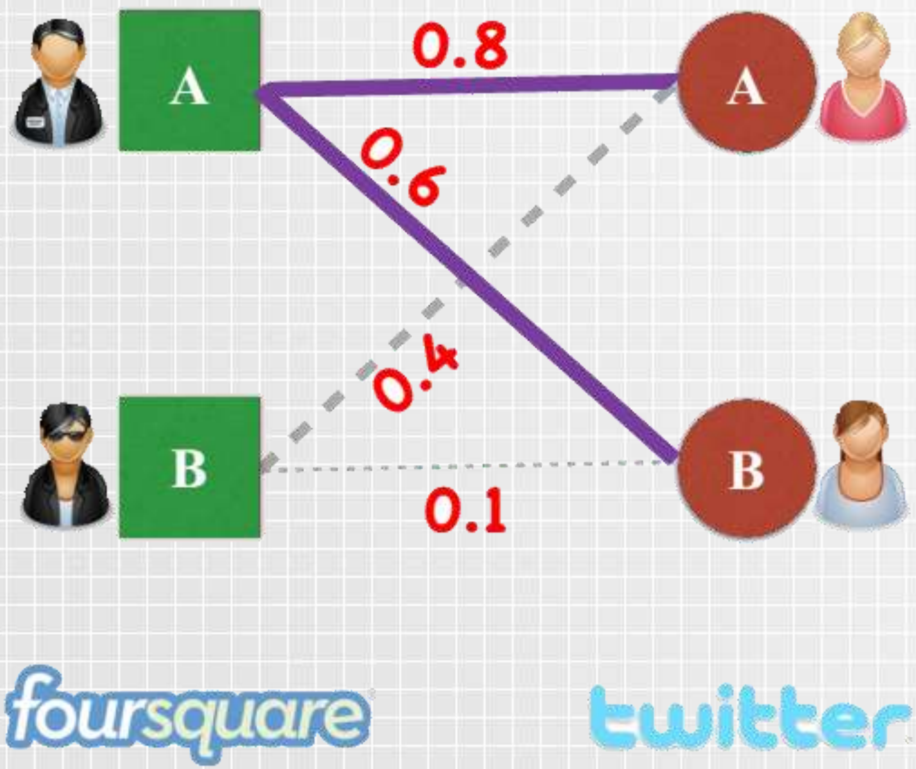
Anchor Link Inference

❖ Inference w.r.t One-to-One Constraint
- Predicted Scores



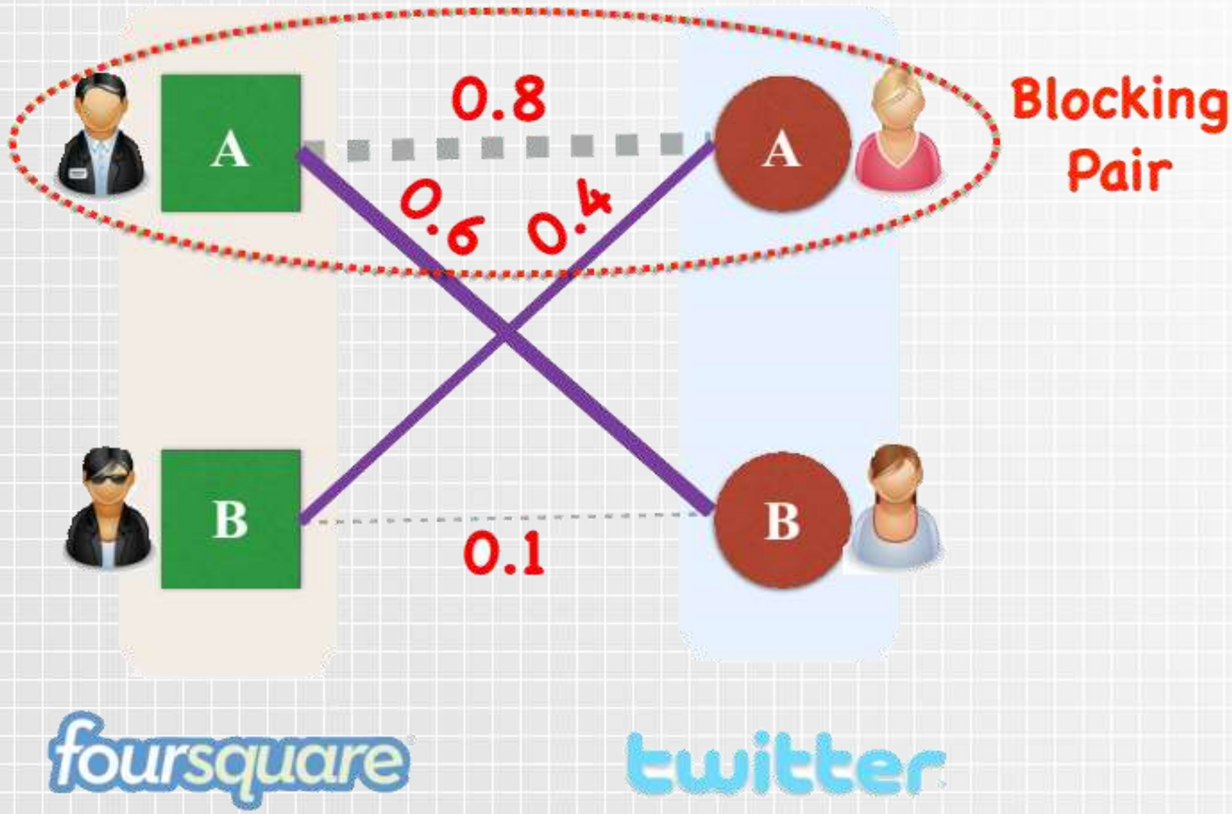
Anchor Link Inference

❖ Inference w.r.t One-to-One Constraint
- Conventional Link Prediction



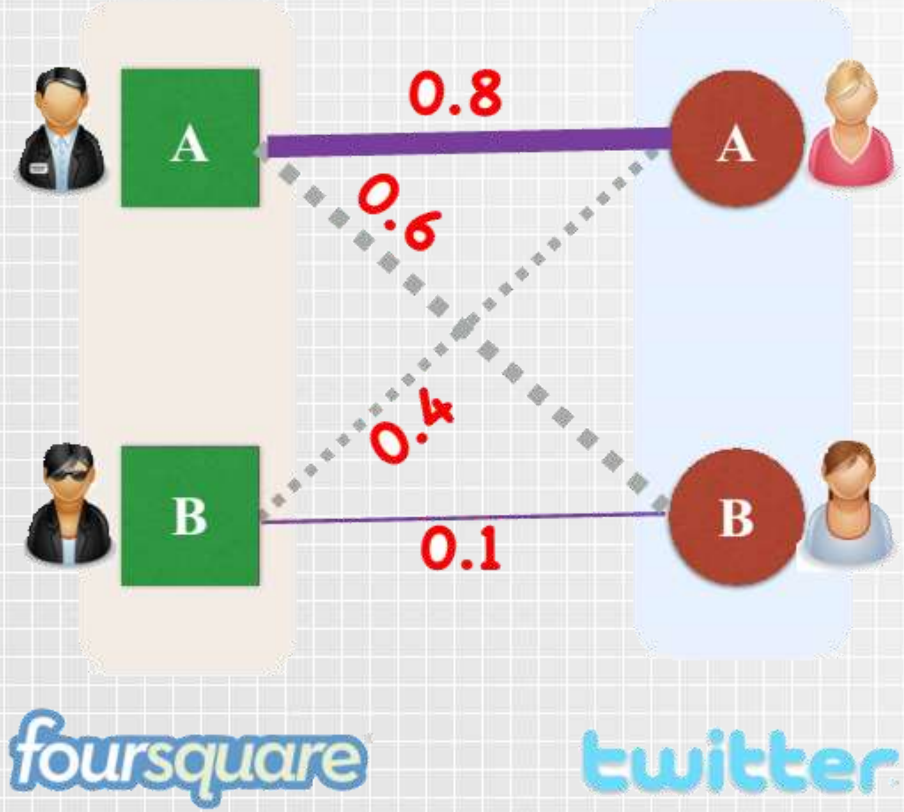
Anchor Link Inference

❖ Inference w.r.t One-to-One Constraint
  - Max sum of scores



Anchor Link Inference

- Inference w.r.t One-to-One Constraint
  - Stable Matching



Outline

- Introduction
- Link Prediction in Homogeneous Networks
- Link Prediction in Heterogeneous Networks
- Link Prediction in Aligned Networks
- Future Works

Future Works

- Anchor Link Formation Prediction
  - Predicting whether a user in the source network will join the target network in the future
  - This depends on the amount of influence he/she receives from the target network
  - An influence model can be learned using the training data
  - Positive-Unlabeled learning can improve the prediction performance

References

[1] L. Lü and T. Zhou, "Link prediction in complex networks: A survey," *Physica A: Statistical Mechanics and its Applications,* vol. 390, pp. 1150-1170, 2011.

[2] M. Al Hasan, V. Chaoji, S. Salem, and M. Zaki, "Link prediction using supervised learning," in *SDM'06: Workshop on Link Analysis, Counter-terrorism and Security*, 2006.

[3] J. Zhang and S. Y. Philip, "Link Prediction across Heterogeneous Social Networks: A Survey," 2014.

[4] Y. Sun, R. Barber, M. Gupta, C. C. Aggarwal, and J. Han, "Co-author relationship prediction in heterogeneous bibliographic networks," in *Advances in Social Networks Analysis and Mining (ASONAM), 2011 International Conference on*, 2011, pp. 121-128.

[5] X. Kong, J. Zhang, and P. S. Yu, "Inferring anchor links across multiple heterogeneous social networks," in *Proceedings of the 22nd ACM international conference on Conference on information & knowledge management*, 2013, pp. 179-188.

Q&A

Thank You!