



Evolving Large Networks

— the case of Bibsonomy —

Ulrik Brandes Jürgen Lerner Uwe Nagel

Department of Computer & Information Science
University of Konstanz

Sunbelt XXIX
San Diego, 10–15 March, 2009





focus question

$n=1.4\text{m}$ tag assignments

(3780 users tagging 500k resources using 100k tags)

how interesting is a tag assignment,
given the history of tagging in Bibsonomy?





focus question

$n=1.4\text{m}$ tag assignments

(3780 users tagging 500k resources using 100k tags)

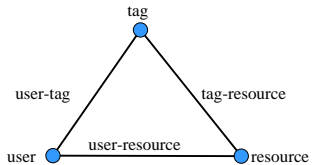
how interesting is a tag assignment,
given the history of tagging in Bibsonomy?

- ▶ \neq spam detection
- ▶ distinguish likely and unlikely tag assignments
- ▶ identify interesting phases
- ▶ identify interesting (groups of) users, tags, and resources
- ▶ classify users, tags, and resources
- ▶ tag recommendation



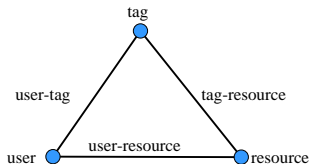


defining interestingness





defining interestingness

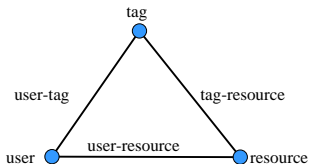


	i	ordinal timepoint $1 \leq i \leq n$
$\text{tas}_i = (u_i, t_i, r_i)$		tag assignment at time i
TAS_i		tag assignments before time i
$s_1(\text{tas}_i \text{TAS}_i)$		#prev. occ. of u_i
$s_2(\text{tas}_i \text{TAS}_i)$		#prev. occ. of t_i
$s_3(\text{tas}_i \text{TAS}_i)$		#prev. occ. of r_i
$s_4(\text{tas}_i \text{TAS}_i)$		#prev. joint occ. of u_i and t_i
$s_5(\text{tas}_i \text{TAS}_i)$		#prev. joint occ. of u_i and r_i
$s_6(\text{tas}_i \text{TAS}_i)$		#prev. joint occ. of t_i and r_i





defining interestingness



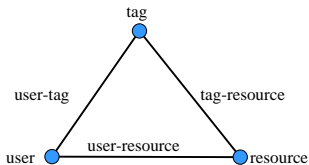
i	ordinal timepoint $1 \leq i \leq n$
$\text{tas}_i = (u_i, t_i, r_i)$	tag assignment at time i
TAS_i	tag assignments before time i
$s_1(\text{tas}_i \text{TAS}_i)$	#prev. occ. of u_i
$s_2(\text{tas}_i \text{TAS}_i)$	#prev. occ. of t_i
$s_3(\text{tas}_i \text{TAS}_i)$	#prev. occ. of r_i
$s_4(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of u_i and t_i
$s_5(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of u_i and r_i
$s_6(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of t_i and r_i

$$P(\text{tas}_i) = \frac{1}{Z(\text{TAS}_i)} \exp \left\{ - \sum_{k=1}^6 \theta_k \cdot s_k(\text{tas}_i | \text{TAS}_i) \right\}$$





defining interestingness



i	ordinal timepoint $1 \leq i \leq n$
$\text{tas}_i = (u_i, t_i, r_i)$	tag assignment at time i
TAS_i	tag assignments before time i
$s_1(\text{tas}_i \text{TAS}_i)$	#prev. occ. of u_i
$s_2(\text{tas}_i \text{TAS}_i)$	#prev. occ. of t_i
$s_3(\text{tas}_i \text{TAS}_i)$	#prev. occ. of r_i
$s_4(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of u_i and t_i
$s_5(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of u_i and r_i
$s_6(\text{tas}_i \text{TAS}_i)$	#prev. joint occ. of t_i and r_i

$$P(\text{tas}_i) = \frac{1}{Z(\text{TAS}_i)} \exp \left\{ - \sum_{k=1}^6 \theta_k \cdot s_k(\text{tas}_i | \text{TAS}_i) \right\}$$

computational problem





avoiding computation of probabilities

S : $n \times 6$ -matrix of statistics

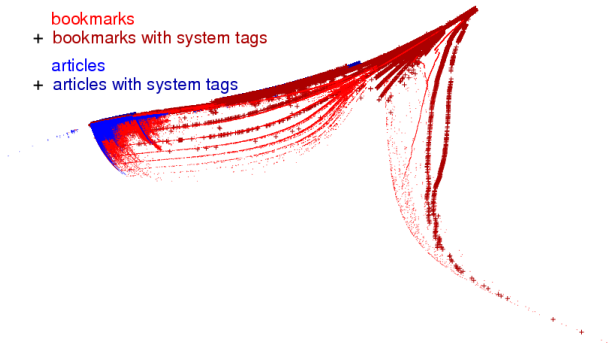




avoiding computation of probabilities

S : $n \times 6$ -matrix of statistics

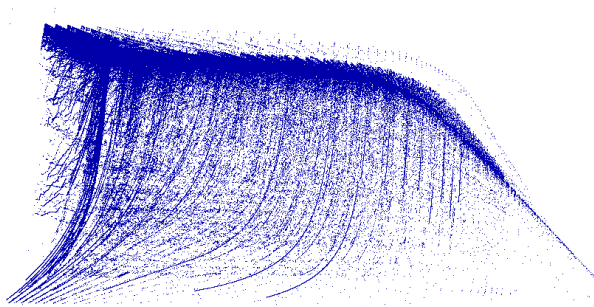
- bookmarks
- + bookmarks with system tags
- articles
- + articles with system tags





avoiding computation of probabilities

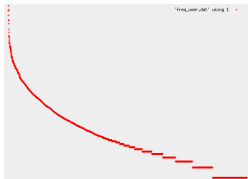
S : $n \times 6$ -matrix of statistics (without system tags)



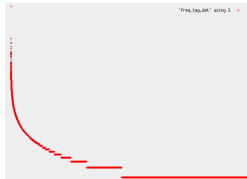


a closer look

users



tags



118606	2518
91196	3017
85445	2703
82564	0
61337	347
44546	1
43989	3215
34536	2778
24246	54
19359	61

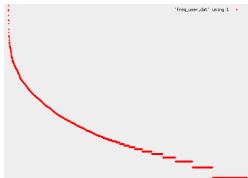
153228	imported
15786	public
12079	system:imported
11839	ZZZ_TO_SORT
9183	indexforum
9153	nn
8267	video
6601	programming
6283	books
6182	software



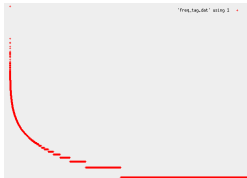


a closer look

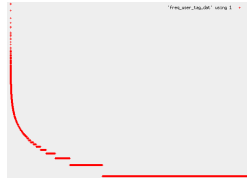
users



tags



user-tag



118606	2518
91196	3017
85445	2703
82564	0
61337	347
44546	1
43989	3215
34536	2778
24246	54
19359	61

153228	imported
15786	public
12079	system:imported
11839	ZZZ_TO_SORT
9183	indexforum
9153	nn
8267	video
6601	programming
6283	books
6182	software

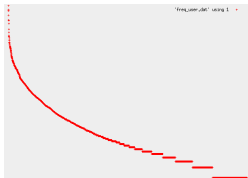
38099	2703	imported
36565	2518	imported
24934	0	imported
15574	0	public
11839	3017	ZZZ_TO_SORT
9183	0	indexforum
8896	1	system:imported
8896	1	imported
6610	0	video



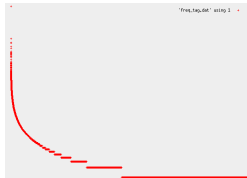


a closer look

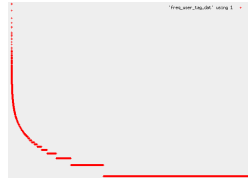
users



tags



user-tag



118606	2518
91196	3017
85445	2703
82564	0
61337	347
44546	1
43989	3215
34536	2778
24246	54
19359	61

153228	imported
15786	public
12079	system:imported
11839	ZZZ_TO_SORT
9183	indexforum
9153	nn
8267	video
6601	programming
6283	books
6182	software

38099	2703	imported
36565	2518	imported
24934	0	imported
15574	0	public
11839	3017	ZZZ_TO_SORT
9183	0	indexforum
8896	1	system:imported
8896	1	imported
6610	0	video
		⋮
1526	2234	juergen
1517	2155	juergen





video

