

Modeling Navigation in Information Networks

Dimitar Dimitrov
GESIS – Leibniz Institute for the Social Sciences
Cologne, Germany
dimitar.dimitrov@gesis.org

ABSTRACT

Navigation in an information space is a natural way to explore and discover its content. Information systems on the Web like digital encyclopedias (e.g., Wikipedia) are interested in providing good navigational support to their users. To that end, navigation models can be useful for estimating the general navigability of an information space and for understanding how users interact with it. Such models can also be applied to identify problems faced by the users during navigation and to improve user interfaces.

Studying navigation on the Web is a challenging task that has a long tradition in our scientific community. Based on large studies, researchers have made significant steps towards understanding navigational user behavior on the Web identifying general usage patterns, regularities, and strategies users apply during navigation. The seminal *information foraging* theory has been developed suggesting that people follow links by constantly estimating their quality in terms of information value and cost associated with obtaining that value by interacting with the environment. Furthermore, models describing the network structure of the Web like the *bow tie model*, and the *small world models* have been introduced. These models contributed valuable insights towards characterizing the underlying network topology on which the users operate and the extent to which it allows efficient navigation. In the context of information networks, researchers have successfully modeled user navigation resorting to *Markov chains* and to *decentralized search*. With respect to the users' navigational behavior and their click activities to traverse a link, researchers have found a valuable source of information in the log files of Web servers. Click data has also been collected by letting humans play navigational games on Wikipedia. With this data, researchers tested different navigational hypotheses; for example, (i) if humans tend to navigate between semantically similar articles, (ii) if they experience a trade-off between following links leading towards semantically similar articles and following links leading towards possibly well-connected articles. For navigation with a particular target in mind, users are found to be greedy with respect to the next click if they are confident to be on the right path, whereas they tend to explore the information network at random if they feel insecure or lost and have no intuition about the next click.

Although these research lines have advanced our understanding of navigational user behavior in information networks, for the goal of the proposed thesis—modeling navigation—related work does not address and cover the following questions: (i) What is the relationship between the user's awareness regarding the structure and the topology of the information network and the efficiency of navigation, i.e., modeled as decentralized search and (ii) How do users interact with the content to explore and discover it, i.e., are there some specific links that are especially appealing and what are their characteristics? My research focuses on modeling navigation in an information space represented as an information network. Regarding the first question, I introduce and apply partially informed decentralized search to model the extent to which a user is exposed to the network structure of the information space and can make informed decisions about her next step towards exploring the content [1]. I test different hypotheses regarding the type and the amount of network structural information used to model navigation. My results show that only a small amount of knowledge about the network structure is sufficient for efficient navigation.

For the second question, I study large-scale click data from the English version of Wikipedia. I observe a focus of the users' attention towards specific links. With this part of the proposal, I want to shed light on a different aspect of navigation and concentrate on the question why some links are more successful than others. In particular, I study the relationship between the link properties and the link popularity as measured by transitional click data. To that end, I formulate navigational hypotheses based on different link features, i.e., network features, semantic features and visual features [2, 3]. The plausibility of these hypotheses is then tested using a Markov chain-based Bayesian hypothesis testing framework. Results suggest that Wikipedia users tend to select links located at the top of the page. Furthermore, users are tempted to select links leading towards the periphery of the Wikipedia network.

To conclude, I believe that the won insights may have impact on system design decisions, i.e., existing guidelines for Wikipedia contributors can be adapted to better reflect the usage of the system.

Keywords: Navigation; Human Click Behavior; Wikipedia

References

- [1] D. Dimitrov, P. Singer, D. Helic, and M. Strohmaier. The role of structural information for designing navigational user interfaces. In *International Conference on Hypertext and Social Media*, 2015.
- [2] D. Dimitrov, P. Singer, F. Lemmerich, and M. Strohmaier. Visual positions of links and clicks on wikipedia. In *International Conference Companion on World Wide Web*, 2016.
- [3] D. Dimitrov, P. Singer, F. Lemmerich, and M. Strohmaier. What makes a link successful on wikipedia? *arXiv preprint arXiv:1611.02508*, 2016.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

WSDM 2017 February 06-10, 2017, Cambridge, United Kingdom

© 2017 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-4675-7/17/02.

DOI: <http://dx.doi.org/10.1145/3018661.3022754>