

Scholarly Impact from a Social Bookmarking Perspective

Daniel Zoller, Stephan Doerfel, Robert Jäschke, Gerd Stumme and Andreas Hotho

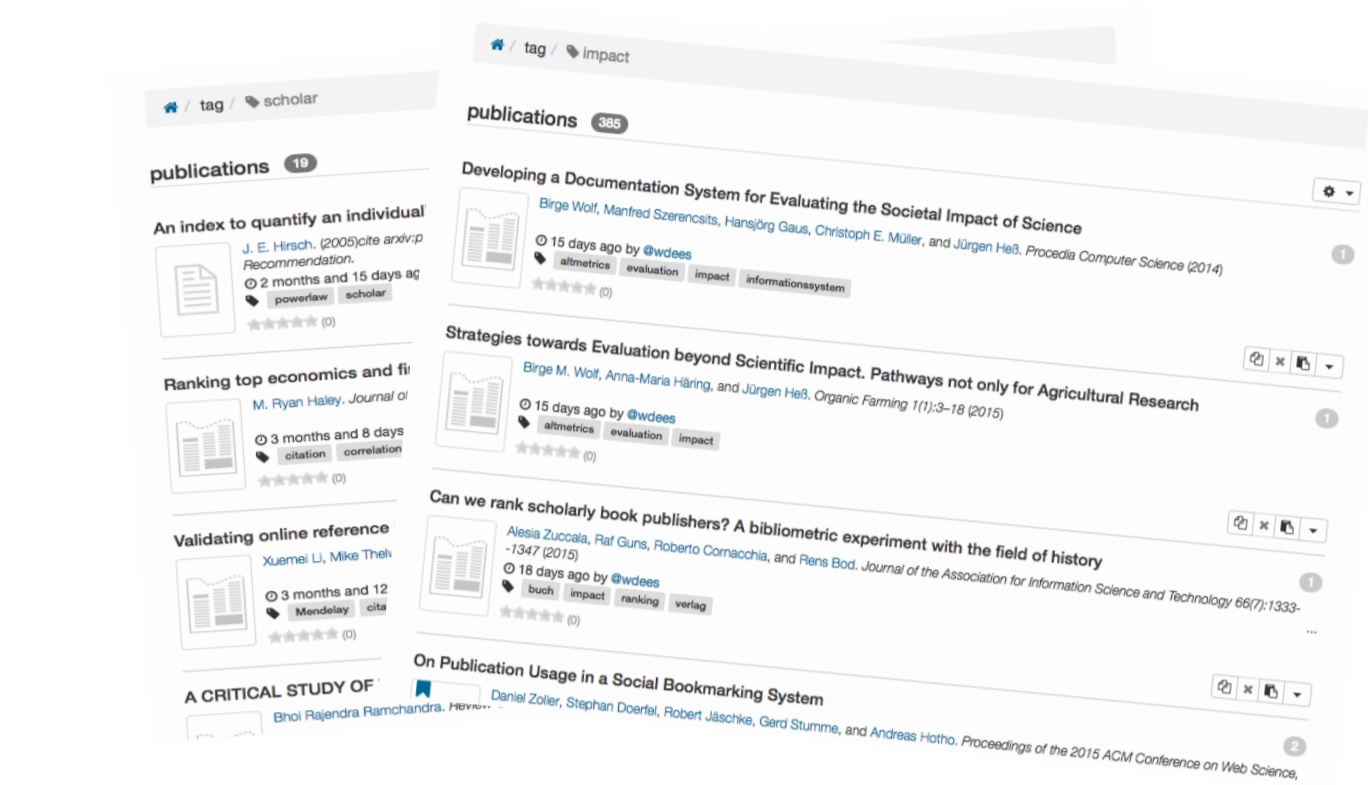
Social Bookmarking

- 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

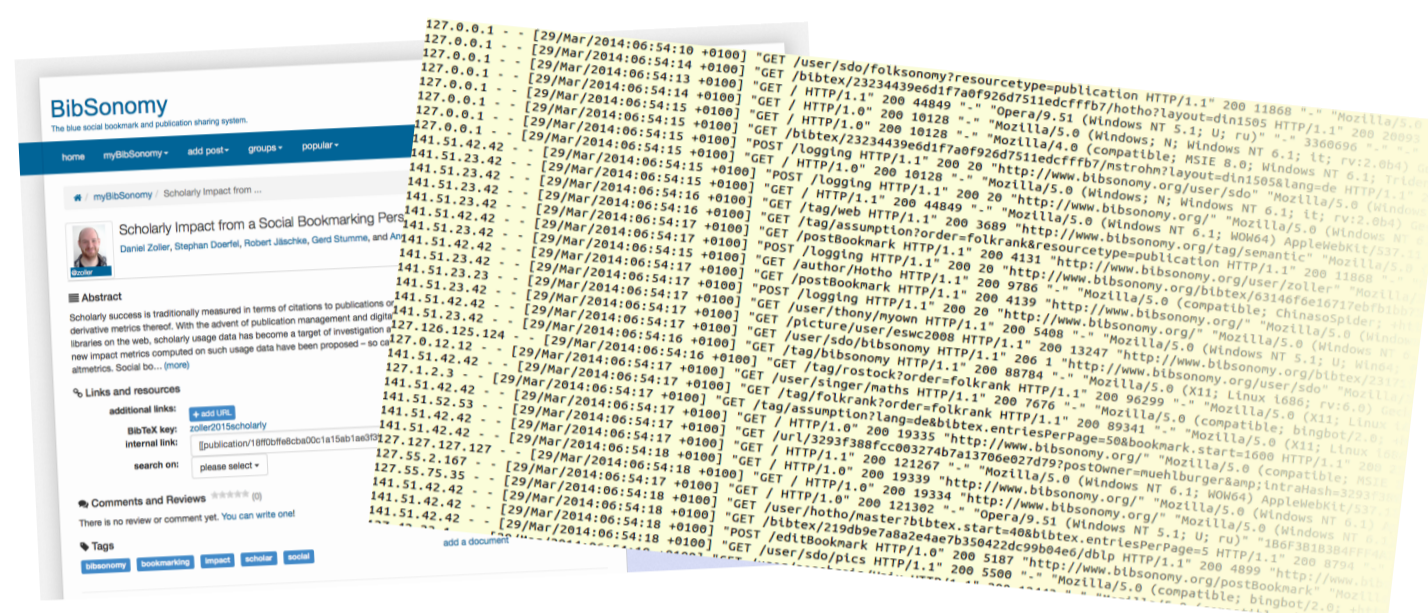
Usage Metrics

post a publication (post)

title: Scholarly Impact from a Social Bookmarking Perspective
 author(s): Zoller, Daniel and Doerfel, Stephan and Jäschke, Robert and Stumme, Gerd and Hotho, Andreas
 year: 2015



search publications by tag (tag)



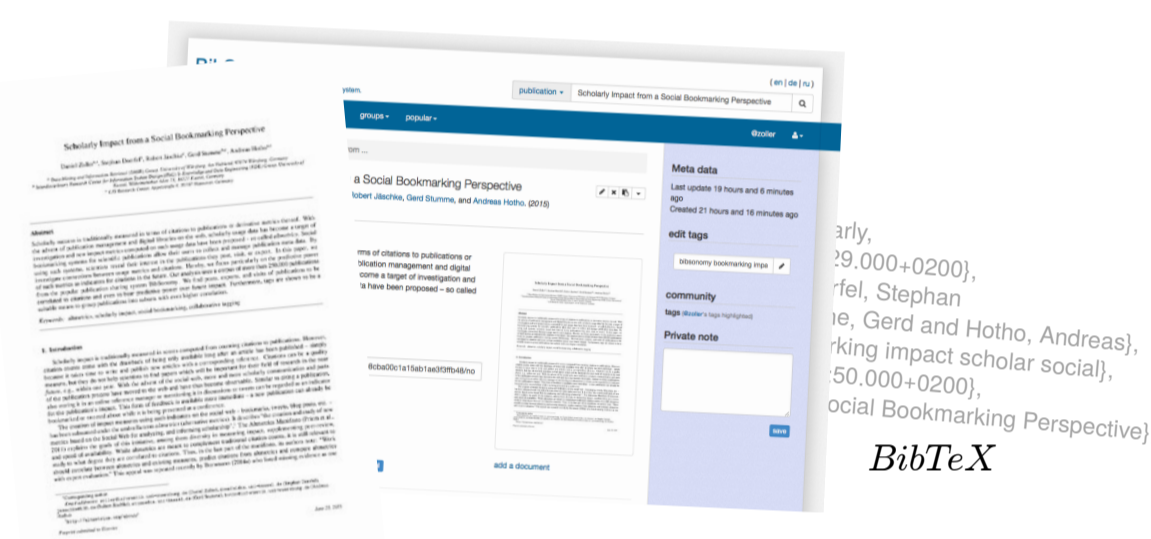
view a publication post (view)

export into various formats, e.g. for citation (exp); only BibTeX (exp_{Bib})

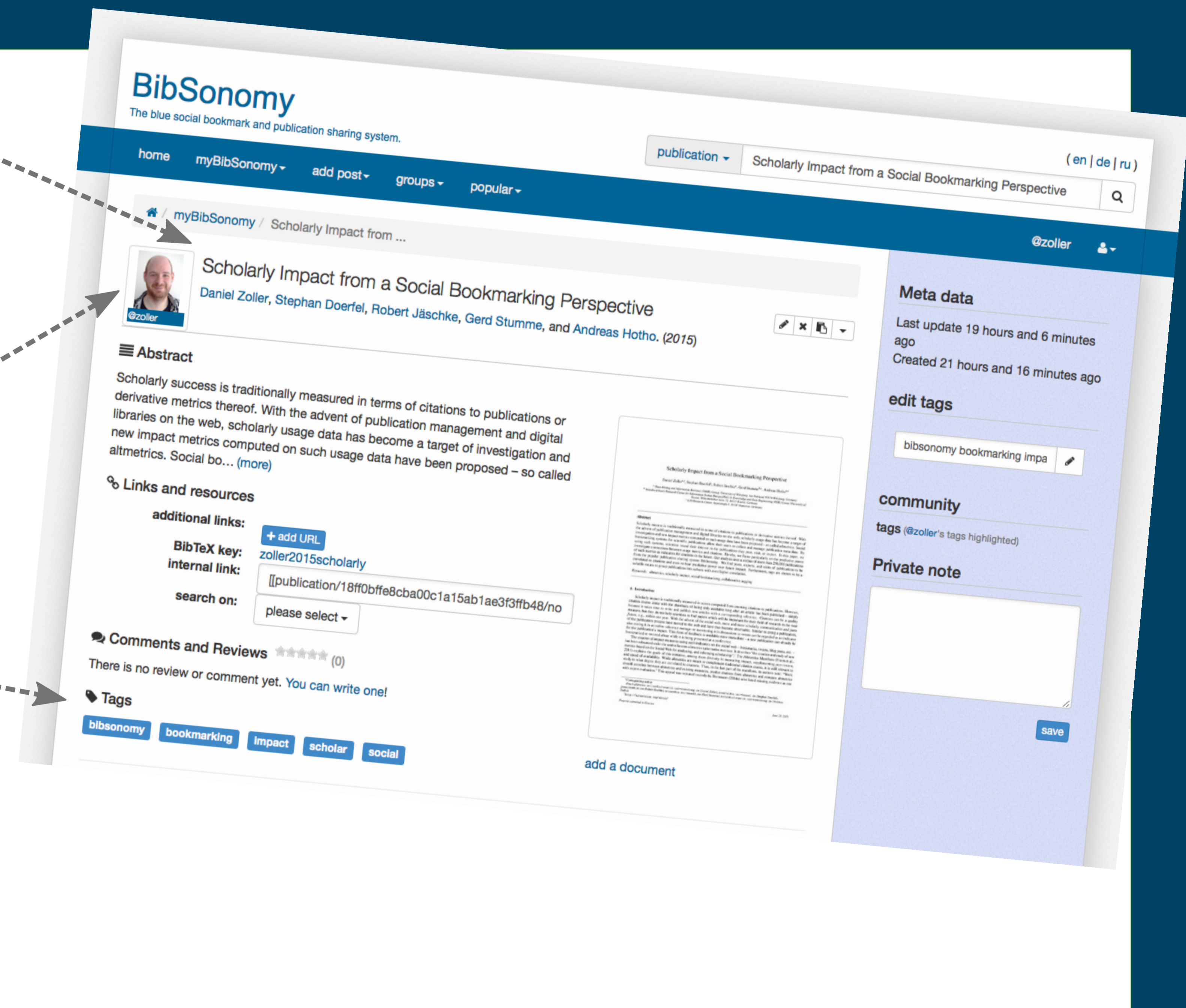
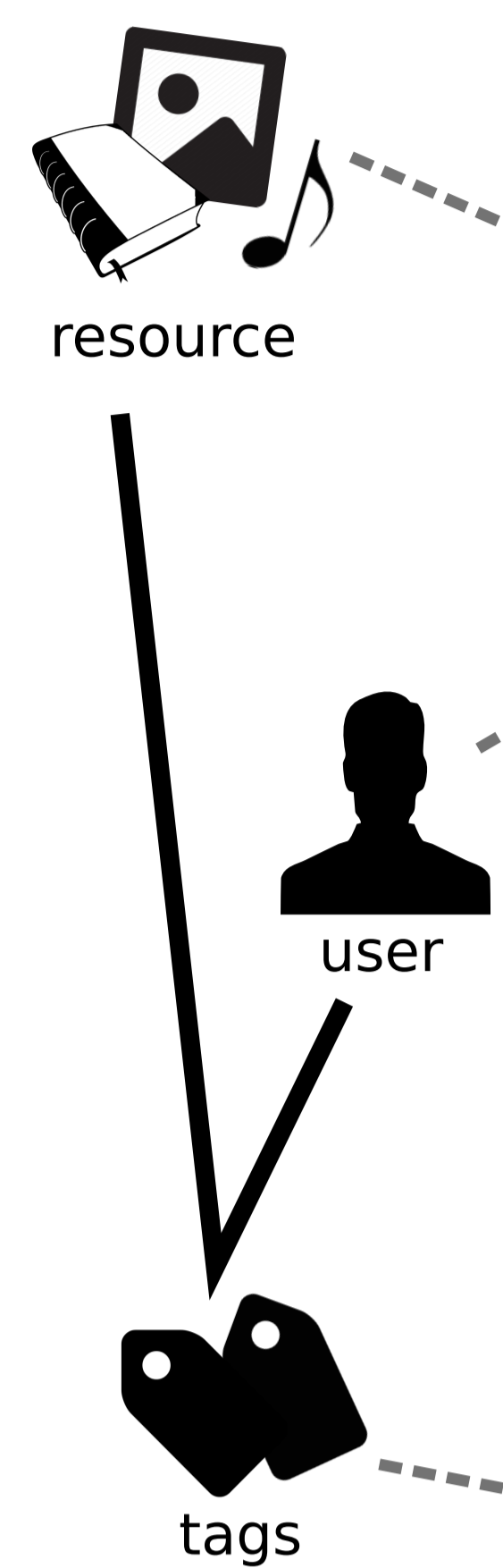
```

@inproceedings{zoller2015Scholarly,
  added-at = {2015-06-23T15:50:29.000+0200},
  author = {Zoller, Daniel and Doerfel, Stephan and Jäschke, Robert and Stumme, Gerd and Hotho, Andreas},
  keywords = {BibSonomy bookmarking impact scholar social},
  timestamp = {2015-06-23T15:00:50.000+0200},
  title = {Scholarly Impact from a Social Bookmarking Perspective},
  year = 2015
}
  
```

BibTeX

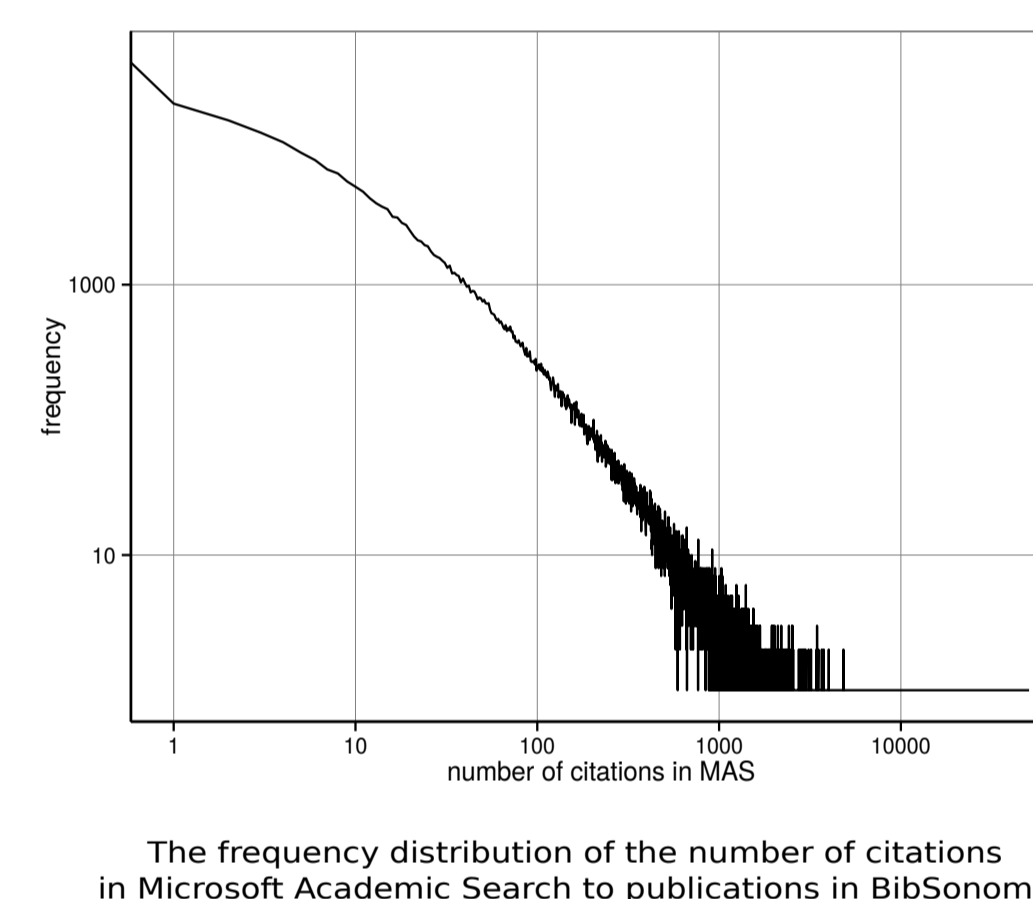


all publication related requests (req)



Citations Meet Social Data

- Citations from Microsoft Academic Search (crawled in 2014)
- Matched to publications in BibSonomy



- fat-tailed distribution - similarly found in previous work
- most of the publications in BibSonomy not cited by any other scientific work
- publications with one citation are the second largest set
- frequency decreases continuously with higher numbers of citations

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?

	post	view	exp	exp _{Bib}	req	tag	cit
post	1	0.644	0.638	0.633	0.446	0.330	0.181
view	0.322	1	0.725	0.705	0.656	0.322	0.091
exp	0.317	0.429	1	0.988	0.742	0.279	0.157
exp _{Bib}	0.328	0.417	0.955	1	0.722	0.277	0.160
req	0.325	0.912	0.663	0.634	1	0.213	0.072
tag	0.277	0.272	0.237	0.242	0.267	1	0.036
cit	0.199	0.098	0.122	0.120	0.098	0.014	1

Correlation between different usage features in BibSonomy and the number of citations of a publication. Upper right triangle Person's r and lower left Spearman's ρ.

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 exp_{Bib} and req and view, none of the usage metrics is strongly correlated to another one => truly *alternative* metrics

Prediction of Future Citations

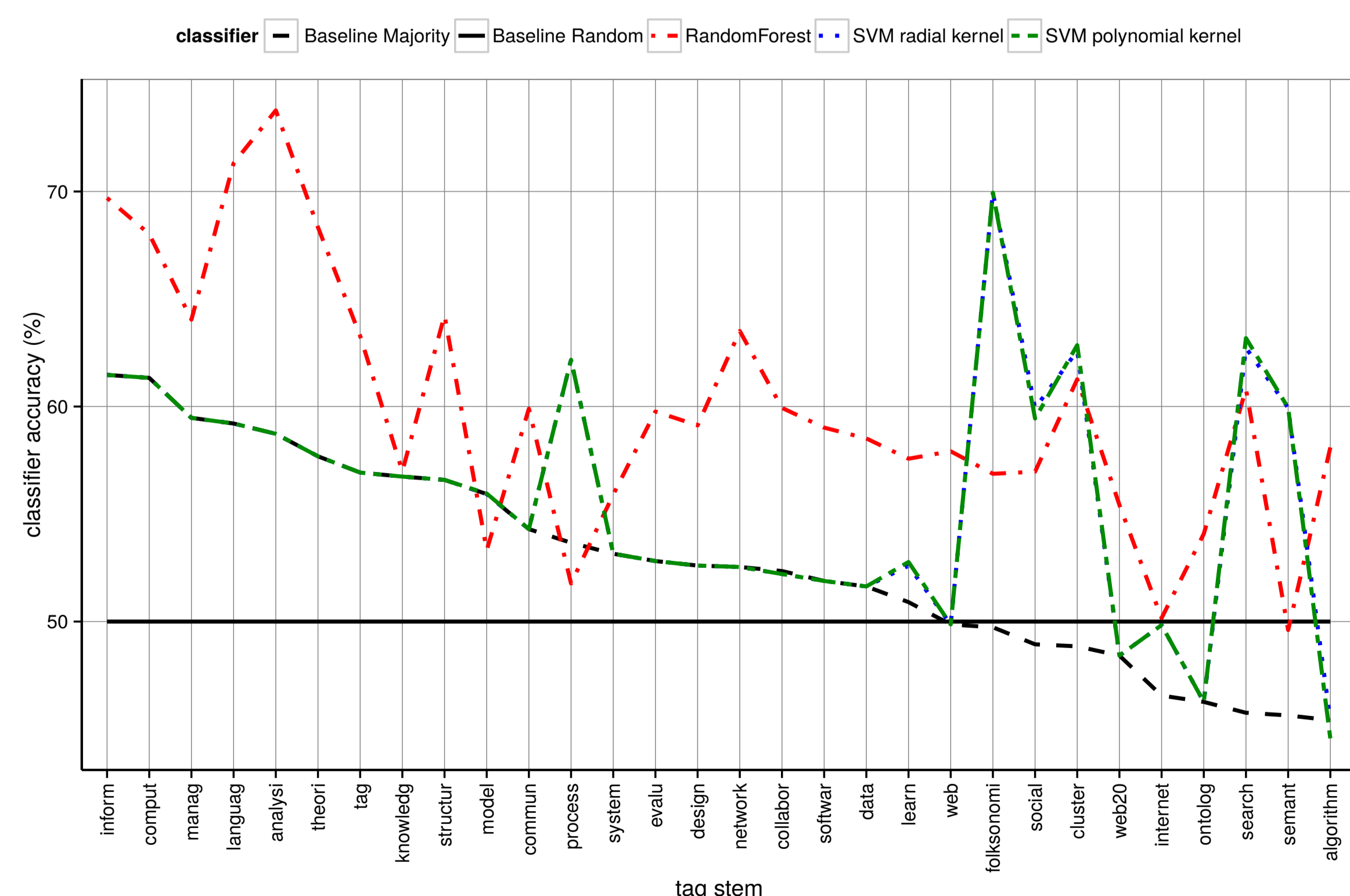
- used classifiers: Random Forest, SVM with radial and polynomial kernel
- select specific publication subsets by tags
- number of citations split into two classes using median in each publication set
- baselines: + Majority (vote for the most frequent class in training set) and + Random (acc = 50%;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



Classification accuracy for the 30 tag-induced publication sets for the three classifiers and two baselines. Note: the diagrams for the two SVMs are almost indiscernible as they often provided the same result.