Co-author Recommender System

Ilya Makarov
Oleg Bulanov
Leonid E. Zhukov
Plan

1. Problem statement
2. Co-authorship building
3. Co-authorship network analysis
4. Recommender system building
5. Recommender system performance
6. Conclusion
Co-authorship network

Co-authorship network – undirected graph which nodes represent authors and edges represent collaborations between authors
Applicability

• University performance analysis
• Inefficient authors/departments detection
• Co-authors recommendations
• Best author of area searching
Problem statement

**Goal:** analyze co-authorship of NRU HSE staff and build a recommender system

**Problems:**
- Build the co-authorship network
- Analyze the network
- Build the recommender system
- Analyze performance of the system
Co-authorship network building

- Publications data
- Staff data
- Journals data

Nodes attributes

Nodes, Edges, Edges attributes

Co-authorship network

Edges attributes
Departments analysis

We calculate:
• Number of authors
• Number of papers
• Number of papers in Q1, Q2

For lecturers, senior lecturers, docents and professors:
• Average number of papers
• Average marks

\[
\begin{align*}
\text{min}(15 \cdot N, 30) & \quad \text{— lecturer, senior lecturer} \\
\text{min}(10 \cdot N, 30) & \quad \text{— docent} \\
\text{min}(6 \cdot N, 30) & \quad \text{— professor}
\end{align*}
\]
Areas interaction
Similarity measures

- **Number of common neighbors.** $sim(v_i, v_j) = |N(v_i) \cap N(v_j)|$, where $N(v)$ — set of neighbors

- **Jacard coefficient** $sim(v_i, v_j) = \frac{|N(v_i) \cap N(v_j)|}{|N(v_i) \cup N(v_j)|}$

- **Adar coefficient** $\sum_{v \in N(v_i) \cap N(v_j)} \frac{1}{\ln |N(v)|}$.

- **Length of shortest path**

- **Norm similarity** $sim(x, y) = \frac{1}{1+||x-y||}$, where $x, y$ characteristics of authors

- **Cosine similarity** $sim(x, y) = \frac{(x,y)}{||x|| ||y||}$

- **Number of common journal areas**
Recommender system learning scheme

- Group nodes
  - Calculate similarity measures for existing and some not existing edges in each group
  - Lasso regression
Group method

\[
group := \text{nodes of department}
\]

\[
group := \text{group } \cup \text{ nodes of similar departments}
\]

\[
group := \text{group } \cup \text{ nodes of authors with similar publications}
\]

Remove not hse authors

\[
group := \text{group } \cup \text{ nodes of clusters that contain nodes from group}
\]

\[
group := \text{hse authors from group}
\]
Making recommendation

1. Select a node(s)
2. Select a group with maximal precision
3. Calculate characteristics of edges from selected vertex to each node in the group
4. Calculate regression
5. Select authors with highest values
Testing recommender system

Power subgraph – subgraph with edges weight greater than some value.

Choose 2 power subgraphs with vertices from group → Select train and test indexes → Learn recommender system → Calculate precision, recall, accuracy, F1 measure, AUC
## Recommender system performance

<table>
<thead>
<tr>
<th></th>
<th>Precision</th>
<th>Recall</th>
<th>Accuracy</th>
<th>F1 measure</th>
<th>AUC</th>
</tr>
</thead>
<tbody>
<tr>
<td>Train</td>
<td>0.916</td>
<td>0.991</td>
<td>0.947</td>
<td>0.950</td>
<td>0.991</td>
</tr>
<tr>
<td>Test</td>
<td>0.901</td>
<td>0.868</td>
<td>0.873</td>
<td>0.870</td>
<td>0.924</td>
</tr>
</tbody>
</table>
Conclusion

• Some patterns of departments efficiency were figured out
• Recommender system was built and showed nice performance
THANK YOU FOR THE ATTENTION!