

# Influences on Tag Choices in del.icio.us

Emilee Rader and Rick Wash  
School of Information, University of Michigan  
ejrader, rwash @ umich.edu

# introduction

# **Folksonomy**

The practice and method of  
**collaboratively** creating and managing  
tags to annotate and **categorize** content.

Delicious

http://delicious.com/

Google

reader

gmail

calendar

docs

scholar

post

CTools

web

experiment app

UM

si calendars

research

stats

dog

weather

movies

financial

delicious

social bookmarking

It's Free!

Join Now

Sign In

The best sites bubble up.  
See the most popular bookmarks for any tag.

Learn More

HIDE INTRO

Search the biggest collection of bookmarks in the universe...

Search Delicious

Search


Popular Bookmarks

Explore Tags

The most popular bookmarks on Delicious right now

See more Popular bookmarks

New bookmarks saved in the last minute 207



The Best Cheat Sheets for Web Developers | Web Resources | WebAppers

SAVE

117


cheatsheet

reference

webdesign

tips

javascript



Script kiddies have awesome tools | dow.ngra.de

SAVE

101


security

wordpress

programming

hacking

web



Barack Obama should be the next president of America | It's time | The Economist

SAVE

85


politics

obama

election

economist

endorsement



Official Google Blog: What we learned from 1 million businesses in the cloud

SAVE

91


google

cloud

gmail

business

cloudcomputing



The Perfect Server - Ubuntu Intrepid Ibex (Ubuntu 8.10) | HowtoForge - Linux Howtos and Tutorials

SAVE

104

ubuntu

server

linux

howto

tutorial

Popular Tags

design

blog

video

software

tools

music

programming

webdesign

reference

tutorial

art

web

howto

javascript

free

linux

web2.0

development

Done

11/05/2008 8:12 AM

Why do people choose some tags over others?

Individual choices → Internet-scale patterns

Users' future tag choices are  
**heavily influenced by tags they  
have chosen** in the past.

main result

# background



**Exact word choices matter** for organizing, re-finding and navigating in tagging systems, but there are **competing hypotheses** in the literature for how people choose tags.

Explore tags on Delicious

reader

gmail

calendar

docs

scholar

post

CTools

web

experiment app

UM

si calendars

research

stats

dog

weather

movies

financial

http://delicious.com/tag

Google

Explore Everyone's Tags

See what's popular, or delve into your own interests.

Tag

type a tag

Click a tag below or type in a word above, and we'll show you the latest bookmarks saved with that tag.

Tag Cloud: Popular

Sort: Alphabetically | [By size](#)

.net

2008

3d

advertising

ajax

and

animation

api

apple

architecture

art

article

articles

artist

audio

blog

blogging

blogs

book

books

browser

business

car

cms

code

collaboration

comics

community

computer

converter

cooking

cool

css

culture

data

database

design

desktop

development

diy

documentation

download

downloads

drupal

ebooks

economics

education

electronics

email

entertainment

environment

fashion

fic

film

finance

firefox

flash

flex

flickr

food

forum

free

freeware

fun

funny

gallery

game

games

geek

google

government

graphics

green

guide

hardware

health

history

home

hosting

house

howto

html

humor

icons

illustration

images

imported

information

inspiration

interactive

interesting

internet

iphone

japan

java

javascript

jobs

jquery

kids

language

learning

library

linux

list

lists

literature

mac

magazine

management

maps

marketing

math

media

microsoft

mobile

money

movie

movies

mp3

music

network

networking

news

online

opensource

osx

people

phone

photo

photography

photos

photoshop

php

plugin

podcast

politics

portfolio

privacy

productivity

programming

psychology

python

radio

rails

realestate

recipe

recipes

reference

religion

research

resources

reviews

rss

ruby

rubyonrails

school

science

search

security

seo

shop

shopping

social

socialnetworking

software

statistics

streaming

teaching

tech

technology

tips

todo

tool

tools

toread

travel

tutorial

tutorials

tv

twitter

typography

ubuntu

usability

video

videos

vim

visualization

web

web2.0

webdesign

webdev

wiki

wikipedia

windows

wishlist

wordpress

work

writing

youtube

What is a tag?

A tag is simply a word you can use to describe a bookmark. Unlike folders, you make up tags when you need them and you can use as many as you like. The result is a better way to organize your bookmarks and a great way to discover interesting things on the Web.

What is a tag cloud?

A list of tags where size reflects popularity.

Done

11/05/2008 8:27 AM

Save a Bookmark on Delicious

http://delicious.com/save?url=http%3A%2F%2Fwww.loc.gov%2Findex.html

**delicious** Save Bookmark

URL  required

TITLE  required

NOTES

1000 characters left

TAGS  reference 42 space separated, 128 characters per tag

☐ Do Not Share

**Tags** People

▼ Recommended

research reference tools library usa book database information  
libraries art education history images search

▼ Popular

government books congress resources loc archive literature

Done

Save a Bookmark on Delicious

http://delicious.com/save?url=http%3A%2F%2Fwww.loc.gov%2Findex.html

**delicious** Save Bookmark

URL  required

TITLE  required

NOTES

1000 characters left

TAGS   42 space separated, 128 characters per tag

☐ Do Not Share

**Tags** **People**

▼ Recommended

research reference tools library usa book database information  
libraries art education history images search

▼ Popular

government books congress resources loc archive literature

Done

Save a Bookmark on Delicious

http://delicious.com/save?url=http%3A%2F%2Fwww.loc.gov%2Findex.html

**delicious** Save Bookmark

URL  required

TITLE  required

NOTES

1000 characters left

TAGS  reference 42 space separated, 128 characters per tag

☐ Do Not Share

**Tags** People

▼ Recommended

research reference tools library usa book database information  
libraries art education history images search

▼ Popular

government books congress resources loc archive literature

Done

Save a Bookmark on Delicious

http://delicious.com/save?url=http%3A%2F%2Fwww.loc.gov%2Findex.html

**delicious** Save Bookmark

URL  required

TITLE  required

NOTES

1000 characters left

TAGS   42 space separated, 128 characters per tag

☐ Do Not Share

**Tags** **People**

▼ Recommended

research reference tools library usa book database information  
libraries art education history images search

▼ Popular

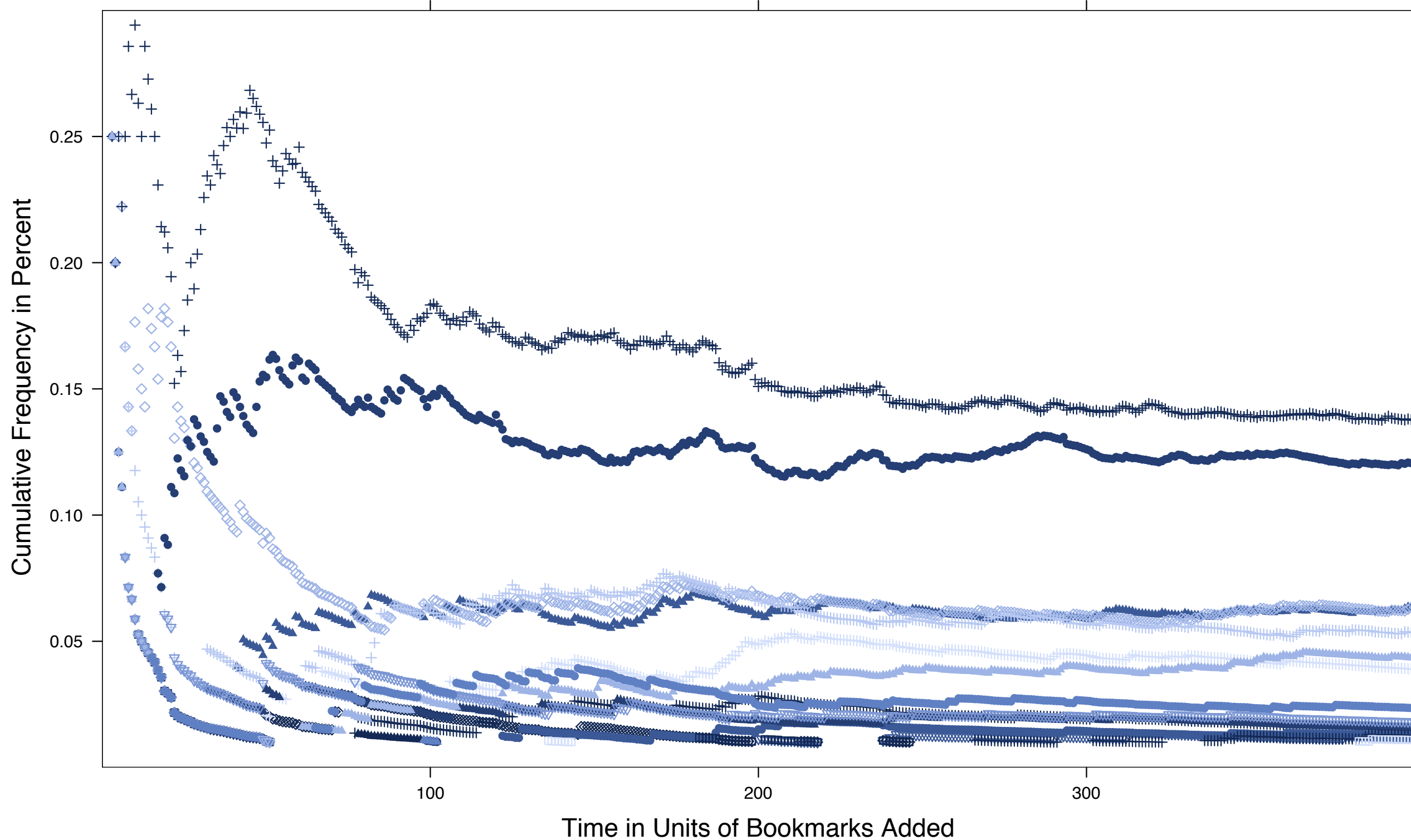
government books congress resources loc archive literature

Done

## **Social Hypothesis**

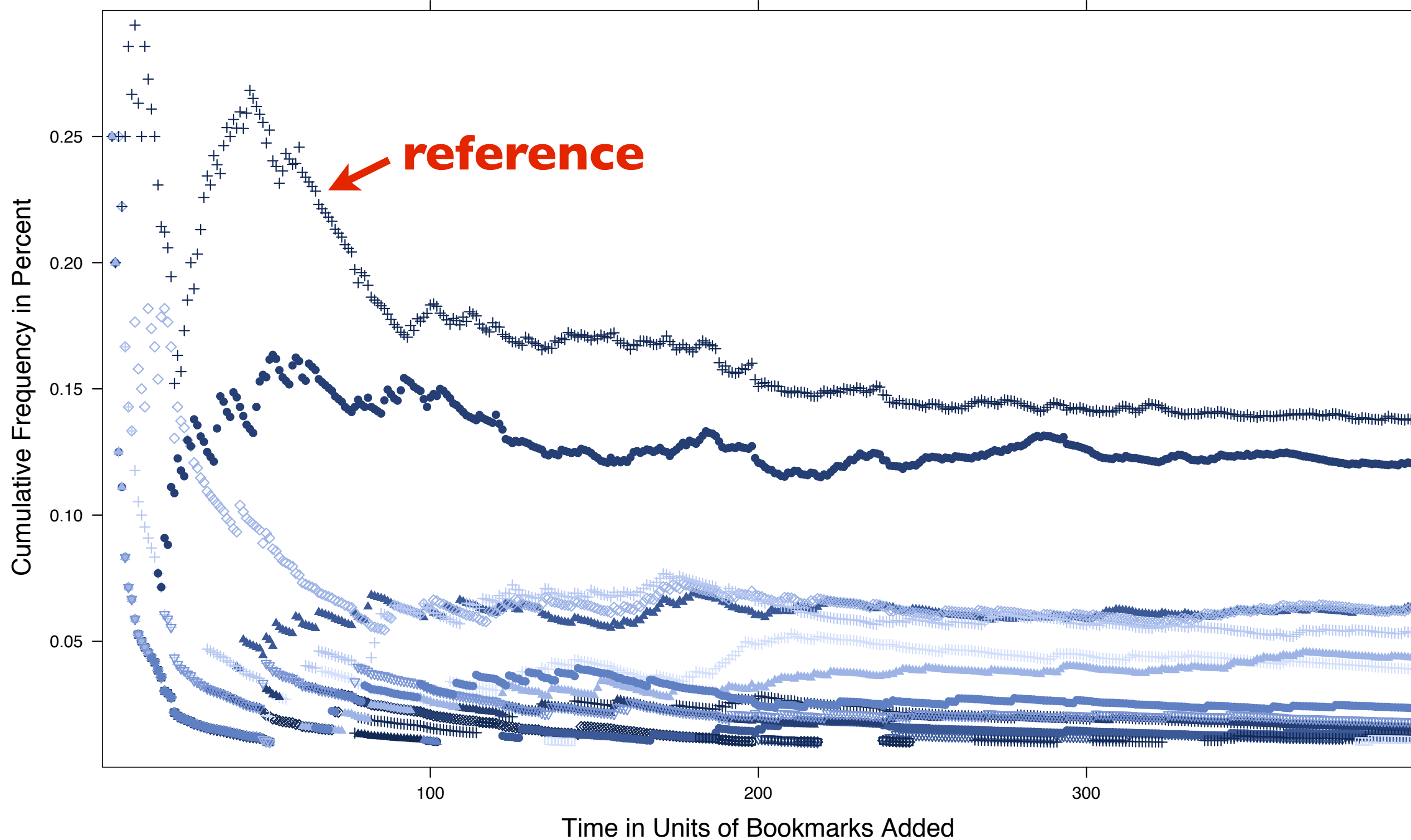
Users' tag choices are influenced  
by the tag choices of others

**Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)**

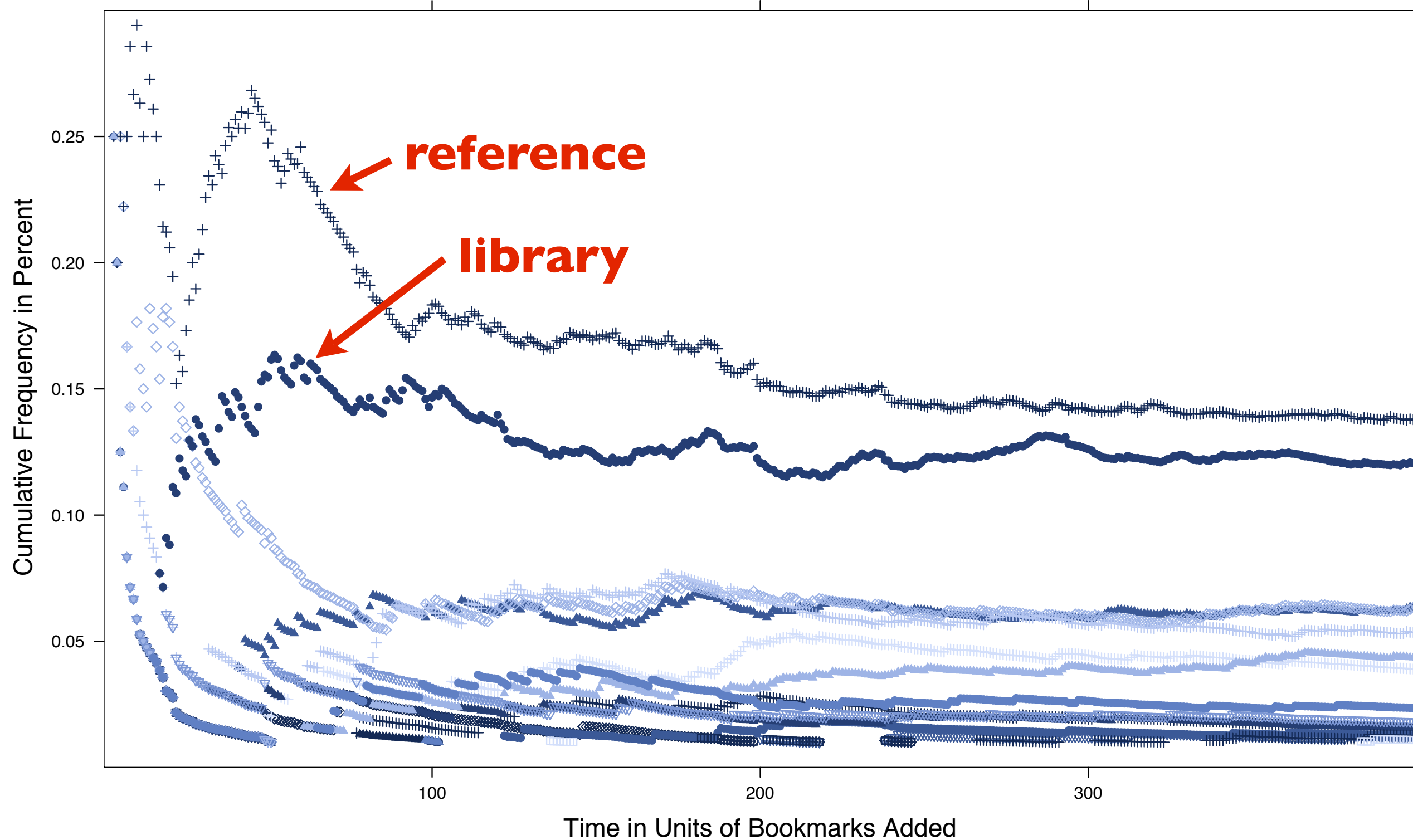




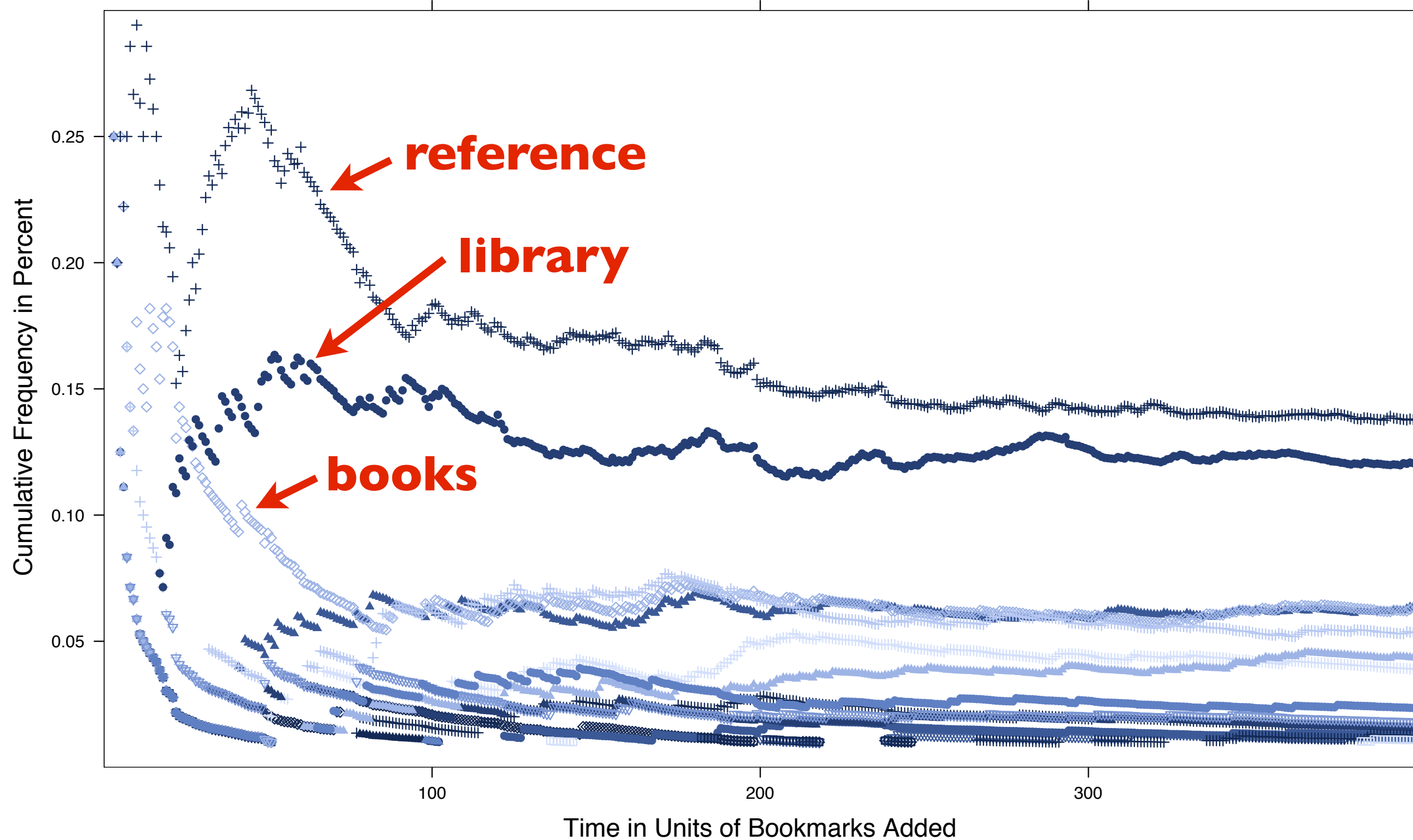
Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)



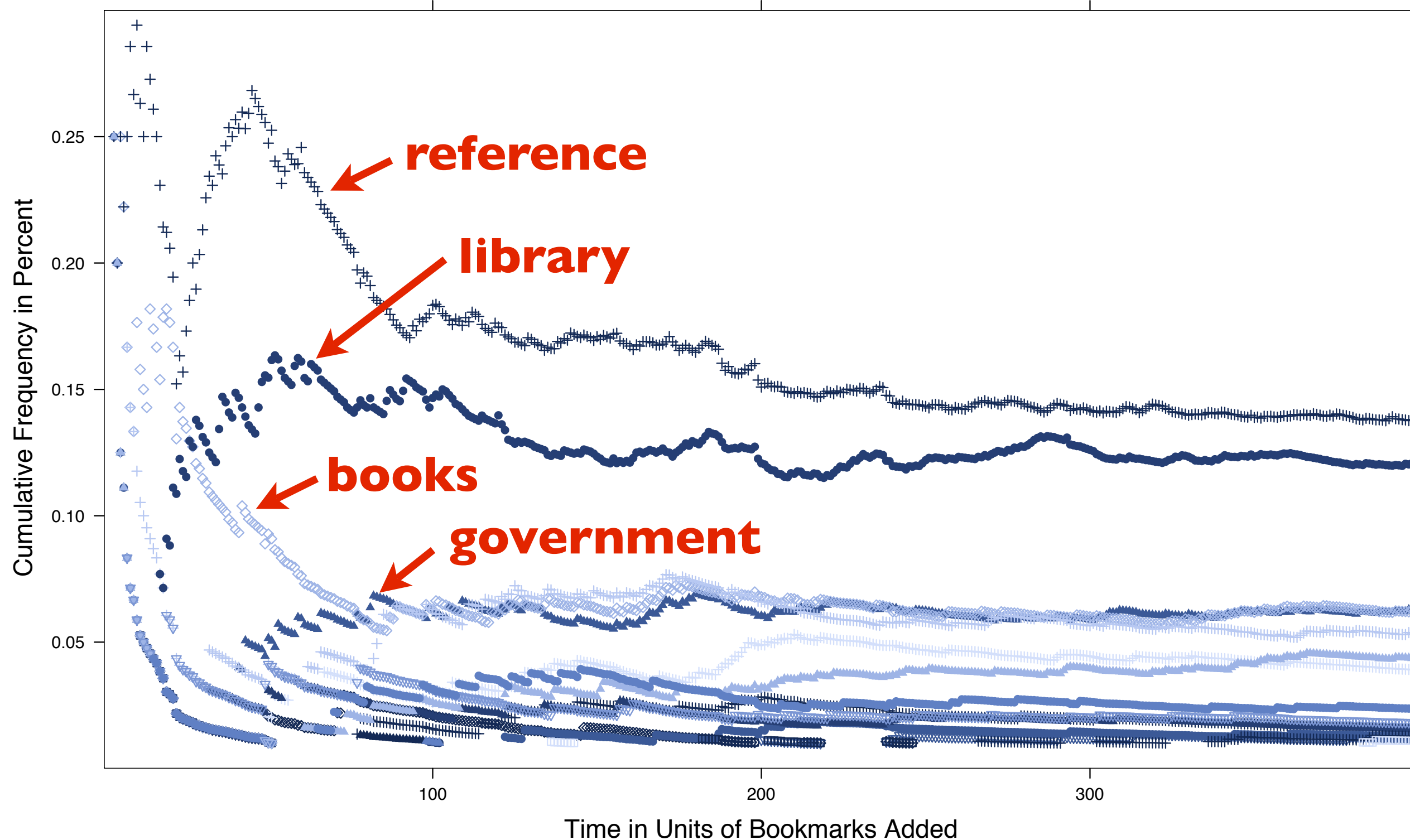
Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)



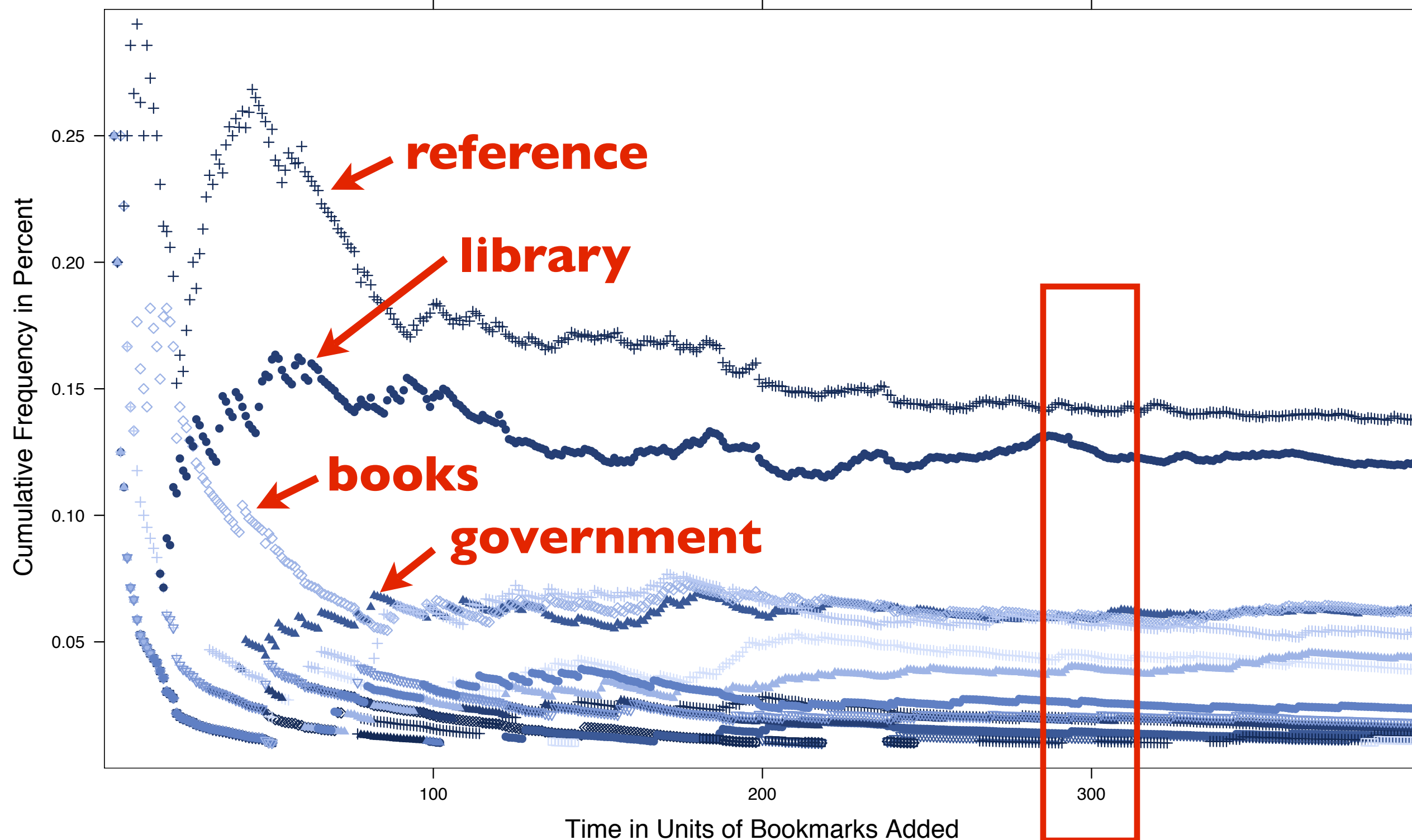
Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)



Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)

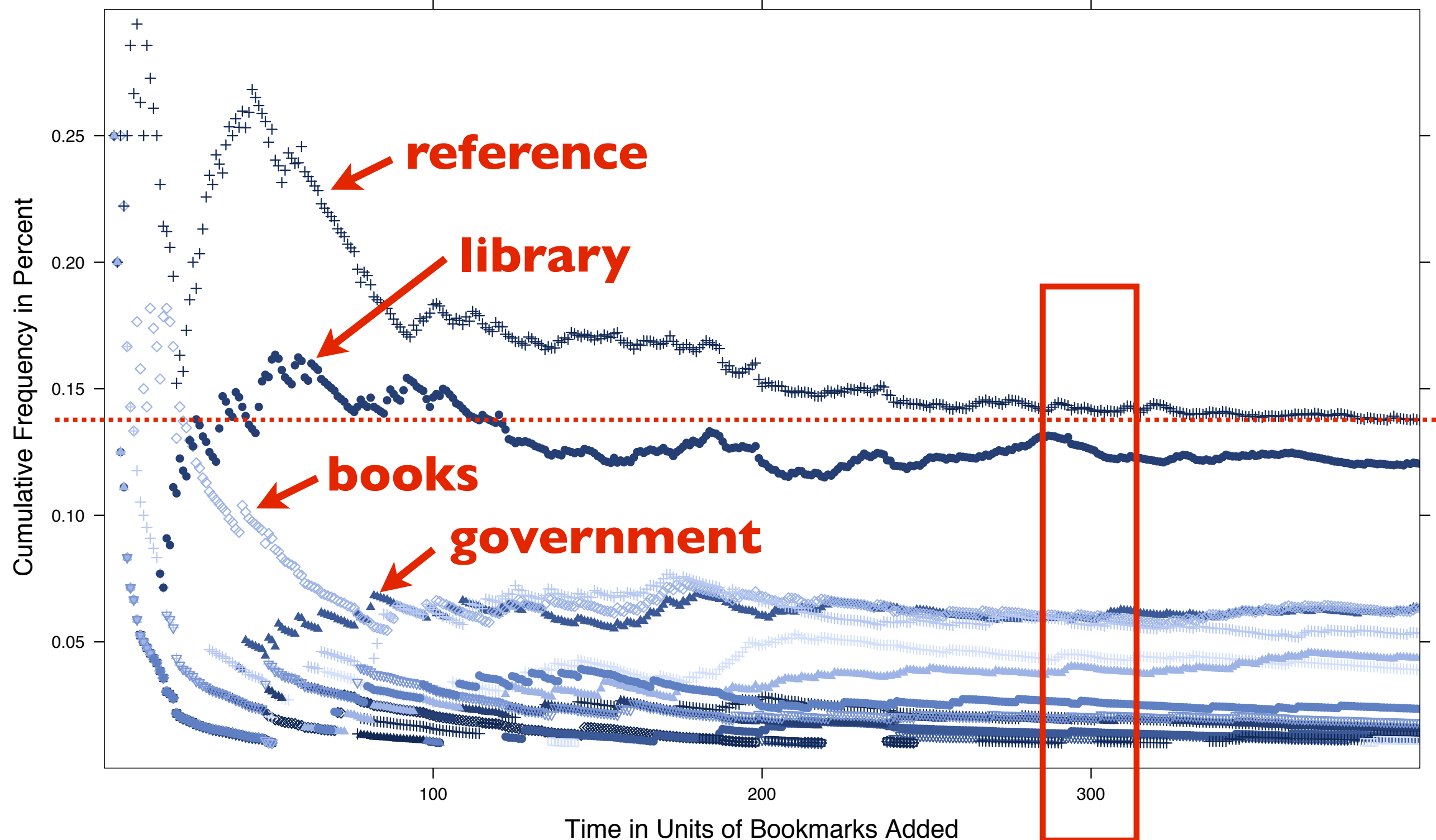


Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)

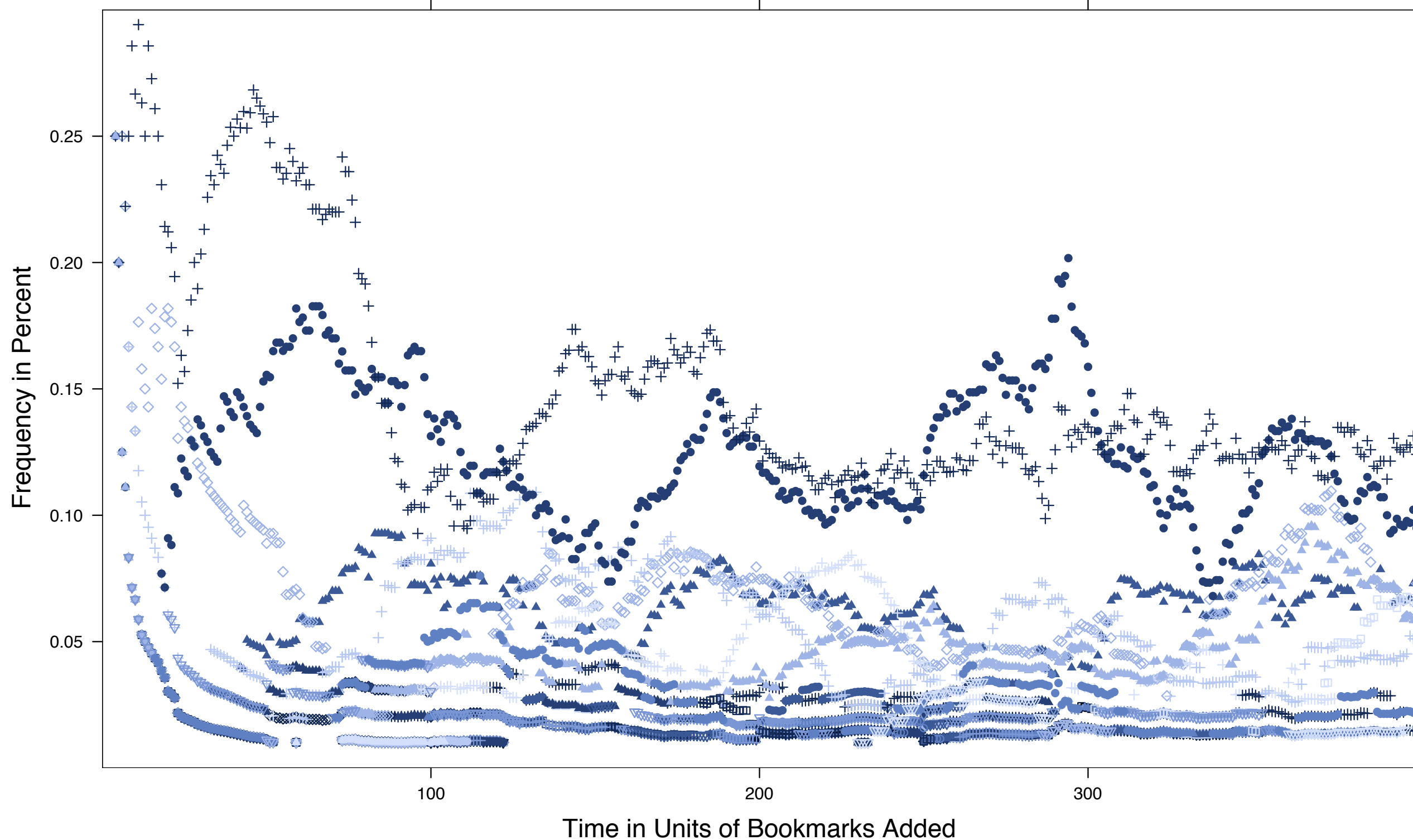




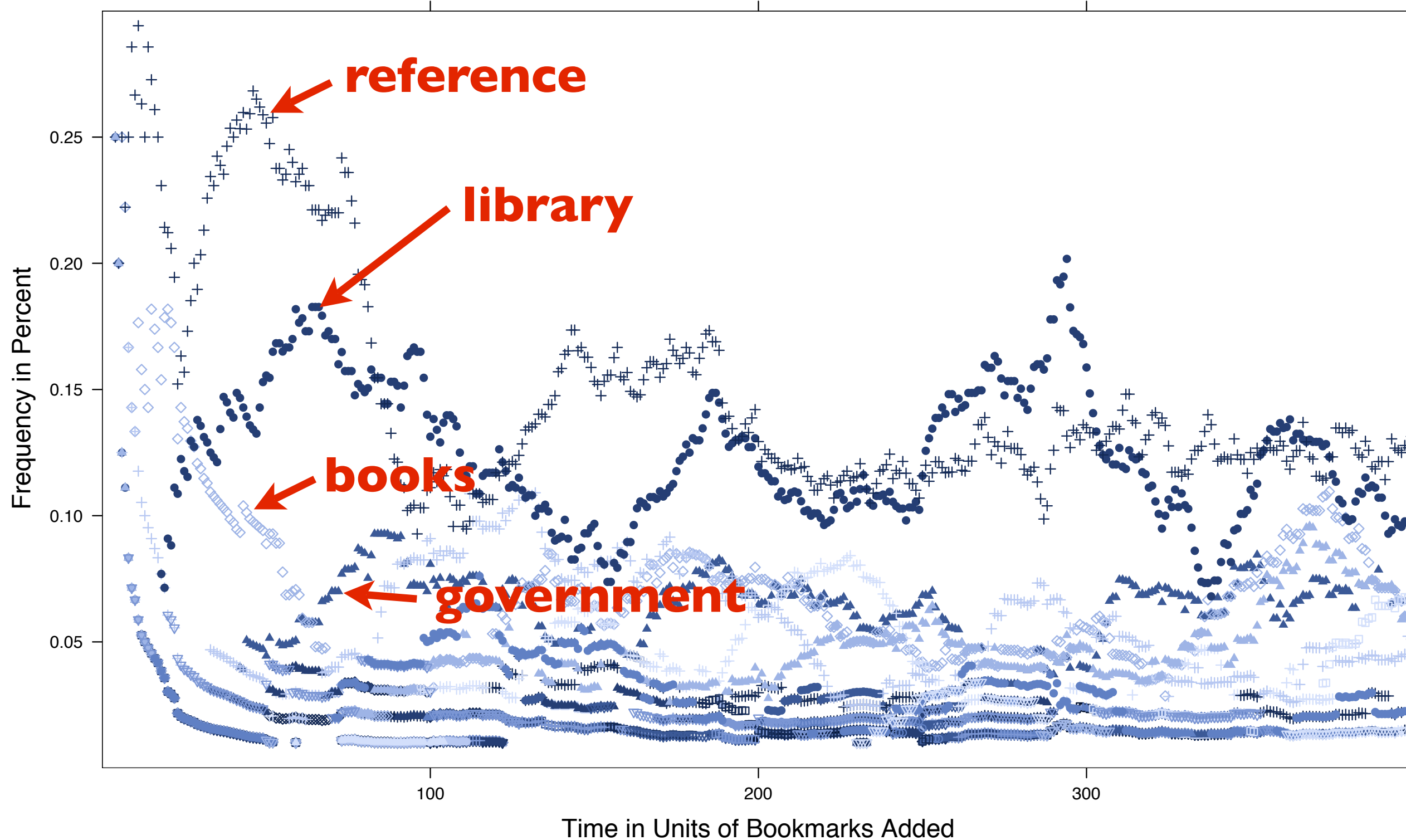
Tag Distribution for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)



**Sliding Window Tag Distribution, for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)**



Sliding Window Tag Distribution, for del.icio.us ID:  
3b3cb60120b337fa373871347e15418c (libcongress)





## **Organizing Hypothesis**

Users' tag choices are personal and idiosyncratic, not influenced by others' tag choices

## Wash and Rader (2007)

- Respondents generally used one or more heuristics for choosing tags:
- Reuse tags they have applied before to other web pages
- Create and adhere to mental rules or definitions for specific tags
- Choose terms they imagine using to re-find bookmarks in the future

We set out to look for a **connection**  
between the **small scale** (individual tag choices)  
and the **large scale** (aggregate patterns)  
for tags on [del.icio.us](http://del.icio.us).

# logistic regression

# Hypotheses

# Hypotheses

- Imitation: Users imitate tags that previous users have applied to a web page

# Hypotheses

- Imitation: Users imitate tags that previous users have applied to a web page
- Organizing: Users re-use tags that they have applied to other web pages

# Hypotheses

- Imitation: Users imitate tags that previous users have applied to a web page
- Organizing: Users re-use tags that they have applied to other web pages
- Recommended: Users choose tags that are suggested via the del.icio.us posting interface



$$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \text{interaction}, \\ \text{tag\_dummies}, \text{random\_effect}(\text{user}))$$

$$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \text{interaction}, \\ \text{tag\_dummies}, \text{random\_effect}(\text{user}))$$

$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \text{interaction},$   
 $\text{tag\_dummies}, \text{random\_effect}(\text{user}))$

$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \underline{\text{interaction}},$   
 $\text{tag\_dummies}, \text{random\_effect}(\text{user}))$

$$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \text{interaction}, \text{tag\_dummys}, \text{random\_effect}(\text{user}))$$

$$\text{tag\_chosen} = f(\text{used.onSite}, \text{used.byUser}, \text{interaction}, \\ \text{tag\_dummies}, \underline{\text{random\_effect(user)}})$$

# Summary of Logistic Regression Results:

	<b>onSite</b>	<b>byUser</b>	<b>Interaction</b>
<b>Mean Coefficient</b>	-0.245	3.439	-0.280

# Summary of Logistic Regression Results:

	onSite	byUser	Interaction
Mean Coefficient	-0.245	3.439	-0.280



## Summary of Logistic Regression Results:

	onSite	byUser	Interaction
Mean Coefficient	-0.245	3.439	-0.280

## Summary of Logistic Regression Results:

	onSite	byUser	Interaction
Mean Coefficient	-0.245	3.439	-0.280

**Organizing hypothesis** is strongly supported

# computer model

# Computer model

# Computer model

- Imitation-Popular: Users might select popular tags presented to them in the del.icio.us interface

# Computer model

- Imitation-Popular: Users might select popular tags presented to them in the del.icio.us interface
- Imitation-Urn: Imitation of other users' tag choices might be achieved using a path-dependent process

# Computer model

- Imitation-Popular: Users might select popular tags presented to them in the del.icio.us interface
- Imitation-Urn: Imitation of other users' tag choices might be achieved using a path-dependent process
- Organizing: Users might favor tags that they had used previously.

# **Two complimentary measures**

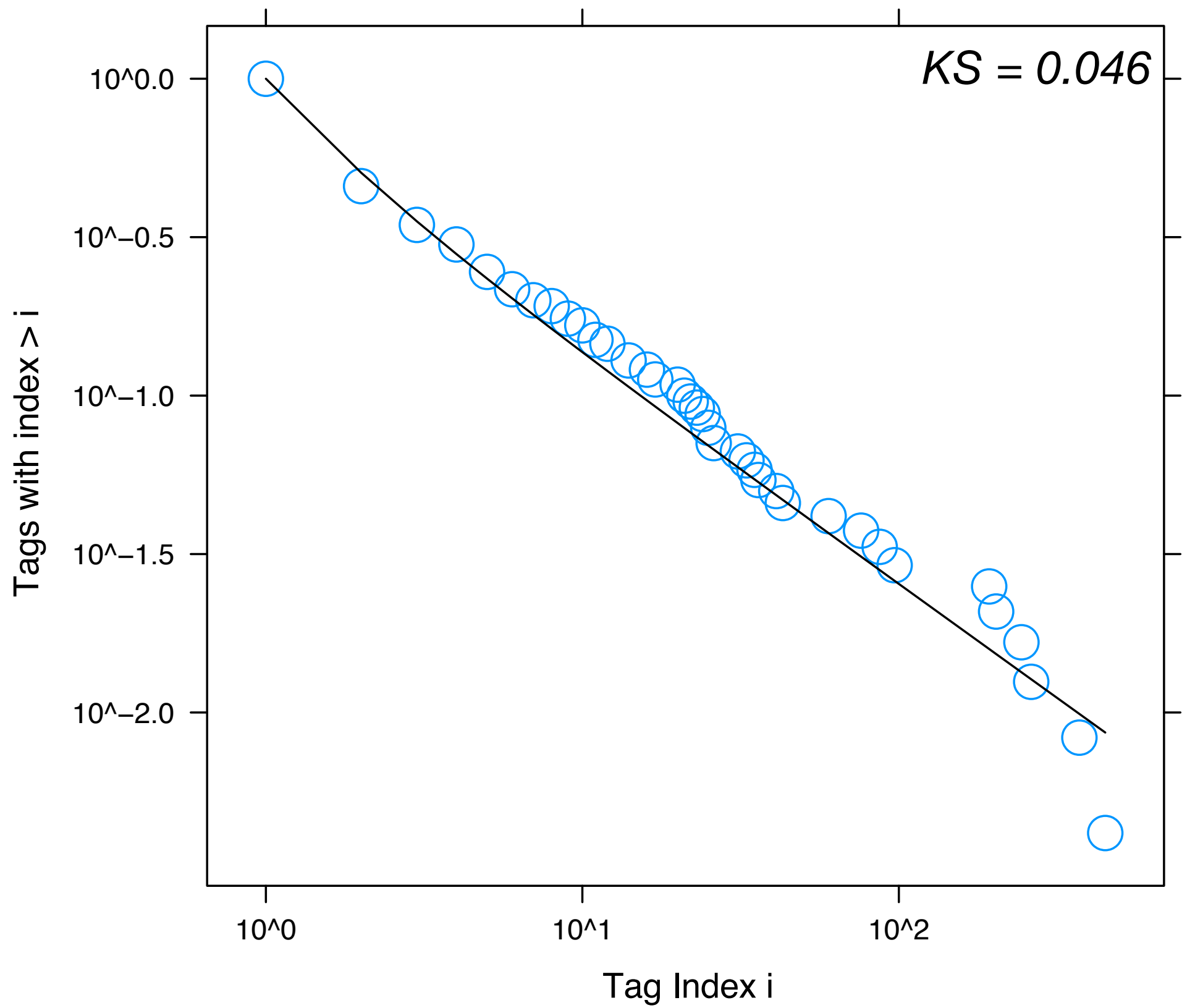


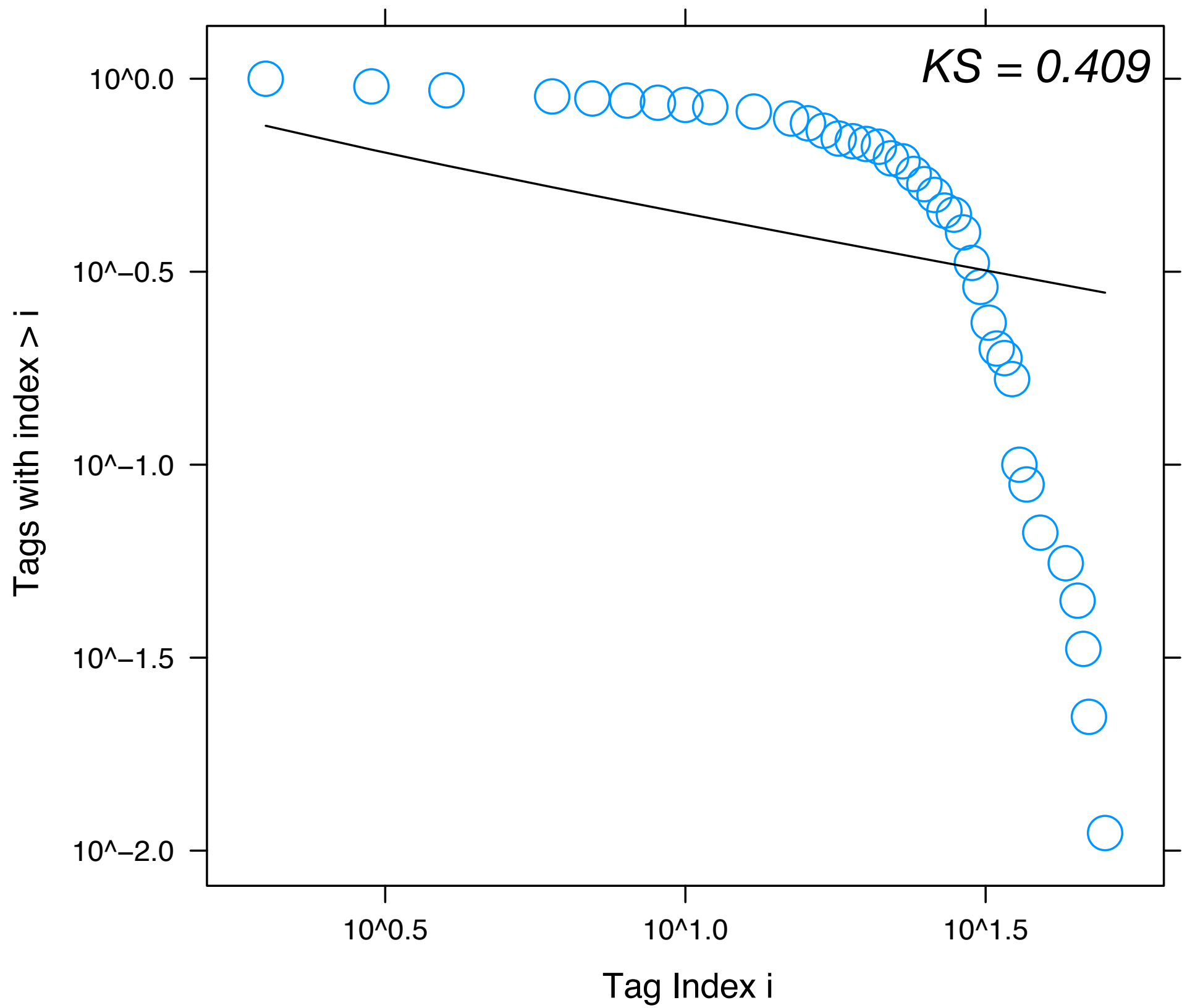
## **Two complimentary measures**

- Distributional Equivalence: Fit to a powerlaw distribution (KS-test)

## **Two complimentary measures**

- Distributional Equivalence: Fit to a powerlaw distribution (KS-test)
- Numerical Identity: Inter-user Agreement (Furnas et. al, 1983)





# Summary of Computer Model Results for **Distributional Equivalence**

	<b>Real World</b>	Imitation- Popular	Imitation- Urn	Organizing
<b>Mean KS statistic</b>	0.069	0.223	0.139	0.084

# Summary of Computer Model Results for **Distributional Equivalence**

	<b>Real World</b>	<b>Imitation- Popular</b>	<b>Imitation- Urn</b>	<b>Organizing</b>
<b>Mean KS statistic</b>	0.069	0.223	0.139	0.084

# Summary of Computer Model Results for **Distributional Equivalence**

	<b>Real World</b>	Imitation- Popular	<b>Imitation- Urn</b>	Organizing
<b>Mean KS statistic</b>	0.069	0.223	0.139	0.084

# Summary of Computer Model Results for **Distributional Equivalence**

	<b>Real World</b>	Imitation- Popular	Imitation- Urn	<b>Organizing</b>
<b>Mean KS statistic</b>	0.069	0.223	0.139	0.084



# Summary of Computer Model Results for **Distributional Equivalence**

	<b>Real World</b>	Imitation- Popular	Imitation- Urn	<b>Organizing</b>
<b>Mean KS statistic</b>	0.069	0.223	0.139	0.084

Rule out **Imitation-Popular** and **Imitation-Urn**

# Summary of Computer Model Results for **Numerical Identity**

# Summary of Computer Model Results for **Numerical Identity**

	<b>Real World</b>	Imitation- Popular	Imitation- Urn	Organizing
<b>Mean IUA</b>	0.144	0.317	0.184	0.182

# Summary of Computer Model Results for **Numerical Identity**

	<b>Real World</b>	<b>Imitation- Popular</b>	<b>Imitation- Urn</b>	<b>Organizing</b>
<b>Mean IUA</b>	0.144	0.317	0.184	0.182

# Summary of Computer Model Results for **Numerical Identity**

	<b>Real World</b>	Imitation- Popular	<b>Imitation- Urn</b>	Organizing
<b>Mean IUA</b>	0.144	0.317	0.184	0.182

# Summary of Computer Model Results for **Numerical Identity**

	<b>Real World</b>	Imitation- Popular	Imitation- Urn	<b>Organizing</b>
<b>Mean IUA</b>	0.144	0.317	0.184	0.182

# Summary of Computer Model Results for **Numerical Identity**

	<b>Real World</b>	Imitation- Popular	<b>Imitation- Urn</b>	<b>Organizing</b>
<b>Mean IUA</b>	0.144	0.317	0.184	0.182

Rule out **Imitation-Popular**

1. Imitation-Popular: **rule out** based on both numerical identity and distributional equivalence
2. Imitation-Urn: **rule out** based on distributional equivalence
3. Organizing: **cannot rule out** (similar to “real world” on all measures)



1. Imitation-Popular: **rule out** based on both numerical identity and distributional equivalence
2. Imitation-Urn: **rule out** based on distributional equivalence
3. Organizing: **cannot rule out** (similar to “real world” on all measures)

1. Imitation-Popular: **rule out** based on both numerical identity and distributional equivalence
2. Imitation-Urn: **rule out** based on distributional equivalence
3. Organizing: **cannot rule out** (similar to “real world” on all measures)

The **Organizing hypothesis** was **strongly supported** by logistic regression, and we could **not rule it out** based on our computer model.

main result

Users' future tag choices are  
**heavily influenced by tags they  
have chosen** in the past.

main result

# conclusion

- **Exact word choices** matter for re-finding and navigation in tagging system, but there are **competing hypotheses** in the literature for how people choose tags.
- We set out to **establish a connection** between the small scale (individual tag choices) and the large scale (aggregate patterns) for tags on del.icio.us
- **Organizing hypothesis** was **strongly supported** by logistic regression, and we could not rule it out based on our computer model

# Thank You!

Emilee Rader and Rick Wash  
School of Information, University of Michigan  
ejrader, rwash @ umich.edu

Online Appendix: <http://bierdoctor.com/papers/cscw08/>