SOCIAL RECOMMENDATION ACROSS MULTIPLE RELATIONAL DOMAINS

Meng Jiang

Joint work with Peng Cui, Fei Wang, Qiang Yang, Wenwu Zhu and Shiqiang Yang
November 1, 2012 – Maui, HI, USA
Recommender Systems

- **Predict missing “user-item” links**

- **Challenge:** Cold-start and extremely high sparsity

![Diagram]

- **web posts**
- **cold-start new item**
- **cold-start new user**
- **high sparsity**

- **users**
OUTLINE

1. Background
2. The Framework
3. HRW Algorithm
4. Experiments
5. Insights
Multiple Domains

• User label domain

Choose < 10 from 200+ labels like ‘iPhone fan’

Peng Cui
Haidian, Beijing
Company: Tsinghua

User labels (5)
Tsinghua, Ph.D., World Wide Web, Social Network, Social Media

Meng Jiang
Haidian, Beijing
University: Tsinghua

User labels (9)
Chinese food, World Wide Web, Social Network, Data Mining, Liverpool Football Club, NBA, Humors, Sports, Ph.D. Candidates
Multiple Domains

• Interest group domain

Interest Groups (2)

Tsinghua University

I love singing!

Interest Groups (3)

Tsinghua University

Social Media & Reputation

World Wide Web Team
Our Goals

- Given: Links on social networks
- Find: A framework that use auxiliary knowledge in multiple domains to best predict “user-item” (target) links when the training set is too small.

Goals:
- G1. Understand link formations on social networks
- G2. A social network framework with multiple domains
- G3. Solve the cold-start problem
Challenges: Multiple Domains

• Relational
  • Within-domain links and cross-domain links

• Heterogeneous
  • Different types of item domains

• Sparse
  • Different sparsity levels
OUTLINE

1. Background
2. The Framework
3. HRW Algorithm
4. Experiments
5. Insights
Reframe Social Networks

- We have user-user, post-post and label-label links (social relation + item similarity).
Reframe Social Networks

• We have user-post and user-label links.
Reframe Social Networks

• No relations between item domains.
• No post-label links in nature.
Reframe Social Networks

• Stronger social relations help collaborate user-item links.
Reframe Social Networks

- More collaborating in user-item links strengthen the social relations.
Star-structured Graph

- Key idea: use “social relation” domain as bridge
3. HRW Algorithm
Star-structured Graph

- Method: Transfer learning + Random walk with restarts
Hybrid Random Walk

- On second-order star-structured graph

- Update cross-domain links

\[
\begin{align*}
\mathbf{P}^{(\mathcal{U})^+} (t + 1) &= \delta \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{U})^+}(t) + (1 - \delta) \mathbf{P}^{(\mathcal{U})^+}(t) \mathbf{R}^{(\mathcal{P})} \\
\mathbf{P}^{(\mathcal{U})^-} (t + 1) &= \delta \mathbf{R}^{(\mathcal{U})}(t) \mathbf{P}^{(\mathcal{U})^-}(t) + (1 - \delta) \mathbf{P}^{(\mathcal{U})^-}(t) \mathbf{R}^{(\mathcal{P})}
\end{align*}
\]
Hybrid Random Walk

- Update within-domain links

\[
R^{(U)}(t + 1) = 
\tau^{(P)}(\mu P^{(UP)^+}(t)P^{(UP)^+}(t)^T + (1 - \mu) P^{(UP)^-}(t)P^{(UP)^-}(t)^T) \\
+ \tau^{(\mathcal{T})} P^{(UT)^+}(t)P^{(UT)^+}(t)^T + \tau^{(U)} R^{(U)}(t)R^{(U)}(t)^T
\]
Hybrid Random Walk

• On high-order star-structured graph

\[
P^{(U \Delta_i^+)}(t+1) = \delta_i R^{(U)}(t) P^{(U \Delta_i^+)}(t) + (1 - \delta_i) P^{(U \Delta_i^+)}(t) R^{(\Delta_i)}
\]
\[
P^{(U \Delta_i^-)}(t+1) = \delta_i R^{(U)}(t) P^{(U \Delta_i^-)}(t) + (1 - \delta_i) P^{(U \Delta_i^-)}(t) R^{(\Delta_i)}
\]
\[
R^{(U)}(t+1) = \sum_{\Delta_i \in \mathcal{D}} \tau_i \mu_i P^{(U \Delta_i^+)}(t) P^{(U \Delta_i^+)}(t)^T
\]
\[
+ \sum_{\Delta_i \in \mathcal{D}} \tau_i (1 - \mu_i) P^{(U \Delta_i^-)}(t) P^{(U \Delta_i^-)}(t)^T
\]
\[
+ \tau^{(U)} R^{(U)}(t) R^{(U)}(t)^T
\]
| 1. Background                      |
| 2. The Framework                  |
| 3. HRW Algorithm                  |
| 4. Experiments                    |
| 5. Insights                       |
## Data Set

- Tencent Weibo (January 2011)

<table>
<thead>
<tr>
<th>Domain</th>
<th>Size</th>
<th>Cross-domain links</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Accept</td>
</tr>
<tr>
<td>User</td>
<td>53.4K</td>
<td>—</td>
</tr>
<tr>
<td>Web post</td>
<td>142K</td>
<td>1.47M (0.02%)</td>
</tr>
<tr>
<td>User label</td>
<td>111</td>
<td>330K (5.57%)</td>
</tr>
</tbody>
</table>
Good to Transfer?

- Comparative Algorithms (RWR)
  - $W^{(P)}$: Use web post similarity?
  - $W^{(U)}$: Use social relation?
  - $R^{(U)}$: Update tie strength?
  - $W^{(T)}$: Use user label similarity?

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>$R^{(U)}$</th>
<th>$W^{(U)}$</th>
<th>$W^{(P)}$</th>
<th>$W^{(T)}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRW</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>BRW-$R_U$-P</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>(TrustWalker)</td>
<td>✓</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>BRW-$W_U$</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>BRW-$W_U$-P</td>
<td>×</td>
<td>✓</td>
<td>✓</td>
<td>×</td>
</tr>
<tr>
<td>BRW-$W_U$</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>(ItemRank)</td>
<td>×</td>
<td>✓</td>
<td>×</td>
<td>×</td>
</tr>
<tr>
<td>BRW-P</td>
<td>×</td>
<td>×</td>
<td>✓</td>
<td>×</td>
</tr>
</tbody>
</table>
Good to Transfer!

- Compare with RWR models

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MAE</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Kendall’s $\tilde{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRW</td>
<td>0.227±1.5e-3</td>
<td>0.711±1.3e-3</td>
<td>0.921±1.4e-3</td>
<td>0.802±1.1e-3</td>
<td>0.792±2.5e-3</td>
</tr>
<tr>
<td>BRW-$R_U$-P (TrustWalker)</td>
<td>0.276±1.1e-3</td>
<td>0.657±7.6e-4</td>
<td><strong>0.935±9.8e-4</strong></td>
<td>0.772±7.6e-4</td>
<td>0.774±1.6e-3</td>
</tr>
<tr>
<td>BRW-$R_U$</td>
<td>0.282±5.3e-3</td>
<td>0.655±4.0e-3</td>
<td>0.921±1.2e-2</td>
<td>0.765±7.7e-3</td>
<td>0.725±2.8e-3</td>
</tr>
<tr>
<td>BRW-$W_U$-P</td>
<td>0.292±1.1e-3</td>
<td>0.666±7.0e-4</td>
<td>0.900±5.2e-4</td>
<td>0.765±6.6e-4</td>
<td>0.725±8.5e-4</td>
</tr>
<tr>
<td>BRW-$W_U$ (ItemRank)</td>
<td>0.318±1.4e-3</td>
<td>0.671±1.5e-3</td>
<td>0.713±2.4e-3</td>
<td>0.691±1.2e-3</td>
<td>0.661±2.2e-3</td>
</tr>
<tr>
<td>BRW-P</td>
<td>0.438±2.6e-4</td>
<td>0.571±3.4e-4</td>
<td>0.499±4.2e-4</td>
<td>0.532±3.2e-4</td>
<td>0.606±2.3e-4</td>
</tr>
</tbody>
</table>

- Compare with Baselines

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>MAE</th>
<th>Precision</th>
<th>Recall</th>
<th>F1</th>
<th>Kendall’s $\tilde{\tau}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>HRW</td>
<td>0.227±1.5e-3</td>
<td>0.711±1.3e-3</td>
<td>0.921±1.4e-3</td>
<td>0.802±1.1e-3</td>
<td>0.792±2.5e-3</td>
</tr>
<tr>
<td>BRW-$R_U$-P (TrustWalker)</td>
<td>0.276±1.1e-3</td>
<td>0.657±7.6e-4</td>
<td><strong>0.935±9.8e-4</strong></td>
<td>0.772±7.6e-4</td>
<td>0.774±1.6e-3</td>
</tr>
<tr>
<td>BRW-$W_U$ (ItemRank)</td>
<td>0.318±1.4e-3</td>
<td>0.671±1.5e-3</td>
<td>0.713±2.4e-3</td>
<td>0.691±1.2e-3</td>
<td>0.661±2.2e-3</td>
</tr>
<tr>
<td>MCF [5]</td>
<td>0.352±2.3e-4</td>
<td>0.592±1.8e-3</td>
<td><strong>0.951±6.0e-4</strong></td>
<td>0.730±1.3e-3</td>
<td>0.582±4.3e-4</td>
</tr>
<tr>
<td>CF [22]</td>
<td>0.506±3.4e-4</td>
<td>0.552±1.5e-3</td>
<td>0.589±7.2e-4</td>
<td>0.570±1.0e-3</td>
<td>0.540±5.2e-4</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td></td>
</tr>
<tr>
<td><strong>OUTLINE</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1.</td>
<td>Background</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>The Framework</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3.</td>
<td>HRW Algorithm</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4.</td>
<td>Experiments</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5.</td>
<td>Insights</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Insights

• If we do transfer (from user-label domain), we need only ~30% to reach the same performance.
• Advice: build more apps for new users to give more info.
Questions?

Meng Jiang
mjiang89@gmail.com
http://www.meng-jiang.com