Uncovering the Formation of Triadic Closure in Social Networks
Zhanpeng Fang and Jie Tang
Department of Computer Science, Tsinghua University

- Group formation is a central research issue in social science. A triad is a group of three people, one of the most basic human groups in social networks.
- Understanding the formation of triads can help reveal the complex and subtle mechanism that governs dynamics of all social networks.
- Decoding triadic closure: for a given closed triad, predict which link was firstly created, which followed and finally which one was the last to close the triad.

Problem Definition

Let \( G = (V, E) \) denote a social network. Associate a time stamp \( t_{ij} \) to each edge.

Definition 1 Closed Triad: For three users \( \Delta = (A, B, C) \), if there is a link between any two users, i.e., \( e_{AB}, e_{BC}, e_{AC} \in E \), then we say that \( \Delta \) is a closed triad.

DeTriad Model

- Synchronous method for correlation factor
  \[
  h(y_i, y_j) = \frac{1}{Z_3} \exp\left(\sum_{k} \mu_k \cdot I_k(y_i, y_j)\right),
  \]
- Asynchronous method for correlation factor
  \[
  h(y_i, y_j) = \frac{1}{Z_4} \exp\left(\sum_{k,j} \mu_{kj} \cdot I_{kj}(y_i, y_j)\right).
  \]
- Local factors: \( f(y_i|x_k) \) represents posterior probability of the decoding result given all features defined for triad \( \Delta \).
- Correlation factor: \( h(y_i, y_j) \) represents the correlation between decoding results of \( \Delta_i \) and \( \Delta_j \).

\[
\frac{\partial \mathcal{L}(\theta)}{\partial \mu_k} = \mathbb{E}_{\Delta}(h(y_i, y_j) | \Delta, X) \frac{\partial \log \mathcal{L}(\theta)}{\partial \mu_k}.
\]

Experimental Results

Dataset: coauthor network from 1995 to 2014 in Arnetminer includes 1,910,979 publications, 1,145,632 authors, 4,322,998 coauthor relationships, and 5,913,455 closed triads.

4 Categories of features: C(Correlation), D(Demography), I(Interaction), S(Social effect)

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>Spearman</th>
<th>Kendall</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rule</td>
<td>0.4604</td>
<td>0.3525</td>
<td>0.3293</td>
</tr>
<tr>
<td>SVM</td>
<td>0.3205</td>
<td>0.2286</td>
<td>0.4121</td>
</tr>
<tr>
<td>Logistic</td>
<td>0.3379</td>
<td>0.2407</td>
<td>0.4830</td>
</tr>
<tr>
<td>DeTriad-A</td>
<td>0.3060</td>
<td>0.2190</td>
<td>0.5550</td>
</tr>
<tr>
<td>DeTriad</td>
<td>0.2716</td>
<td>0.1935</td>
<td>0.5964</td>
</tr>
</tbody>
</table>

Factor Contribution:
Each category of features contributes improvement to the performance.

Effect of Correlation Factors:
The improvement of DeTriad by adding correlation factors is significantly higher than that of logistic regression by adding correlation features.

Training/Test Ratio:
Improvements can be obtained by DeTriad when increasing the size of training data.