

FolkTrails

Interpreting Navigation Behavior in a Social Tagging System



Thomas Niebler¹, Martin Becker¹, Daniel Zoller¹, Stephan Doerfel², Andreas Hotho^{1,3}

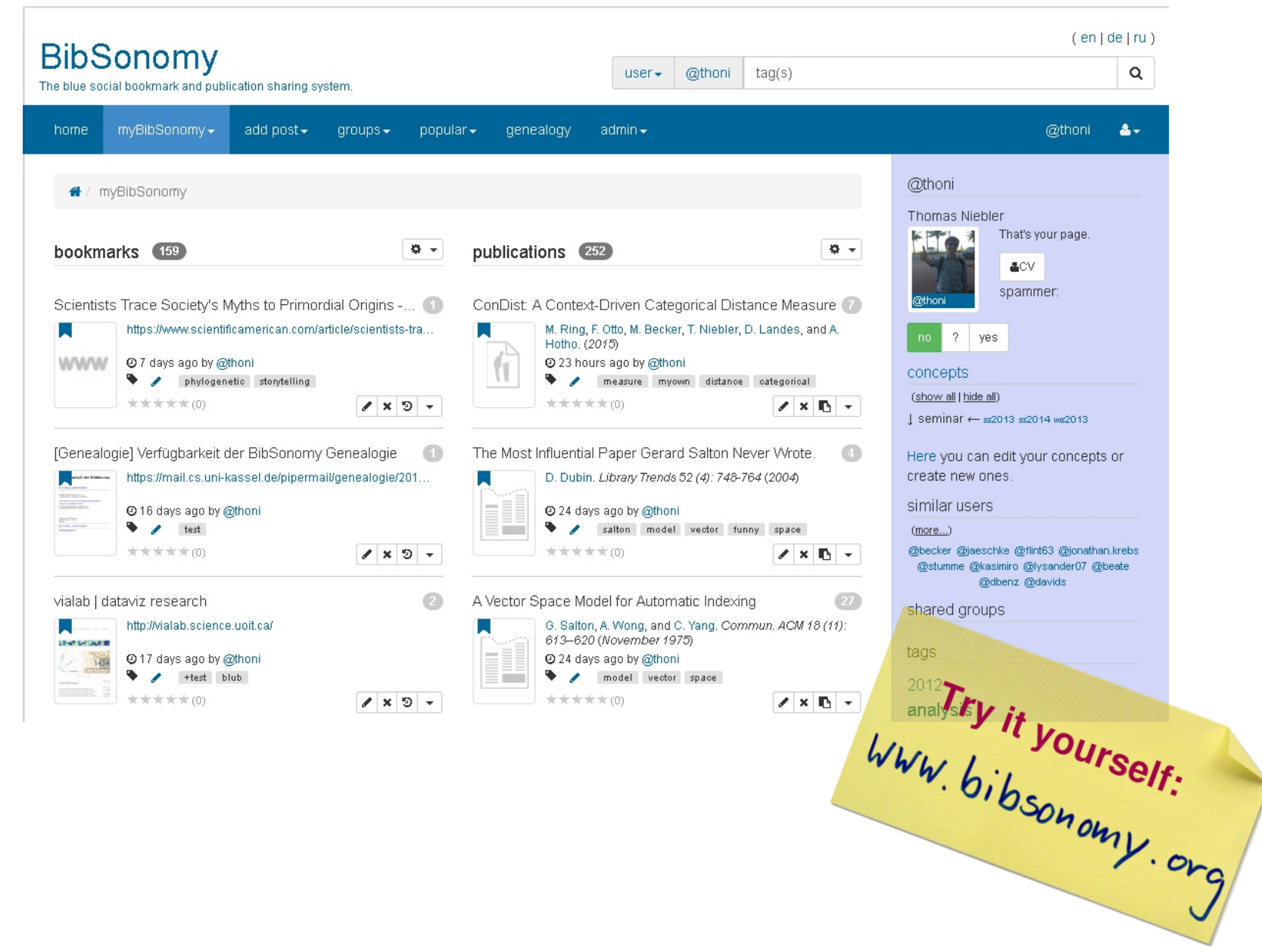
*Data Mining and Information Retrieval Group, University of Würzburg, Germany⁽¹⁾,
KDE Group, University of Kassel, Germany⁽²⁾, L3S Research Center, Hanover, Germany⁽³⁾*

Motivation

The Navigation Behaviour of users on Web 2.0 systems is largely unexplored. This is however important to effectively analyze and e.g. improve website structures for better content provision.

In this work, we focus on the **analysis of navigation in Social Tagging Systems**, specifically in BibSonomy. Social Tagging Systems have been in the focus of research for many years now and interesting results about the generated content entities have been presented, exhibiting the great semantic exploitability of such systems. Still, **how users navigate** such systems is **largely unknown**.

We provide and analyze several hypotheses about the incentives of user navigation in the Social Tagging System BibSonomy, thus **shedding light on how users consume the provided content**.



Technical Background

HypTrails is a Bayesian approach for comparing hypotheses about human trails on the web. It models transition data in first-order Markov chains and represents a transition dataset as a matrix of transition counts.

Folksonomies are the tripartite structures behind social tagging systems and are comprised of users, resources and tags as well as a set of tag assignments.

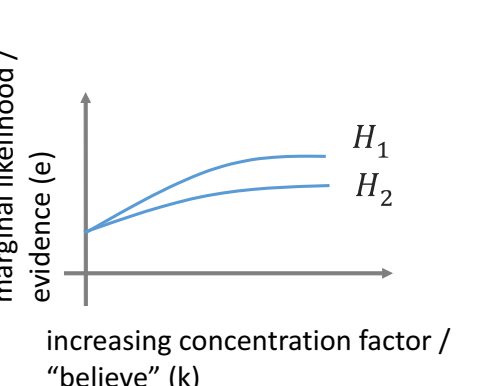
Marginal likelihood

$$P(D|H) = \int P(D|\theta)P(\theta|H)d\theta$$

Bayes factor

$$B_{1,2} = \frac{P(D|H_1)}{P(D|H_2)}$$

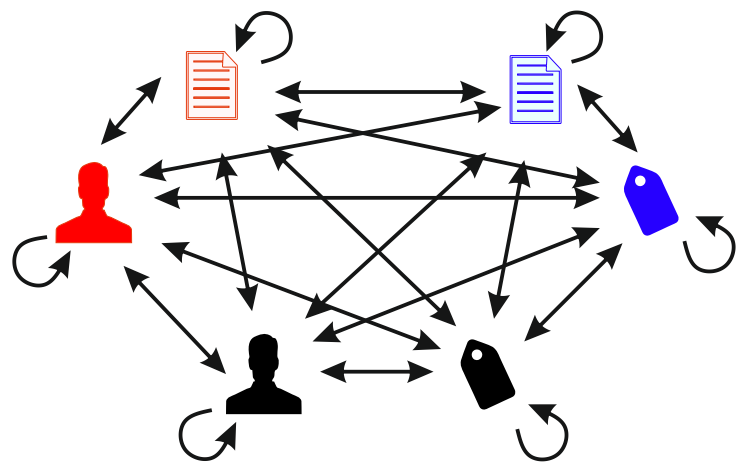
Marginal likelihood / evidence curves



Hypotheses about Navigation

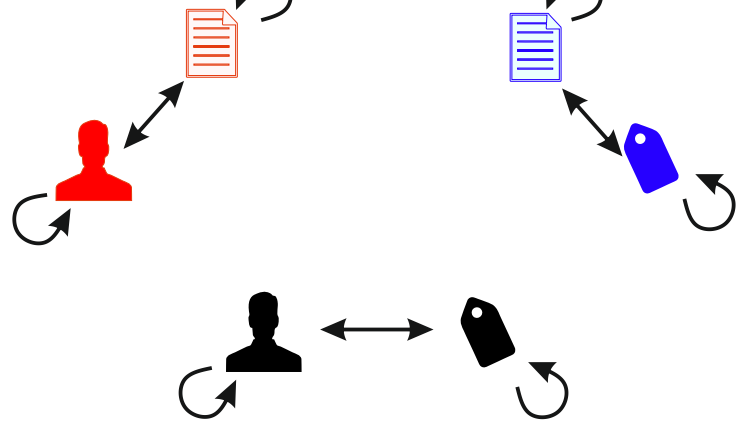
Uniform Hypothesis (uniform)

The Uniform Hypothesis serves as the baseline. It assumes that users **navigate randomly**.



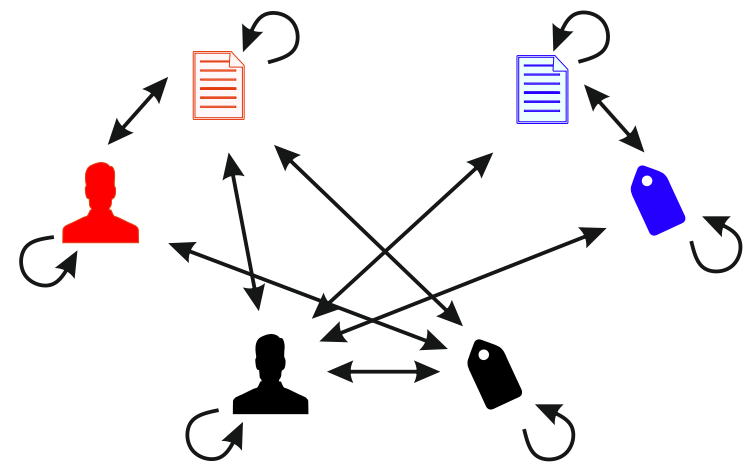
User Consistent Hypothesis (user)

Many people make use of BibSonomy as a place to save their resources. Thus, they navigate mainly on their **own pages**.



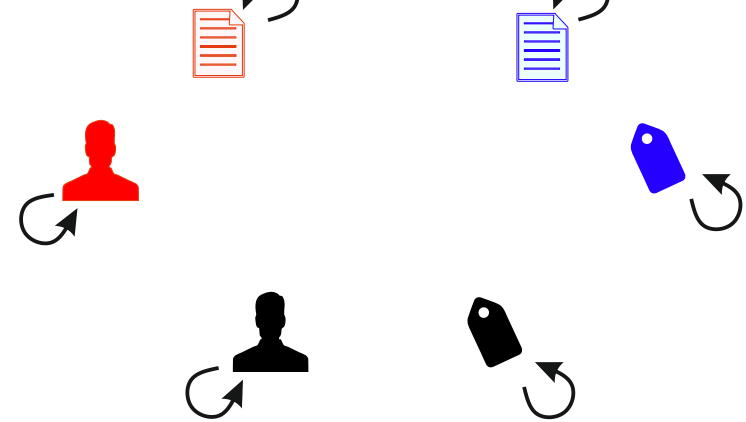
Folksonomy Consistent Hypothesis (folk)

Because BibSonomy is based on the Folksonomy structure, we expect users to only follow links provided by the Folksonomy.



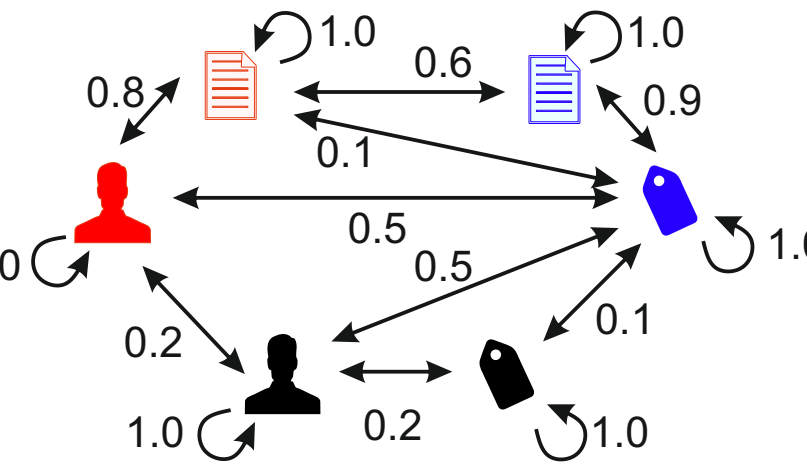
Page Consistent Hypothesis (page)

In previous studies it was found that users often **stay on the same pages**, e.g. due to pagination effects.



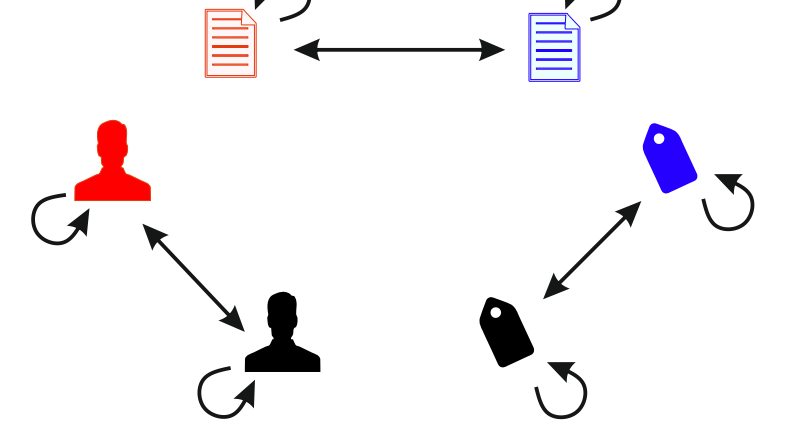
Semantic Navigation Hypothesis (tfidf)

Because tagging data contain a lot of semantic information, we expect users to prefer pages, which are **semantically stronger related**.



Category Consistent Hypothesis (cat)

A relaxation of the page consistent hypothesis, we assume that users tend to **stay on pages with the same category**, e.g. only on user pages.



We additionally investigated **combinations of the above hypotheses**, such as user-tfidf, where users navigate semantically on their own resources.
Legend: 1. Colored items indicate ownership. 2. : user page; : resource page; : tag page; an arrow indicates allowed navigation.

Datasets

We analyze a large set of content and log data from the Social Tagging System BibSonomy, spanning from 2006 to 2012.

User and Content Dataset

We use the Folksonomy data from nonspammers with their respective resources and tags. **17,932 users** were explicitly classified as nonspammers. They created **456,777 bookmark posts** and **2,410,844 publication posts** using **65,228 distinct tags**, which have been at least used twice.

Request Log Dataset

The BibSonomy log files include all HTTP requests to the system. After filtering, the remaining dataset contains **103,415 distinct visited content pages**, i.e. pages which show at least one resource. We recorded **327,060 transitions** between these pages. **123,452 transitions were self-transitions** (i.e., transitions from a page to itself) and **261,300 were transitions**, where the logged in **user owns both the source and the target page**.

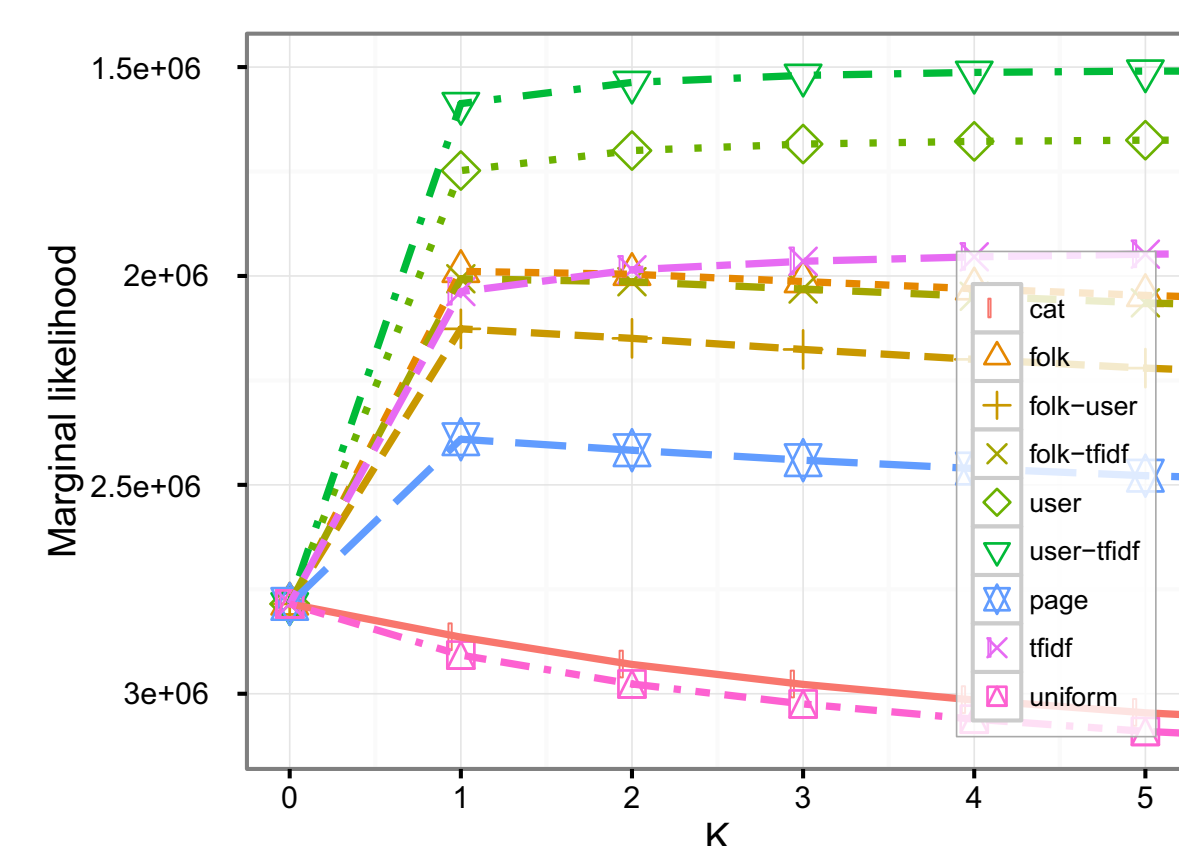
Request Log Subsets

We additionally investigated **outside navigation**, i.e., where users navigate outside their own pages. The subset of those requests counts **42,193 requests**, and effects of usage continuity. The subset of requests by **short-term users**, i.e. who used the system less than half a year, is comprised of **48,221 requests**.

Results

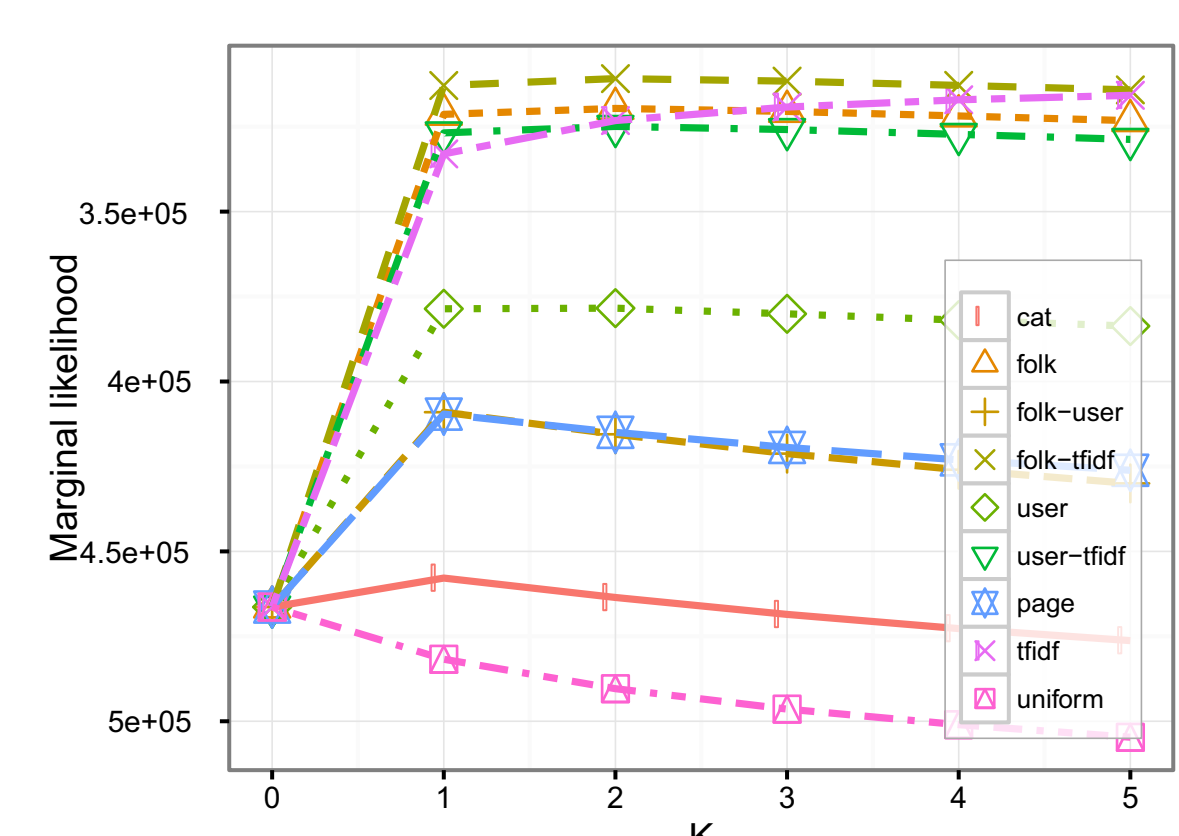
We analyzed navigation on the overall request log dataset and on two subsets: Outside navigation and short-term usage.

Overall Request Log Dataset



- **All of the basic hypotheses** explain the observed transitions **better than the baseline**, which indicates at least some structural properties explaining the observed transitions.
- The good performance of the semantic hypotheses indicates that **semantic similarity of pages is a strong factor for navigation** in BibSonomy. However, users tend to mostly navigate on their **own pages**.
- Overall, the **combination of the user consistent and the semantic hypothesis** performs best, indicating that navigation on BibSonomy can mainly be explained by **semantic navigation within the resources of a specific user**.

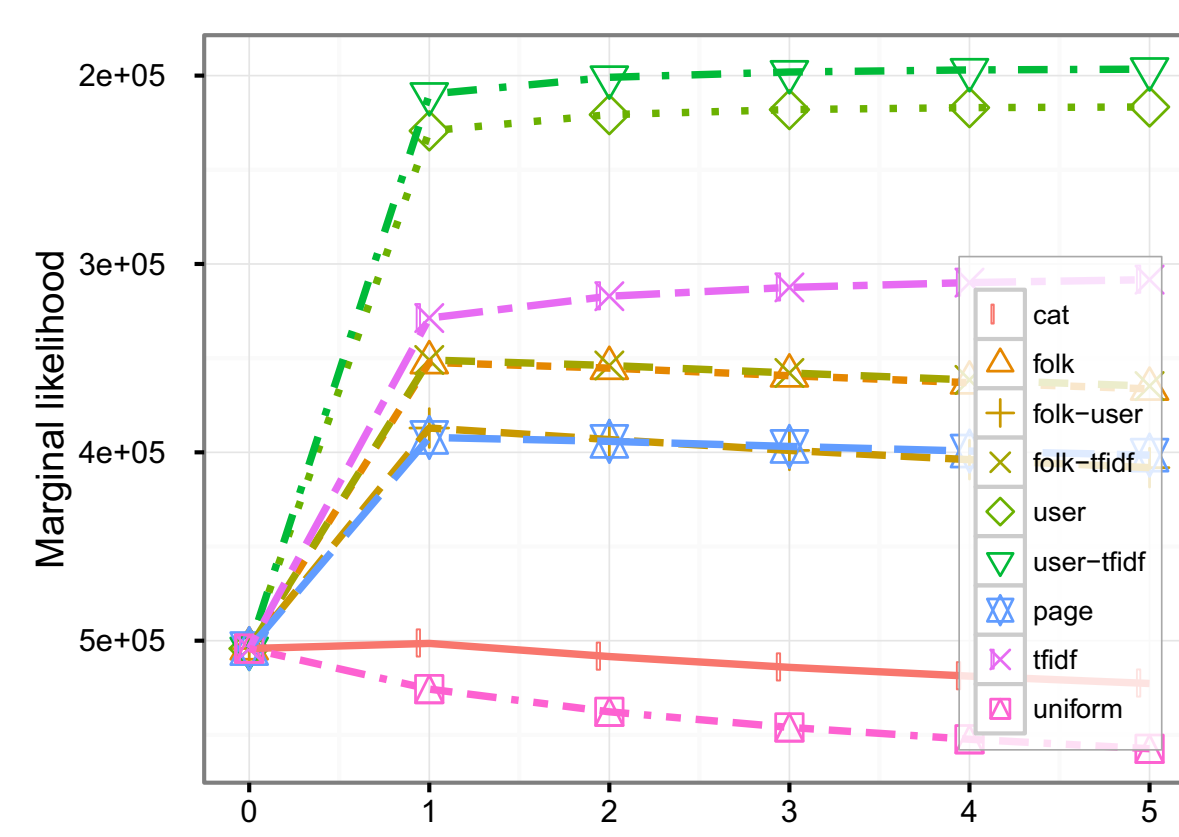
Request Subset: Inside vs. Outside Navigation



Motivated by the fact that users often navigate on their own pages, we investigate whether users **behave differently** when they are browsing the folksonomy **outside of their own pages**.

- The hypothesis which best explains the behaviour outside one's own resources is the **folksonomy consistent & semantic navigation hypothesis**.
- This indicates that users utilize the folksonomy structure combined with a semantic incentive, i.e. rather following **semantically related** pages, when discovering new resources.

Request Subset: Usage Continuity



Since we expect users to adapt to systems they are using, we investigate if their navigation behaviour changes **over time**.

- The **semantic hypothesis** performs significantly better than the folksonomy hypothesis. We may thus observe a **learning process**: Short-term users are not as adapted to the folksonomy structure as long-term users and thus rely on their semantic intuition.
- The **page consistent** and **folksonomy consistent & user consistent** hypotheses explain navigation equally well. This might be due to **increased use of pagination** as well as a **lack of personal resources**.

Conclusion

- In this work, we studied a **large dataset of webserver logs** from the social tagging system **BibSonomy** in order to **analyze and understand** components of **user navigation in social tagging systems**.
- Our results show that there is a **strong semantic component** inherent in user navigation. On subsets of the set of request logs, we can also show that **users navigating outside their own resources largely follow the folksonomy structure**.
- Overall, we were able to gain **new insights into the underlying processes of navigation** in tagging systems, which can be extended and leveraged in the future, for example, by considering new hypotheses, **improving navigation experience** or **extracting the latent semantic information**.