Social Bookmarking

- social bookmarking tools allow users to annotate resources (e.g. photos, music, publications) with tags
- in BibSonomy, users can store and retrieve publications and website bookmarks

**Usage Metrics**

- post a publication (post)
- search publications by tag (tag)
- view a publication post (view)
- export into various formats, e.g. for citation (exp); only BibTeX (expBib)

**Altmetrics**

- alternative metrics, based on Social Web
- impact measures other than citation counts
- e.g. tweets, mentions, likes, blog posts, etc.
- in our case: based on Social Bookmarking usage

Research Question: Do Altmetrics (Usage Metrics) Correlate with Citations?

<table>
<thead>
<tr>
<th></th>
<th>post</th>
<th>view</th>
<th>exp</th>
<th>expBib</th>
<th>req</th>
<th>tag</th>
<th>cit</th>
</tr>
</thead>
<tbody>
<tr>
<td>post</td>
<td>1</td>
<td>0.844</td>
<td>0.538</td>
<td>0.634</td>
<td>0.546</td>
<td>0.330</td>
<td>0.181</td>
</tr>
<tr>
<td>view</td>
<td>0.322</td>
<td>1</td>
<td>0.725</td>
<td>0.705</td>
<td>0.656</td>
<td>0.322</td>
<td>0.091</td>
</tr>
<tr>
<td>exp</td>
<td>0.317</td>
<td>0.429</td>
<td>1</td>
<td>0.988</td>
<td>0.742</td>
<td>0.279</td>
<td>0.157</td>
</tr>
<tr>
<td>expBib</td>
<td>0.328</td>
<td>0.417</td>
<td>0.955</td>
<td>1</td>
<td>0.722</td>
<td>0.277</td>
<td>0.160</td>
</tr>
<tr>
<td>req</td>
<td>0.323</td>
<td>0.912</td>
<td>0.663</td>
<td>0.634</td>
<td>1</td>
<td>0.213</td>
<td>0.072</td>
</tr>
<tr>
<td>tag</td>
<td>0.277</td>
<td>0.272</td>
<td>0.237</td>
<td>0.242</td>
<td>0.267</td>
<td>1</td>
<td>0.036</td>
</tr>
<tr>
<td>cit</td>
<td>0.199</td>
<td>0.098</td>
<td>0.122</td>
<td>0.120</td>
<td>0.098</td>
<td>0.014</td>
<td>1</td>
</tr>
</tbody>
</table>

Correlations

- significant (0.01 level) positive correlations between usage features and number of citations
- noticeable bias between citations and both posting (post) and exporting (exp)
- no real correlation between tag metric and citations (tag can occur in many posts)
- apart from exp and expBib and req and view, none of the usage metrics is strongly correlated to another one => truly alternative metrics

Prediction of Future Citations

- used classifiers: Random Forest, SVM with radial and polynomial kernel
- select specific publication subsets by tags
- number of citations split into two classes using median in each publication set
- baselines: + Majority (vote for the most frequent class in training set) and + Random (acc = 50%; random guessing)

Random Forest:
- outperforms random baseline on 29 out of 30 datasets
- outperforms majority baseline on 28 out of 30 datasets
- sign test and Wilcoxon signed-rank test confirm significant differences

SVM:
- less successful than Random Forest on average
- in 25 out of 30 cases better than random baseline (sign and Wilcoxon signed test corroborate significant differences)
- on average: positive improvements over the majority baseline