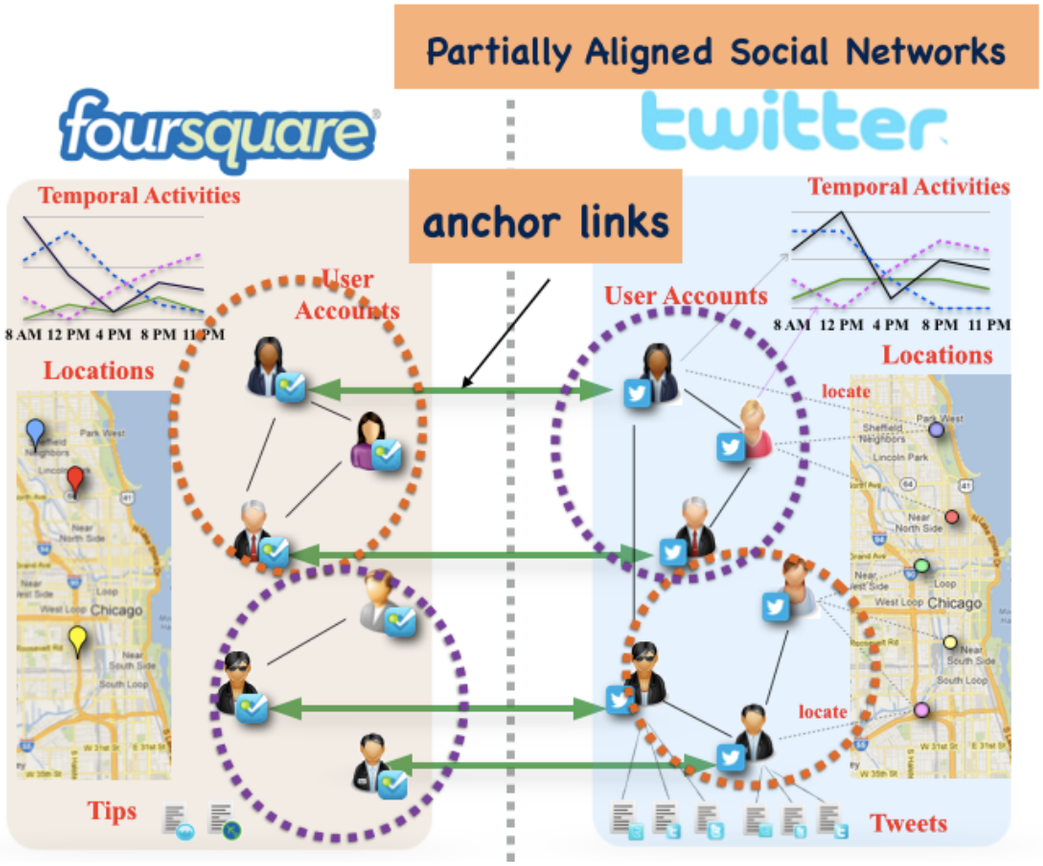
MCD: Mutual Clustering across Multiple Social Networks

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

New York City, USA
June 27 - July 2, 2015

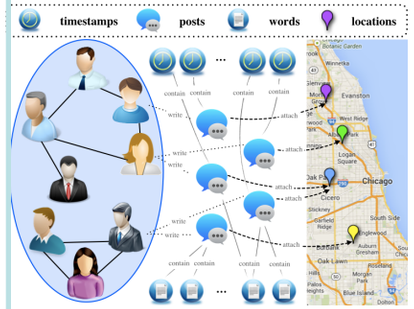


Background Knowledge:
(1)users are involved in multiple online social networks simultaneously
(2)social networks containing common users share similar structures

Problem Studied: Mutual Clustering across Multiple Aligned Social Networks

Challenges
Challenge 1: **Similarity measure** among users with heterogeneous information in social networks
Challenge 2: **Community detection in each network**
Challenge 3: **Mutual community detection**

Solution to Challenge 1: Meta Path based Similarity Measure



ID	Notation	Heterogeneous Network Meta Path
1	U → U	User $\xrightarrow{follow}$ User
2	U → U → U	User $\xrightarrow{follow}$ User $\xrightarrow{follow}$ User
3	U → U ← U	User $\xrightarrow{follow}$ User $\xrightarrow{follow^{-1}}$ User
4	U ← U → U	User $\xrightarrow{follow^{-1}}$ User $\xrightarrow{follow}$ User
5	U → P → W ← P ← U	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Word $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write^{-1}}$ User
6	U → P → T ← P ← U	User $\xrightarrow{write}$ Post $\xrightarrow{contain}$ Time $\xrightarrow{contain^{-1}}$ Post $\xrightarrow{write^{-1}}$ User
7	U → P → L ← P ← U	User $\xrightarrow{write}$ Post $\xrightarrow{attach}$ Location $\xrightarrow{attach^{-1}}$ Post $\xrightarrow{write^{-1}}$ User

e.g., similarity score between x and y based on meta path 1-7

$$\mathrm{Sim}(x,y)=\sum_i \omega_i \left(\frac{|\mathcal{P}_i(x\rightsquigarrow y)|+|\mathcal{P}_i(y\rightsquigarrow x)|}{|\mathcal{P}_i(x\rightsquigarrow \cdot)|+|\mathcal{P}_i(y\rightsquigarrow \cdot)|}\right)$$

Solution to Challenge 2: Normalized-Cut based Isolated Community Detection

let $\mathcal{C}=\{U_1,U_2,\ldots,U_k\}$ be the community structures detected from G.

$$cut(\mathcal{C})=\frac{1}{2}\sum_{i=1}^{k}S(U_i,\overline{U_i})=\frac{1}{2}\sum_{i=1}^{k}\sum_{u\in U_i,v\in\overline{U_i}}S(u,v),\quad Ncut(\mathcal{C})=\frac{1}{2}\sum_{i=1}^{k}\frac{S(U_i,\overline{U_i})}{S(U_i,\cdot)}=\sum_{i=1}^{k}\frac{cut(U_i,\overline{U}_i)}{S(U_i,\cdot)},$$

Solution to Challenge 3: Normalized-Discrepancy based Mutual Community Detection

Let u_i and u_j be two anchor users in the network,

Definition 2 (Discrepancy): The discrepancy between the clustering results of u_i and u_j across aligned networks $G^{(1)}$ and $G^{(2)}$ is defined as the difference of confidence scores of u_i and u_j being partitioned in the same cluster across aligned networks.

$$d_{ij}(\mathcal{C}^{(1)},\mathcal{C}^{(2)})=\left(\mathbf{h}_i^{(1)}(\mathbf{h}_i^{(1)})^T-\mathbf{h}_i^{(2)}(\mathbf{h}_j^{(2)})^T\right)^2$$

$$d(\mathcal{C}^{(1)},\mathcal{C}^{(2)})=\sum^{n^{(1)}}\sum^{n^{(2)}}d_{ij}(\mathcal{C}^{(1)},\mathcal{C}^{(2)}),$$

$$Nd(\mathcal{C}^{(1)},\mathcal{C}^{(2)})=\frac{d(\mathcal{C}^{(1)},\mathcal{C}^{(2)})}{\left(|A^{(1,2)}|\right)\left(|A^{(1,2)}|-1\right)}.$$

Joint Objective Function

$$\arg\min_{\mathcal{C}^{(1)},\mathcal{C}^{(2)}}\ \alpha\cdot Ncut(\mathcal{C}^{(1)})+\beta\cdot Ncut(\mathcal{C}^{(2)})+\theta\cdot Nd(\mathcal{C}^{(1)},\mathcal{C}^{(2)})$$

1 University of Illinois at Chicago
2 Tsinghua University

psyu@cs.uic.edu, jzhan9@uic.edu