

NAVTAG - A Network-Theoretic Framework to Assess and Improve the Navigability of Tagging Systems

Christoph Trattner

Graz Technical University of Technology
Knowledge Management Institute and
Institute for Information Systems and Computer Media
Inffeldgasse 21a/16c
A-8010 Graz
ctrattner@iicm.edu

Abstract. This paper presents NAVTAG - A network theoretical framework to assess and improve the navigability of tagging systems. The framework provides the developer of a tagging system with a simple to use and scalable tool to assess the navigability of a given tag network or a tag network that is generated by the NAVTAG framework using different tag cloud and resource list generation algorithms. To the best of our knowledge this framework is the first approach of a tool that is able to assess and improve the navigability of a given tagging system from a network-theoretic perspective.

Key words: tagging systems, navigability, tag networks, network theory

1 Introduction

Recently tagging systems [2] gained tremendously in popularity. While tagging systems were in the past typically associated with online bookmarking systems such as Delicious or CiteULike, the term is nowadays also connected with modern Web 2.0 web applications such as Amazon or LastFM. Basically, a tagging systems is a tool that allows the user to apply light-weighted key words - the so-called tags - to the resources of system. On LastFM for instance people apply tags to resources such as videos, photos or music. On Amazon as an other example of a modern tagging system people apply tags to product items such as books etc. Typically, developers provide the user with tagging functionality to allow the user to organize or describe the resources of a system. By visualizing the tags into the so-called tag clouds the users are then able to navigate to the resource of the system via tags.

While recent research has studied navigation in tagging systems from user interface [7], [8], [6] and information-theory [1] perspectives, the unique focus

of our work is the network-theoretic analysis of tagging systems. In previous research it was observed that different tag cloud or resource list calculation algorithms [4], [10], different tag taxonomy induction algorithms [3] or also different types of tags [9] influence the navigability of a tagging system significantly. Basically, a navigable tagging system is defined as a system where the underlying tag network has a low diameter bounded by $\log(N)$, where N is the number of nodes in the network, and an existing giant component, i.e. a strongly connected component containing almost all resource of the tagging system [5]. An efficiently navigable tagging system is defined as a navigable network (see previous definition) and a network for which a decentralized searcher exists that is able to search the network in $\log(N)$ [5]. To measure these network properties a network-theoretic framework was developed. The framework is called *NAVTAG* – A Network-Theoretic Framework to Assess and Improve the Navigability of Tagging Systems. The novelty of the tool is on the one hand it’s uniqueness regarding the possibility to measure the navigability of tagging systems, to our knowledge there are no other tools that offers such a feature, and on the other hand it’s uniqueness to do this on a network-theoretic level.

2 Approach & Implementation

Basically, the NAVTAG framework consist of three different modules: A tag network generation module, a tag-resource taxonomy generation module and a tag network analysis module. In the following sections the functionality and the implementation of the modules are described.

2.1 Tag Network Generation Module

The tag network generation module takes as input a given tagging dataset, a given tag cloud calculation algorithm and a given resource list generation algorithm and generates as output a tag network based on the given input tag dataset and algorithms. The module is implemented in Java and consists of the following three sub-modules:

Input-Reader: The input-reader basically stores the provided tagging data into a database module implemented with the Apache Lucene Search Engine.

Network Generator: This module generates the tag network. As input parameter the module takes a given tag cloud algorithm and/or a given resource list algorithm. The tag cloud algorithm implements the following interfaces routines $getTagCloud(r, n)$ where r is the currently processed resource and n the maximum tag cloud size and $getResourceList(r, t, k)$, where r is the currently processed resource, t the currently processed tag and k the maximum resource list size. Due to reasons of flexibility, generic interface routines are also provided.

Output-Writer: The output-writer module provides different output formats. Preferable, the module stores the generated tag network as bipartite or as tripartite tag network into a file. Additionally, the module outputs the resource-resource network and a tag-tag network of the tag network as well.

2.2 Tag-Resource Taxonomy Generation Module

Additionally to the network generation module NAVTAG provides a tag-resource taxonomy generation module. The module allows to generate different tag or resource taxonomies based on the tagging dataset provided by tag network generation module. Since tag-resource taxonomies are a popular approach to improve the navigability of tagging systems [4], the module implements popular tag taxonomy induction algorithms such as Heymann, Deg/Cos and Deg/Cooc [4]. As well as the tag network generation module, this module is implemented in Java.

2.3 Tag Network Analysis Module

The tag network analysis module is an extension of the Stanford *SNAP*¹ network library. Due to reasons of performance and scalability, this module is implemented in *C++*. By default the module provides the functionality to measure network properties such as in- and out-degree of a given network as well as navigability properties such as the size of largest strongly connected component or the efficient diameter of given network. To measure the efficiency of a tag network, a hierarchical decentralized searcher was implemented [3]. The searcher is able to measure whether a network is searchable in $\log(N)$ or not, i.e. the searcher conducts if the given tag network is also efficiently navigable.

As output the module generates a “navigability”-statistic file containing detailed information about the size of the largest strongly connected component, the efficient diameter of the tag network and the average number of hops of the searcher to reach a given number of nodes pairs in the tag network. The results are printed to a file in plain text format. Additional figure representations of the results in PS format are produced.

3 Results

All in all, NAVTAG was successfully deployed for a number of projects [4] [10], [3], [9] related to the field of tagging systems. For instance in [3], [4] the framework was successfully used to assess the navigability of different tagging networks (billion order) such BibSonomy, CiteULike, Delicious, LastFM or Flickr. The memory consumed by NAVTAG in the case of Delicious, the largest tag network on the web today was less than one GB. In [9] the framework was used to investigate and develop novel tag cloud and resource list generation algorithms. In [9] the framework was used to investigate the navigational utility of the so-called Google query tags compared to tags generated by users. In our latest research regarding the navigability of tagging systems the framework was deployed to assess the navigational utility of different tag taxonomy induction algorithms [3].

¹ <http://snap.stanford.edu/>

4 Conclusions and Future Work

In this work, NAVTAG - A network theoretical framework to assess and improve the navigability of tagging systems was presented. The framework provides the developer of a tagging system with a simple to use and scalable tool to assess the navigability of a given tag network or a tag network that is generated by the NAVTAG framework using different tag cloud and resource list generation algorithms. The framework was successfully deployed for a couple of major projects related to the field of tag navigation. To the best of our knowledge this framework is the first approach of tool that is able to assess and improve the navigability of a given tagging system. Future work will include further improvements towards the functionality of the framework and the release through the Google code project service.

Acknowledgments: This work is funded by - *BMVIT - the Federal Ministry for Transport, Innovation and Technology*, program line *Forschung, Innovation und Technologie für Informationstechnologie*, project NAVTAG – *Improving the navigability of tagging systems*.

References

1. E. H. Chi and T. Mytkowicz. Understanding the efficiency of social tagging systems using information theory. In *HT '08: Proceedings of the nineteenth ACM conference on Hypertext and hypermedia*, pages 81–88, New York, NY, USA, 2008. ACM.
2. T. Hammond, T. Hannay, B. Lund, and J. Scott. Social bookmarking tools (i): A general review. *D-Lib Magazine*, 11(4), 2005.
3. D. Helic, M. Strohmaier, C. Trattner, M. Muhr, and K. Lerman. Pragmatic evaluation of folksonomies. In *Proceedings of the 20th international conference on World wide web, WWW '11*, pages 417–426, New York, NY, USA, 2011. ACM.
4. D. Helic, C. Trattner, M. Strohmaier, and K. Andrews. Are tag clouds useful for navigation? a network-theoretic analysis. *International Journal of Social Computing and Cyber-Physical Systems*, 2011.
5. J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems (NIPS) 14*, page 2001, Cambridge, MA, USA, 2001. MIT Press.
6. C. S. Mesnage and M. J. Carman. Tag navigation. In *SoSEA '09: Proc. of the 2nd international workshop on Social software engineering and applications*, pages 29–32, New York, NY, USA, 2009. ACM.
7. A. W. Rivadeneira, D. M. Gruen, M. J. Muller, and D. R. Millen. Getting our head in the clouds: toward evaluation studies of tagclouds. In *Proc. of the SIGCHI conference on Human factors in computing systems, CHI '07*, pages 995–998, New York, NY, USA, 2007. ACM.
8. J. Sinclair and M. Cardew-Hall. The folksonomy tag cloud: when is it useful? *Journal of Information Science*, 34:15, 2008.
9. C. Trattner and D. Helic. Linking related documents: combining tag clouds and search queries. In *Proceedings of the 10th international conference on Web engineering, ICWE'10*, pages 486–489, Berlin, Heidelberg, 2010. Springer-Verlag.
10. C. Trattner, D. Helic, and M. Strohmaier. On the construction of efficiently navigable tag clouds using knowledge from structured web content. *j-jucs*, 17(4):565–582, 2011.