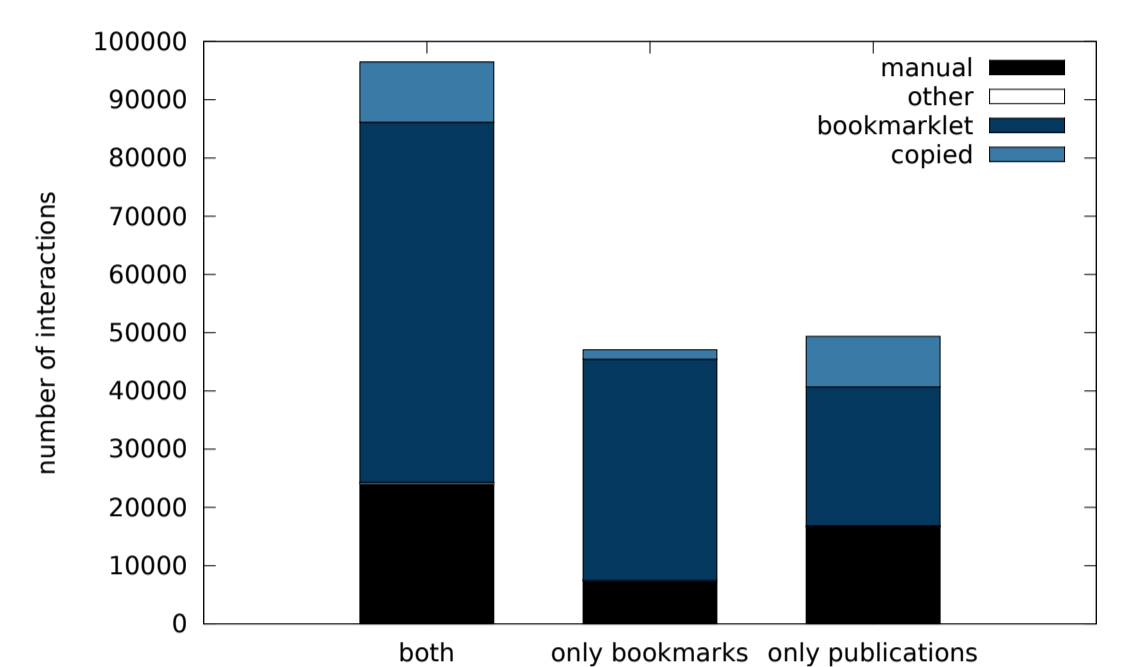


Assumptions about Social Tagging

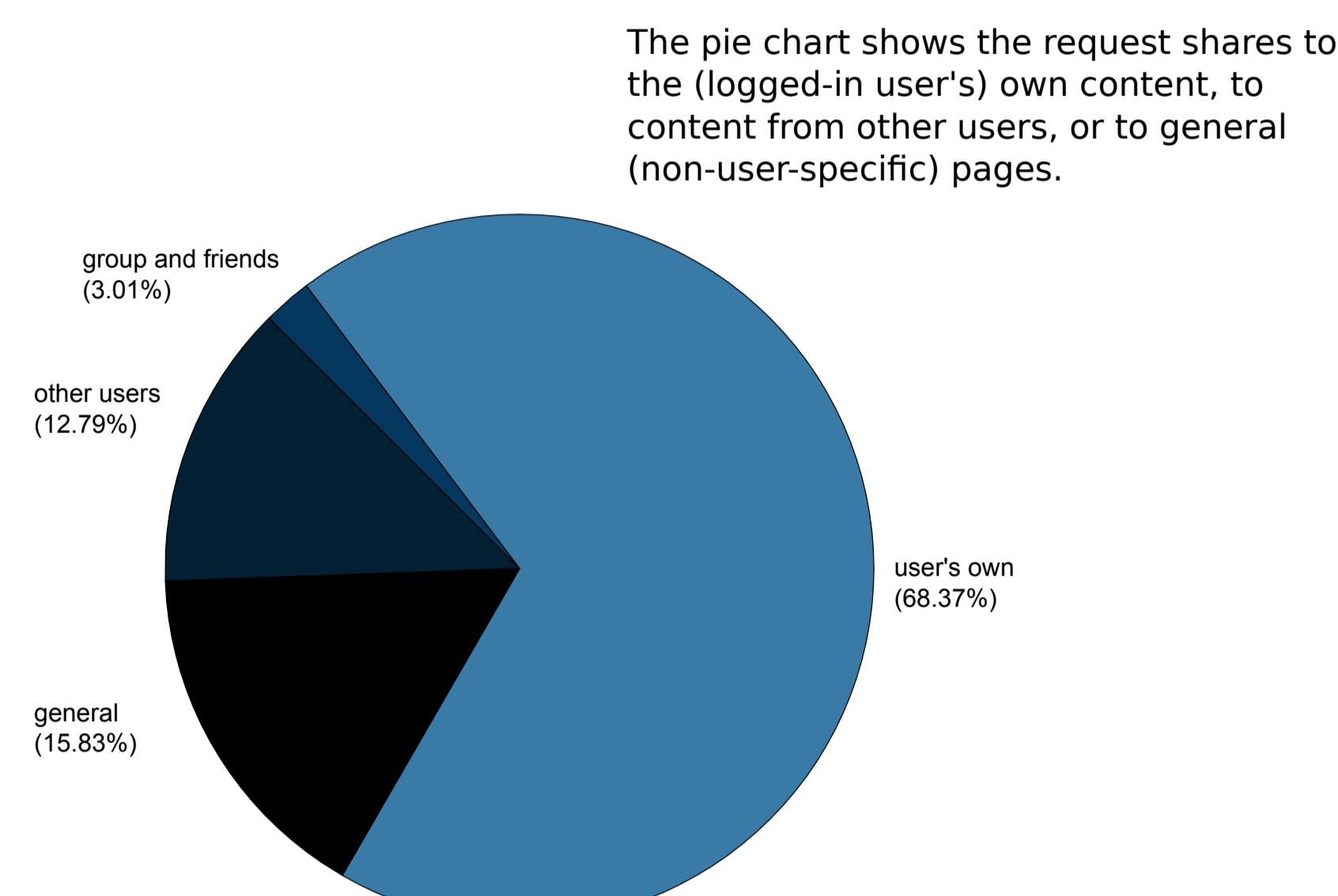
Stephan Doerfel, Daniel Zoller, Philipp Singer, Thomas Niebler, Andreas Hotho and Markus Strohmaier

Social

How **social** is social tagging? Do users **(re)use resources** that have been shared and tagged **by others**?



The histogram shows how resources are posted: through manual input, through a bookmarklet, through copying posts, or through other methods (extremely small share).



The pie chart shows the request shares to the (logged-in user's) own content, to content from other users, or to general (non-user-specific) pages.

Results:

Visit

- in a large number of requests (32%) users target content of others
- the majority of requests (68%) target a user's own pages

Copied Posts

- 11% of all posts have been copied from other users
- if a post's resource had already been shared publicly, in 42% of the cases, that post was created as a copy

Copied Tags

- tags are adopted frequently (at least one tag in 42% of all copies)
- 44% of those copied posts, where no tag of the original post was used, had only special tags (e.g. "imported")
- user's own vocabulary was used in 87% of all copy requests

Summary

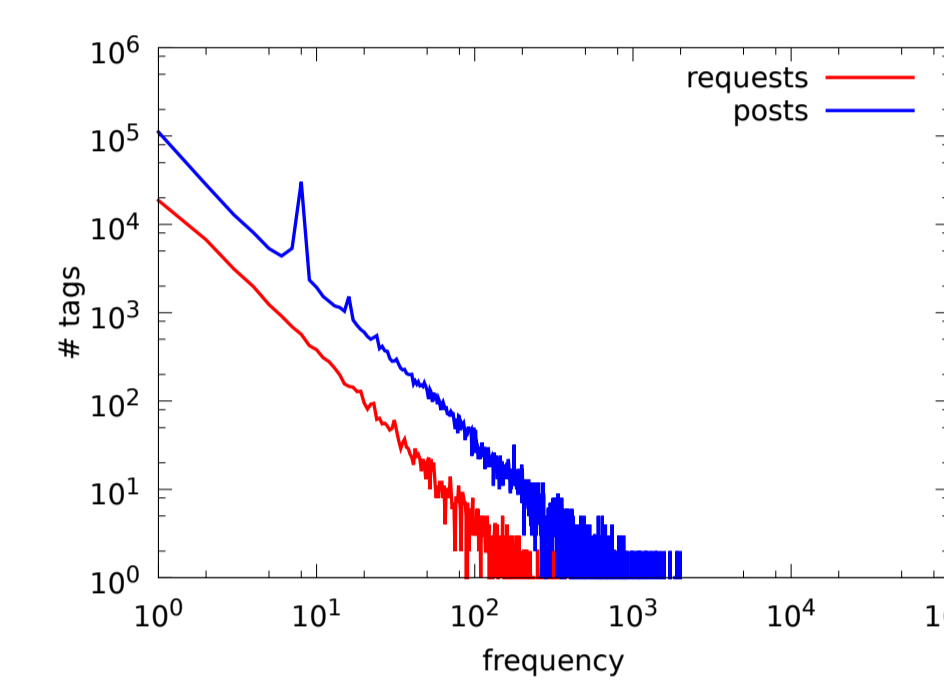
- visit and copy requests reveal a strong interest in other users' content
- a larger part of system interaction happens with the own collection

Popularity

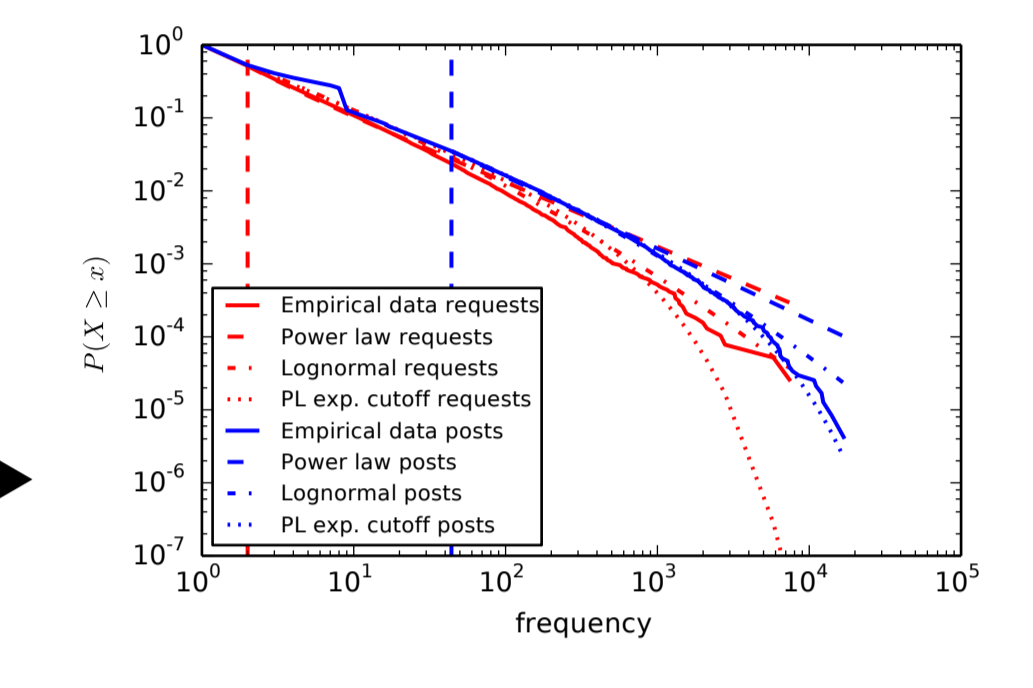
Does the popularity of a user/resource/tag in **posts** match its popularity in **requests**?

Analysis Assumption Behavior
BibSonomy blog bookmark
bookmarking citation correlation equality
FolkRank Folksonomy fun java keyword
LogFiles myown Popularity
programming retrieval Sharing Social
Tagging User Web WWW

Popularity in posts is used
- in recommendations,
- in rankings, and
- in tag clouds.

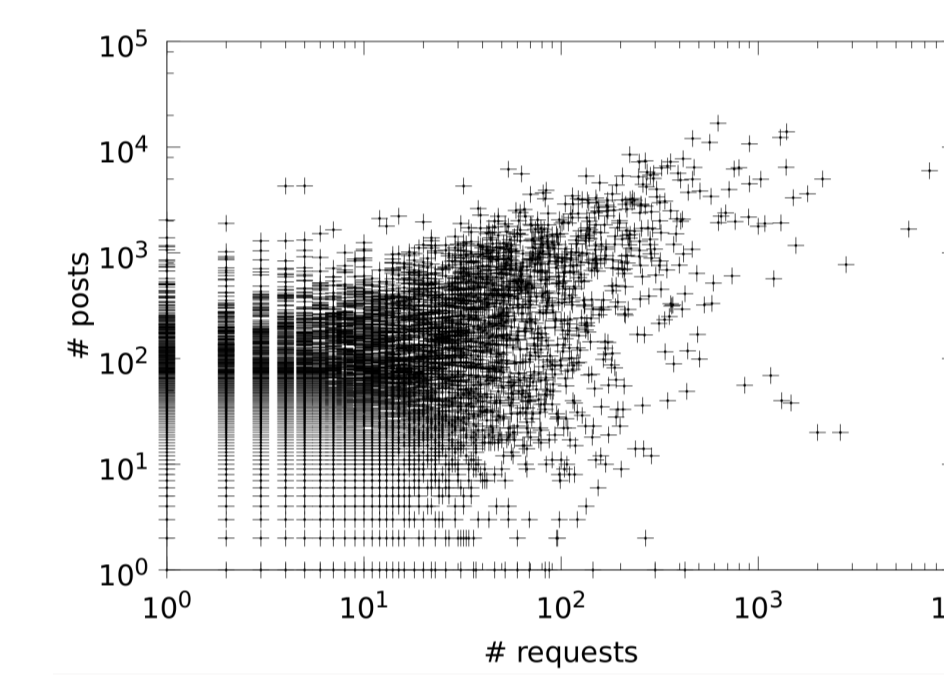


Distribution of tag frequencies in posts and resources.



Fits of candidate distributions to the cumulative tag frequency distributions.

Tags with occurrences in requests and occurrences in posts.



Results:

- similar results for resources, users, and tags
- frequency distributions are heavy-tailed and similar in posts and requests
- on the level of individual tags we observe only weak or moderate correlation between occurrences in posts and in requests
- especially for tags it is surprising that there is no stronger correlation, given that tags are presented to the user in clouds reflecting popularity

Analyzing the Bookmarking System BibSonomy

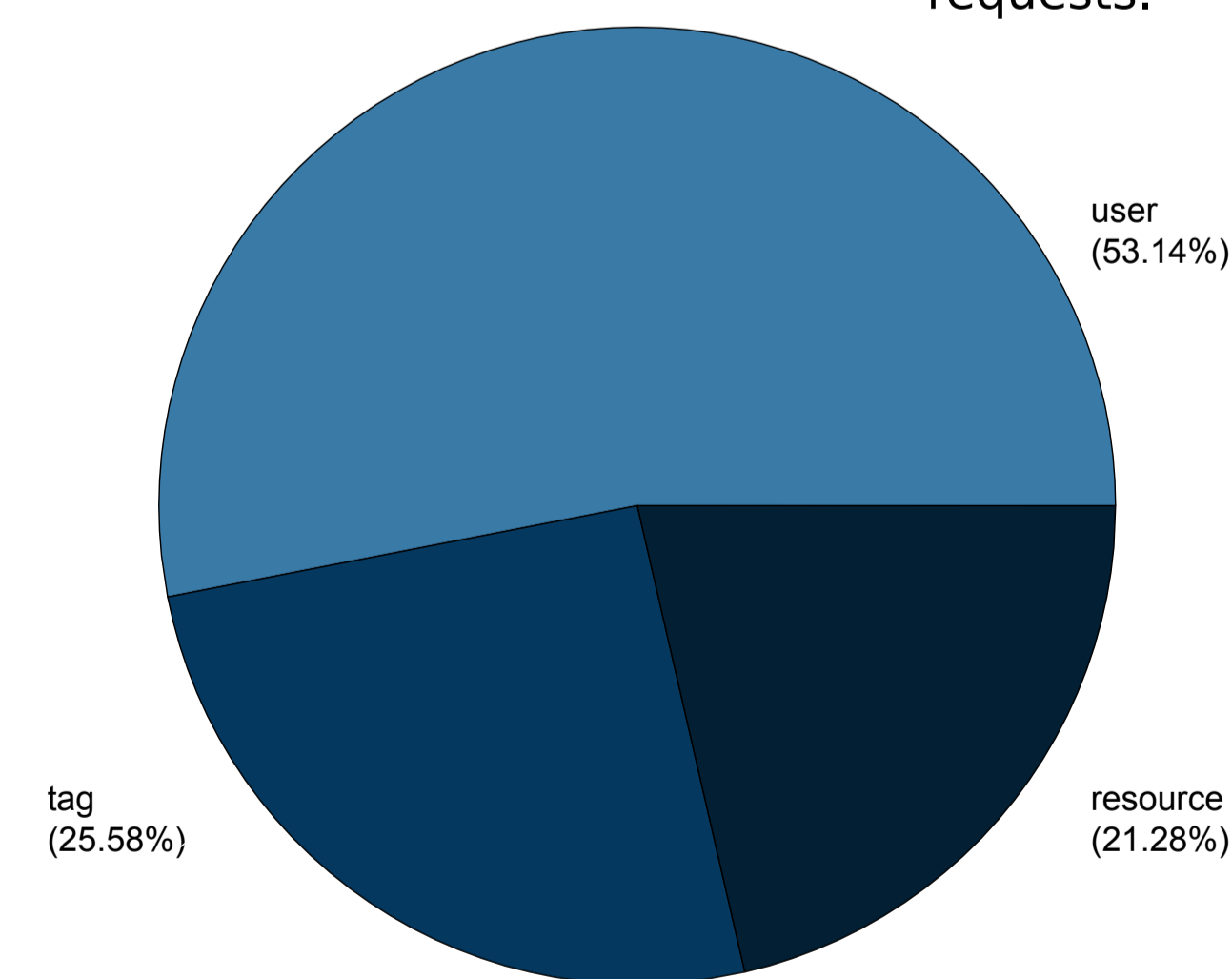
```
127.0.0.1 - [29/Mar/2014:06:54:10 +0100] "GET /user/sdo/folksonomy/resource-type-publication HTTP/1.1" 200 11800 "Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:14 +0100] "GET /bibtex/23234439eedf7a8926d7516dcff7b7/hochschul-layout-dl1565 HTTP/1.1" 200 20000 "Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:14 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (compatible; MSIE 8.0; Windows NT 6.1; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:15 +0100] "GET /bibtex/23234439eedf7a8926d7516dcff7b7/hochschul-layout-dl1565/language HTTP/1.1" 200 20000 "Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:15 +0100] "POST /logging HTTP/1.1" 200 20 "http://www.bibsonomy.org/user/sdo/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:15 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:15 +0100] "POST /logging HTTP/1.1" 200 20 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:16 +0100] "GET /tag/web HTTP/1.1" 200 3689 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:16 +0100] "GET /tag/assumption/for/der/folksonomy/resource-type-publication HTTP/1.1" 200 13247 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "POST /logging HTTP/1.1" 200 20 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /user/hotho HTTP/1.1" 200 9786 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /post/bookmark HTTP/1.1" 200 4139 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /user/thony/myown HTTP/1.1" 200 5488 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /jctvuser/essence2008 HTTP/1.1" 200 13247 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /user/sdo/bibsonomy HTTP/1.1" 200 20 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.0.1 - [29/Mar/2014:06:54:16 +0100] "GET /tag/folksonomy HTTP/1.1" 200 9629 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.12.12 - [29/Mar/2014:06:54:17 +0100] "GET /user/singer/math HTTP/1.1" 200 7676 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:17 +0100] "GET /tag/folksonomy HTTP/1.1" 200 9629 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.12.12 - [29/Mar/2014:06:54:17 +0100] "GET /tag/assumption/lang-debibtex.entriesPerPage=588bookmar HTTP/1.1" 200 13247 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:17 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:17 +0100] "GET /url/329338f8cc082274b1a17060e2d793postCher-muelli HTTP/1.1" 200 13247 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.23.42 - [29/Mar/2014:06:54:18 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.12.12 - [29/Mar/2014:06:54:18 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.12.12 - [29/Mar/2014:06:54:18 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
127.0.12.12 - [29/Mar/2014:06:54:18 +0100] "GET / HTTP/1.0" 200 10128 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:18 +0100] "GET /user/hotho/master70bibtex.start=466bibtex.entriesPerPage=588bookmar HTTP/1.1" 200 13247 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:18 +0100] "POST /edit/bookmark HTTP/1.0" 200 5187 "http://www.bibsonomy.org/" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
141.51.42.42 - [29/Mar/2014:06:54:18 +0100] "GET /user/sdo/pics HTTP/1.1" 200 5500 "-" Mozilla/5.0 (Windows NT 6.0; Win64; x64; rv:10.0.1) Gecko/20100101 Firefox/35.0"
```

Using Posts

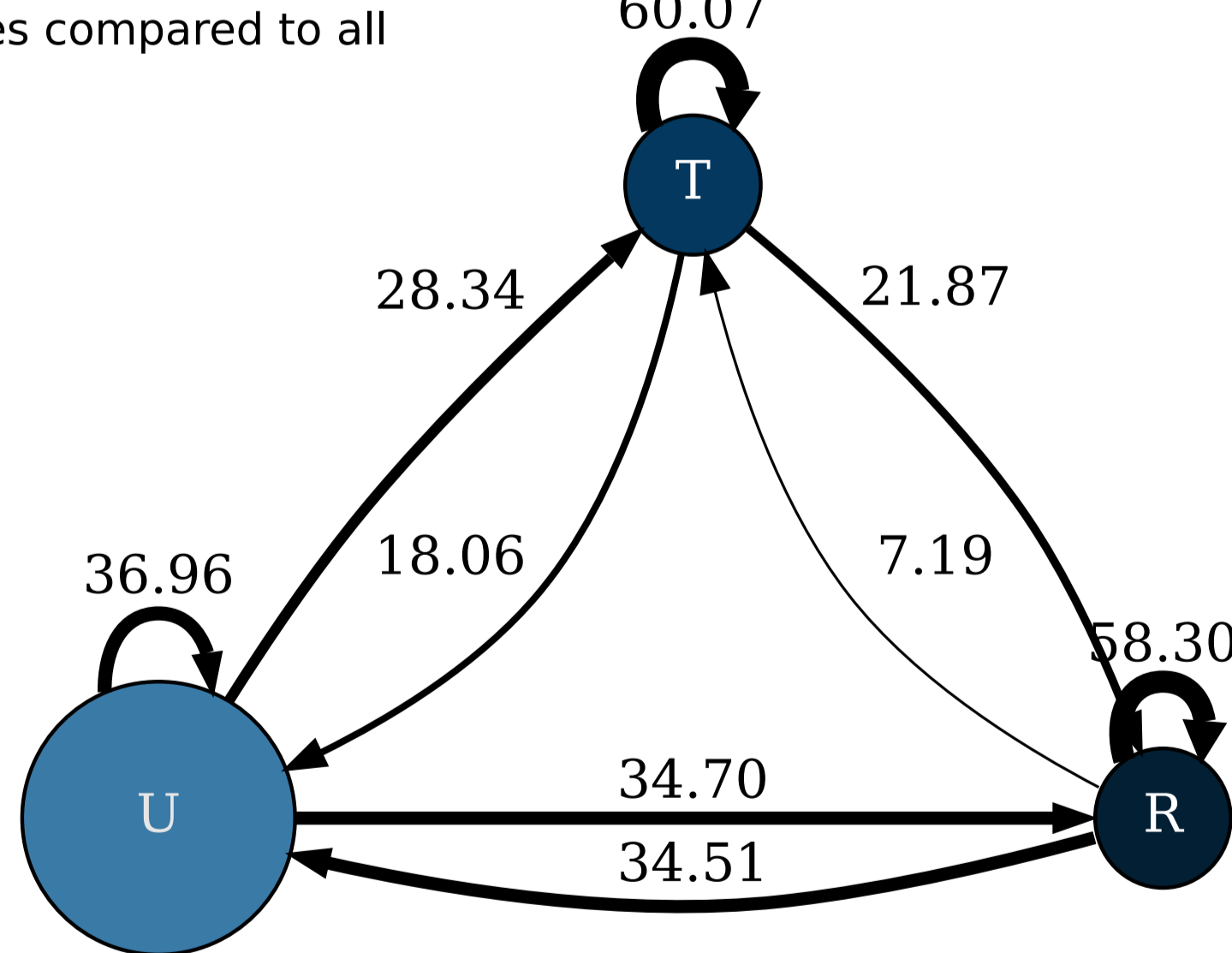


Equality

Are users, resources, and tags **equally important**?



The pie diagram shows the shares of requests to user pages, to resource pages and to tag pages compared to all requests.



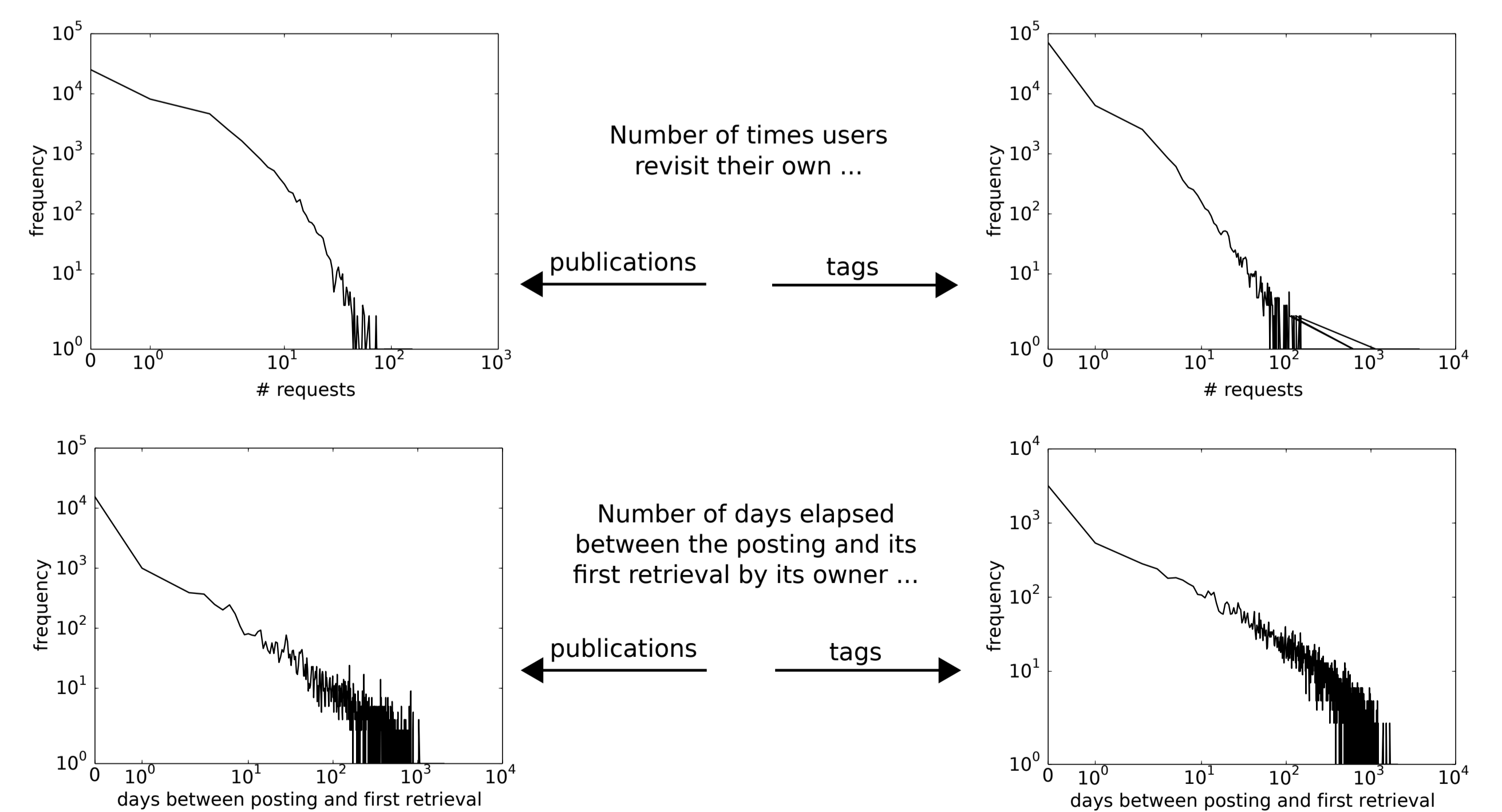
The transition graph shows that self-transitions are dominant, whereas intercategory transitions occur very rarely and are also distributed unequally.

Results:

- the numbers of requests to tags and resources are similar
- requests to users dominate requests to resources and tags
- there are far less users than resources or tags, thus this imbalance in requests is even more surprising

Retrieval

Do users **retrieve** their posts and tags **later on**?



Results:

- about 49% of all publications are revisited by their owner at least once
- only 17% of used tags are queried at least once
- publications and tags are mostly revisited only once
- revisits often take place shortly (within a day) after a resource/tag has been posted



Try it yourself:
www.bibsonomy.org