A Naïve Bayes model based on overlapping groups for link prediction in online social networks

Jorge Valverde-Rebaza
and
Alneu de Andrade Lopes

Laboratory of Computational Intelligence (LABIC)
University of São Paulo (USP)

Brazil

April 2015
Social Networks

- Structure made up of a set of actors (individuals or organizations) and social relations between them.

- SNA is an interesting research field in graph and complex network theory, data mining, machine learning and other areas.

- Rise of online social networks (OSN).
Groups detection

- Real networks are characterized by high concentration of links within special groups of vertices and low concentrations of links among these groups.

- Online social networks (OSNs) offer a wide variety of possible (overlapping) groups: families, working and friendship circles, artistic or academic preferences, towns, nations, etc.
Link Prediction (LP) process
Presence of groups
Presence of overlapping groups
Presence of overlapping groups
Presence of overlapping groups
Presence of overlapping groups

A NB model on overlapping groups for link prediction in OSN
Link Prediction in the presence of overlapping groups
LP measures

**Traditional** [Lü and Zhou, 2011]
- Common Neighbors (CN)
- Adamic Adar (AA)
- Jaccard (Jac)
- Resource Allocation (RA)
- Preferential Attachment (PA)
- Others

Based on the Naïve Bayes Model
- Local Naïve Bayes (LNB)
- CN with Local Naïve Bayes (LNB-CN)
- AA with Local Naïve Bayes (LNB-AA)
- RA with Local Naïve Bayes (LNB-RA)

Based on Overlapping Groups
- CN Within and Outside of Common Groups (WOCG)
- CN of Groups (CNG)
- CN with Total and Partial Overlapping of Groups (TPOG)

Our proposals
- Group Naïve Bayes (GNB)
- CN with Group Naïve Bayes (GNB-CN)
- AA with Group Naïve Bayes (GNB-AA)
- RA with Group Naïve Bayes (GNB-RA)
LP measures

**Traditional** [Lü and Zhou, 2011]
- Common Neighbors (CN)
- Adamic Adar (AA)
- Jaccard (Jac)
- Resource Allocation (RA)
- Preferential Attachment (PA)
- Others

**Based on the Naïve Bayes Model** [Liu et al., 2011]
- Local Naïve Bayes (LNB)
- CN with Local Naïve Bayes (LNB-CN)
- AA with Local Naïve Bayes (LNB-AA)
- RA with Local Naïve Bayes (LNB-RA)

**Based on Overlapping Groups**
- CN Within and Outside of Common Groups (WOCG)
- CN of Groups (CNG)
- CN with Total and Partial Overlapping of Groups (TPOG)

Our proposals
- Group Naïve Bayes (GNB)
- CN with Group Naïve Bayes (GNB-CN)
- AA with Group Naïve Bayes (GNB-AA)
- RA with Group Naïve Bayes (GNB-RA)
LP measures

**Traditional** [Lü and Zhou, 2011]
- Common Neighbors (CN)
- Adamic Adar (AA)
- Jaccard (Jac)
- Resource Allocation (RA)
- Preferential Attachment (PA)
- Others

**Based on the Naïve Bayes Model** [Liu et al., 2011]
- Local Naïve Bayes (LNB)
- CN with Local Naïve Bayes (LNB-CN)
- AA with Local Naïve Bayes (LNB-AA)
- RA with Local Naïve Bayes (LNB-RA)

**Based on Overlapping Groups Information** [Valverde-Rebaza and Lopes, 2014]
- CN Within and Outside of Common Groups (WOCG)
- CN of Groups (CNG)
- CN with Total and Partial Overlapping of Groups (TPOG)
LP measures

**Traditional**  [Lü and Zhou, 2011]
- Common Neighbors (CN)
- Adamic Adar (AA)
- Jaccard (Jac)
- Resource Allocation (RA)
- Preferential Attachment (PA)
- Others

**Based on the Naïve Bayes Model**  [Liu et al., 2011]
- Local Naïve Bayes (LNB)
- CN with Local Naïve Bayes (LNB-CN)
- AA with Local Naïve Bayes (LNB-AA)
- RA with Local Naïve Bayes (LNB-RA)

**Based on Overlapping Groups Information**  [Valverde-Rebaza and Lopes, 2014]
- CN Within and Outside of Common Groups (WOCG)
- CN of Groups (CNG)
- CN with Total and Partial Overlapping of Groups (TPOG)

**Our proposals**
- Group Naïve Bayes (GNB)
- CN with Group Naïve Bayes (GNB-CN)
- AA with Group Naïve Bayes (GNB-AA)
- RA with Group Naïve Bayes (GNB-RA)
Definitions

Given the network $G(V, E)$ with $M > 1$ groups identified by different group labels $g_1, g_2, \ldots, g_M$. Each node $x \in V$ belongs to a set of node groups $G_\alpha = \{g_a, g_b, \ldots, g_p\}$ with size $P > 0$ and $P \leq M$.

When a node $x$ belongs to a set of node groups $G_\alpha$, this node is represented as $x^G_\alpha$.

The overlapping groups neighborhood of a node:
$$\Gamma^G(x) = \{y^G_\beta \mid ((x^G_\alpha, y^G_\beta) \in E \lor (y^G_\beta, x^G_\alpha) \in E) \land G_\alpha \cap G_\beta \neq \emptyset\}.$$  

The overlapping groups degree of a node: $k^G(x) = |\Gamma^G(x)|$.

The set of common neighbors of groups is defined as:
$$\Lambda^G_{x, y} = \Gamma^G(x) \cap \Gamma^G(y).$$

We define the overlapping groups clustering coefficient of a node:
$$C^G_x = \frac{\Delta^G_x}{\Delta^G_x + \Lambda^G_x},$$ where $\Delta^G_x$ and $\Lambda^G_x$ are respectively the number of connected and disconnected pair of nodes whose common neighbors of groups include $x$. 

Jorge Valverde-Rebaza
We denote by $L_{x,y}$ and $\overline{L}_{x,y}$ the class variables of link existence and nonexistence, respectively. Thus, the posterior probability of connection and disconnection of the pair $(x, y)$ given its set of common neighbors of groups are:

$$P(L_{x,y} \mid \Lambda^g_{x,y}) = \frac{P(L_{x,y})P(\Lambda^g_{x,y} \mid L_{x,y})}{P(\Lambda^g_{x,y})} \quad P(\overline{L}_{x,y} \mid \Lambda^g_{x,y}) = \frac{P(\overline{L}_{x,y})P(\Lambda^g_{x,y} \mid \overline{L}_{x,y})}{P(\Lambda^g_{x,y})}$$

We define the ratio between these equations define the likelihood score $s_{x,y}$. Decomposing $P(\Lambda^g_{x,y} \mid L_{x,y}) = \prod_{z \in \Lambda^g_{x,y}} P(z \mid L_{x,y})$ and $P(\Lambda^g_{x,y} \mid \overline{L}_{x,y}) = \prod_{z \in \Lambda^g_{x,y}} P(z \mid \overline{L}_{x,y})$, we have:

$$s_{x,y} = \frac{P(L_{x,y})}{P(\overline{L}_{x,y})} \prod_{z \in \Lambda^g_{x,y}} \frac{P(L_{x,y})P(L_{x,y} \mid z)}{P(\overline{L}_{x,y})P(L_{x,y} \mid z)}$$

Considering that $P(L_{x,y} \mid z) = C^g_z$ and $P(\overline{L}_{x,y} \mid z) = 1 - C^g_z$, we define the group naïve Bayes (GNB) measure as:

$$s_{x,y}^{GNB} = \prod_{z \in \Lambda^g_{x,y}} \Omega^{-1} N^g_z$$

where $N^g_z = \frac{\Lambda^g_z + 1}{\Lambda^g_z + 1}$ and $\Omega = \frac{P(L_{x,y})}{P(\overline{L}_{x,y})}$. 
Group Naïve Bayes Forms

- From the GNB equation, we add an exponent \( f(k^G(x)) \) to \( \Omega^{-1} N_z^G \), where \( f \) is a function of overlapping groups degree. Using Log function on both sides, we obtain the next linear equation:

\[
s_{x,y}^{GNB'} = \sum_{z \in \Lambda_{x,y}^G} f(k^G(z)) \log(\Omega^{-1} N_z^G)
\]

- Here we consider three forms of function \( f \): \( f(k^G(x)) = 1 \), \( f(k^G(x)) = \frac{1}{\log(k^G(x))} \) and \( f(k^G(x)) = \frac{1}{k^G(x)} \), which are corresponding to the group naïve Bayes form of CN, AA and RA, respectively:

\[
s_{x,y}^{GNB-CN} = |\Lambda_{x,y}^G| \log(\Omega^{-1}) + \sum_{z \in \Lambda_{x,y}^G} \log(N_z^G)
\]

\[
s_{x,y}^{GNB-AA} = \sum_{z \in \Lambda_{x,y}^G} \frac{1}{\log(k^G(z))} (\log(N_z^G) + \log(\Omega^{-1}))
\]

\[
s_{x,y}^{GNB-RA} = \sum_{z \in \Lambda_{x,y}^G} \frac{1}{k^G(z)} (\log(N_z^G) + \log(\Omega^{-1}))
\]
Outline

1. Introduction
2. Proposal
3. Experiments
4. Conclusions
## Datasets

### Table: Topological features of social networks

<table>
<thead>
<tr>
<th></th>
<th>Flickr</th>
<th>LiveJournal</th>
<th>Orkut</th>
<th>Youtube</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of nodes</td>
<td>1,846,198</td>
<td>5,284,457</td>
<td>3,072,441</td>
<td>1,157,827</td>
</tr>
<tr>
<td>Number of links</td>
<td>22,613,981</td>
<td>77,402,652</td>
<td>223,534,301</td>
<td>4,945,382</td>
</tr>
<tr>
<td>Avg. degree</td>
<td>12.24</td>
<td>16.97</td>
<td>106.1</td>
<td>4.29</td>
</tr>
<tr>
<td>Fraction of symmetric links</td>
<td>62.0%</td>
<td>73.5%</td>
<td>100.0%</td>
<td>79.1%</td>
</tr>
<tr>
<td>Avg. path length</td>
<td>5.67</td>
<td>5.88</td>
<td>4.25</td>
<td>5.10</td>
</tr>
<tr>
<td>Diameter</td>
<td>27</td>
<td>20</td>
<td>9</td>
<td>21</td>
</tr>
<tr>
<td>Avg. clust. coef.</td>
<td>0.313</td>
<td>0.330</td>
<td>0.171</td>
<td>0.136</td>
</tr>
<tr>
<td>Assortativity coef.</td>
<td>0.202</td>
<td>0.179</td>
<td>0.072</td>
<td>-0.033</td>
</tr>
<tr>
<td>Number of groups</td>
<td>103,648</td>
<td>7,489,073</td>
<td>8,730,859</td>
<td>30,087</td>
</tr>
<tr>
<td>Avg. of groups which a user belongs to</td>
<td>4.62</td>
<td>21.25</td>
<td>106.44</td>
<td>0.25</td>
</tr>
<tr>
<td>Avg. group size</td>
<td>82</td>
<td>15</td>
<td>37</td>
<td>10</td>
</tr>
<tr>
<td>Avg. group clust. coef.</td>
<td>0.47</td>
<td>0.81</td>
<td>0.52</td>
<td>0.34</td>
</tr>
<tr>
<td>Avg. overlapping groups degree</td>
<td>9.65</td>
<td>6.19</td>
<td>50.85</td>
<td>0.42</td>
</tr>
<tr>
<td>Avg. overlapping groups clust. coef.</td>
<td>0.06</td>
<td>0.13</td>
<td>0.18</td>
<td>0.02</td>
</tr>
</tbody>
</table>
Experimental setup: unsupervised strategy

For a network $G$, the set $E$ is divided into the training set $E^T$ and the probe set $E^P$. We randomly select the links for these sets considering just the links formed by nodes whose number of neighbors is two times greater than the average degree per node.

For each pair of nodes from $E^T$, the connection likelihood is calculated based on the link direction, choosing the highest score between its $in$ and $out$ scores as the final and unique score, e.g., by vertex pair $(x, y)$ if $s_{x,y}^{out} > s_{x,y}^{in}$ then $s_{x,y} = s_{x,y}^{out}$, otherwise, $s_{x,y} = s_{x,y}^{in}$.
We use decision tree (J48), naïve Bayes (NB), multilayer perceptron with backpropagation (MLP) and support vector machine (SMO) classifiers.

For each network, we compute a set of feature vector formed by randomly selected pair of nodes from $E^T$. If the pair of nodes taken from the predicted links list from $E^T$ is also in $E^P$ then the feature vector formed by this pair of nodes takes the positive class (existent link), otherwise takes the negative class (nonexistent link).
Experimental setup: supervised strategy

**Table: Number of instances by class**

<table>
<thead>
<tr>
<th>Network</th>
<th>Existent</th>
<th>Non-existent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>7,100</td>
<td>35,500</td>
<td>42,600</td>
</tr>
<tr>
<td>LiveJournal</td>
<td>4,500</td>
<td>22,500</td>
<td>27,000</td>
</tr>
<tr>
<td>Orkut</td>
<td>16,000</td>
<td>80,000</td>
<td>96,000</td>
</tr>
<tr>
<td>Youtube</td>
<td>2,700</td>
<td>13,500</td>
<td>16,200</td>
</tr>
</tbody>
</table>

**Table: Data sets created for each network**

<table>
<thead>
<tr>
<th>Data set</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>VLocal</td>
<td>CN, AA, Jac, RA and PA</td>
</tr>
<tr>
<td>VGroups</td>
<td>WOCG, CNG and TPOG</td>
</tr>
<tr>
<td>VLN B</td>
<td>LNB, LNB-CN, LNB-AA and LNB-RA</td>
</tr>
<tr>
<td>VGN B</td>
<td>GNB, GNB-CN, GNB-AA and GNB-RA</td>
</tr>
<tr>
<td>VLocal-Groups</td>
<td>VLocal and VGroups</td>
</tr>
<tr>
<td>VLocal-GNB</td>
<td>VLocal and VGNB</td>
</tr>
<tr>
<td>VLN B-Groups</td>
<td>VLN B and VGroups</td>
</tr>
<tr>
<td>VLN B-GNB</td>
<td>VLN B and VGNB</td>
</tr>
<tr>
<td>VGroups-GNB</td>
<td>VGroups and VGNB</td>
</tr>
<tr>
<td>VTotal</td>
<td>VLocal, VGroups, VLN B and VGN B</td>
</tr>
</tbody>
</table>
### Results: unsupervised evaluation

**Table: The prediction results measured by AUC**

<table>
<thead>
<tr>
<th>Method</th>
<th>Flickr</th>
<th>Livejournal</th>
<th>Orkut</th>
<th>Youtube</th>
<th>Avg. rank</th>
<th>Method</th>
<th>Flickr</th>
<th>Livejournal</th>
<th>Orkut</th>
<th>Youtube</th>
<th>Avg. rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>CN</td>
<td>0.674</td>
<td>0.582</td>
<td>0.572</td>
<td>0.834</td>
<td>10.50</td>
<td>WOCG</td>
<td>0.637</td>
<td>0.596</td>
<td>0.649</td>
<td>0.434</td>
<td>10.75</td>
</tr>
<tr>
<td>AA</td>
<td>0.656</td>
<td>0.580</td>
<td>0.620</td>
<td>0.928</td>
<td>8.25</td>
<td>CNG</td>
<td>0.728</td>
<td>0.611</td>
<td>0.621</td>
<td>0.723</td>
<td>9.63</td>
</tr>
<tr>
<td>Jac</td>
<td>0.431</td>
<td>0.624</td>
<td>0.575</td>
<td>0.217</td>
<td>12.50</td>
<td>TPOG</td>
<td>0.728</td>
<td>0.665</td>
<td>0.651</td>
<td>0.555</td>
<td>8.63</td>
</tr>
<tr>
<td>RA</td>
<td>0.616</td>
<td>0.565</td>
<td>0.566</td>
<td>0.892</td>
<td>11.00</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PA</td>
<td>0.566</td>
<td>0.542</td>
<td>0.602</td>
<td>0.917</td>
<td>10.00</td>
<td>-</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LNB</td>
<td>0.860</td>
<td>0.880</td>
<td>0.446</td>
<td>0.872</td>
<td>7.25</td>
<td>GNB</td>
<td>0.857</td>
<td>0.853</td>
<td>0.525</td>
<td>0.800</td>
<td>10.0</td>
</tr>
<tr>
<td>LNB-CN</td>
<td>0.859</td>
<td>0.877</td>
<td><strong>0.706</strong></td>
<td>0.873</td>
<td>4.50</td>
<td>GNB-CN</td>
<td>0.861</td>
<td>0.855</td>
<td>0.639</td>
<td>0.808</td>
<td>6.25</td>
</tr>
<tr>
<td>LNB-AA</td>
<td><strong>0.884</strong></td>
<td><strong>0.883</strong></td>
<td>0.342</td>
<td>0.890</td>
<td>5.75</td>
<td>GNB-AA</td>
<td>0.875</td>
<td>0.862</td>
<td>0.572</td>
<td>0.807</td>
<td>6.75</td>
</tr>
<tr>
<td>LNB-RA</td>
<td><strong>0.890</strong></td>
<td>0.880</td>
<td>0.333</td>
<td><strong>0.896</strong></td>
<td>5.75</td>
<td>GNB-RA</td>
<td>0.874</td>
<td>0.856</td>
<td>0.539</td>
<td>0.790</td>
<td>8.50</td>
</tr>
</tbody>
</table>

**Figure: Post-hoc test for results from AUC results**
Results: unsupervised evaluation

Figure: Precision results on four social networks. Different values of $L$ are used to select the top-$L$ highest scores for predicting links.
## Results: supervised evaluation

### Table: Classifiers results measured by AUC

<table>
<thead>
<tr>
<th>Network</th>
<th>Data set</th>
<th>J48</th>
<th>NB</th>
<th>SMO</th>
<th>MLP</th>
<th>Network</th>
<th>Data set</th>
<th>J48</th>
<th>NB</th>
<th>SMO</th>
<th>MLP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Flickr</td>
<td>VLocal</td>
<td>0.774</td>
<td>0.746</td>
<td>0.583</td>
<td>0.778</td>
<td>Livejournal</td>
<td>VLocal</td>
<td>0.808</td>
<td>0.829</td>
<td>0.658</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>VGroups</td>
<td>0.761</td>
<td>0.728</td>
<td>0.504</td>
<td>0.734</td>
<td></td>
<td>VGroups</td>
<td>0.767</td>
<td>0.768</td>
<td>0.607</td>
<td>0.777</td>
</tr>
<tr>
<td></td>
<td>VLNB</td>
<td>0.748</td>
<td>0.664</td>
<td>0.501</td>
<td>0.685</td>
<td></td>
<td>VLNB</td>
<td>0.732</td>
<td>0.776</td>
<td>0.547</td>
<td>0.800</td>
</tr>
<tr>
<td></td>
<td>VGBN</td>
<td>0.737</td>
<td>0.502</td>
<td>0.501</td>
<td>0.516</td>
<td></td>
<td>VGBN</td>
<td>0.775</td>
<td>0.503</td>
<td>0.503</td>
<td>0.510</td>
</tr>
<tr>
<td></td>
<td>VLocal-Groups</td>
<td>0.789</td>
<td>0.776</td>
<td>0.585</td>
<td>0.778</td>
<td></td>
<td>VLocal-Groups</td>
<td>0.802</td>
<td>0.826</td>
<td>0.654</td>
<td>0.854</td>
</tr>
<tr>
<td></td>
<td>VLocal-GNB</td>
<td>0.796</td>
<td>0.725</td>
<td>0.583</td>
<td>0.780</td>
<td></td>
<td>VLocal-GNB</td>
<td>0.807</td>
<td>0.828</td>
<td>0.660</td>
<td>0.852</td>
</tr>
<tr>
<td></td>
<td>VLNB-Groups</td>
<td>0.792</td>
<td>0.723</td>
<td>0.504</td>
<td>0.753</td>
<td></td>
<td>VLNB-Groups</td>
<td>0.783</td>
<td>0.806</td>
<td>0.612</td>
<td>0.835</td>
</tr>
<tr>
<td></td>
<td>VLNB-GNB</td>
<td>0.769</td>
<td>0.642</td>
<td>0.502</td>
<td>0.688</td>
<td></td>
<td>VLNB-GNB</td>
<td>0.804</td>
<td>0.767</td>
<td>0.550</td>
<td>0.798</td>
</tr>
<tr>
<td></td>
<td>VGBN Groups</td>
<td>0.796</td>
<td>0.698</td>
<td>0.505</td>
<td>0.736</td>
<td></td>
<td>VGroups-GNB</td>
<td>0.768</td>
<td>0.772</td>
<td>0.609</td>
<td>0.781</td>
</tr>
<tr>
<td></td>
<td>VTotal</td>
<td>0.793</td>
<td>0.747</td>
<td>0.586</td>
<td>0.782</td>
<td></td>
<td>VTotal</td>
<td>0.799</td>
<td>0.825</td>
<td>0.664</td>
<td>0.858</td>
</tr>
<tr>
<td>Orkut</td>
<td>VLocal</td>
<td>0.883</td>
<td>0.862</td>
<td>0.629</td>
<td>0.873</td>
<td>Youtube</td>
<td>VLocal</td>
<td>0.836</td>
<td>0.801</td>
<td>0.551</td>
<td>0.808</td>
</tr>
<tr>
<td></td>
<td>VGroups</td>
<td>0.829</td>
<td>0.870</td>
<td>0.626</td>
<td>0.863</td>
<td></td>
<td>VGroups</td>
<td>0.734</td>
<td>0.671</td>
<td>0.662</td>
<td>0.726</td>
</tr>
<tr>
<td></td>
<td>VLNB</td>
<td>0.823</td>
<td>0.837</td>
<td>0.558</td>
<td>0.859</td>
<td></td>
<td>VLNB</td>
<td>0.832</td>
<td>0.687</td>
<td>0.507</td>
<td>0.739</td>
</tr>
<tr>
<td></td>
<td>VGBN</td>
<td>0.816</td>
<td>0.500</td>
<td>0.500</td>
<td>0.532</td>
<td></td>
<td>VGBN</td>
<td>0.802</td>
<td>0.506</td>
<td>0.501</td>
<td>0.499</td>
</tr>
<tr>
<td></td>
<td>VLocal-Groups</td>
<td>0.880</td>
<td>0.872</td>
<td>0.644</td>
<td>0.871</td>
<td></td>
<td>VLocal-Groups</td>
<td>0.822</td>
<td>0.819</td>
<td>0.579</td>
<td>0.825</td>
</tr>
<tr>
<td></td>
<td>VLocal-GNB</td>
<td>0.857</td>
<td>0.862</td>
<td>0.629</td>
<td>0.876</td>
<td></td>
<td>VLocal-GNB</td>
<td>0.851</td>
<td>0.800</td>
<td>0.551</td>
<td>0.812</td>
</tr>
<tr>
<td></td>
<td>VLNB-Groups</td>
<td>0.872</td>
<td>0.869</td>
<td>0.634</td>
<td>0.861</td>
<td></td>
<td>VLNB-Groups</td>
<td>0.822</td>
<td>0.720</td>
<td>0.662</td>
<td>0.755</td>
</tr>
<tr>
<td></td>
<td>VLNB-GNB</td>
<td>0.828</td>
<td>0.830</td>
<td>0.558</td>
<td>0.858</td>
<td></td>
<td>VLNB-GNB</td>
<td>0.835</td>
<td>0.683</td>
<td>0.509</td>
<td>0.738</td>
</tr>
<tr>
<td></td>
<td>VGBN Groups</td>
<td>0.830</td>
<td>0.856</td>
<td>0.626</td>
<td>0.863</td>
<td></td>
<td>VGroups-GNB</td>
<td>0.820</td>
<td>0.681</td>
<td>0.562</td>
<td>0.723</td>
</tr>
<tr>
<td></td>
<td>VTotal</td>
<td>0.861</td>
<td>0.873</td>
<td>0.644</td>
<td>0.873</td>
<td></td>
<td>VTotal</td>
<td>0.823</td>
<td>0.768</td>
<td>0.578</td>
<td>0.821</td>
</tr>
</tbody>
</table>
Results: supervised evaluation

Figure: Post-hoc test for classification results.
Outline

1. Introduction
2. Proposal
3. Experiments
4. Conclusions
Conclusions

- Based on a naïve Bayes model, four new link prediction measures were proposed considering the actual scenario of online social networks where users participate in overlapping groups.

- Individually the local naïve Bayes model and the overlapping groups naïve Bayes model measures outperform those based only on overlapping group information and local information. Moreover, when local measures are combined with measures based on overlapping groups and on overlapping groups naïve Bayes model, the link prediction accuracy improves.

- Our results suggest that using overlapping groups information improves the link prediction accuracy.


Thank you

Jorge Valverde-Rebaza
jvalverrr@icmc.usp.br