Dipl.-Ing. Daniel Lamprecht, BSc

Analyzing, Modeling and Improving Navigability in Networked Information Systems

DOCTORAL THESIS

to achieve the university degree of
Doktor der technischen Wissenschaften
submitted to

Graz University of Technology

Supervisor
Univ.-Prof. Dipl-Ing. Dr.techn. Markus Strohmaier

Institute of Interactive Systems and Data Science
Faculty of Computer Science and Biomedical Engineering
Graz University of Technology
Graz, Austria

GESIS—Leibniz Institute for the Social Sciences
Cologne, Germany

Department of Computer Science
University of Koblenz-Landau
Koblenz, Germany

Graz, April 2017

AFFIDAVIT

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly indicated all material which has been quoted either literally or by content from the sources used. The text document uploaded to TUGRAZonline is identical to the present doctoral thesis.

Date

Signature

Abstract

Much of human knowledge and expertise resides in information networks such as webpages, social media, or scientific citation networks. Being able to successfully retrieve information in these complex and evolving systems has become a critical skill. The two main strategies for information retrieval are search (query-based access) and navigation (browsing). Navigation requires less cognitive effort than search and is the technique of choice when an information need cannot be explicitly expressed. This thesis first presents an empirical investigation of the factors influencing link selection during navigation. A case study of Wikipedia compares a range of potential influences, such as generality of an article, the page organization, or the semantic similarity to a target article of a navigation scenario. The results show that page organization, i.e., the layout of a webpage and the position of its links, is the dominant factor influencing link selection. The central contribution of this thesis is a novel evaluation method, which builds on decentralized search and models navigation as a greedy process based on local knowledge. The presented method is able to automatically evaluate the state of navigability in an information network, can provide valuable insights into the dynamics of navigation, and measure retrieval performance. Finally, this thesis provides suggestions to improve navigability in real-world information networks based on the analysis of the influencing factors and the evaluation method. For recommender systems, this thesis shows evidence that navigability can be improved by introducing diversified recommendations. For Wikipedia, relocating links to sections where users are more likely to view them increases the number of articles users are able to reach. Both approaches keep the link targets relevant to a page's topic and the number of links constant, thus avoiding cognitive overload of users. The evaluation method and the empirical findings of this thesis provide a powerful tool to analyze the state of navigability of information networks, and are relevant for both researchers and practitioners. By automatically evaluating the ramifications of changes to the network, the method enables website operators to efficiently evaluate ways to make their networks more accessible.

Kurzfassung

Weite Teile des Wissens der Menschheit befinden sich in Informationsnetzwerken, wie etwa auf Websites, in soziale Medien oder in Zitationsnetzwerken. Die Fähigkeit des erfolgreichen Auffindens von Informationen in komplexen und veränderlichen Systemen ist zu einem wesentlichen Bestandteil des alltäglichen Lebens geworden. Die zwei Hauptstrategien für das Auffinden von Information sind Suche (mittels Suchanfragen an ein System) und Navigation (Browsing mittels Links). Navigation erhebt davon die geringeren kognitiven Anforderungen als Suche, und ist das Mittel der Wahl, wenn ein Informationsbedürfnis nicht explizit ausgedrückt werden kann. Diese Dissertation stellt zunächst eine empirische Untersuchung der Einflussfaktoren auf Linkauswahl in Navigationsprozessen vor. In einer Fallstudie werden anhand von Navigationsdaten von Wikipedia eine Reihe von potentiellen Einflussfaktoren verglichen, nämlich die Generalität von Artikeln, die inhaltliche Gliederung (anhand des Layouts und der Positionen von Links) sowie die Ähnlichkeit zu einem Zielartikel eines Navigationsszenarios. Das Ergebnis zeigt, dass die inhaltliche Gliederung von Artikeln der bestimmende Einflussfaktor für Linkauswahl darstellt. Den Hauptbeitrag dieser Dissertation stellt eine neuartige Evaluierungsmethode dar, die auf dezentraler Suche aufbaut und Navigation als einen greedy Algorithmus basierend auf lokalem Wissen modelliert. Die vorgestellte Methode ist in der Lage, automatisch den Zustand der Navigierbarkeit eines Systems zu evaluieren und gewährt Einblicke in die Navigationsdynamiken sowie die navigationale Performance. Aufbauend auf der Analyse der Einflussfaktoren sowie der Evaluierungsmethode werden im letzten Teil dieser Dissertation praxisorientierte Vorschläge für die Verbesserung von Navigierbarkeit vorgestellt. Für Empfehlungssysteme zeigt sich, dass durch das Ersetzen nur weniger Empfehlungen durch diversifizierte Empfehlungen die Navigierbarkeit deutlich verbessert werden kann. Für Wikipedia kann durch das Verschieben ausgewählter Links in besser sichtbare Bereiche von Artikeln die Anzahl der gegenseitig durch Navigation erreichbaren Artikeln wesentlich erhöht werden. Beiden Ansätzen ist gemein, dass die Linkziele relevant bleiben und gleichzeitig die Anzahl der Links auf einer Seite konstant gehalten wird. Die Evaluierungsmethode und die empiri-

schen Ergebnisse dieser Dissertation stellen ein leistungsfähiges Instrument dar, um den Zustand von Navigierbarkeit in einem Informationsnetzwerk zu untersuchen, und sind für die Wissenschaft und die Praxis gleichsam bedeutend. Durch die automatische Evaluierung der Auswirkungen von Änderungen eines Netzwerks können Websitebetreibende effizient Verbesserungsmöglichkeiten für Navigation untersuchen und damit Websites einem breiteren Publikum besser zugänglich machen.

Acknowledgments

The verb *navigate* stems from the Latin word for ship, *navis*, and originally stood for sailing and steering clear of obstacles at sea. The voyage towards this PhD thesis would not have been possible without navigational advice from numerous people, and they deserve my gratitude.

First and foremost, I am grateful to my supervisor Markus Strohmaier. Markus taught me the bits and bytes of doing research, and his inspiration, patience, and feedback always encouraged me in my work. Markus does a great job of keeping the big picture in mind, coming up with new ideas and finding the right pitch, even though his passion for soccer was constantly underappreciated in our Skype meetings.

In the same way, I am obliged to Denis Helic, whose mathematical skills, infallible memory, and creativity in finding solutions to seemingly impossible problems helped me overcome many obstacles. Denis always had an open door to delve into the nitty-gritty of mathematics, algorithms, or Star Wars, and our meetings and discussions were always fruitful.

I would like to thank Kristina Lerman, whom I had the pleasure of working with during an internship at ISI in Marina del Rey. Kristina has a very sharp mind, and our collaboration showed me a new perspective of doing research through her support and our many interesting discussions.

The journey towards my PhD was also guided by our research group at TU Graz, and I am thankful for having been a part of it. I am obliged to all members of our group for listening to my presentations, providing feedback and motivation, and setting a creative atmosphere. Special gratitude must go to Florian Geigl for teaching me about pandas, machine learning, working out, nutrition and Arnold Schwarzenegger, for his camaraderie, and his overall positive attitude. I would like to thank Simon Walk for always providing fruitful and constructive feedback, having an open mind for new ideas, keeping up our spirits and for providing us with cat pictures and novel board games, Patrick Kasper for his support with wikigames, panda gifs, hiding Easter eggs in the office, and teaching us proper British English, and Thomas Wurmitzer for being the go-to expert for Linux in particular, everything else in general and for never saying no when a

problem occurred. I am likewise grateful to all who have been members of our research group at various stages, namely Lukas Eberhard, Rainer Hofmann-Wellenhof, Tomas Karas, Florian Klien, Philipp Koncar, Tiago Teixeira dos Santos, and Massimo Vitiello. I am looking forward to more game nights playing board games and *Achtung, die Kurve!* crammed together on a single keyboard.

I would also like to thank the extended part of our group at GESIS in Cologne: Dimitar Dimitrov for his enthusiasm in Wikipedia research, his good spirit and discussions over Glühwein, Lisa Posch for our conversations, coffee breaks and for proofreading this thesis, Claudia Wagner for always asking the right questions in research meetings and Philipp Singer for putting up with me suddenly squatting his office when he was still at TU Graz, without which I would likely not have started my PhD studies.

I am very grateful for the group of close friends I have in Graz, both the ones I've known for close to twenty years and all the additions to our group that we've picked up along the way. Thanks to Stephi, with whom I've shared most of my time during this work and without whose support I would have not made it to this point.

More than to anyone else, I would like to send my gratitude and appreciation to my parents Erhard and Romana and my sister Stefanie, who have always supported and encouraged me, and without whom I would not be who I am today.

Institutional Acknowledgments. This thesis was partially funded by the Austrian Science Fund (FWF) under grant P 24866 (Navigability of Decentralized Information Networks). I am grateful to Graz University of Technology and the Institute for Interactive Systems and Data Science (formerly the Knowledge Technologies Institute) for giving me the opportunity to write this thesis and for letting me freely pursue my ideas, and to the Information Sciences Institute of the University of Southern California for my internship.

Contents

1 Introduction

1.1 Motivation

In the 1960s, Stanley Milgram and Jeffrey Travers were faculty members of the psychology department at Harvard, researching social networks. Specifically, they were interested in the connectivity of large social networks, such as the one spanning the entire United States. However, long before the emergence of online social networks or even personal computers, studying large social networks posed a difficult task. Not being able to conduct a study by conventional means, Milgram and Travers devised a plan to probe social networks by means of letters and conducted a series of *small-world experiments* [Milgram, 1967; Travers and Milgram, 1969].

The experiments aimed at forwarding a letter to a target person in the Boston area of Massachusetts, in geographical proximity to Harvard. Milgram and Travers randomly selected participants in Wichita, Kansas and Omaha, Nebraska—two cities that embodied a great distance to Massachusetts both geographically and socially. These participants acted as starting persons and received a letter with a few bits of information about the target person (viz. his name, address, and profession). The letter also contained instructions that requested all starting persons to forward the letter to the target person in case they knew him personally. Otherwise, they were asked to forward the letter to someone among their friends and acquaintances who they thought would bring the letter closest to the target person. Anyone receiving such a letter was then asked to repeat the same procedure.

After only a few days, letters started to arrive to the target person. As a result, chains of letters emerged, spanning the (at the time) 200 million inhabitants of the United States. Of the successful chains, the length averaged at 5.2 intermediaries—a surprisingly small number, given the enormous dimension of the network. These experiments established for the first time that entities in the social network of the United States could be connected by surprisingly short chains even across large cultural and geographic distances, and that humans were able to identify these chains.

Today, more than half a century later, rapid advances in information technology have radically changed the virtues of research on large networks. Powerful computing technology has enabled the analysis of networks on different orders of magnitude, making entirely new research questions accessible. In addition to social networks, even larger information networks have emerged, connecting millions of pieces of information. The vast majority of information is now stored digitally, and the pervasiveness of computers and sensors has led to an extreme growth of the rate of information production. In 2010, Eric Schmidt (then-CEO of Google) estimated that at the current rate it took just two days to produce the same amount of information as from the dawn of humanity until the present [Siegler, 2010].

This deluge of information has significantly facilitated information retrieval in many cases, such as for quick fact checking. However, information needs to be interpreted in a context, and being able to fulfill complex information needs has become an essential skill for most areas of life. The advent of the Web has enabled the rise of large-scale information networks, and the ability to successfully retrieve information in them has become crucial. Full-text search engines have made retrieval easy for large corpora of documents, but while searching is readily accessible, navigating to retrieve information remains more challenging. Navigation is required for a large variety of use cases, especially when an information need cannot be explicitly expressed in words. When searching on sites such as YouTube, which possesses only little textual information about videos, navigating based on recommendations (which rely on rating data to establish connections) becomes vital.

This thesis investigates *navigability*, which is the ability of a networked information system to support information retrieval by means of following hyperlinks. This process works in much the same as letter forwarding in the small-world experiments by Milgram and Travers. However, instead of the letter being moved between humans to get to a target location, in information systems the human participant generally remains the same and the location (e.g., the webpage) changes. Support for navigability can be measured (i) statically by looking at network properties (such as path lengths or clustering) or (ii) dynamically (by simulating navigation scenarios). This thesis studies both aspects, investigates influencing factors, and suggests methods to improve navigability (and thereby accessibility) in information networks.

1.2 Information Seeking in Networked Information Systems

Fundamentally, information seeking in networked information system can be described by two strategies, namely *navigation* (also known *browsing*) and *search* [Hearst, 2009, Chapter 3.5.3]. The first hypertext information systems in the 1960s introduced cross-references between documents with hyperlinks, clickable pieces of text that redirected the interface to the referenced document. Navigation is the strategy of retrieving information in an information network by following these hyperlinks. This strategy takes advantage of the human ability to recognize a piece of information without having the need to explicitly describe it, as well as the ability to skim large volumes of data and quickly identify desired pieces of information [Hertzum and Frøkjær, 1996].

Later generations of hypertext systems commonly included a central index tracking all documents. This index was commonly made available to users of the system in the form of a query-based interface, that allowed the immediate retrieval of all documents relevant to an entered query term, usually ranked by their relevance. This enabled search as the second information seeking strategy. Notably, the WWW does not have a central index and therefore does not include a built-in possibility to search. This lack of functionality was, however, later amended by search engine providers which indexed the Web with automated crawlers and then made the results accessible through external search engines.

The retrieval of specified pieces of information enabled by search engines can be quicker and more direct than retrieval via navigation. However, search comes at the cost of requiring knowledge of the search space and increasing memory load for users, who have to explicitly specify their information need [Budiu, 2014]. Generally, humans find it easier to progress in an information retrieval process in small steps, allowing them to view information in context and being able to recognize concepts instead of expressing them [Hearst, 2009, Chapter 3.5.3].

Information seeking can be further classified based on three motives [Toms, 2000]:

(i) retrieval of familiar items (e.g., *"Show me information about Graz!"*),

(ii) retrieval of items that cannot be explicitly described but will be recognized on sight (e.g., *"What's the name of that restaurant we went to last week?"*), and

(iii) serendipitous discovery (e.g., *"What's a good movie I could watch?"*).

Information seeking is usually supported by networked information systems by multiple techniques. Retrieval of familiar items (i) is for the most part covered by a full-text search function that permits the immediate retrieval of pieces of information specified by key words. The retrieval of items that cannot be explicitly described but will be recognized on sight (ii) may also be supported by a search function (e.g., users could search for semantically similar terms), but is supported primarily by navigation (e.g., browsing through restaurants in a given category to find a specific one). Serendipitous discovery (iii) requires results that are novel and diverse, and is therefore more difficult to support by a search function. It is hence better supported by relating documents to similar documents via links, such as by cross-referencing documents, by organization into categories (e.g., based on cuisines or genres), or by recommendations. All of these require following links (i.e., navigation) to retrieve information. The following section describes navigation in more detail and provides examples.

1.3 Navigation in Networked Information Systems

Navigation is the information seeking strategy consisting of following links to retrieve pieces of information. Navigation generally demands less cognitive capacity than explicitly expressing search terms for query-based retrieval and therefore places less burden on users. Knowledge gained during the navigation process puts information into context and helps with learning and decision-making [Resnick and Varian, 1997; Marchionini, 2006]. For reason such as these, some users prefer navigation to search even if they know what they are looking for [Teevan et al., 2004].

Navigation is also critical for discovering novel content. In systems such as Stack Overflow, Quora, Netflix, or YouTube [Davidson et al., 2010], where no clear structuring of items exists, users extensively rely on navigation to explore items. Flickr users predominately discover new images through *social browsing* [Lerman and Jones, 2007].

The links in information systems are frequently generated dynamically, which in turn affects the navigability of the system. The following subsections introduce navigation in recommender systems, where links are generated by recommendation algorithms, and Wikipedia, where links are placed at the discretion of the community of editors.

1.3.1 Navigation in Recommender Systems

Recommender systems help users retrieve information by providing recommendations of related, supplementary, or novel items. As such, they help to filter information. This becomes especially vital in large networked information systems, where users can easily lose track of the structure of the underlying information space. In the style of sales clerks who react to their customers' needs and give advice, recommender systems also adapt their output based on the recorded preferences of individual users.

One of the most prominent algorithms for providing recommendations is based on ratings that users assign to items, and is referred to as collaborative filtering. This algorithm recommends unrated items that other users with similar tastes have rated favorably. Other approaches aim to identify

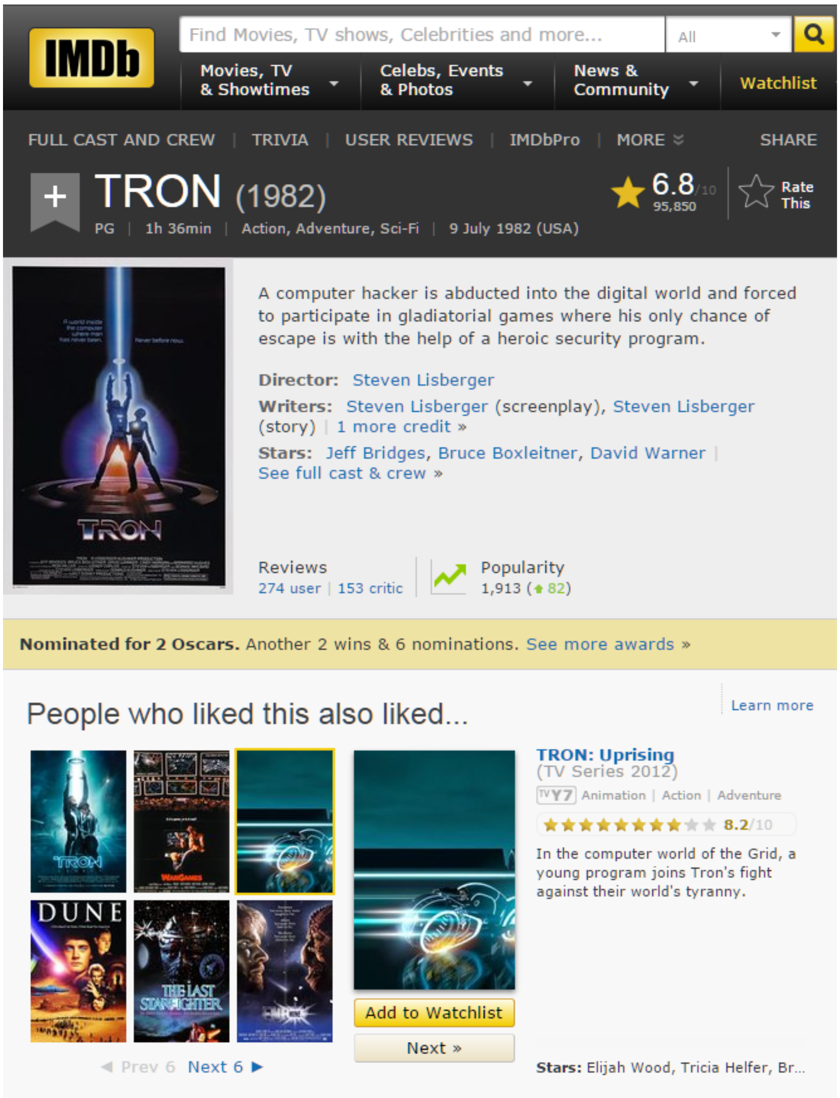


Figure 1.1: **Example of Movie Recommendations.** The website of IMDb.com features recommendations for a large number of titles. In the section entitled *People who liked this also liked...*, this subpage associated with the title *Tron* recommends six related titles (*Dune*, *Wargames*...). Six more recommendations are available by clicking a link at the bottom of the section.

items to recommend based on content features, knowledge bases, or use a hybrid approach.

Many recommender systems produce a list of recommendations to be presented with items of a collection. For example, IMDb.com associates a list of items as *People who liked this also liked...* (cf. Figure 1.1). As such, the items and their associated recommendations form a *recommendation network*, and example of which is shown by Figure 1.2. Recommendations are, in general, not symmetric. For example, a popular movie might be recommended by many other movies, but can only reference a small number of other movies itself. The links in the recommendation network are therefore directed.

By connecting items, recommendations help users to discover new content [Lerman and Jones, 2007; Davidson et al., 2010]. The knowledge gained along the navigation path helps to understand items in their context and aids with decision-making [Resnick and Varian, 1997; Marchionini, 2006]. Since the choice of recommendation algorithm and its parameters can be set and tuned by website operators, recommendations constitute a prime example of the how navigability in a system can be dynamically shaped by algorithms.

The development of recommendation algorithms initially focused mainly on the accuracy of predictions [Shani and Gunawardana, 2011]. However, this focus may lead to a bias towards popular items or a filter bubble effect [Celma and Herrera, 2008; Nguyen et al., 2014]. More recent recommendation approaches therefore also evaluate other properties such as diversity, novelty or coverage. However, the support of a recommendation algorithm for navigation cannot be evaluated by any of the methods proposed so far.

1.3.2 Navigation on Wikipedia

Wikipedia[1] was established in 2001 based on the idea of an online encyclopedia that anyone could edit. It quickly rose to be the most popular website of its kind by a large margin. As of 2017, Wikipedia ranks as the

[1] `http://www.wikipedia.org`

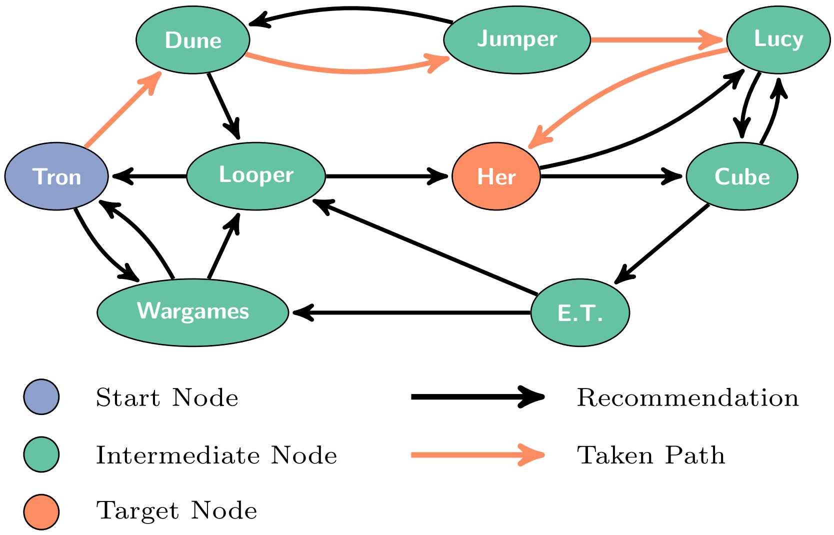


Figure 1.2: **Example of Navigation in a Movie Recommendation Network.** A website that associates recommendations with every item in its collection implicitly forms a *recommendation network*, that is, a view of the network where items are nodes and recommendations are directed edges between nodes. Generally, an item points to a small and constant number of recommended items. This leads to a network with fixed outdegree and variable indegree. The network in this figure shows an example with two recommendations pointing away from every item, and a possible navigation path from *Tron* to *Her*.

5th most-popular website worldwide [Alexa, 2017] and is available in over 250 languages, all while being maintained by volunteer contributors.

Since its inception, the Wikipedia community has established a comprehensive set of policies governing the content of its articles as well as the conduct between volunteers. Wikipedia articles need to follow a definite structure with a lead section and a table of contents, and should be segmented into sections. Many articles also contain an infobox, a tabular summary of the key facts of the article. Figure 1.3 shows an example of a Wikipedia article. Facts present in a recurring layout (such as in

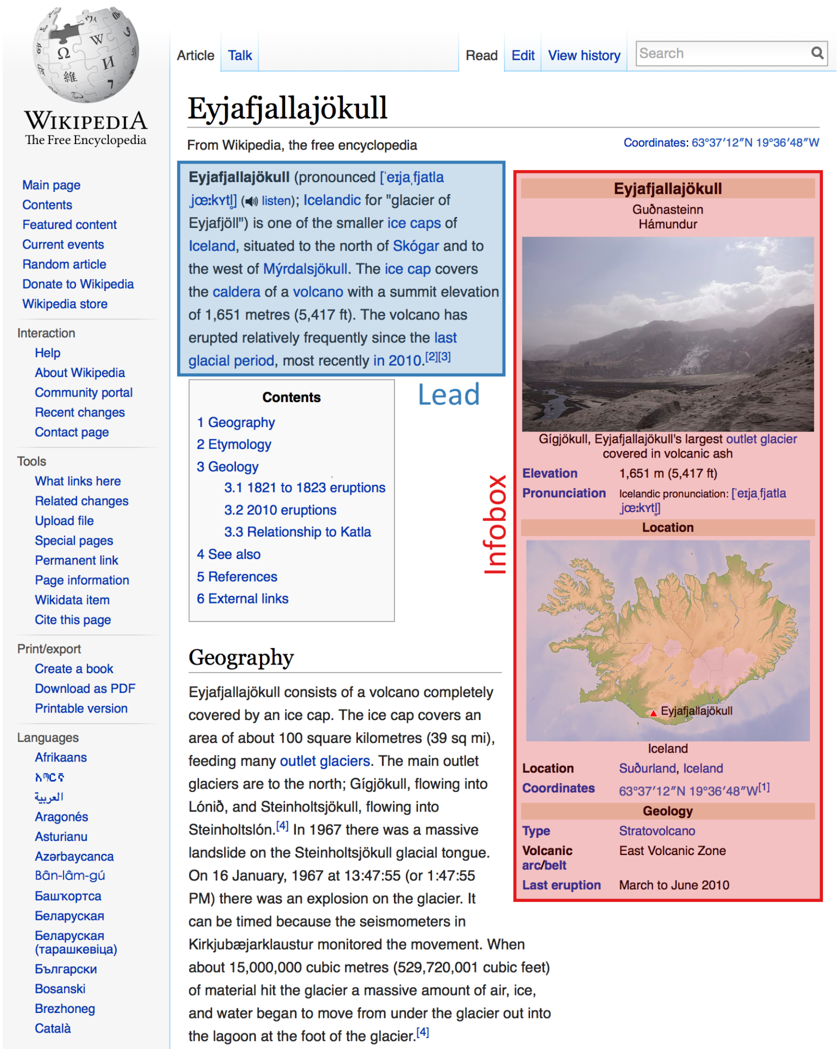


Figure 1.3: **Example of a Wikipedia Article.** The links in Wikipedia articles are introduced and maintained by the community of volunteer editors. The figure shows the article for the Icelandic volcano *Eyjafjallajökull*, and points out two sections, namely the *lead* (highlighted in blue) and the *infobox* (highlighted in red). These sections receive a large share of the user attention during navigation.

infoboxes) are regularly automatically extracted and made available in a machine-readable format, as exercised by Wikidata[2].

The articles and the links of the English Wikipedia span a vast information network, consisting of over five million articles, with around 900 new articles being created every day [Wikipedia, 2017d]. Most of the articles are mutually reachable by following links [Kamps and Koolen, 2009].

Data from a 2015 clickstream [Wulczyn and Taraborelli, 2015] shows that 70% of all article views are referred to from websites external to Wikipedia, many among which search engines. This indicates that many users are able to immediately satisfy their information needs on the encyclopedia. However, there also exists a large variety of use cases for navigating Wikipedia that make up the remainder of 30%. These include, for example, researching a topic, reading articles to pass time, or retrieving something that is on the tip of one's tongue.

1.4 Problem Statement, Objectives and General Approach

Problem Statement. Being able to find information in complex and evolving information networks has become a critical part of modern-day information retrieval skills. However, information is not always readily accessible, for reasons such as a confusing interface or required knowledge about a site. To increase navigability of a website, operators need to have a comprehensive understanding of user behavior and some method to automatically evaluate the effects of adaptations.

Navigating information networks has been the key method for retrieving information since the very beginning of the Web and allows for both goal-directed and unfocused information seeking. Present-time information networks connect large collections of items that before the digital era were only loosely tied together by means of cross references or footnotes. The topology of an information network, (i.e., the arrangement of its nodes and links) is generally not static and predefined, but can be adapted as needed

[2] http://www.wikidata.org

by website operators in line with the use cases of a site. For example, when a website is required to enable customers to find products efficiently, it will require an adequate link topology to this end, such as categories and recommendations. A website that is used to generate revenue from users remaining on the site, however, (such as YouTube), needs to set its links in accordance to that objective.

For Wikipedia, links are added and maintained by the community of volunteer editors with the objective of helping users to satisfy their information needs as efficiently as possible. All links need to adhere to the rules of the style guide and the linking policies [Wikipedia, 2017b]. Most importantly, links in an article are limited to target articles that are helpful to better understand the source article and provide background information.

On websites featuring recommendation systems, the link structure is generally even more dynamic: a website operator may apply any recommendation algorithm that fits the needs of their site, and may also further fine-tune the parameters of the algorithm. Recommendation systems can help websites to fulfill a variety of goals: help to filter an information space to identify the right products, cross-sell products (e.g., recommend additional accessories when a customer buys a laptop) or keep users engaged and entertained to retain them on the site for longer periods (e.g., on an entertainment website).

The dynamic adaptability of an information network's topology means that an understanding of the needs of users and the influences on the navigational process can be directly translated into actionable advice for websites to enable more efficient navigational structures. A better understanding of user behavior would allow the automated modeling of the effects of adaptations to an information network and allow to quickly decide what actions to take.

Objectives. The proposal of adaptations to improve navigability requires both a comprehensive understanding of the process of navigation and a method for quickly evaluating the ramifications of changes. The objectives of this thesis are hence threefold. Firstly, this thesis aims at identifying influencing factors in Web navigation, such as the generality of a document's topic, the semantic similarity, or the influences of any cognitive biases (such as a preference to look at the top of pages). Secondly, this thesis aims to model user navigation based on three established information seeking scenarios with decentralized search as a navigation method, and to provide this framework for automated simulations of navigation. Thirdly and lastly, this thesis has the objective of presenting improvements to navigability that website operators can address to make their websites more navigable in line with their objectives.

Overall, this thesis aims to combine the three above steps into an approach that allows to execute them in a coherent manner. The final result is a comprehensive framework for evaluating the support for navigation, which can be used to automatically assess the effects of adaptations to the structure of an information network and allows for quick and easy testing of modifications. This framework can then form the basis to present specific recommendations to improve navigability for users of a system.

General Approach. To build a coherent evaluation framework for navigability in information networks, this thesis first investigates influences on navigational decisions. To this end, several influences are modeled with probabilistic click models and then compared to the ground truth click distribution collected from user data. This method is applied to click data from a Wikipedia clickstream (containing aggregated click data) and wikigames (containing goal-directed navigation).

Next, this thesis evaluates navigability in an information network based on its topological characteristics, namely components, the bow-tie model and path lengths. This analysis is conducted on the eight largest Wikipedia versions and the recommender systems of IMDb. Building on these results, navigability is then modeled with the help of decentralized search and used to evaluate the navigation dynamics in a network. To this end, several different information seeking scenarios are automatically evaluated by means of simulations. This approach is applied to recommendations generated with three rating datasets as well as the recommender systems of IMDb.

Finally, the results from these analyses are translated into advice to render information networks more navigable. The results of the evaluations of navigability for recommender systems suggest that diversification can be an effective means to reduce path lengths and increase the success rates in information seeking. Moreover, this thesis presents a method to suggest specific links on Wikipedia that editors could move into more visible sections of articles to better connect Wikipedia's information network.

1.5 Main Publications

This cumulative thesis consists of the following publications:

- **Article 1 [Lamprecht et al., 2016b]:** Lamprecht, D., Lerman, K., Helic, D. and Strohmaier, M. (2016). How the Structure of Wikipedia Articles Influences User Navigation. *New Review of Hypermedia and Multimedia*, 23(1):29–50.

- **Article 2 [Lamprecht et al., 2016c]:** Lamprecht, D., Strohmaier, M. and Helic, D. (2016). A Method for Evaluating the Navigability of Recommendation Algorithms. In *Proceedings of the 5th International Workshop on Complex Networks and their Applications*, pp. 247—-259.

- **Article 3 [Lamprecht et al., 2015a]:** Lamprecht, D., Geigl, F., Karas, T., Walk, S., Helic, D. and Strohmaier, M. (2015). Improving Recommender System Navigability Through Diversification: A Case Study of IMDb. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pp. 21:1–21:8.

- **Article 4 [Lamprecht et al., 2016a]:** Lamprecht, D., Dimitrov, D., Helic, D. and Strohmaier, M. (2016). Evaluating and Improving Navigability of Wikipedia: A Comparative Study of Eight Language Editions. In *Proceedings of the 12th International Symposium on Open Collaboration*, pp. 17:1–17:10.

1.6 Further Publications

- **Article 5 [Geigl et al., 2015]:** Geigl, F., Lamprecht, D., Hofmann-Wellenhof, R., Walk, S., Strohmaier, M. and Helic, D. (2015). Random Surfers on a Web Encyclopedia. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pp. 5:1–5:8.
- **Article 6 [Lamprecht et al., 2015b]:** Lamprecht, D., Helic, D. and Strohmaier, M. (2015). Quo Vadis? On the Effects of Wikipedia's Policies on Navigation In *Proceedings of the Workshop on Wikipedia, a Social Pedia at the Ninth International AAAI Conference on Web and Social Media*, pp. 64–66.

1.7 Research Questions

This thesis strives to investigate the state of navigability on the Web and to identify approaches to improve it, and does so in three steps. The first step to this end is to understand the factors influencing humans in the process of navigating webpages. The second step investigates how to model human navigation behavior with simulations based on a framework using decentralized search. Building on the conclusions of these steps, the third step then aims at suggesting adaptations to facilitate navigation for users. The following introduces the three research questions that this thesis investigates. Figure 1.4 shows the relations between the research questions and the publications associated with them.

RQ1: What factors influence the navigational choices of users?

Problem. To be able to suggest improvements to better support users in Web navigation, it is important to first identify the factors influencing the navigational process. Information retrieval is generally modeled as an adaptive and dynamic process, for example by Berrypicking [Bates, 1989], Orienteering [O'Day and Jeffries, 1993], or Information Foraging [Pirolli,

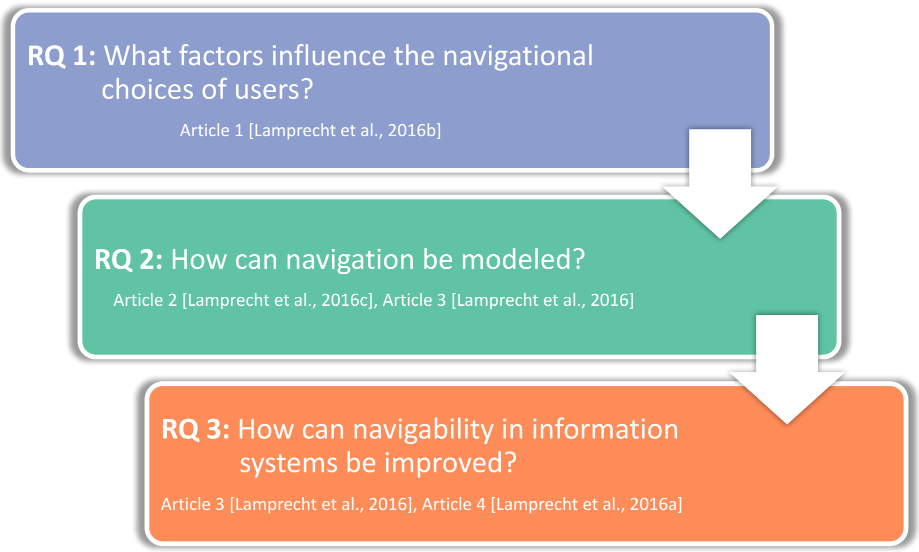


Figure 1.4: **Structure of the Research Questions of this Thesis.** This thesis strives to answer three research questions. Research question 1 studies the influences on navigational choices of users, such as page organization or the generality of a topic. The findings of this question form the foundation for research question 2, which investigates how navigation can be modeled to automatically evaluate the ramifications of changes to the navigational interface. Based on this, research question 3 investigates possible improvements to navigability in networked information systems.

2007]. In these models, users constantly adapt their strategies based on their expectations, intuitions and the current web page.

Generally, humans are subject to cognitive biases such as a confirmation bias, or assigning higher values to items located at the top of a list [Lerman and Hogg, 2014]. When analyzing factors influencing humans in their navigational decisions, it is therefore important to take potential biases into account.

An important influencing factor in navigation is the organization of a webpage. Overall, users scan webpages in an F-shaped pattern [Nielsen, 2006] and focus much of their attention on the areas at the top and on the left [Buscher et al., 2009]. This indicates that users have become used to recurring regularities in webpage organization, such as navigation bars being located at the top. For large websites with a rigidly regular structure, such as Facebook or Wikipedia, users will likely have prior expectation about where to find content and links.

Another influencing factor is the semantic similarity to the navigation target in the case of goal-directed navigation. In wikigames, navigational games played on Wikipedia, users have been shown to be drawn to articles with a large number of incoming links in the first steps (an indicator of the generality of an article) and to the semantic similarity to the target article in the final clicks [West and Leskovec, 2012b].

A better understanding of the influences in navigation would help website operators to make their content more suitable to navigation as an information retrieval method.

Approach. This thesis compares the impact of influencing factors by evaluating click distributions based on a range of influences, such as the generality of a link target's topic or the position of the link text on screen. To do so, the influences are modeled by click probabilities. A model assigns a click probability to every link based on an influencing factor, and the aggregate of all click probabilities is then evaluated to yield a click distribution among all links. This approach is used to evaluate and compare the influence of a range of factors in Wikipedia navigation, namely structure-based influences (i.e., the influence of links located in specific sections of Wikipedia articles), semantic similarity of link targets, and the generality of articles. The resulting distributions are compared to the ground-truth click distribution of user clicks with the Kullback-Leibler divergence, which measures the difference between two probability distributions based on entropy.

Contributions. Overall, the results show that users dedicate substantially more attention to areas located towards the top and left of articles (cf. Section 3.2 and Publication 1 [Lamprecht et al., 2016b]). These biases are so substantial that they are in fact able to explain navigational choices better than the other investigated influences for most cases, and they hold for both free-form and goal-directed navigation on Wikipedia. The lead and infobox sections, located at the beginning of articles, are of particular relevance. For goal-directed navigation, the influence of the semantic similarity between a potential link and the navigation target is also of considerable importance. Overall, the generality of articles matters more in the first steps of navigation, and the semantic similarity becomes important towards the end, as users home in towards the navigation target. This confirms previous evidence [Juvina and van Oostendorp, 2008; West and Leskovec, 2012b] that both semantic and structural knowledge is vital for navigation. However, the remainder of steps is best explained by the influence of page organization. These findings enable us to better characterize the navigation behavior of users based on structural influences and open up ways to address page organization as a means to better support navigability.

RQ2: How can navigation be modeled?

Problem. The question of what facilitates navigation in an information network has received a great deal of attention. Research has shown that topological characteristics, such as strong clustering and short average path lengths, are good predictors of the navigability of a network [Watts and Strogatz, 1998; Boguñá et al., 2009]. These properties indicate requirements for efficient navigation, and ensure that a network is efficiently navigable by a decentralized search algorithm [Kleinberg, 2000a], given certain conditions are met.

Likewise, research on information retrieval has produced a wide spectrum of models explaining choices and reasoning during the navigation process, such as berrypicking [Bates, 1989], information foraging [Pirolli and Card, 1999] and orienteering [Teevan et al., 2004].

To enable and support navigation, a variety of techniques exists, and a wide range of interfaces offer navigational support for users. An effective method to expose the actual navigation dynamics on a network is to conduct a user study probing the usability and effectiveness of the navigational aids. However, involving humans as test subjects is a very expensive method, rendering it infeasible in all but a very limited number of cases.

A method to obtain information about the state of navigability in a system quickly and automatically, would allow for a speedy evaluation of adaptations. Knowing more about the effects of changes to the network structure or the interface on the navigational choices of users would enable website operators to adjust interfaces to better support users for a wider variety of information retrieval scenarios.

Approach. To be able to model navigation in a networked information system, this thesis presents a simulation framework. This framework investigates and assesses practical navigability by conducting simulations of information retrieval based on three different information seeking scenarios.

These simulations are based on decentralized search, which models navigation as a graph search problem. Decentralized search is a greedy search algorithm operating only on local knowledge. At every step during the navigation process, the algorithm decides to follow one of the locally available links (i.e., it selects one of the links leading away from the current page). The decision of what link to follow is based on some notion of intuition (or background knowledge). For example, links could be selected based on how similar the description of the link target is to the navigation goal. In a typical information seeking setting, users navigate in much the same way:

without global knowledge of the network, users select links among the ones available at a page based on their intuitions and contextual information such as images or text.

The framework presented by this thesis makes use of decentralized search to model navigation based on three different information seeking scenarios. Firstly, navigation is modeled based on point-to-point search. This is the original scenario proposed by Watts, Kleinberg and others [Watts and Strogatz, 1998; Kleinberg, 2000a] and represents a focused, goal-directed navigation scenario aimed at finding a single target page. Secondly, the framework models navigation as a dynamic, adaptive process, based on the berrypicking model proposed by Marcia J. Bates [Bates, 1989]. Berrypicking models information seeking like picking berries scattered on bushes, picking which requires constant adjustment of the retrieval strategy. Thirdly, the framework models navigation based on information foraging [Pirolli and Card, 1999], which takes from research on optimal foraging theory in nature and translates it to human information seeking. Information foraging describes foraging of patches of information, and the framework models the retrieval of multiple items from a patch.

The core method of the presented framework, namely decentralized search, has been established as a method to simulate navigation and leads to characteristics comparable to human navigation behavior [Trattner et al., 2012; Lamprecht et al., 2015c].

Contributions. The result of this approach is a framework that is generally applicable to evaluating the navigability of an arbitrary networked information system. The framework extends traditional, one-hop-based evaluation measures of networks, such as clustering or degree, towards a dynamic, path-based evaluation (cf. Section 3.3 and Publication 2 [Lamprecht et al., 2016c]). The evaluation method is demonstrated on top-N recommender systems and their implicitly associated recommendation networks.

The results show that more complex recommendation algorithms, such as matrix factorization, can lead to better navigability in a recommender system. Moreover, navigability substantially improves when increasing the number of recommendations present. A further study on the recommendation networks of IMDb corroborates this finding. Both recommender systems of IMDb apply comparatively simple recommendation approaches, and consequently the evaluation shows that the resulting networks are poorly navigable (cf. Section 3.4 and Publication 3 [Lamprecht et al., 2015a]).

The presented evaluation framework supplies website operators with a powerful tool to automatically evaluate the navigational ramifications of changes to their website structures and allows to quickly gather facts for decision-making.

RQ3: How can navigability in information systems be improved?

Problem. Users of networked information systems often rely on links to connect pieces of information and access knowledge. By making sure that their website is readily accessible by navigation, website operators can help users to efficiently retrieve information, discover new content, or research a topic.

The simplest method to improve navigability in a system lies in the introduction of novel links, which directly increases the number of mutually reachable items. However, cognitive attention limits make it difficult for humans to process and make sense of a larger number of links [Bollen et al., 2010; Lerman, 2016]. For this reason, methods that leave the number of links untouched but instead change some of the link targets are often preferred in practice, as they do not change the overall number of links present.

Rewiring a small share of links to random targets in an otherwise regular graph already suffices to substantially bring down the average path lengths in a network [Watts and Strogatz, 1998]. However, link targets selected at random are likely irrelevant to a page. Humans navigating a network generally do not possess global information about connectivity and are only familiar with a fraction of the network nodes [Dimitrov et al., 2015]. To be actually useful to humans navigating a network, rewired links therefore need to take the relevance of the link targets for the current page into account.

Approach. This thesis investigates two empirical approaches that increase navigability while leaving the overall number of links constant. The first approach targets recommender systems. Based on previous results [Lamprecht et al., 2016c], simpler recommendation algorithms, such as textbook collaborative filtering, may lead to less navigable recommendation networks. To overcome this, a more complex recommendation algorithm could be applied. However, the increase in computational complexity may render them infeasible for large systems. A simpler adaptation consists of the diversification of recommendations [Ziegler et al., 2005]. To this end, this thesis presents an approach to switch out a small number of recommendations for still relevant, yet more diverse recommendations.

The second approach investigates links on Wikipedia. When considering the entirety of all links present on Wikipedia's articles, the resulting article network is very well connected. However, when restricting the links to sections that receive the majority of user attention (especially the lead section), navigability plummets substantially. To address this issue, the presented approach selects single links from the remainder of the article to bring forth to the top. This method has the advantage of only selecting links that have already been shown to be useful to the article.

The two approaches are evaluated by measuring the size of the largest set of mutually reachable articles (i.e., the largest strongly connected component in the article network). The diversification approach for recommender systems is additionally evaluated in terms of path lengths and success rates in simulations of navigation with decentralized search.

Contributions. Diversifying some of the recommendations on the recommendation networks of IMDb leads to substantial improvements of navigability. The success rate, which measures the number of reached navigation targets per navigation session, shows an up to threefold increase when introducing diversified, yet relevant recommendations (cf. Section 3.3 and Publication 3 [Lamprecht et al., 2015a]). This corroborates the findings from Publication 2 [Lamprecht et al., 2016c] showing that more complex navigation algorithms, which generally include a higher diversification, can lead to better navigability in a recommendation network.

On Wikipedia, the links between articles enable a densely-connected article network for all of the encyclopedia's eight largest language versions, where close to 100% of all articles are mutually reachable. However, this does not hold when limiting the links to the ones that are actually used by visitors, as taking only the links in the lead section into account brings the number of mutually reachable articles in the article network down to around a third. As a remedy for this, links could be selected based on their navigational effects (cf. Section 3.5 and Publication 4 [Lamprecht et al., 2016a]). The application of this approach to Wikipedia shows that by recommending links in order of their navigational potential and introducing them to the lead, the share of mutually reachable articles can be increased. This method has the advantage of only selecting links relevant to the article, as they have been introduced by an editor at a previous point in time. The recommendation algorithm could be applied to suggest links and show their navigational potential, while still leaving it up to editors whether to take these considerations into account.

In summary, these findings give website operators tools to make their websites more easily accessible for users in a wide range of information-seeking scenarios. The proposed changes allow users to navigate networks that are better connected and make a larger number of documents accessible.

1.8 Contributions and Implications

In summary, this thesis makes both empirical and methodological contributions to the fields of network science, recommender systems and Wikipedia research. The major contributions are threefold.

- First, this thesis presents an empirical investigation of the influences on navigational choices of users during both goal-directed and free-form navigation. It presents a novel technique to model the effects of influencing factors and shows that page organization exerts the largest influence on link selection of all investigated factors.

- Second, this thesis contributes a method to model navigation with decentralized search based on three information retrieval models. The method is able to comprehensively evaluate navigability by conducting extensive simulations of navigation and assessing the success rates of navigation scenarios. The results of these simulations are able to provide insight into the practical navigation dynamics on arbitrary information networks.

- Third, this thesis provides empirical suggestions to improve navigability in networked information systems. For recommender systems, this thesis shows evidence that navigability can be improved by introducing diversified recommendations. By this approach, the number of links stays constant (avoiding cognitive overload of users), while the links are kept relevant. For Wikipedia, navigability can be improved by selecting links to bring forth to sections of an article that are more frequently viewed by users, such as the lead section.

The methods and empirical analyses presented by this thesis help to improve the navigation experience for users of networked information systems. The investigation of the factors influencing navigation assists in understanding the reasons behind clicks. Together with an automated analysis method of navigability, the results of this thesis help to test and implement navigational improvements for information systems across multiple information retrieval scenarios.

1.9 Structure of this Thesis

The remainder of this thesis is structured as follows.

Chapter 2 reviews related work. Section 2.1 introduces the field of navigability of networks, its history in social network analysis, navigation on the Web and navigation models. The model of decentralized search is especially relevant for this thesis and is discussed in Section 2.2. Section 2.3 discusses navigability in the context of recommender systems and Wikipedia.

Chapter 3 consists of the publications that form the main body of this cumulative thesis and answer the research questions posed in Section 1.7. Section 3.1 details the contributions to these publications by the author of this thesis and all collaborators.

Chapter 4 concludes this thesis. Section 4.1 summarizes the answers to the research questions, the results and the contributions, Section 4.2 states the implications of this work, Section 4.3 discusses the limitations, and Section 4.4 outlines avenues for future work.

2 Related Work

Information is rarely useful on its own: pieces of information are generally shaped by the surrounding context, and the connection to other pieces of information is often what makes them meaningful. Historically, information could be connected by a discussion of a topic in a piece of text, cross-references between books, or index cards in libraries catalogs. As technology advanced, ideas to facilitate connections emerged. In 1945, Vannevar Bush envisioned electronic aids to connect pieces of information stored on microfilms [Bush, 1945]. Bush anticipated an information storage unit named *Memex* that enabled indexing, searching and linking of content.

The 1960s saw the emergence of the concept of *hypertext*—text that includes *hyperlinks*, clickable elements that cross-reference other documents and make them immediately accessible. Hypertext therefore enables non-sequential reading of text. More broadly speaking, the term *hypermedia* is a generalization of hypertext to include other types of documents such as images, sound clips, or videos. Both terms were coined by Ted Nelson to describe a file structure for complex and changing pieces of information. Nelson developed *Project Xanadu*, one of the first hypertext system [Nelson, 1965], which included advanced features such as bidirectional hyperlinks and access rights management.

The years and decades following the 1960s saw the emergence of a large number of hypertext and hypermedia systems, such as the Hypertext Editing System at Brown University [Carmody et al., 1969], Hypercard by Apple, Notecards at Xerox PARC [Halasz et al., 1987], Microcosm at the University of Southampton [Fountain et al., 1990], and the NLS by Douglas Engelbart at Stanford [Engelbart, 1988].

All of these systems shared the concept of some kind of associative text (i.e., hypertext) to connect related documents. However, the systems differed in their implementation details for the display, creation and editing of hypertext, were mostly incompatible, and sometimes required extensive training to operate. Most systems were research projects, and while all embraced the idea of hypertext, most branched off into a direction that added novel and unique functionality. The increasing popularity of these systems led to the creation of the ACM Hypertext conference, held for the first time in 1987 at the University of North Carolina at Chapel Hill [Nielsen, 1988]. The conference quickly emerged as the main venue to discuss hypertext-related research. Most of the ideas implemented by early hypertext systems anticipated the functionality of today's WWW. However, due to their complexity and incompatibility they were predominantly used by a core group of dedicated users.

In the late 1980s, Peter Halasz and his collaborators worked on a hypertext system called Notecards [Halasz et al., 1987] at Xerox PARC. Based on the design and usage of the system, Halasz reflected on its shortcomings and described seven critical issues that the next generation of hypermedia systems needed to address [Halasz, 1988]. These issues comprised the need for full-text search functions, composites of nodes (i.e., inclusion of separate and independent parts), dynamic computation at retrieval time and the need for simple extensibility to fit custom requirements.

The next major step in the history of hypertext took place at CERN in Geneva, where high-energy physics researchers routinely worked with large quantities of data and increasingly faced the difficulty of how to share their data and analyses with their collaborators. To address this issue, Tim Berners-Lee proposed a hypertext system to exchange information among CERN researchers [Berners-Lee and Cailliau, 1990]. The system soon became immensely popular at the institution, and Berners-Lee and his collaborator Robert Cailliau soon developed the first widely available server and browser for the new system, which they named the *World Wide Web*, or WWW for short. Over time, it became to be known simply as the *Web*.

Berners-Lee and Cailliau proposed their work to the 1991 Hypertext conference. However, the conference committee was not impressed and would only allow for a poster and demo [Berners-Lee and Cailliau, 1990]. Their work was considered far too simplistic: links were only unidirectional (leading to dead links) and there was no central index or any of the other, more sophisticated features of the state-of-the-art hypertext systems at the time. Critically, none of the seven issues by Peter Halasz were addressed by WWW.

However, WWW quickly grew larger than any other hypertext system in terms of users and servers. The much-criticized simplicity turned out to be its major strength: usage of the system only required the installation of a browser, and adding content was as easy as publishing a document on a web server and required little effort. The 1993 release of Mosaic, a graphical web browser, further helped the WWW to explode in popularity [National Center for Supercomputing Applications (NCSA), 2017].

At the same time, several systems competed with the WWW. One example of this was Hyper-G, a system developed at Graz University of Technology [Andrews et al., 1996]. Hyper-G had addressed most of the shortcomings of other hypertext systems: it included bidirectional links, was designed for multi-lingual use, had access rights management, powerful search facilities, and a uniform user interface. However, Hyper-G required more effort to set up and to add content to, which in the end proved to be detrimental to its success.

WWW quickly left its competitors behind and ended up winning the hypertext race. However, other hypertext systems lived on as specialized systems, such as intranets or knowledge management systems. Hyper-G was rebranded as HyperWave and used internally by a number of large corporations [Maurer, 1996]. Some of the features of these systems still occur on systems on the Web today. An example for this is Wikipedia, which features forward and backward links as well as indications for dead links.

WWW, while simple in its original conception, was gradually extended with more complex concepts. In 2001, Peter Halasz revisited the seven issues from his 1988 paper [Halasz, 2001], and found that most of them had been addressed on the WWW in some way. However, some of them had not been built into the core of WWW but were rather added by external parties—a result of the easy extensibility of the WWW. A prime example of this are search engines: wile other hypertext systems included a central index, search on the WWW was enabled only by search engine providers that continuously crawled and indexed the Web. However, this fact relieved the burden of keeping an index from the infrastructure of the WWW and outsourced it to external parties, which made adding content to the Web easier. Other issues were addressed by the advancement of Web standards, such as the dynamic inclusion of content via HTML iframes or server-side technologies. Of the seven original issues, only the versioning of content never caught on, and was only introduced in specialized systems such as Wikipedia or HyperWave.

In just little over a decade, the advent of the Web has radically changed the way humanity interacts with information, and a vast amount of information is now accessible in networks of interconnected pieces. The invention of hypertext has led to a paradigm shift in terms of how text is read. In a very short time, the Web has become the largest and most complex system humanity has ever created. However, many aspects of the Web have not been researched or explored. Gaining a better understanding of the dynamics of human behavior on the Web would allow us to continue to improve the system and make it more easily accessible for a wider audience.

Being able to find and connect pieces of information in networks has become a critical skill in today's information society. This thesis investigates navigation as a method to retrieve information on the Web. Based on an analysis of the factors influencing navigational choices, it aims to propose a way to model navigation scenarios and show up ways to improve navigability.

This chapter is structured as follows. Section 2.1 gives an overview of the history of research on network navigability, introduces navigation on the Web and surveys web navigation models. Section 2.2 describes decentralized search, the web navigation model central to this thesis.

Finally, section 2.3 outlines the state-of-the art of research on navigation dynamics on recommender systems and Wikipedia.

2.1 Navigability of Networks

The term *navigability* refers to the degree to which a network is accessible by means of following edges, such as hyperlinks on the Web. Evaluating the navigability of a system reveals how accessible a system is in terms of using its links to retrieve information. The following subsection introduces small-world networks, the class of networks which started research on navigation in networks, and which can be found in many real-world networks. The second subsection surveys navigation on the Web, followed by a subsection on models of human navigation for the Web.

2.1.1 Small-World Networks

Research on the navigability of networks first became the focus of research attention by the small-world experiments of Milgram and Travers in the 1960s [Milgram, 1967; Travers and Milgram, 1969]. Conducted long before the era of the Web, these experiments probed the navigability of the social network of the entire United States by means of letter forwarding. Participants were given a description of a target person—a stockbroker working in Boston—and were asked to forward a letter to him if they knew him on a first-name basis. Otherwise, they were requested to forward it to an acquaintance that they thought would either know the target person or get the letter closer to him. The results of these experiments established for the first time that entities in a vast social network could be connected by very short chains of around six hops. The expression "six degrees of separation" has since entered the English language. The accuracy of the chain lengths has been subject of much debate, and criticism has targeted the methodology such as the selection of participants via newspaper advertisements [Erickson, 1979; Schnettler, 2009]. However, the experiments were later successfully repeated with e-mail forwarding [Dodds et al., 2003], and for the social graph of Facebook, where the average distance between users was determined to be 4.74 [Backstrom et al., 2012].

The insights from the small-world experiments spawned subsequent research on the network properties effecting efficient navigation. Watts and Strogatz proposed *small-world networks* as a class of network with high clustering and low characteristic path length [Watts and Strogatz, 1998; Watts et al., 2002]. They located these networks at an area in the spectrum from random graphs to regular lattices and proposed generative models based on rewiring to introduce long-range links into otherwise regular networks. This type of network has been identified in a large number of real-world cases, such as actor collaboration networks, power grids and the neural networks of C. elegans, a small roundworm. However, the proposed generative small-world models lack the power-law degree distributions usually found in real-world networks.

Subsequently, Jon Kleinberg showed that for the Watt-Strogatz small world networks no decentralized search algorithm operating on local knowledge only could exist that was effective, i.e., could find paths between arbitrary nodes in sublinear time [Kleinberg, 2000a,b]. Such an algorithm is termed *decentralized* because it forwards the search problem to another node presumably closer to the target and relies only on information about the current node's neighbors, as well as the grid structure underlying the network. Kleinberg then generalized the model of small-world networks and showed that an effective algorithm exists for a specific model that rewires links probabilistically: a link is rewired to a node at distance d with probability $d^{-\alpha}$, where α is the dimension of the underlying lattice. The probability of rewiring therefore follows a power-law based on distance in the grid, and this setting allows a decentralized search algorithm to reduce the distance to the target by an order of magnitude every few steps. The expected number of steps is then polylogarithmic in the number of nodes, which holds for exactly this class of small-world networks.

These findings were, however, limited by the fact that very few real-world networks are based on a regular lattice structure. For applications of these findings, researchers have attempted to use other means of informing a decentralized search algorithm about the network structure. In peer-to-peer networks, where little regularity is present, the assumption of an underlying lattice structure is violated in particular. For these networks, forwarding to high-degree nodes was shown to constitute an efficient search algorithm [Adamic et al., 2001]. In this manner, very little information about the underlying structure is required, which was exploited by decentralized file-sharing applications such as Gnutella, where search queries were forwarded by nodes in a decentralized manner until a target was found.

Another option of informing a decentralized search algorithm is by means of a hierarchy associated with a network. In the generative model proposed by Kleinberg [Kleinberg, 2001], the nodes of a network correspond to the leaf nodes of a hierarchy. To construct a network, these nodes are taken as the nodes of a separate network and connected with probability proportional to their distance in the hierarchy. The resulting network is then efficiently navigable with a decentralized search algorithm given the information about the hierarchy. For real-world applications, this means that if the nodes of a network can be classified by an associated hierarchy (e.g., a taxonomy), the network is likely to be efficiently navigable.

These findings were applied to the e-mail network of HP labs, where hierarchies based on organizational structure or physical office proximity worked best for search algorithms [Adamic and Adar, 2005]. For the Web, the hierarchy could be taken from a web directory [Menczer, 2002]. Later models showed that the hierarchy does not need be an exact tree, and that it suffices to provide information about characteristics of nodes [Watts et al., 2002; Lattanzi et al., 2011].

Kleinberg's notion of a hierarchy as the background knowledge to inform navigation was then extended to *hidden metric spaces*, which have been shown to explain the scale-invariance in degree distributions and the strong clustering of real-world networks [Serrano et al., 2008; Krioukov et al., 2010]. These findings have been exploited to design more efficient search algorithms in networks. In particular, mapping the Internet to a hyperbolic space instead of a geography-based space can improve navigation performance [Boguñá et al., 2010]. The hyperbolic metric space can be thought of as the hierarchy associated with a network in Kleinberg's terms.

When nodes are connected with probability proportional to their distance in the metric space, power-law degree distributions and strong clustering emerge naturally, and the metric space is able to explain the existence of effective search algorithms. When using geography as the hidden metric space for flights, there are distinct zoom-out and zoom-in phases in airport routing (when selecting the next airport greedily as the one closest to destination airport). These phases are enabled by hub nodes (for the zoom out) and clustering (for the zoom in) [Boguñá et al., 2009].

These findings show that efficient navigation is theoretically possible in most real-world networks, which usually satisfy the small-world property. Search with a decentralized search algorithm has been shown to likely lead to a target in a small number of steps, which further increases the likelihood that a real-world network is efficiently navigable—depending on the kind of background knowledge that an agent applying a decentralized search algorithm has at its disposal.

2.1.2 Navigation on the Web

Web navigation refers to the process of information retrieval by means of following links on webpages. The shape of the Web has been described as a bow-tie structure, with a 30–50% of websites mutually reachable in a *core* component, 20–30% of websites in an *IN* component from which the core is reachable, but not vice versa, and 6–20% of nodes in an OUT component, that is reachable from the core but from which no paths back to the core exist [Broder et al., 2000; Meusel et al., 2014]. The number of in-links and the component sizes follow a heavy-tail distribution. The length of a path between a pair nodes, given that it exists, is 16 hops—a surprisingly low number for a large information network such as the Web.

Generally, navigation is a more useful retrieval strategy than search if multiple pages along the navigation trail are of interests to the user and can help to make the context clearer [Marchionini, 2006]. This is the case when learning about a new topic step by step (such as on Wikipedia) or on entertainment sites such as YouTube, where all the pages along a path are potentially consumed by a visitor[Davidson et al., 2010]. The ability to retrieve information by a query to a search engine also comes at the cost of requiring knowledge of the search space and increasing memory load for users, who have to explicitly specify their information needs [Budiu, 2014]. Overall, humans find it easier to progress in an information retrieval process in small steps, allowing them to view information in context and being able to recognize concepts instead of expressing them [Hearst, 2009, Chapter 3.5.3]. Consequently, some users prefer navigation as an information-seeking strategy even when they know exactly what they are looking for [Teevan et al., 2004].

Eye tracking studies have shown that users inspect webpages in an F-shaped pattern [Nielsen, 2006] and focus on the area located near the top left corner [Buscher et al., 2009]. This region, also known as the *golden triangle*, receives around 45% of all clicks. Additionally, 76.5% of all clicks are made *above the fold*, that is, in the area visible without scrolling down [Weinreich et al., 2006]. Web users have been found to very quickly make a decision as to whether a page is relevant to them: around half of all page visits are shorter than ten seconds [Weinreich et al., 2006]. These findings indicate that many pages are only skimmed, and few are read in detail.

User decisions on webpages, such as where to click, are influenced by both semantic and structural knowledge [Juvina and van Oostendorp, 2008]. The semantic similarity between webpages affects navigational choices, and can be exploited to predict clicks [Kaur and Hornof, 2005]. Similarly, click paths by users on Wikipedia can be used to compute the similarity between concepts [West et al., 2009a].

The structural knowledge in terms of the organization of webpages also has a substantial influence on the interaction behavior of users. Many cognitive biases, such as the influence of presentation order [Salganik et al., 2006] occur on the Web [Lerman and Hogg, 2014]. Structural information about a network, such as degree or clustering coefficients can be used to inform a decentralized search algorithm, and only a small amount of this information suffices to efficiently guide the algorithm [Dimitrov et al., 2015].

The interaction of users with web search engines has been the subject of a great amount of research. Search user interfaces were initially complex and required the specification of several search fields and Boolean operators. As such, they tended to be used mostly by specialists, such as librarians. However, over time interfaces have become simpler and easier to use for a wider audience [Hearst, 2009, Chapter 1.1]. While today's search user interfaces still allow for advanced syntax such as Boolean operators and parentheses, these features are rarely used [Jansen et al., 2000]. Further techniques have been applied to reduce memory load on users, such as a search history and suggestions [Hearst, 2009, Chapter 1.7].

Queries to search engines can be classified into three categories: *navigational* ("Show me the website for Graz!"), *informational* ("What is Skyr?") or *transactional* ("Show me travel information for Iceland!"). Roughly half of all search engine queries on the Web are informational, a quarter is navigational and another quarter transactional [Broder, 2002]. Most queries tend to be simple, agreeing with the tendency towards simpler search user interfaces [Jansen et al., 2000].

Search engine result pages (or *SERPs* for short) usually consist of a vertical arrangement of search results, each with a brief summary of the corresponding webpage consisting of the title and an excerpt of its content. The results are commonly ranked by their relevance to the entered query. How often a result is clicked depends largely on its position on the SERP, as users frequently only inspect the first few results [Pan et al., 2007; Buscher et al., 2009; González-Caro and Marcos, 2011]. This is likely an artifact of the good perceived quality of ranking algorithms. Users often only skim the results at first to decide if they are worth looking at [Liu et al., 2014]. However, if few relevant hits are present among the top results, search engines users do explore the remainder of the results [Salmerón et al., 2013]. Click trails on search engine result pages can be used to improve ranking by viewing interactions as implicit endorsements [Bilenko and White, 2008].

Apart from the position of a result, the click probability is also influenced by the presentation of the result. Factors contributing to a higher click probability are the briefness and conciseness of the summary, a title containing many query terms and a short URL [Clarke et al., 2007]. The embedding of multimedia results in a search results page can improve click probabilities, but only if they are placed unobtrusively just above the fold [Hotchkiss et al., 2010].

User interactions with search engines have received a great amount of research attention since the very beginning. However, comparatively little is known about interactions with navigational structures, such as the more recently established recommender systems and collaborative knowledge production systems such as Wikipedia. This thesis aims to fill in this gap by investigating factors influencing navigation in these systems, and by providing a model for navigation in them.

2.1.3 Web Navigation Models

Web navigation models capture the dynamics of human navigation, such as navigation targets, background knowledge and click decisions, with algorithmic descriptions and specifications. These models can be applied to automatically evaluate the navigability of information systems and to study the factors shaping navigation. Simulations of human behavior with navigation models can allow to improve server load [Bestavros, 1995] and facilitate evaluating the ramifications of interface changes [Dimitrov et al., 2015].

The most simplistic model of navigation is a random walk, a stochastic process in which the next node is selected uniformly at random among the current node's neighbors. Despite the modesty of this approach, random walks have led to a variety of applications such as graph generation [Blum et al., 2006], detecting community structure in networks [Rosvall and Bergstrom, 2008], or ranking of search results based on network centrality. The latter was popularized by Page and Brin's PageRank algorithm [Brin and Page, 1998]. PageRank adopts a random surfer model, which combines a random walk with teleportation that enables jumps to arbitrary nodes at random intervals, and ranks pages based on the visit probability of the random surfer (which corresponds to the stationary distribution of the underlying stochastic process).

In addition to a random walk, PageRank can also be viewed as a simple Markov chain. A Markov chain is a stochastic process for which the next node selection depends only on a finite number of previously visited states. PageRank corresponds to a zero-order Markov chain, meaning that the next state depends only on the current state. Navigation models with Markov chains based on previously visited pages in web navigation have been applied extensively (e.g., [Li and Tian, 2003; Sen and Hansen, 2003; Chierichetti et al., 2012]). Zero-order (i.e., memoryless) or first-order Markov chains have been found to model web navigation best [Singer et al., 2014].

The model of decentralized search, tracing back to the small-world experiments described in Section 2.1, is generally also memoryless: every node along the search path decides where to forward the problem to independently and based only on local information as well as some sort of background knowledge or intuitions. Decentralized search has been applied extensively to model human navigation [Trattner et al., 2012; Helic et al., 2013; Lamprecht et al., 2015c]. As decentralized search forms an integral part of this thesis, it will be covered in more detail in the next section (cf. Section 2.2).

Another effective means of modeling web navigation are computational cognitive models. The perhaps best-known of these models is information foraging [Pirolli and Card, 1999]. This model takes from research on optimal foraging theory in nature, which represents animal as making optimal use of their resources in foraging behavior. Food sources can be modeled as occurring in patches, such as a tree full of cherries. Animals feeding on cherries must decide what tree to select and at what time to move to another tree (e.g., if most of the cherries on a tree have been consumed). The model of information foraging translates these concepts to information retrieval and models humans as *informavores* foraging on information. Humans are assumed to be guided by information scent [Chi et al., 2001], a concept that represents an intuition a user might have about a webpage based on clues such as its title. Informavores are then guided by the same mechanisms as animals in nature and need to decide what patch of information to feed on.

The concept of information scent gave rise to the models of SNIF-ACT and CoLiDeS. SNIF-ACT (Scent-Based Navigation and Information Foraging in the ACT Cognitive Architecture) [Pirolli and Fu, 2003] models navigation choices between webpages and predicts the link selection of users. CoLiDeS (Comprehension-Based Linked Model of Deliberate Search) [Kitajima et al., 2000] models navigational choices within webpages. As such, they are considered complimentary models [Kitajima et al., 2007]. CoLiDeS+ [Juvina et al., 2005] extends CoLiDeS to include the information on previously visited webpages in the link selection process.

Another cognitive model of the information retrieval process is berrypicking [Bates, 1989], which models information-seeking like picking berries on bushes. In general, berries are scattered and do not occur in bunches, which requires gatherers to adapt their strategy and constantly decide where to look next. In much the same way, information retrieval can be described dynamically, as users adapt their information needs based on pieces of information. A similar approach is orienteering [O'Day and Jeffries, 1993], which describes the retrieval process as a series of dependent search sessions, where users examine a topic from different perspectives. Like in the sport of orienteering, users need to inspect their current information landscape to determine where to go next.

The next section provides a detailed description of the navigation model of decentralized search, which is of essential importance to this thesis.

2.2 Decentralized Search

Decentralized search is a graph search algorithm which takes its name from its decentralized decision making based on the current node along the search path. Stanley Milgram's small world experiments applied decentralized search to pass a message towards a target by forwarding a letter. The decisions about the message passing (i.e., where to forward the search problem) were therefore always made in a decentralized manner: participants had no information about the past decisions, except for the knowledge of who forwarded the letter to them. In addition to the decentralized decision making, the decentralized search algorithm was also restricted to local knowledge (namely about the current participant's acquaintances) as well as some background knowledge (the occupation and geographic location of the target).

Efficient decentralized search algorithms with local knowledge were later formalized by Watts and Strogatz [1998] and Kleinberg [2000a] in the context of their small-world network models. In the simplest case, a network consists of a regular lattice structure with additionally introduced long-range links. The next node to pass the search problem to can then be selected greedily among the current node's neighbors based on their lattice distance to the target node. Kleinberg showed that when certain conditions are met (cf. Section 2.1), this enables efficient decentralized search algorithms with an expected delivery time sublinear in the number of network nodes.

Kleinberg also suggested the applicability of the decentralized search model to the Web, such as for web crawlers or human web navigation [Kleinberg, 2001]. The translation of decentralized search from its origins in social networks to information networks entails several adaptations [Helic et al., 2013]. Most importantly, the search problem in information networks is controlled by the same entity for the entire search process (i.e., a single user of an information system). However, the knowledge available at each node is still local—users generally know only about the links leading away from a webpage and may have additional intuitions about the structure of a website. Further, the consultation of candidate nodes is comparably cheap—little effort is needed to click on a link leading to an irrelevant page and track back.

Algorithm 1 presents the basic recipe for decentralized search with local knowledge. The required input consists of a graph, a start and a target node, a selection function and a distance function. The basic algorithm is then much like a recursive implementation of depth-first search: at each step, the next node is chosen by the selection function among $\Gamma(n)$, the set of neighbors of the current node n. The selection function *get_next* makes use of a distance function d that measures the similarity between pairs of nodes.

Algorithm 1 The Basic Algorithm for Decentralized Search with Local Knowledge in Web Navigation. The input to the algorithm consist of a graph G, a start node s from which navigation starts, a target node t, a function *get_next* that selects the next node in navigation and a distance function $d(i, j)$ that provides a heuristic distance between pairs of nodes i and j. The algorithm starts from the node s and continuously selects the next node based on the *get_next* function until the target node is found. This algorithm also requires the specification of *get_next* and d. The paths produced by the algorithm can be evaluated to analyze the navigation dynamics, for example the distribution of node visits.

Require: Graph $G(V, E)$, start node s, target node t, selection function *get_next*, distance function d

```
function DECENTRALIZED NAVIGATION(G, s, t, get_next, d)
    n ← s
    OUTPUT n
    while n ≠ t do
        n ← get_next(n, Γ(n), t, d)
        OUTPUT n
    end while
end function
```

This algorithm requires the specification of two parts: a) the selection function *get_next* and b) the distance function d (i.e., the background knowledge). For example, a) could be implemented by a greedy selection strategy and b) by the semantical similarity of the text associated with nodes in the graph. The following subsections describe possible implementations for these parts in more details.

2.2.1 Node Selection

The nodes in decentralized search are selected based only on local knowledge in the network. At each step during the search in Algorithm 1, the function *get_next* takes as parameters the current node n, its neighbors $\Gamma(n)$, the target node t and a distance function d. The distance function d takes two nodes as arguments and returns a numeric distance value representing the distance between them. The function can take on a number of forms, as described in subsection 2.2.2. The function *get_next* may evaluate the function d with parameters as $d(j,t)$ for all $j \in \Gamma(n)$ and t equal to the search target, but not for arbitrary pairs of nodes in the graph.

The information resulting from the distance measurements can then be exploited in a variety of ways as described by Helic et al. [2013]:

Randomly. Select the next node uniformly at random. This leads to a random walk that does not take any background knowledge into account. Algorithm 2 shows an implementation for the *get_next* function for this approach.

Greedy (deterministic). Select the node with the smallest distance to the target. This leads to a greedy and deterministic search path. Algorithm 3 shows an implementation for the *get_next* function for this approach.

ϵ-greedy. Select the next node uniformly at random with a small probability ϵ. Otherwise, select greedily. This models a degree of uncertainty in the search process and yields a stochastic process (i.e., search paths are not deterministically determinable).

Decaying ϵ-greedy. Proceed as in the ϵ-greedy approach, but decay ϵ with the number of steps. For example, use exponential decay: Let $\epsilon(t) = \epsilon_0 \lambda^{-t}$, where ϵ_0 is the is the initial value, λ is the decaying factor and t is the time (i.e., the number of hops). This approach can be used to model the higher uncertainty in the decision making of users in the initial stages of navigation which has been found to decrease as navigation progresses [Helic et al., 2013].

Softmax. Choose a candidate node j with probability

$$p(j) = \frac{\exp(cf(j))}{\sum_{i \in \Gamma(n)} \exp(cf(i))},$$

Algorithm 2 Random Walk Implementation of a random *get_next* selection function.

Require: current node n, set of neighbors $\Gamma(n)$, target node t, distance function d

1: **function** GET_NEXT$(n, \Gamma(n), t, d)$
2: $\quad n \leftarrow$ a node selected among uniformly at random among $\Gamma(n)$
3: $\quad$ **return** n
4: **end function**

Algorithm 3 Greedy Decentralized Navigation. Implementation of a greedy deterministic *get_next* selection function.

Require: current node n, set of neighbors $\Gamma(n)$, target node t, distance function d

1: **function** GET_NEXT$(n, \Gamma(n), t, d)$
2: $\quad$ **return** $\operatorname*{argmin}_{j \in \Gamma(n)} d(j, t)$
3: **end function**

where $f(j)$ is a fitness function computed from $d(j, t)$ and c is a tunable confidence parameter. A larger value for c leads to a higher probability of selecting the maximum, whereas a smaller value leads to a selection proportionally more closely to $f(j)$.

Inverse Distance Rule. Select a node with probability

$$p(j) = \frac{f(j)^{-c}}{\sum_{i \in \Gamma(n)} f(i)^{-c}},$$

where, as for softmax, $f(j)$ is a fitness function and c is a confidence parameter.

Stochastic adaptations have been found to model human behavior better in goal-directed web navigation [Helic et al., 2013], where users tend to explore more in the initial steps of navigation and make more deterministic decisions as they progress.

2.2.2 Background Knowledge

The function *get_next* makes use of a distance function d to evaluate the distance between pairs of nodes in the network. The knowledge contained in the distance function is also referred to as the *background knowledge*. In the Milgram experiments, the description of the target person listed his occupation and location. The background knowledge therefore consisted of the knowledge about geography and professions that the participants had. Based on this information, participants could use their intuitions to forward the letter to someone living closer to the target person or working in the same profession. This is similar to the notion of information scent, where users are guided by their intuitions based on what information they think a target page contains [Chi et al., 2001].

The background knowledge can be represented by a hierarchy [Kleinberg, 2001]. The distance function can then measure the distance in the hierarchy or the height of the lowest common ancestor. Hierarchies as background knowledge have been applied to studying decentralized search in social tagging systems, where tag clouds are displayed as a navigational aid. The resulting bipartite tag-resource graphs are efficiently navigable in theory, but are limited by user interface decisions [Helic et al., 2010]. From these tag clouds, hierarchies can be extracted by a number of algorithms. These hierarchies are called *folksonomies* (a portmanteau of folk-generated taxonomies) and can be used as a hierarchical background knowledge for decentralized search in social tagging systems and to evaluate the navigability of folksonomy extraction algorithms [Helic and Strohmaier, 2011; Helic et al., 2011]. Broad folksonomies (where many users tag the same resource, such as del.ici.ous) have been found more useful for navigation than narrow folksonomies (where a single user tags a resource, such as on Flickr) [Helic et al., 2012]. Hierarchies extracted from the same network that navigation takes place on result in navigational trails closer to human behavior than hierarchies extracted from external data [Trattner et al., 2012]. Ontologies have also been found useful as background knowledge and lead to results comparable to human navigation patterns [Lamprecht et al., 2015c].

The background knowledge may consist of knowledge about groups associated with the network [Kleinberg, 2001] or, more generally, may be represented by an arbitrary distance function $d(i, j)$ that provides information about some measure of distance between nodes i and j. This function need not compute a distance strictly in the mathematical sense (e.g., it need not satisfy the triangle inequality) but can be any function mapping pairs of nodes to real positive numbers. As an example, the TF-IDF similarity between the text associated with nodes in a network can be used to inform navigation [Lamprecht et al., 2015a], as can the cardinality of the set of overlapping neighbors of two nodes, or the information about the degree of a neighboring node.

Additionally, the background knowledge does not need contain complete information about all pairs of nodes, but may be restricted to partial information. For structural information (degrees and clustering coefficients of nodes), only a small amount is required for successful decentralized search [Dimitrov et al., 2015]. For small networks, the cosine similarity has been found more important than the degree information [Geigl and Helic, 2014].

2.3 Applications of Navigation

There is a good body of knowledge about the interaction of users with search engines in general and search engine result pages in particular. However, comparatively little is known about interactions with kinds of webpages that have blossomed in the recent history, such as Wikipedia and recommender systems. A common characteristic of all these system is the dynamic nature of the link structure, which is shaped by algorithms (such as recommendation algorithms) or a large group of human editors (such as Wikipedia). The following subsections survey the current state of research on navigation in these systems.

2.3.1 Navigation in Recommender Systems

Recommender systems help users cope with information overload, make it easier to find relevant content and provide personalized advice. They can help users to answer questions such as "What movie should I watch?" or "What did other customers find a useful addition to this item?" [Herlocker et al., 2004; Ge et al., 2010]. For e-commerce, recommender systems have mostly replaced advice otherwise provided by salespersons, and personalized recommendations act much like recommendations given by a close friend and have proven to be substantially more useful than standard advertising [Linden et al., 2003].

This section is divided into three parts. The first surveys algorithms to compute recommendations, the second describes evaluation measures for these algorithms, and the last part characterizes the relevant work on reachability and navigability on websites employing recommendation systems.

Recommendation Algorithms

Mining data about co-purchases has a long-standing tradition in marketing. In market-basket analysis, the data about items in shopping baskets can be exploited to yield association rules stating what products are likely to be co-purchased [Agrawal and Srikant, 1994]. Modern recommender systems are a relatively young concept from the 1990s [Resnick and Varian, 1997] that are built on these methods, and can make use of content, ratings or a knowledge base to generate recommendations [Jannach et al., 2010]. Most current recommender systems make heavy use of implicit and explicit rating data.

One of the key recommendation algorithms is collaborative filtering [Resnick et al., 1994]. This algorithm filters information by exploiting the collective wisdom in the ratings of a user base. For a given user and an unrated item, the collaborative filtering algorithm first identifies similar users that have already rated the given item by measuring, e.g., the cosine similarity between the rating vectors of users. Then, the algorithm predicts a rating for the item based on the weighted average of the rating of the other users, weighted by their similarity.

The development of recommendation algorithms was substantially spurred on by the Netflix Prize [Koren et al., 2009], an open competition to improve the recommendation accuracy for Netflix, a media streaming and DVD rental service. The prize money of a million dollars led to the development of a number of novel rating-based recommendation algorithms. In particular, the Netflix Prize led to methods applying interpolation weights and matrix factorization for rating predictions.

Interpolation weights [Bell and Koren, 2007] are weights representing the similarities between users. Whereas for collaborative filtering, these similarities are evaluated with a predefined measure (e.g., the cosine similarity) between the rating vectors of users, interpolation weights are learned from the rating data with the help of machine learning.

Matrix factorization [Koren et al., 2009] is a latent factor model that represents users and items by their association to a small number of latent factors. The algorithm factorizes the user-item rating matrix into two lower-dimensional matrices based on machine learning and uses the resulting users-to-factors and items-to-factors matrices to predict ratings. Both approaches lead to significantly better prediction accuracy. More recently, deep learning approaches have started to emerge [Covington et al., 2016].

Evaluation of Recommendation Algorithms

The prediction accuracy, as used in the Netflix Prize, has traditionally been the key factor in the evaluation of recommender systems [Gunawardana and Shani, 2015]. Prediction accuracy is measured by withholding a part of the given ratings and computing the deviation of the algorithmic predictions by the root-mean-square error or a similar metric. However, the focus on accuracy has been identified to neglect many other relevant aspects of recommendations. As an example, recommending a list of Star Trek movies to a Trekkie is likely highly accurate, but also likely of little use. Focusing on prediction accuracy can lead to a popularity bias [Celma and Herrera, 2008; Su et al., 2016] and a filter bubble effect [Nguyen et al., 2014] in recommender systems. To combat these effects, additional evaluation techniques have been developed:

Diversity measures the dissimilarity between the items in a list of recommendations [Boim et al., 2011; Castells et al., 2015]. A list of Star Trek movies would therefore receive a low diversity score. Diversification in recommendation lists has been found to increase user satisfaction with recommender systems [Ziegler et al., 2005].

Novelty measure the unfamiliarity with recommended items [Castells et al., 2015; Gunawardana and Shani, 2015]. For example, a recommendation list consisting only of popular items would receive a low score for novelty. A related concept, **serendipity**, or fortunate happenstance, measures the degree to which recommendations are both novel and relevant [Herlocker et al., 2004; Ge et al., 2010].

Reachability and Navigability

The **reachability** of a recommender system measures the degree to which items in the system are reachable by following recommendation links. A recommender system can be viewed as a *recommendation network*, where items form nodes and recommendations form directed links. Evaluation techniques from network science can then be applied to these networks to measure reachability and navigability.

A lower bound for reachability is **coverage**, which measures how many items of a recommender systems can be recommended to users [Herlocker et al., 2004; Ge et al., 2010; Gunawardana and Shani, 2015]. However, coverage does not take disconnected components in a recommendation network into account. Computing the size of the largest strongly connected component in a recommendation network helps to answer this question more clearly, as it states the cardinality of the largest set of mutually reachable items.

Web users frequently use browsing to discover new items [Lerman and Jones, 2007]. More generally, browsing an item collection without wanting to buy is an activity users find pleasant [Herlocker et al., 2004]. Even with search engines, the complete click trails have a higher utility for users and improve coverage, novelty, and diversity in web search [White and Huang, 2010]. Similarly, some users prefer navigation to search, even when they know exactly where they want to go [Teevan et al., 2004].

Reachability and navigability for recommendation networks have been extensively studied for music recommender systems. Music recommendation networks generally exhibit network properties enabling efficient navigation, namely heavy-tail degree distributions, strong clustering, and small-world properties [Cano et al., 2006]. However, many items are never recommended at all [Seyerlehner et al., 2009a], which could be improved by selecting recommendations based on their effects on the reachability of the resulting recommendation network [Seyerlehner et al., 2009b]. For movie recommender systems, the recommendation networks of IMDb have been shown to exhibit heavy-tail degree distributions [Grujić, 2008].

While these approaches have investigated the static navigability of recommendation networks and shown some of the properties enabling efficient navigability, the applied navigation dynamics have never been thoroughly studied. This thesis aims at uncovering the navigation dynamics in recommendation networks in real-world scenarios.

2.3.2 Navigation on Wikipedia

Wikipedia[1] is a large online general-purpose encyclopedia based on a wiki, a collaborative editing system. Since its launch in 2001, Wikipedia has grown to be the largest encyclopedia on the Web and has risen to be one of the top ten most popular websites worldwide [Alexa, 2017], available in more than 250 languages. Wikipedia is operated by the Wikimedia foundation and maintained predominantly by volunteers. The community of Wikipedians has established roles such as administrators, which are chosen in public elections. Furthermore, detailed policies and guidelines [Wikipedia, 2017c] have been developed. A detailed manual of style describes how the content of articles should be structured [Wikipedia, 2017a].

[1] http://www.wikipedia.org

Hyperlinks between articles are also placed by editors and should be introduced to establish context and help to understand the article at hand better [Wikipedia, 2017b]. When articles are viewed as nodes and links as directed edges, an article network emerges. The article network of Wikipedia grows by preferential attachment and features a gigantic strongly connected component that contains the vast majority of all articles [Capocci et al., 2006; Kamps and Koolen, 2009]. In other words, most of the articles on Wikipedia are mutually reachable by following links in the article text. Authors who contribute to multiple topics help to bridge the knowledge gap by contributing to both sides and writing integrative articles [Halatchliyski et al., 2010]. Compared to other websites and the Web in general, Wikipedia's article network is more densely connected [Kamps and Koolen, 2009].

This section is divided into two parts. The first part describes information retrieval on Wikipedia, and the second part surveys application areas of navigational information gained from Wikipedia.

Information Retrieval on Wikipedia

Researchers have long been interested in how users retrieve information on the encyclopedia based on its access logs. In general, access log data for Web information systems is important because it allows a nonreactive study of human navigation behavior and can be used for a variety of applications. However, data on Wikipedia visits is difficult to obtain, as the Wikimedia foundation does not publish detailed access logs for its Wikipedia projects due to privacy concerns. Nevertheless, a few alternative ways of obtaining data have been exploited.

Firstly, web toolbars can be mined by their operators to filter only Wikipedia-related clicks. Based on this data, the PageRank teleportation factor in Wikipedia was established at 60–70% (i.e., users follow links in only 30–40% of cases and otherwise directly jump to articles by, e.g., using the search function or manually typing in an address). This is substantially higher than for general web navigation [Gleich et al., 2010], where the original PageRank formulation assumed a teleportation factor of 15%.

This indicates that many users visit Wikipedia for quick lookup or are perhaps referred by a search engine, brought to the right landing page and are able to satisfy their information need immediately. However, there exist many use cases for Wikipedia, and not all queries can be directly answered by lookup but many also need researching and link-following. Only more recently, the publication of a clickstream dataset for the English Wikipedia [Wulczyn and Taraborelli, 2015] has enabled new research avenues [Dimitrov et al., 2016]. However, this dataset only comprises click-data in aggregated form.

To obtain data about complete navigation sessions including starts, targets and information about whether the navigation goal has been reached, researchers have resorted to studying the log files of Wikipedia games such as the Wiki Game[2] or Wikispeedia[3]. In these games, users start on a predefined article with the objective of reaching a given target article solely by following links in the article text (without using the search function or manually typing in URLs). Log files from these games allow researchers to study complete goal-directed navigation paths with defined start and target articles as well as information about navigational success.

Analysis of wikigames has shown that human wayfinding on Wikipedia is generally very efficient and the user click paths are only a few clicks longer than the shortest possible paths in the network [West and Leskovec, 2012b]. Humans click the back button on average once every other game, and most commonly backtrack to high-degree hubs when they get lost [Scaria et al., 2014]. Even though human paths are only slight longer than the shortest paths, their structure is different. Successful wikigame paths exhibit distinct zoom-out and home-in phases, in which humans click to high-degree articles (hubs) first and then home in towards the target article. This is similar to research on hidden metric spaces, where the same phenomena have been found to occur in greedy airport routing [Boguñá et al., 2009]. For wikigames, human navigation has been found to be strongly influenced by high-degree hubs in the beginning and by semantic similarity with the target article in the end phase [West and Leskovec, 2012b]. Interestingly, human-like intelligence does not seem to be required

[2] http://www.thewikigame.com
[3] http://www.wikispeedia.net

for finding short paths in Wikipedia. Automatic navigation based on simple heuristics such as degree or TF-IDF cosine similarity outperforms humans in terms of path length [West and Leskovec, 2012a], but produces paths of very different characteristics.

Applications Areas of Navigational Data from Wikipedia

Access data from Wikipedia can be mined to inform a variety of applications. A prime example of this is the inference of semantic distance between concepts from knowledge contained in Wikipedia click trails. While the entirety of all links in the article network can be exploited to this end, many of the links are not semantic connections but rather link concepts needed to understand a topic. In goal-directed Wikipedia games, however, links can be weighted by their selection frequency to measure the semantic similarity between concepts [West et al., 2009a; Singer et al., 2013].

Another application area is the field of suggesting new links on Wikipedia. On the English Wikipedia alone, around 900 new articles are created every day [Wikipedia, 2017d], and keeping the article network up to date poses a daunting task. A range of automated approaches have been proposed to this end. Most approaches parse the article texts and suggest links based on identifying links from keyword extraction and word sense disambiguation [Mihalcea and Csomai, 2007], based on machine learning [Milne and Witten, 2008] or based on clustering related documents and identifying missing links within clusters [Adafre and de Rijke, 2005]. Other approaches make use only of the article network and its links and learn new links through dimensionality reduction [West et al., 2009b]. This has the inherent advantage of not needing to parse and understand text or disambiguate word senses. Likewise, link suggestions can be made from click trails. The basic idea is as follows: if visitors frequently visit an article A, subsequently follow a link from it and visit one or more other articles and then visit another article B, the assumption can be made that a link leading directly from A to B would be useful. This approach has been exploited based on wikigames [West et al., 2015] and on click trails from the English Wikipedia (in cooperation with the Wikimedia Foundation) [Paranjape et al., 2016].

Click trails from Wikipedia games can also be compared against navigational models. The model of decentralized search has been applied to simulate human navigation in wikigames and to identify the influence of a range of parameters. Biomedical ontologies as background knowledge have been found to lead to navigation paths comparable to human navigation behavior [Lamprecht et al., 2015c]. Stochastic adaptations in decentralized search have been identified to model human behavior better, as users tend explore more initially and exploit their background knowledge in the end phases [Helic et al., 2013].

3 Publications

This chapter contains the research publications that make up this thesis. The first section details the author contributions to them, and the subsequent sections contain the publications themselves.

3.1 Contributions to the Publications

- **Article 1 [Lamprecht et al., 2016b]:** Lamprecht, D., Lerman, K., Helic, D. and Strohmaier, M. (2016). How the Structure of Wikipedia Articles Influences User Navigation. *New Review of Hypermedia and Multimedia*, 23(1):29–50.

The ideas for this publication stem from discussions with Kristina Lerman while I was visiting her at ISI, and were refined in discussions with Markus Strohmaier and Denis Helic. Kristina and I designed the experiments with the Wikipedia click data for the characteristics of the sections of Wikipedia articles, their links, and the influencing factors on link selection. Denis and I developed the ideas for the stepwise evaluation of influencing factors in goal-directed Wikipedia navigation. I designed and wrote the code for the experiments. All authors discussed the results and wrote the paper.

- **Article 2 [Lamprecht et al., 2016c]:** Lamprecht, D., Strohmaier, M. and Helic, D. (2016). A Method for Evaluating the Navigability of Recommendation Algorithms. In *Proceedings of the 5th International Workshop on Complex Networks and their Applications*, pp. 247—-259.

The ideas and the evaluation method for recommendation networks for this publication were developed and refined in discussions between Markus

Strohmaier, Denis Helic and me. I conducted the experiments and designed and implemented the code for the evaluation method framework, which is available as an open-source software package[1]. All authors designed the experiments and drew conclusions from the results. All authors contributed to the interpretation of the results and the writing of this paper.

- **Article 3 [Lamprecht et al., 2015a]:** Lamprecht, D., Geigl, F., Karas, T., Walk, S., Helic, D. and Strohmaier, M. (2015). Improving Recommender System Navigability Through Diversification: A Case Study of IMDb. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pp. 21:1–21:8.

The ideas for this paper were developed in discussions with Tomas Karas, Florian Geigl, Simon Walk, Denis Helic and Markus Strohmaier. Tomas wrote the code to obtain the IMDb rating dataset and managed the database. Tomas and I wrote the code for the recommendation algorithms, and I wrote the code for the evaluation framework of recommendation algorithms to assess the static and dynamic dimensions of recommendation networks, and for the diversification approaches. I led the writing of the paper, and all authors contributed to discussion of the results and provided feedback for the paper.

- **Article 4 [Lamprecht et al., 2016a]:** Lamprecht, D., Dimitrov, D., Helic, D. and Strohmaier, M. (2016). Evaluating and Improving Navigability of Wikipedia: A Comparative Study of Eight Language Editions. In *Proceedings of the 12th International Symposium on Open Collaboration*, pp. 17:1–17:10.

The ideas for the navigational views of Wikipedia for this paper stem from discussions between Markus Strohmaier and me. We designed the experiments together with Denis Helic and Dimitar Dimitrov, and extended them with a recommender to demonstrate the findings. Dimitar and I performed the experiments and wrote the code for the evaluations. All authors were involved in the discussion of the results and the writing of the paper.

[1] https://github.com/lamda/RecNet

3.2 How the Structure of Wikipedia Articles Influences User Navigation

This first article of this thesis addresses the first research question, namely what factors influence the navigational choices of users. To this end, the article makes use of two datasets of Wikipedia click logs. A Wikipedia clickstream dataset of all clicks within the English Wikipedia over the duration of an entire month provides the opportunity to study free-form navigation, that is, an aggregation over all occurring navigation scenarios. A second dataset of logs from a wikigame allows to study focused, goal-directed navigation, and contains complete click paths.

The article first investigates the characteristics of the organization of Wikipedia articles and the placement of hyperlinks, which are guided by a set of policies established by the community of Wikipedia. The results show that sections located closer to the top of articles display substantially different characteristics, as the lead and infobox sections contain links to more general topics. The analysis is complicated by the ambiguity of links, as links that occur multiple times on the same article are indistinguishable in the log files. To circumvent this issue, the article proposes probabilistic link selection models based on several influencing factors, including degree, view counts, semantic similarity to navigation target, and position in the article. The results of the link selection models are then compared to the ground truth distribution of the links that were actually selected by users.

For free-form Wikipedia navigation, the results show that the influencing factor that best models click choices is a bias towards page organization. For goal-directed navigation, the results show that degree is the best explanation for the first step and similarity to the target article is the best explanation for the last step. The remainder of steps is again best modeled by a bias towards page organization.

The findings of this article highlight the importance of page organization for the navigability of an information network, and open up possibilities to better connect webpages by restructuring content.

New Review of Hypermedia and Multimedia
Vol. 00, No. 00, Month 2009, 1–21

RESEARCH ARTICLE

How the structure of Wikipedia articles influences user navigation

Daniel Lamprecht [a,*] Kristina Lerman [b], Denis Helic [a], and Markus Strohmaier [c,d]

[a]*Knowledge Technologies Institute, Graz University of Technology, Austria*
[b]*Information Sciences Institute, University of Southern California, USA*
[c]*Department of Computer Science, University of Koblenz-Landau, Germany*
[d]*GESIS - Leibniz Institute for the Social Sciences, Cologne, Germany*
(*Received 00 Month 200x; final version received 00 Month 200x*)

In this work we study how people navigate the information network of Wikipedia and investigate (i) free-form navigation by studying all clicks within the English Wikipedia over an entire month and (ii) goal-directed Wikipedia navigation by analyzing wikigames, where users are challenged to retrieve articles by following links. To study how the organization of Wikipedia articles in terms of layout and links affects navigation behavior, we first investigate the characteristics of the structural organization and of hyperlinks in Wikipedia and then evaluate link selection models based on article structure and other potential influences in navigation, such as the generality of an article's topic. In free-form Wikipedia navigation, covering all Wikipedia usage scenarios, we find that click choices can be best modeled by a bias towards article structure, such as a tendency to click links located in the lead section. For the goal-directed navigation of wikigames, our findings confirm the zoom-out and the homing-in phases identified by previous work, where users are guided by generality at first and textual similarity to the target later. However, our interpretation of the link selection models accentuates that article structure is the best explanation for the navigation paths in all except these initial and final stages. Overall, we find evidence that users more frequently click on links that are located close to the top of an article. The structure of Wikipedia articles, which places links to more general concepts near the top, supports navigation by allowing users to quickly find the better-connected articles that facilitate navigation. Our results highlight the importance of article structure and link position in Wikipedia navigation and suggest that better organization of information can help make information networks more navigable.

Keywords: Wikipedia, Navigation, Article Structure, Generality

*Corresponding author. Email: daniel.lamprecht@tugraz.at

ISSN: 1361-4568 print/ISSN 1740-7842 online

DOI: 10.1080/1361456YYxxxxxxxx
http://www.informaworld.com

1. Introduction

Much of human knowledge and expertise resides in networks, such as the World Wide Web, Wikipedia, scientific citation networks, and, increasingly, user-generated content on social media. Successfully finding relevant information in these networks, even as they become larger and more complex, is key to our continued ability to innovate, grow, and prosper. While search engines have drastically facilitated information-seeking, not every information need is directly satisfiable. In situations when a query cannot be expressed in an explicit fashion, navigation and exploration are necessarily the information retrieval techniques of choice. Information needs are generally dynamic and evolving (e.g., as modeled by Berrypicking (Bates 1989) or Information Scent (Chi *et al.* 2001)), and knowledge gained during the navigation process can put information in context and help with decision-making (Marchionini 2006). Some users prefer navigation over search even when they know what they are looking for (Teevan *et al.* 2004). For a large encyclopedia such as Wikipedia, possible navigation scenarios generally can span a large range, from goal-directed navigation to following a link to learn more about a certain concept, to explorative search and many more.

Problem. Many questions about information networks, and how people navigate them to find relevant information, have not yet been fully answered. For example, how should webpages be structured to facilitate navigation and searchability? How does an individual's familiarity with the knowledge contained in the network influence navigation? Answering question such as these will help to create efficient navigation structures in massive information networks. Generally, more attention is paid to items that are displayed at the top of the screen or the top of a list of items, even when no ranking is present (Payne 1951). For webpages, users scan the pages in an f-shaped pattern (Nielsen 2006) and dedicate more attention to the top and left (Buscher *et al.* 2009).

On Wikipedia, articles are subject to a common page organization. For example, the first section usually introduces the article in more general and more broadly accessible terms, and the infobox summarizes the main facts. We therefore hypothesize that the accessibility of the topmost section and this clear organization helps users to more easily find relevant links and to successfully navigate Wikipedia. Hence, we study how the organization of Wikipedia articles in terms of sections, infoboxes and link positions affects navigation in Wikipedia. Previous work has shown both semantic and structural knowledge to influence Web navigation (Juvina and van Oostendorp 2008). For goal-directed Wikipedia navigation, users have been found to select links based on both semantic similarity and overlap of link titles (Salmerón *et al.* 2015). In this work, we compare the influence of textual similarity of articles to the influence of structural elements. Specifically, we examine the following research question:

Research question: To what extent does article structure affect Wikipedia navigation?

Approach. We address this questions by studying how people use Wikipedia to find information. Our approach is two-fold:

(i) We analyze an entire month of all clicks within the English Wikipedia. This permits us to gain insight into unrestricted free-form clicking behavior on Wikipedia, covering all usage scenarios.
(ii) We analyze wikigames, which challenge people to navigate from a source Wikipedia article to a given target article solely by using the existing links in the text. Wikigames allow us to inspect goal-directed Wikipedia navigation from

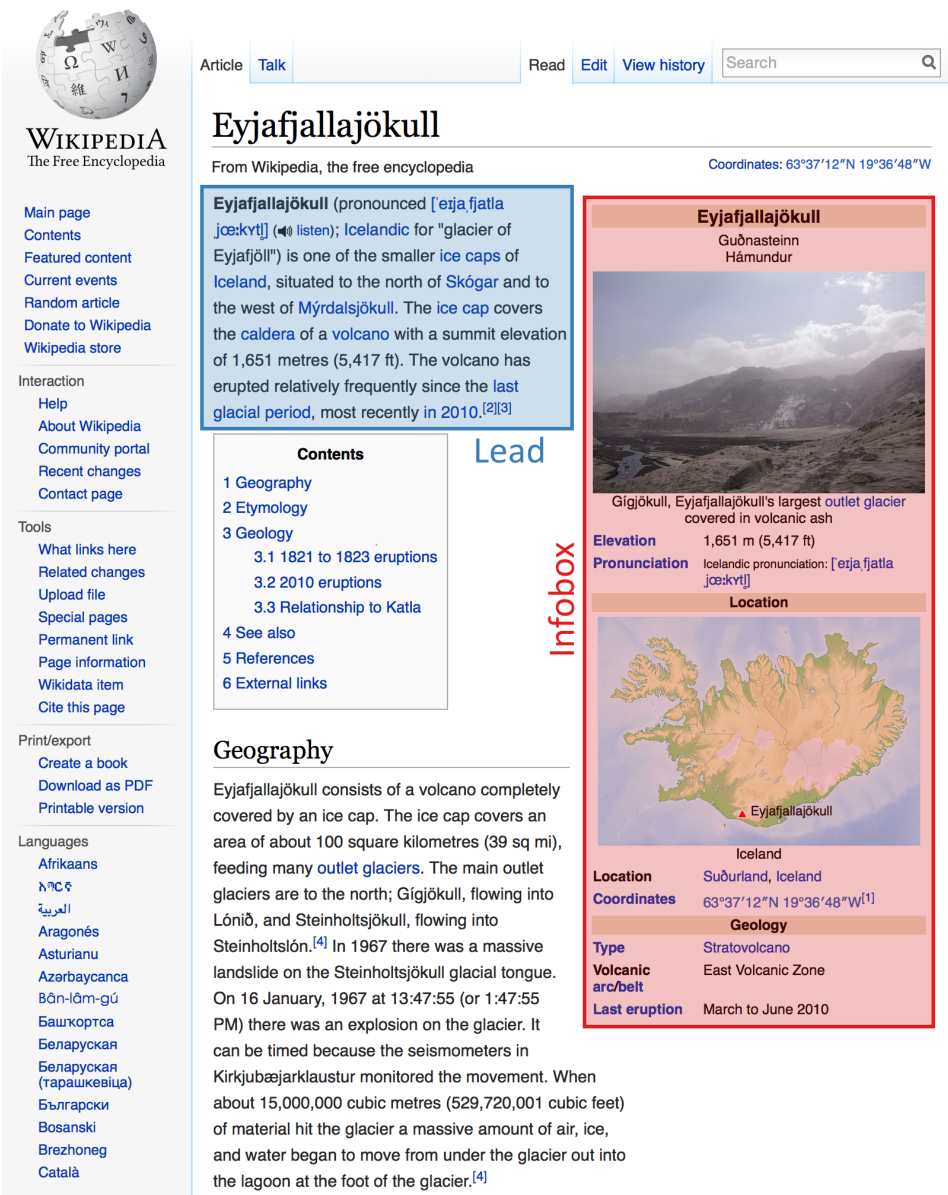


Figure 1.: **Example of the structure of a Wikipedia article.** The image shows a Wikipedia article with the lead section (blue shading) and the infobox (red shading), followed by the table of contents and the start of the main content. In this paper we show that users focus their attention on the lead sections and the infoboxes.

the perspective of very focused navigation set up as a game, for which we have access to very detailed log files and can investigate clicking behavior step by step.

We use these two datasets to study the structure of Wikipedia article in terms of lead sections and infoboxes (see Figure 1 for an example). We first investigate the characteristics of these sections and of hyperlinks in Wikipedia and then analyze potential influences on navigation by simulating link selection models based on a range of influence factors. Finally, we consider the detailed logs of goal-directed wikigames and study link selection behavior step by step.

Contributions. Our results suggest that article structure has a strong influence on navigation. We find evidence that a large share of user clicks are to links in the lead section or an infobox. For free-form Wikipedia navigation, navigation decisions can be best explained by a bias towards the article structure, favoring links located near the top of the article. For the goal-directed navigation of wikigames, our findings confirm the zoom-out and the homing-in phases identified therein by previous work (West and Leskovec 2012b), where users are guided by generality at first and textual similarity to the target towards the end. However, the outcomes of the link selection models accentuate that article structure is the best explanation for wikigames in all except these initial and final stages. Our results highlight the importance of article structure and link position in Wikipedia navigation and suggest that better organization of information can help make information networks more navigable.

2. Related Work

2.1. *Navigation in social networks*

Research on navigation in networks was brought into being in the 1960s by Stanley Milgram's influential letter forwarding experiments. These experiments established that participants were able to find short chains between unrelated individuals in the social network of the entire United States (Milgram 1967) with a decentralized search approach (i.e., the search problem was forwarded with a letter and not controlled by a centralized instance).

In the 1990s, Watts and Strogatz demonstrated that many social, technological and biological networks exhibited the *small-world property* (high clustering and a small diameter), which ensured that most pairs of nodes were reachable in only a few hops (Watts and Strogatz 1998). Jon Kleinberg subsequently showed what properties made these networks efficiently navigable with decentralized search algorithms (Kleinberg 2000, 2001).

2.2. *Navigation in information networks*

Whereas in Milgram's experiments navigation was conducted in social networks, the focus for navigation research has since shifted to information networks. Information networks imply different characteristics in navigation: in contrast to social networks, navigation is only conducted by a single agent who can more easily explore larger parts of the network (Helic *et al.* 2013). In terms of decentralized search, this signifies that even though the user conducting the search stays the same over the entire duration, the decision of what link to click is taken independently at each step.

Web navigation can be effectively modeled by computational cognitive models. One of the most prominent models is information foraging (Pirolli and Card 1999), which models information-seeking behavior based on optimal foraging theory in biology. In this model, *information scent* (Chi *et al.* 2001) guides users to patches of information. Just as animals are thought to maximize their benefit gained from foraging, information seekers are thought to make optimal use of their resources to gain information. Based on the notion of information scent, the cognitive model of SNIF-ACT has been developed to explain navigation choices in navigating between webpages (Pirolli and Fu 2003). Another model based on information scent is CoLiDeS (Comprehension-based Linked Model of Deliberate Search) (Kitajima *et al.* 2000) that explains navigational choices of users within webpages. This model was later extended as CoLiDeS+, which includes the previously visited webpages in the link selection process (Juvina *et al.* 2005). SNIF-ACT and CoLiDeS are considered complimentary models (Kitajima *et al.* 2007). Whereas SNIF-ACT models information patches as entire websites, CoLiDeS models the link selection decisions within regions of a webpage. Similar to CoLiDeS, in this paper we study the link selection behavior for different areas of Wikipedia articles.

Information network navigation has also been found useful as an evaluation method for information systems. For medical documents, navigation with decentralized search was found to be comparable to human navigation (Lamprecht *et al.* 2015b) and was used to point out differences in folksonomy generation algorithms (Helic *et al.* 2011). Seyerlehner et al. used navigability to examine recommender systems and found top-N collaborative filtering to be inherently poorly navigable (Seyerlehner *et al.* 2009). Lamprecht et al. later confirmed this finding for the recommendation networks of the Internet Movie Database (IMDb) and suggested to use diversification to make networks more navigable (Lamprecht *et al.* 2015a).

2.3. *Navigation on Wikipedia*

Log data for many Web information systems are chronological—they form trails or traces of user activity. The ultimate and concrete user goals, however, are rarely known directly from log data. In click-trails from logs, goals can occur at any part of a path (if at all) and are not distinguishable from other clicks. To overcome this issue, researchers have resorted to studying clicktrails of navigational games such as wikigames. These games (for example Wikispeedia[1] or the Wiki Game[2]) challenge players to reach a predetermined target article only by following links within the body text. Log files from these games equip researchers with concrete start-target scenarios for navigation and allow for a more detailed investigation. Wikipedia logs have been found useful to find semantically similar Wikipedia articles (West *et al.* 2009, Singer *et al.* 2013) and discover missing links (West *et al.* 2015).

Wikigames have been extensively studied on the Wikispeedia dataset (West *et al.* 2009). This wikigame is played on the Wikipedia for Schools 2007 selection, a subset of around $4,600$ articles chosen based on the UK National Curriculum. Players of Wikispeedia have been found very efficient at finding goals (West and Leskovec 2012b). Paths found by players showed a tendency to navigate (and backtrack (Scaria *et al.* 2014)) to a high-degree hub first and then home in on the target based on content similarity (West and Leskovec 2012b). These characteristics were different from shortest paths or paths found by search algorithms. Surprisingly, simple search algorithms (e.g., based on textual similarity) were even more efficient than humans, proving that no high-level reasoning skills were necessary to find targets (West and Leskovec 2012a).

2.4. *Influence of webpage organization*

Computational cognitive models such as CoLiDeS model link selection based on the semantic similarity to the target (Kitajima *et al.* 2000). However, structural knowledge has been shown to have an impact navigation performance (Juvina and van Oostendorp

[1] `www.wikispeedia.net`
[2] `www.thewikigame.com`

2008). In this paper, we compare the influences of textual similarity to the influence of webpage structure.

The organization of a webpage can exert a significant influence on viewing and click decisions. In this paper we study these influences in the context of Western culture and its left-to-right writing direction. Generally, humans are known to be biased by presentation order even in the absence of explicit ranking. In multiple-choice questions, people more frequently select answers located closer to the top (Payne 1951, Blunch 1984). The same effects have been identified in cultural markets (Salganik *et al.* 2006) and recommender systems (Lerman and Hogg 2014). For websearch, users have been found to predominantly focus their attention on the first few results (Pan *et al.* 2007, Craswell *et al.* 2008, Buscher *et al.* 2010, González-Caro and Marcos 2011). Eye-tracking has shown that humans scan webpages in an F-shaped pattern (Nielsen 2006) and generally focus on the top and left areas of webpages (Buscher *et al.* 2009).

Web users have been found to quickly decide whether a page is worth their interest. Weinreich *et al.* (2006) report that users stay on most pages only for a short time span and that 52% of all visits are shorter than 10 seconds. Web users frequently skim a page at first to determine its relevancy (Liu *et al.* 2010, 2014). This behavior also shows in the analysis of click locations: in a study, 76.5% of clicks where made in the area visible without scrolling and 45% on links located near top left corner (Weinreich *et al.* 2006).

These findings suggest that users mostly skim content and immediately decide whether to stay or to move on. However, when information needs are not satisfied by the top links, users can adapt and dedicate more attention to all of the search results (Salmerón *et al.* 2013). This suggests that users attribute certain characteristics to specific parts of webpages but adjust them based on the actual content.

3. Datasets

In this work we look at Wikipedia navigation from two distinct vantage points: First, we analyze data from an entire month of all clicks within the English Wikipedia. This allows us to gain insight into unconstrained free-form navigation on the encyclopedia. Second, we study goal-directed Wikipedia navigation based on wikigames, which permits us to inspect complete navigation paths with explicit navigation targets.

3.1. *Wikipedia clickstream*

We investigate free-form click behavior in the English Wikipedia based on a dataset recording all clicks to the Desktop version of the English Wikipedia within the month of February 2015 (Wulczyn and Taraborelli 2015). Visitors use Wikipedia in many different ways, such as looking up specific facts, trying to learn about a concept, reading articles to pass time, and many more. The Wikipedia clickstream datasets records clicks from all types of visits in aggregate in the form of link click counts for links in articles. As a consequence, the dataset does not reveal any user sessions, and potential navigation sessions or navigation targets cannot be identified. The Wikipedia clickstream therefore allows us to study unconstrained free-form navigation obtained non-reactively (i.e., not subject to potential behavioral changes due to the fact that users knew they were being recorded).

For this work, we only consider clicks between pairs of Wikipedia articles and exclude any external webpages. To allow for a fair comparison with the wikigame dataset

Table 1.: **Sample entries of the Wikispeedia dataset.** The figure shows a part of the log files of successful wikigames. The path column lists the visited pages from the start article (in bold) to the target article (also in bold), separated by semicolons. A < character indicates a back click, which we resolved to the previous page.

hashed IP address	timestamp	path
1d11d305144df277	1233667617	**Wake_Island**;Guam;<;Pacific_Ocean;Peru;Brazil;**Rio_de_Janeiro**
36dabfa133b20e3c	1249525912	**14th_century**;China;Gunpowder;**Fire**
051611353cd98688	1260476499	**Commodore_64**;United_States;India;**New_Delhi**
473d6ac602c2b198	1322605407	**Asteroid**;Jupiter;Roman_mythology;<;<;Comet;Denmark;**Viking**

described in the following, we restrict our analysis to the same roughly 4,600 articles available in Wikispeedia. However, we analyze all clicks from these articles to any article in the English Wikipedia (including all those not contained in the selection). This leaves us with 56,961,992 clicks to study.

3.2. *Wikigames*

For many Web information systems, the log data are chronological and form trails of user activity. The ultimate and concrete user objectives, however, are rarely known relying solely on log data. In click-trails, the exact user goals can be present at arbitrary points in the trails, as there exists no clear distinction of goals within paths.

Navigational games played on Wikipedia (such as Wikispeedia[1] or the Wiki Game[2]) allow us to circumvent this problem. In these games, the objective is to reach a given target article without using the search function or any external information. These navigation tasks are conducted as follows: Starting from a *start article*, users aim to find a given *target article* by following links in the text. For this work, we use data from *Wikispeedia* (West *et al.* 2009), a wikigame played on the Wikipedia for Schools 2007 selection. This selection is a subset of the English Wikipedia of around 4,600 articles produced for educational purposes by the charity SOS Children. The articles are chosen based on the UK National Curriculum to illustrate educational topics and not necessarily for their link structure and navigability.

In Wikispeedia, players are challenged to play a mission, which by default consists of randomly selected start and target articles. Players can also opt to manually select a start and target article for a mission. Before starting and at any time during the game, players may inspect the full content of the target article (but not its inlinks or related articles). We use the log files of roughly 75,000 wikigames and 475,000 clicks. Table 1 shows a sample of the log files. As a preprocessing step, we resolved all clicks on the back button (logged as "<") to the corresponding articles. These log files equip us with concrete start-target scenarios for navigation and allow for a more detailed investigation.

[1] `www.wikispeedia.net`
[2] `www.thewikigame.com`

4. Structure of Wikipedia articles

As the first step of our analysis, we study the characteristics of structural organization and hyperlinks. The structural requirements, which Wikipedia articles are expected to follow, are laid out in the encyclopedia's manual of style (Wikipedia 2016a). Articles should generally start with a *lead section* (or *lead* for short), which is the first section before the table of contents and the first heading. The lead should serve as an easy-to-understand introduction to the article and establish context. Similar to an abstract for a scientific article, the lead should not be divided into any sections (Wikipedia 2016c).

Articles can optionally contain an *infobox*—a tabular description of the article's most important facts (Wikipedia 2016b) (e.g., *scientific classification*, *country*, *area* or *date of birth*). In our datasets, infoboxes are present in 73% of the 4,600 articles from the Wikipedia clickstream and 55% of the articles used in Wikispeedia. Infoboxes generally appear next to the lead section in the top right of an article. Figure 1 gives an example of an article used in Wikispeedia and shows both the lead section and an infobox.

The lead section is followed by the *content* of the article, which is usually divided into a number of sections. Wikipedia guidelines do not specify how to structure the content, and this decision is left to the Wikipedia editors (Wikipedia 2016a). The content may be followed by appendices (such as references and external links) and footers (such as navigation templates or categories). Due to the fact that the structure of the content part is not regulated and humans have been found to dedicate more attention to the top of lists and webpages (Buscher *et al.* 2009, Lerman and Hogg 2014), we will focus our analysis on the structure at the top of Wikipedia articles and its effects on navigation.

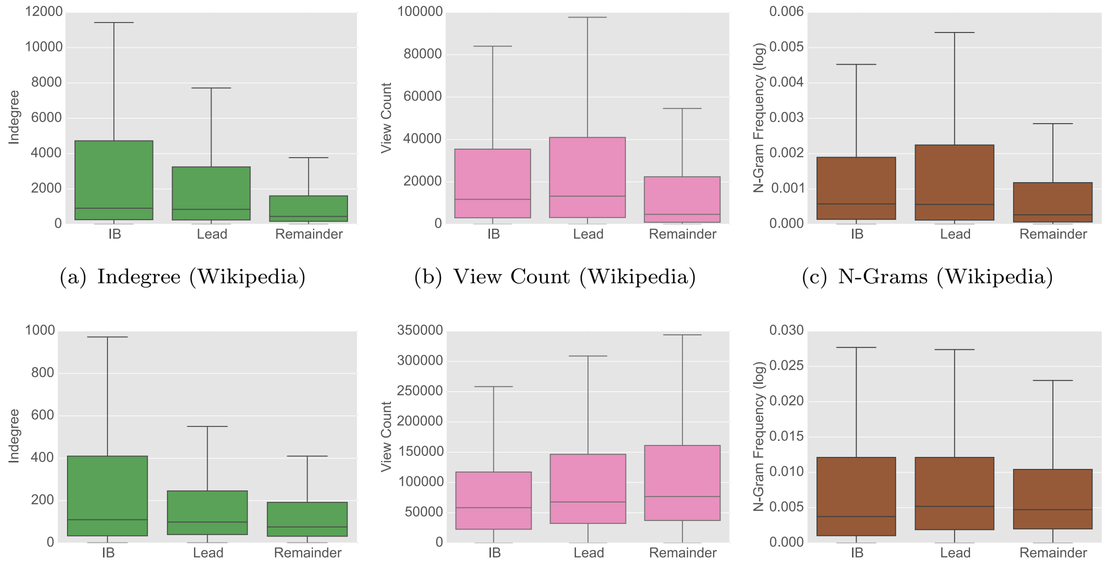
4.1. *Characteristics of lead and infobox*

In this section, we explore the characteristics of the structures located near the top of a Wikipedia article. Wikipedia guidelines describe an article's lead section as follows: *The lead serves as an introduction to the article and a summary of its most important contents. [...] The opening sentence should provide links to the broader or more elementary topics that are important to the article's topic or place it into the context where it is notable* (Wikipedia 2016c). Moreover, the number of links in the lead should be restricted to what is required (Wikipedia 2016d). These guidelines suggest the presence of more general links in the lead sections of articles.

Approach. To assess whether the top of an article contains more general links, we compare the generality of links in the lead, the infoboxes and the remainder of the article. To measure generality, we make use of the usage frequency, which has been identified as a good proxy measure for generality (Benz *et al.* 2011). We use the following three measures:

- **Indegree:** The indegree of an article is the number of links pointing to it from within Wikipedia. Indegree measures the *navigational quality* of a node as well as *generality* (Gabrilovich and Markovitch 2009, p. 450). Indegree captures the generality as viewed by Wikipedia editors, who placed these links.
- **View count:** We use the number of views [1] that Wikipedia articles received in February 2015 (the same time that the Wikipedia clickstream was collected). The

[1]Retrieved from `http://stats.grok.se`



(a) Indegree (Wikipedia) (b) View Count (Wikipedia) (c) N-Grams (Wikipedia)

(d) Indegree (Wikispeedia) (e) View Count (Wikispeedia) (f) N-Grams (Wikispeedia)

Figure 2.: **Indegree, view count and search query n-gram frequency for lead, infoboxes and the remainder of articles.** The figure shows the values for the link targets in these sections (outliers are excluded for the sake of clarity). Articles linked in the lead and infobox have a higher indegree, which holds for both Wikipedia and Wikispeedia. For Wikipedia we also find that articles linked in the lead have a higher view count and higher n-gram frequencies. These results show that links towards the top of an article tend to lead to more general articles.

number of times an article is visited measures the popularity and generality as seen by Wikipedia visitors.

- **Search query n-gram frequency (n-grams):** We measure the familiarity and generality of a term in the active vocabulary of users of a websearch engine by the number of occurrences of article titles in search queries to Microsoft Bing (from the Microsoft Web N-gram corpus (Wang *et al.* 2010)).

Results. Articles linked in the lead and infobox mostly lead to pages with a higher generality (cf. Figure 2). For the Wikipedia clickstream, links in the lead sections and infoboxes indeed lead to articles with a higher generality in all of the three measures. This confirms our hypothesis that these sections contain links to more general articles.

For the articles used in Wikispeedia, we find that links in the lead and infobox also lead to articles with a higher indegree. However, this does not hold true in terms of view count and only for the lead section in terms of search query n-gram frequency. One possible explanation for this is the limited number of possible link targets in Wikispeedia, where links are restricted to articles within the subset of Wikipedia for schools. This may introduce a bias, since a large number of the more specific articles are not available as link targets, increasing the generality of links outside the lead and infobox. However, when considering the network-intrinsic measure of indegree, links in the lead and infobox do lead to more general articles. These results show that links towards the top of an article tend to lead to more general articles.

4.2. *Links in Wikipedia articles*

Hyperlinks in Wikipedia articles are meant to establish context, help to understand the content better and provide references to more background information (Wikipedia 2016e). As a general rule, *a link should appear only once in an article, but if helpful for readers, may be repeated in infoboxes, tables [...] and at the first occurrence after the lead* (Wikipedia 2016d). Specifically, the article should be complete even without reading the infobox (Wikipedia 2016b).

This indicates that links can occur multiple times within the same article. And indeed, for the Wikipedia clickstream, 49.0% of link targets are linked multiple times within the same article, and these ambiguous link targets received 75.8% of all clicks in the dataset. For Wikispeedia, 28.5% of all link targets on an article are linked multiple times and received 43.6% of all clicks.

To the best of our knowledge, there exist no public Wikipedia datasets which contain the information of the exact position of clicked links. We therefore measure the influences of lead and infobox by comparing link selection models, as described in the following section.

5. Influences on aggregated Wikipedia navigation

To establish the influence of article structure on navigation choices on Wikipedia, we now turn our attention to link selection models. These models allow us to obtain the degree of influence of article structure, which we then compare to other potential influences on navigation. We study a range of influences on aggregated Wikipedia click data in the form of (i) the clickstream (which is the aggregate of many sessions) and (ii) the aggregate of all clicks from the Wikispeedia logs. To this end, we combine all clicks from all Wikispeedia games in Wikispeedia and count the clicks on each link for each article. This aggregation therefore excludes all information about game paths and allows us to compare both datasets in the same form.

5.1. *Influencing factors*

We investigate the following influences on navigation:

Article structure: Wikipedia articles follow specific guidelines that lead to structural regularities. We examine the influence of
- **Lead:** The first section of an article
- **Infobox:** An optional tabular description of the article's main facts

Generality: We investigate three factors for the generality of articles (cf. Section 4.1):
- **Indegree:** generality as seen by Wikipedia editors
- **View count:** generality and popularity as seen by Wikipedia visitors
- **Search query n-gram frequency (n-grams):** generality and familiarity as seen by search engine users

TF-IDF similarity to the target: We measure the cosine term frequency-inverse document frequency (TF-IDF) similarity between an article and the target article of goal-directed navigation (only evaluated for wikigames, where targets are explicitly known).

5.2. *Link selection models*

As a large share of articles on Wikipedia includes repeated links to the same target (cf. Section 4.2), we cannot resort to simply counting the number of clicks in each section. Therefore, to model link selection, we assign each link a probability based on the influencing factors. We then evaluate the models and compare them to the clicks made by users. This approach implies a memoryless navigation process, which has been found to be a good fit for human navigation of Wikipedia (Singer *et al.* 2014).

Approach. We investigate the influence of infoboxes and the lead sections as follows. First, we set a probability p for the section (e.g., $p = 0.6$). We then distribute this probability uniformly over all links within the section and remaining probability of $1-p$ (e.g., 0.4) over the links in the remainder of the article. We investigate values for $p \in \{0.01, 0.02, \dots 0.99\}$.

For indegree, view count and search query n-grams, we set the link selection probability to be proportional to the value of these factors for the link target. For example for indegree, we first compute the sum s of the indegrees d_i of all link targets i reachable from a given article and then assigned each link target i a probability of $\frac{d_i}{s}$.

To model the influence of textual similarity to the navigation target, we model the TF-IDF cosine similarity between the text of the current and the target article and assign each link a probability proportional to the TF-IDF similarity of the link target to the navigation target. This is possible only for Wikispeedia, where navigation targets are explicitly known.

Finally, we use a uniform model as the baseline, where we assign each link l a weight of $\frac{l}{L}$, where L is the number of links in the article (and hence assign each link a uniform click probability).

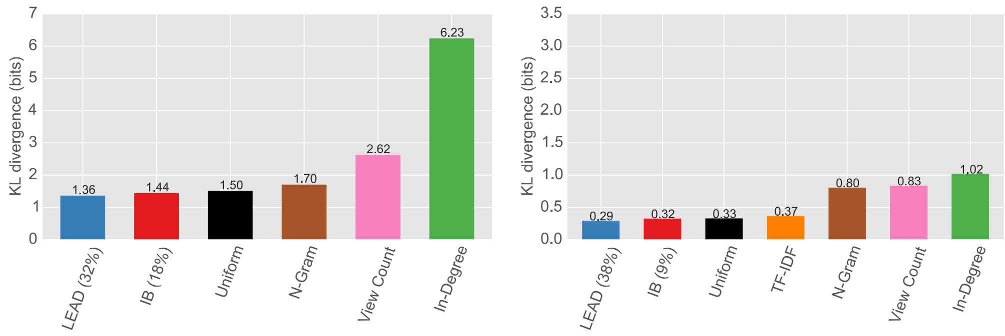
Kullback-Leibler divergence. To compare the models to the ground truth (i.e., the clicks made by users, we use the Kullback-Leibler divergence. Let p and q be two discrete probability distributions on $\mathcal{X}$. The Kullback-Leibler divergence of q from p is then

$$D(p\,||\,q) = \sum_{x \in \mathcal{X}} p(x) \log_2 \frac{p(x)}{q(x)}. \tag{1}$$

The Kullback-Leibler divergence measures the distance of the distribution q from p and states the expected number of additional bits needed to code samples from p when an optimal code for q is used instead. This gives us a measure for how well a distribution can be used to approximate another. We use a Laplace smoothing of 0.0001 for all values to avoid any problems with zero entries.

Model evaluation. We evaluate the effects of the influencing factors as follows:

(1) We count the number of all user clicks going away from an article of the dataset. We denote these as the *outclicks*.
(2) We use the influence model to compute the link selection probability of each link in an article. We then multiply these probabilities by the number of *outclicks* registered from that article. For example, for a link with selection probability of 0.2 for a model and 8 registered *outclicks* on the article, this would result in a value of 1.6 for the target article.
(3) We count the sum of values received this way by each article from any other article. We denote these as the *inclicks* of articles.
(4) Finally, we normalize the *inclicks* over all articles. We then compare this distribu-



(a) Wikipedia Clickstream (b) Wikispeedia

Figure 3.: **Kullback-Leibler (KL) divergences for link selection models when substituting for the ground truth distribution of clicks.** The figure shows the KL divergence when substituting the distribution of clicks to articles with a range of link selection models. For the models regarding infobox (IB) and lead, only the models with the smallest KL divergence are shown. Models based on article structure explain navigation choices best, followed by the uniform model. Generality-based models are not a good fit.

tion of *inclicks* to articles to the normalized ground truth from the dataset (i.e., the number of times articles were clicked by users) by computing the Kullback-Leibler divergence for substituting the ground truth with the result of the link selection models.

5.3. *Results*

Figure 3 shows the results of the comparison of Kullback-Leibler divergences of the distributions when substituting for the ground truth distributions.

Article structure best explains navigation choices. The models with the lowest Kullback-Leibler divergence to the ground distribution are the ones placing more importance on lead or infobox. The resulting models also represent the importance of these sections: Wikipedia is best fit with 32% of all clicks made in the lead or 18% in the infobox. This is substantially more than the link counts would suggest: For Wikipedia, the lead contained 17% and the infobox contained 4% of the links on articles. This indicates that these section are considerably more important for users than the number of present links would suggest. The same holds true for Wikispeedia, where the infobox (4% of all links) and the lead (6% of links) are again better fitted with higher importance weights.

Generality is not a good explanation for navigation choices. Indegree, view count and search query n-grams all lead to distributions with large differences to the ground truth. A tendency to click on more general articles is therefore not a good fit for explaining the aggregate of all clicks.

Selecting links uniformly is a good fit. The baseline model, selecting links uniformly at random across the entire article, is able to fit navigation choices comparatively well. This indicates that navigation uniformly at random (such as the one used to compute PageRank) can serve as a useful model for explaining the aggregate of all clicks to Wikipedia.

TF-IDF similarity to the target is a fairly good fit. The TF-IDF similarity to the target leads to a larger divergence than the uniform model, but still considerably better than the generality models. This suggests that textual similarity plays an important role in the goal-directed navigation of Wikigames.

Both datasets show similar characteristics. The Wikipedia clickstream and the Wikispeedia datasets we investigated consist of data from different navigational situtations: While the Wikispeedia dataset is restricted to focused, goal-directed navigation, the Wikipedia clickstream dataset comprises a large range of forms of navigation. Despite this difference, the navigational influences for both datasets are very much alike. For both datasets, the order in goodness of fit for the resulting models is exactly the same. This indicates that, in their aggregate form, Wikispeedia is subject to the same influences as the Wikipedia in its entirety. The most notable difference is the larger Kullback-Leibler divergence for Wikispeedia for N-Gram and View Count. However, this is likely due to the restriction to a subset of articles (cf. Section 4.1).

Overall, article structure best explains Wikipedia navigation when aggregating over all clicks and navigation scenarios. However, we see that TF-IDF similarity to the target is also a fairly good explanation for navigational choices in Wikispeedia. In the light of these findings, we now investigate the influence of structure in a step-by-step analysis of goal-directed navigation.

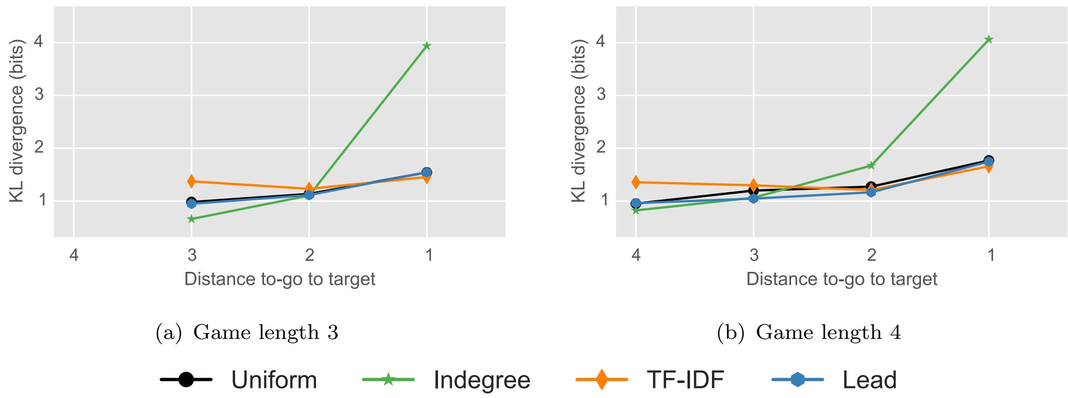
6. Influences on goal-directed Wikipedia navigation

We now shift our analysis to goal-directed Wikipedia navigation in the form of wikigames. Wikigames challenge users to retrieve articles by following links in articles without using any external help or the search function. As such, they are examples of goal-directed navigation: starting from a given article, the aim is to find a target article with as few clicks as possible. Log files from these games have the inherent advantage of explicitly specifying the navigation target at all times, therefore permitting us to examine navigation step-by-step.

Two navigational phases have been identified for wikigames: (i) an initial zoom-out phase where players navigate to high-degree nodes, and (ii) a final home-in phase to the target, in which players navigate based on textual similarity to the target (West and Leskovec 2012b). In what follows, we aim to identify the impact of article structure on the goal-directed navigation of wikigames.

Approach. To quantify the impact of different influences on goal-directed navigation, we conduct the simulation and analysis of link selection models as described in Section 5.2. However, we now evaluate the models step by step. To study both the initial and the final stage of the game, we restrict our analysis to successful games, for which users were able to find the target article. These make up around two thirds of the dataset (around $50,000$ out of $75,000$ games). Our approach was as follows:

(1) We split up all games by (a) shortest possible solution and (b) game length (i.e., clicks it took the user to find the goal). This partitions games into classes, e.g., all paths of length eight for games with a shortest solution of three clicks.
(2) For each class of games, we evaluate the link selection models separately for every step. We restrict the ground truth clicks to the ones performed in one step of the class and thus model each step in every partition class on its own.



(a) Game length 3 (b) Game length 4

Figure 4.: **Comparison of navigation strategies.** The figures show the Kullback-Leibler (KL) divergences of the link selection models to the user clicks for the top three models and the random baseline for games with a shortest possible solution of three clicks for which users found solutions of (a) three and (b) four clicks. In the beginning of games, the indegree model performs best. Towards the end, the TF-IDF model has the lowest KL divergence, while in between these phases, article structure is a good fit. This shows that the influence of article structure is notable in all but the first and last clicks.

(3) We compare the models to the ground-truth models for each step by computing the Kullback-Leibler divergence.

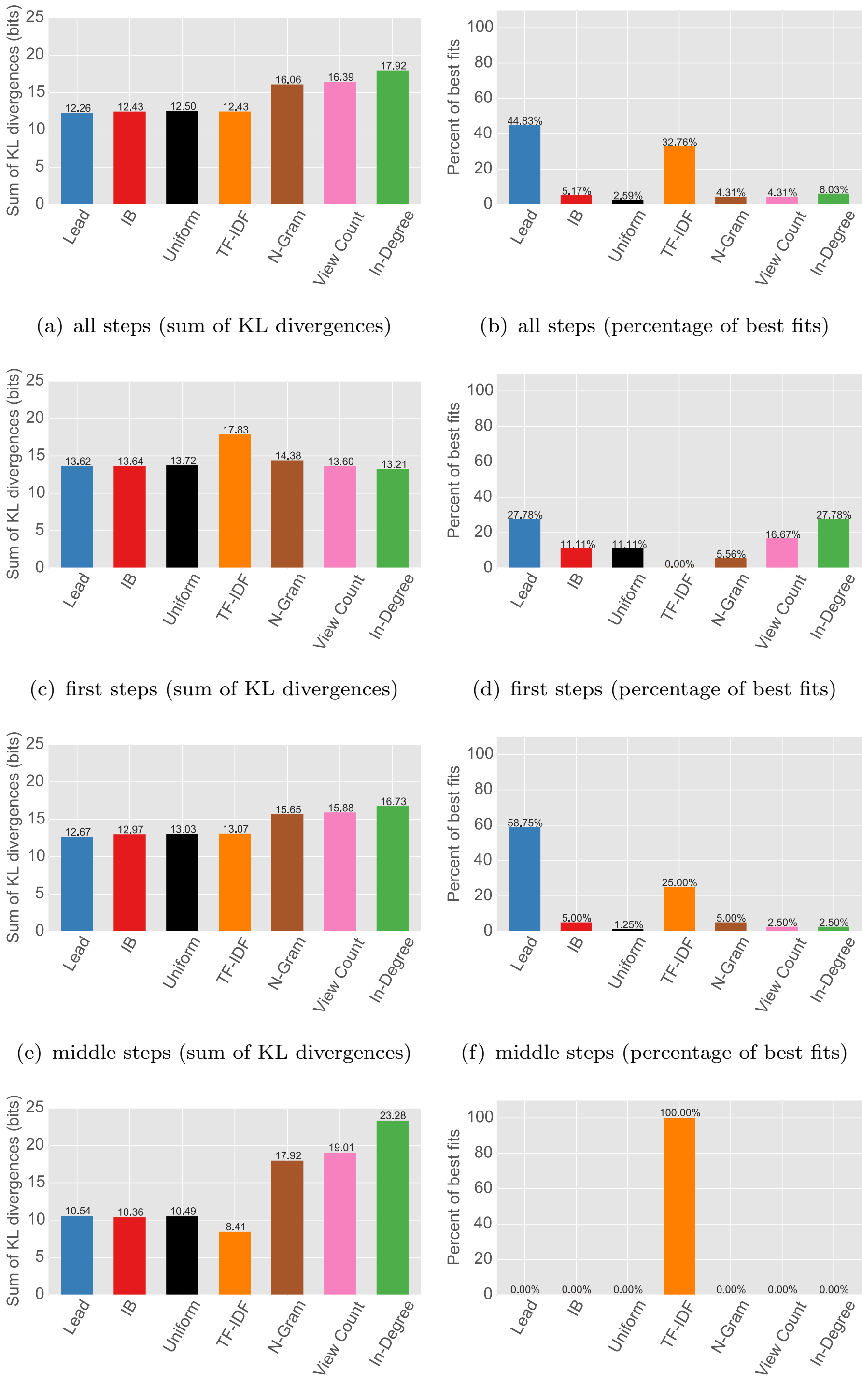
This approach leaves us with models for every step in every class of the partitioned games and allows us to compare the corresponding best fits.

Results. Figure 4 shows the results of the application of the link selection models to two classes of (short) games and displays the three best-performing models, the uniform model and their resulting Kullback-Leibler divergences to the user click distribution. The results show that in the beginning of games, the indegree model performs best. Towards the end, the TF-IDF model has the lowest Kullback-Leibler divergence. These correspond to the zoom-out and the home-in phases in line with the work of West and Leskovec (2012b). All other steps (in the central stages of the games), however, are best modeled based on article structure.

We now analyze all successful games with shortest possible solutions of 3, 4 and 5 clicks and user paths with up to 10 clicks, which comprises a total of $23,717$ games and $76,874$ clicks. Figure 5 shows the results of this step-by-step analysis of a total of 146 link selection models (one for every step in every class of games).

Generality explains the initial phase. The first steps in games are best explained by generality-based measures (indegree, view count and n-grams), which together make up half of the best fits for all first steps. This confirms that in the initial phase of navigation, users tend to select a link leading to a general and high-degree article—the zoom-out phase. A likely explanation for this phase would be that users start out from an unfamiliar article and try to get to a landmark serving as a point of orientation.

TF-IDF similarity explains the endgame. The last clicks in games are overwhelmingly best fit by TF-IDF similarity models. This again confirms previous work, which found that users tend to navigate by textual similarity in the endgame.



(a) all steps (sum of KL divergences)

(b) all steps (percentage of best fits)

(c) first steps (sum of KL divergences)

(d) first steps (percentage of best fits)

(e) middle steps (sum of KL divergences)

(f) middle steps (percentage of best fits)

(g) last steps (sum of KL divergences)

(h) last steps (percentage of best fits)

Figure 5.: **Step-by-step analysis of successful games.** The figure shows the analysis of successful navigation paths. Games were partitioned into classes by (a) shortest possible solution (3, 4 and 5 clicks) and (b) game length (i.e., clicks it took the user to solve it, maximum 10). E.g., all paths of length eight for games with a shortest solution of three clicks. We then evaluated click models to every step in these partitioned games, leaving us with 146 models. The left column shows the sum of KL divergences for these models. The right column shows the fraction of stepwise results for which a model was the best fit. Article structure is able to explain navigational choices best in all except the first and last steps.

Overall, article structure best fits the largest fraction of steps. For combination of the models for all steps, the lead model has the smallest Kullback-Leibler divergence to the distribution of user clicks. Together, the lead and the infobox models are the best fit for half of the steps in the games, followed by the TF-IDF similarity to the target. When we exclude the initial and final phases of games, article-based models best fit around 64% of all steps. This indicates that in all except the initial and final phases of games, the bias to article structure is actually substantially stronger than biases to generality or TF-IDF similarity.

Influence of game design. The step-by-step analysis of navigation paths shows that the first and last steps are substantially different from the remainder of clicks. This is possibly due to the influence of game design on these paths: The first click frequently serves to get to a landmark article serving as a point of orientation. If we compare this to regular Web use, where powerful search engines are available, it appears less likely that the first step would be necessary in a setting where users find it easy to start information retrieval with a query to a search engine.

Regarding the last step, we notice that links in the Wikipedia for Schools selection of articles (on which Wikispeedia is based) occur proportionally to the TF-IDF similarity between articles (West and Leskovec 2012b). It therefore appears likely that this causes the last link in navigation paths to be best explainable by TF-IDF similarity.

If we exclude the first and last steps and examine the combination of the remainder of steps, the models are in the exact same order as the aggregate of all steps for the Wikipedia clickstream as well as the aggregate of all clicks in successful and unsuccessful games for Wikispeedia (cf. Figure 3). We therefore argue that the first and last steps are outliers in a sense—the reasons for the click decision in these steps could well be due to the way the wikigame is set up and not reveal the true navigational needs of users. As such, we believe that a bias towards article structure is able to explain the largest fraction of navigational decisions in goal-directed navigation.

In summary, these results suggest that the characteristic first and last steps of wikigames are subject to different conditions. Outside these two phases, however, navigational choices can best be modeled by focusing on the article structure as well as the textual similarity to the navigation target.

7. Discussion

In this paper we studied human navigation of Wikipedia by investigating logs of click data from two perspectives: (i) an combination of many usage scenarios (the Wikipedia clickstream) and (ii) focused goal-directed navigation (Wikispeedia). We investigated the following research question:

Research question: To what extent does article structure affect Wikipedia navigation?

We addressed this question by looking at influences on Wikipedia navigation from the perspectives of aggregated navigation and of goal-directed navigation.

Aggregated navigation. The aggregated navigation data combines measurements from a large number of visits to Wikipedia. For this data, our results show that the structure of Wikipedia articles has a substantial influence on human navigation. We compared a range of potential navigational influences and found that, overall, article structure has a larger influence on navigational choices than generality. The influence of TF-IDF similarity was substantial, but slightly less strong than the influence of article structure. This confirms previous work that has found both semantic and structural knowledge to influence navigation performance (Juvina and van Oostendorp 2008). Our analysis showed that links occurring in the lead section or the infobox play an especially important role for navigation. The best-fitting models for the lead and infobox placed a weight on these sections that was substantially higher than the number of links in them would lead to expect. These results hold true for the aggregate of all clicks recorded for the Wikipedia clickstream as well as for wikigames. In fact, the order of best fits of the navigational influences for wikigames exactly matches the one for Wikipedia. This suggests that the aggregate clicks of goal-directed wikigames are very similar in nature to free-form Wikipedia navigation.

Goal-directed navigation. A more detailed analysis of goal-directed navigation in the form of wikigames showed that the navigational decisions in the navigation paths are subject to very similar influences as the aggregated clicks. If we exclude the first and last steps—which have been found to be special cases and might well be caused by the specific setup of the games—the influences are ranked in exactly the same order as for the aggregate of all clicks. Our analysis suggests that the elaborate structure of Wikipedia articles, with a lead-section at the start of the article linking to broader, better known concepts, helps users in the navigation task by making high-degree hubs easier to find. Thus, even if a user were completely lost, clicking randomly on the links appearing on their screen (which tend to be the links in the article lead and infobox), is likely to bring them to a familiar, broader concept, which the user can then use as a point of orientation in the information space. In effect, it is the structure of information in the article that guides navigation, and not necessarily to the network structure.

Comparison of free-form and goal-directed navigation. The Wikipedia clickstream covers all possible usage scenarios: looking up facts, learning about concepts, reading to pass time, and many more. The Wikispeedia dataset is more specific and is limited to goal-directed navigational games, which also make up an unknown fraction of the clicks in the clickstream. As such, we expected the navigational influences to differ between the datasets. However, our results have shown that the aggregated clicks from Wikispeedia led to the same order of influence model results as the Wikipedia clickstream. This could be due to two reasons: It could mean that a large fraction of the clicks in the clickstream stems from goal-directed navigation and therefore the link selection models hold true for them as well. This would imply a confirmation of our results on a second dataset. If, on the other hand, the majority of clicks from the clickstream stems from navigation scenarios other than goal-directed navigation, this would mean that the navigational influence models we presented are valid for the combined clicks of other scenarios as well. This in turn would imply that our models generalize to the aggregate of all Wikipedia usage scenarios. Due to the limited availability of Wikipedia click logs, we are unable to answer investigate this further at this point. However, should more detailed log data of Wikipedia become available, it would be fruitful to expand our work to it.

Limitations and future work. The setup as of Wikispeedia as a game as well as the reactive approach to collecting this data (users knew that their log data would be used for evaluations) could potentially introduce biases. However, our step-by-step analysis has shown that while the first and last clicks in Wikispeedia show different characteristics than the Wikipedia clickstream, we could find no substantial difference for the remainder of clicks. As such, we believe that wikigames, despite of their game setup, can provide us with valuable insight into real-world navigation behavior in large information networks. As mentioned, it would certainly be worth repeating this analysis on navigation paths collected from real-life navigation behavior on Wikipedia. This would also open up the opportunity to compare the navigation behavior of different types of users and different usage scenarios, which would deepen our understanding of the navigation dynamics on Wikipedia.

The structure-based link selection models were based on assigning a section of the article a higher weight. The simplicity of this approach raises the questions if more detailed position models could be able to even better explain click selection strategy. A possible extension of this work would be a combination of click models for several sections or influences. Another follow-up study could be to use two separate versions of Wikipedia articles restructured according to different information architectures and observe users navigate on them.

The step-by-step analysis of navigational influences could be used to assess performance levels of users. For example, the navigational influences at the current step could be compared to the typical influences on well-performing users at that step. This could be valuable to, for example, intelligent systems relying on data mining performed by its users. A first step in this direction could be the analysis of unsuccessful wikigames.

Finally, our analysis focuses on the desktop view of Wikipedia for both datasets. In future work, it would be interesting to repeat this analysis for the Wikipedia view for mobile devices, which display Wikipedia in a different design (e.g., infoboxes appear before the lead section instead of next to it as in the desktop view).

8. Conclusion

We have shown that the decentralized organization of Wikipedia leads to elaborate structure in the articles of the encyclopedia, greatly facilitating navigation. This structure and regularity substantially helps user to navigate the information network. Our results clearly demonstrate the navigational importance of links in terms of their position in an article. We have found evidence that a large share of clicks to articles in the English Wikipedia are to links in the lead and the infobox, which in turn suggests that the majority of visitors focus their attention on these sections. This finding is relevant for the shaping of Wikipedia policies and suggests that the upper part of articles should receive the most attention, e.g., in terms of fact checking or monitoring for vandalism. Our results also suggest that attention needs to be paid to edits that change an article only by moving text to another location: given our results, it seems likely that placing sections containing criticism before other content can drastically alter a user's impression of an article without even changing any of the words in it.

Our work also helps to understand the impact of the partitioning of Wikipedia articles into sections. Information from the lead section or the infobox is frequently used in external sites (e.g., by search engines). Our results show that this in fact frequently matches what users actually see when they look at an article itself. In terms of Wikipedia administration, special attention should therefore be paid to the neutrality and balance of the lead section.

9. Acknowledgments

We thank Xiaoran Yan and Patrick Kasper for valuable input. This research was supported in part by a grant from the Austrian Science Fund (FWF) [P24866].

References

Bates, M.J., 1989. The Design of Browsing and Berrypicking Techniques for the Online Search Interface. *Online Information Review*, 13 (5), 407–424.

Benz, D., *et al.*, 2011. One Tag to Bind Them All: Measuring Term Abstractness in Social Metadata. *The Semantic Web: Research and Applications*, 360–374.

Blunch, N.J., 1984. Position Bias in Multiple-Choice Questions. *Journal of Marketing Research*, 21 (000002), 216.

Buscher, G., Cutrell, E., and Morris, M.R., 2009. What Do You See When You're Surfing? Using Eye Tracking to Predict Salient Regions of Web Pages. *In*: J. Dan R Olsen, R.B. Arthur, K. Hinckley, M.R. Morris, S. Hudson and S. Greenberg, eds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* New York, NY: ACM, 21–30.

Buscher, G., Dumais, S.T., and Cutrell, E., 2010. The Good, the Bad, and the Random: an Eye-Tracking Study of Ad Quality in Web Search. *In*: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* ACM, 42–49.

Chi, E.H., *et al.*, 2001. Using Information Scent to Model User Information Needs and Actions on the Web. *In*: J. Jacko and A. Sears, eds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* New York, NY: ACM, 490–497.

Craswell, N., *et al.*, 2008. An Experimental Comparison of Click Position-Bias Models. *In*: M.N. Microsoft, A. Broder and S. Chakrabarti, eds. *Proceedings of the First International Conference on Web Search and Data Mining* ACM, 87–94.

Gabrilovich, E. and Markovitch, S., 2009. Wikipedia-Based Semantic Interpretation for Natural Language Processing. *Journal of Artificial Intelligence Research*, 34 (1), 443–498.

González-Caro, C. and Marcos, M.C., 2011. Different Users and Intents: An Eye-Tracking Analysis of Web Search. *Proceedings of the 4th ACM International Conference on Web Search and Data Mining*, 9–12.

Helic, D., *et al.*, 2013. Models of Human Navigation in Information Networks Based on Decentralized Search. *In*: G. Stumme and A. Hotho, eds. *Proceedings of the 24th ACM Conference on Hypertext and Social Media* New York, NY: ACM, 89–98.

Helic, D., *et al.*, 2011. Pragmatic Evaluation of Folksonomies. *In*: S. Sadagopan, K. Ramamritham, A. Kumar, M.P. Ravindra, E. Bertino and R. Kumar, eds. *Proceedings of the 20th International Conference on World Wide Web* New York, NY: ACM, 417–426.

Juvina, I. and van Oostendorp, H., 2008. Modeling Semantic and Structural Knowledge in Web Navigation. *Discourse Processes*, 45 (4-5), 346–364.

Juvina, I., *et al.*, 2005. Toward Modeling Contextual Information in Web Navigation. *In*: *Proceedings of the 27th Annual Meeting of the Cognitive Science Society* Lawrence Erlbaum Associates, 1078–1083.

Kitajima, M., Blackmon, M.H., and Polson, P.G., 2000. A Comprehension-Based Model of Web Navigation and its Application to Web Usability Analysis. *In*: *People and Computers XIV — Usability or Else!.*, 357–374 Springer London.

Kitajima, M., Polson, P.G., and Blackmon, M.H., 2007. CoLiDeS and SNIF-ACT: Complementary Models for Searching and Sensemaking on the Web. *In*: *Human Computer Interaction Consortium 2007 Winter Workshop* Human Computer Interaction Consortium.

Kleinberg, J.M., 2000. Navigation in a Small World. *Nature*, 406 (6798), 845.

Kleinberg, J.M., 2001. Small-World Phenomena and the Dynamics of Information. *In*: T.G. Dietterich, S. Becker and Z. Ghahramani, eds. *Advances in Neural Information Processing Systems 14* Cambri9dge, MA: MIT Press, 431–438.

Lamprecht, D., *et al.*, 2015a. Improving Recommender System Navigability Through Diversification: A Case Study of IMDb. *In*: *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies* ACM, 21:1–21:8.

Lamprecht, D., *et al.*, 2015b. Using Ontologies to Model Human Navigation Behavior in Information Networks: A Study Based on Wikipedia. *Semantic Web*, 6 (4), 403–422.

Lerman, K. and Hogg, T., 2014. Leveraging Position Bias to Improve Peer Recommendation. *PLOS ONE*, 9 (6), e98914.

Liu, C., White, R.W., and Dumais, S., 2010. Understanding Web Browsing Behaviors Through Weibull Analysis of Dwell Time. *In*: *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval* ACM, 379–386.

Liu, Y., *et al.*, 2014. From Skimming to Reading: A Two-Stage Examination Model for Web Search. *In*: *Proceedings of the 23rd ACM International Conference on Information and Knowledge Management* ACM, 849–858.

Marchionini, G., 2006. Exploratory Search: From Finding to Understanding. *Communications of the ACM*, 49 (4), 41–46.

Milgram, S., 1967. The Small World Problem. *Psychology Today*, 1 (2), 60–67.

Nielsen, J., 2006. F-Shaped Pattern For Reading Web Content. [online] `www.nngroup.com/articles/f-shaped-pattern-reading-web-content` [February 17, 2016].

Pan, B., *et al.*, 2007. In Google We Trust: Users' Decisions on Rank, Position, and Relevance. *Journal of Computer-Mediated Communication*, 12 (3), 801–823.

Payne, S.L., 1951. *The Art of Asking Questions*. Princeton, NJ: Princeton University Press.

Pirolli, P. and Card, S., 1999. Information Foraging. *Psychological Review*, 106 (4), 643.

Pirolli, P. and Fu, W.T., 2003. SNIF-ACT: A Model of Information Foraging on the World Wide Web. *In*: *9th International Conference on User Modeling*. Springer.

Salganik, M.J., Dodds, P.S., and Watts, D.J., 2006. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311 (5762), 854–856.

Salmerón, L., Cerdán, R., and Naumann, J., 2015. How Adolescents Navigate Wikipedia to Answer Questions. *Infancia y Aprendizaje*, 38 (2), 435–471.

Salmerón, L., Kammerer, Y., and García-Carrión, P., 2013. Searching the Web for Conflicting Topics: Page and User Factors. *Computers in Human Behavior*, 29 (6), 2161–2171.

Scaria, A.T., *et al.*, 2014. The Last Click: Why Users Give up Information Network Navigation. *In*: B. Carterette, F. Diaz, C. Castillo and D. Metzler, eds. *Proceedings of the 7th ACM International Conference on Web Search and Data Mining* New York, NY: ACM, 213–222.

Seyerlehner, K., Flexer, A., and Widmer, G., 2009. On the Limitations of Browsing top-N Recommender Systems. *In*: *Proceedings of the 3rd ACM Conference on Recommender systems* ACM, 321–324.

Singer, P., *et al.*, 2014. Detecting Memory and Structure in Human Navigation Patterns

Using Markov Chain Models of Varying Rrder. *PLOS ONE*, 9 (7), e102070.

Singer, P., *et al.*, 2013. Computing Semantic Relatedness from Human Navigational paths: A Case Study on Wikipedia. *International Journal on Semantic Web and Information Systems*, 9 (4), 41–70.

Teevan, J., *et al.*, 2004. The Perfect Search Engine is not Enough: A Study of Orienteering Behavior in Directed Search. *In*: E. Dykstra-Erickson and M. Tscheligi, eds. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* New York, NY: ACM, 415–422.

Wang, K., *et al.*, 2010. An Overview of Microsoft Web N-gram Corpus and Applications. *In*: C.P. Rosé, ed. *Proceedings of 11th Annual Conference of the North American Chapter of the Association for Computational Linguistics (Demonstration Session)* New York, NY: Association for Computational Linguistics, 45–48.

Watts, D.J. and Strogatz, S.H., 1998. Collective Dynamics of Small-World Networks. *Nature*, 393 (6684), 440–442.

Weinreich, H., *et al.*, 2006. Off the Beaten Tracks: Exploring Three Aspects of Web Navigation. *In*: L. Carr, D.D. Roure, A. Iyengar, C. Goble and M. Dahlin, eds. *Proceedings of the 15th International Conference on World Wide Web* New York, NY: ACM, 133–142.

West, R. and Leskovec, J., 2012a. Automatic Versus Human Navigation in Information Networks. *In*: J. Breslin, N. Ellison, J.G. Shanahan and Z. Tufekci, eds. *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media* Palo Alto, CA: AAAI.

West, R. and Leskovec, J., 2012b. Human Wayfinding in Information Networks. *In*: A. Mille, F. Gandon, J. Misselis, M. Rabinovich and S. Staab, eds. *Proceedings of the 21st International Conference on World Wide Web* New York, NY: ACM, 619–628.

West, R., Paranjape, A., and Leskovec, J., 2015. Mining Missing Hyperlinks from Human Navigation Traces: A Case Study of Wikipedia. *In*: A. Gangemi, S. Leonardi and A. Panconesi, eds. *Proceedings of the 24th International Conference on World Wide Web* New York, NY: ACM, 1242–1252.

West, R., Pineau, J., and Precup, D., 2009. Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts. *In*: H. Kitano, ed. *Proceedings of the 21st International Joint Conference on Artificial Intelligence* San Francisco, CA: Morgan Kaufmann Publishers Inc., 1598–1603.

Wikipedia, 2016a. Wikipedia: Help:Infobox. [online] `https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Layout` [February 17, 2016a].

Wikipedia, 2016b. Wikipedia: Help:Infobox. [online] `https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Infoboxes` [February 17, 2016b].

Wikipedia, 2016c. Wikipedia: Manual of Style/Lead Section. [online] `http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section` [February 17, 2016c].

Wikipedia, 2016d. Wikipedia: Manual of Style/Linking. [online] `http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Linking` [February 22, 2016d].

Wikipedia, 2016e. Wikipedia: Writing better articles. [online] `https://en.wikipedia.org/wiki/Wikipedia:Writing_better_articles` [February 17, 2016e].

Wulczyn, E. and Taraborelli, D., 2015. Wikipedia Clickstream. [online] `http://dx.doi.org/10.6084/m9.figshare.1305770` [February 17, 2016].

3.3 A Method for Evaluating the Navigability of Recommendation Algorithms

This article aims to answer the second research questions, that is, how navigation can be modeled. To this end, the article presents an evaluation framework to model navigation and demonstrates it on the example of recommender systems.

Recommender systems are applied by a wide range of websites to automatically construct the link structure among items in a collection, such as movies, books, or questions. The evaluation of the fitness of recommendation algorithms has so far mostly centered on the accuracy of the predictions, as for example measured by the error between predicted ratings and the real user ratings. The presented evaluation technique extends the arsenal of evaluation methods with a technique to assess path-based metrics of recommendation algorithms by modeling navigation behavior based on decentralized search.

This article presents a comprehensive evaluation framework for navigation on websites with recommendations. The framework is based on three established information seeking scenarios, namely point-to-point search, berrypicking and information foraging. The article demonstrates the feasibility of the application of the framework by evaluating it on three datasets and four non-personalized recommendation algorithms. Furthermore, the article demonstrates the applicability to personalized recommendation algorithms.

The presented evaluation framework adds a powerful, novel method to the repertoire of evaluation techniques for recommendation algorithms and allows for a multi-click analysis of the effects of automatically generated recommendations on the structure of an information network. The results can help website operators to select an appropriate recommendation algorithm based on any desired navigability criteria. The method is general and applicable to arbitrary information networks.

A Method for Evaluating the Navigability of Recommendation Algorithms

Daniel Lamprecht, Markus Strohmaier, and Denis Helic

Abstract Recommendations are increasingly used to support and enable discovery, browsing and exploration of large item collections, especially when no clear classification of items exists. Yet, the suitability of a recommendation algorithm to support these use cases cannot be comprehensively evaluated by any evaluation measures proposed so far. In this paper, we propose a method to expand the repertoire of existing recommendation evaluation techniques with a method to evaluate the navigability of recommendation algorithms. The proposed method combines approaches from network science and information retrieval and evaluates navigability by simulating three different models of information seeking scenarios and measuring the success rates. We show the feasibility of our method by applying it to four non-personalized recommendation algorithms on three datasets and also illustrate its applicability to personalized algorithms. Our work expands the arsenal of evaluation techniques for recommendation algorithms, extends from a one-click-based evaluation towards multi-click analysis and presents a general, comprehensive method to evaluating navigability of arbitrary recommendation algorithms.

1 Introduction

Websites with large collections of items need to support three ways of information retrieval: (i) retrieval of familiar items (ii) retrieval of items that cannot be explicitly described but will be recognized once retrieved and (iii) serendipitous discovery [32]. For a website with a large collection of items, such as an e-commerce website, (i) can be enabled with a full-text search function. For (ii) and (iii), however, a search

Daniel Lamprecht and Denis Helic
KTI, Graz University of Technology, e-mail: `daniel.lamprecht@tugraz.at, dehlic@tugraz.at`

Markus Strohmaier
GESIS and University of Koblenz-Landau, e-mail: `strohmaier@uni-koblenz.de`

function is generally not sufficient. These types of information retrieval are therefore often supported by recommendations that connect items and enable discovery and navigation.

Users have been found to enjoy perusing item collections such as e-commerce sites or recommender systems without the immediate intention of making a purchase [15]. More generally, some users prefer navigation to direct search even when they know the target [31]. For platforms where users immediately consume content, such as YouTube or Quora, recommendations serve the use case of *unarticulated want*, and are therefore a crucial part of the user experience [10]. In item collections that do not associate descriptions or metadata with content (such as videos) frequently no clear structuring of items exists, and recommendations play a vital role in the user interfaces. It is therefore critical for these systems to support discovery via links.

When a website provides recommendations along with each item, the items and the associated recommendations form a *recommendation network*—an implicit view of a recommender system where items are nodes and recommendations are edges. This type of recommendations are frequent on e-commerce websites, such as Amazon ("customers who bought this also bought"). Many websites associate a fixed number of recommendations with each item, which leads to a constant outdegree and a varying indegree for each node in the network

Knowing more about recommendation networks would give web-site operators the possibility to assess the effects of recommendations and help to produce recommendations that make it easier for users to discover and explore items. While a few studies have already looked at recommendation networks and provided first important insights into the nature and structure of these networks [6, 8, 20, 29], there is no systematic approach to evaluating the navigability of recommendation algorithms.

This paper presents a general method to evaluate the practical navigability of arbitrary recommendation networks by using simulations based on three navigation models established in the literature, namely *point-to-point navigation* [16], *navigation via berrypicking* [2] and *navigation via information foraging* [28]. The combination of established techniques from the fields of network science and information retrieval allows us to present a novel method that extends common evaluation measures towards a path-based evaluation and expands the arsenal of existing recommendation evaluation techniques.

We show the feasibility of this method by applying it to four non-personalized recommendation algorithms on three datasets and investigate their properties. We also illustrate the general suitability of our method to personalized recommendations and report initial results for a sample configuration.

2 Related Work

Initially, recommender systems were mostly evaluated in terms of prediction accuracy [12]. However, the focus on accuracy has been found to neglect other import applications of recommender systems such as support for the discovery of novel items, browsing, or diversified recommendations, and may lead to a bias towards popular items [8, 30] or a filter bubble effect [25]. For these reasons, a series of evaluation metrics for additional properties of recommender systems has been developed. These metrics include diversity [4, 7], novelty [7, 12], serendipity and coverage [11, 12, 15] and are considered orthogonal to prediction accuracy.

The evaluation method presented in this paper is rooted in Stanley Milgram's small world experiments [24], which laid the foundation for *decentralized search*. Kleinberg [17] and Watts [34] later formalized the property that a navigable network requires short paths between all (or almost all) nodes. Kleinberg also found that an *efficiently navigable* network possesses certain structural properties that make it possible to design efficient decentralized search algorithms that only have local knowledge of the network [16]. The delivery time of such algorithms is then sublinear in the number of network nodes. In this paper, we investigate the efficient navigability of recommendation networks through the simulation of navigation models based on decentralized search.

The static topology of recommendation networks has been extensively studied for the case of music recommenders [8, 29]. Their corresponding recommendation networks have been found to exhibit heavy-tail degree distributions and small-world properties [6], implying that they are efficiently navigable with local search algorithms. A first study [20] has already explored the reachability and navigability of the recommender systems of IMDb. The corresponding recommendation networks were shown to lack support for navigation scenarios. However, the use of diversified recommendations was able to substantially improve this and lead to more navigable recommendation networks. A similar methodology has been applied to suggest links to improve navigability on Wikipedia [19].

3 Evaluation Method

Navigation is at the core of exploration and browsing, which are important use cases of a recommender system, as many users find browsing pleasant [15], use it to discover novel content [22] or consume the content along the browsing path (e.g., on YouTube). A defining property of online navigation is that the knowledge about a website is mostly local: users only perceive the links emanating from the current page and generally only have intuitions about where those links might lead, but lack global knowledge about the system. In the case of a top-N recommender system, users are generally only aware of the recommendations provided with the current item.

The evaluation method we propose makes use of greedy decentralized search to simulate navigation in recommender systems and measures the success rate. This model has been used in previous work to analyze navigation dynamics in networks [13, 14] and has been found to produce comparable results to human navigation patterns [21, 33]. At each step, this algorithm evaluates a heuristic for every present link and greedily selects the one maximizing that heuristic. We take the heuristic to

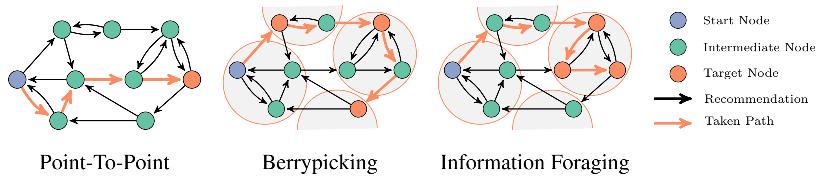


Fig. 1: **Information Seeking Scenarios.** We use three information seeking scenarios to study navigability of recommendation networks. The objective in point-to-point navigation is to find a single goal item. For berrypicking, we cluster the networks and set the goal of finding any one item in four clusters (shown in gray). For information foraging, the goal is to find multiple items in a single cluster.

represent vague intuitions about navigation that users might gain from looking at the descriptions of recommendation targets. For example, if a user was looking for a new science-fiction movie, they might be tempted to follow recommendations to other science fiction movies based on the title, a brief textual description or the displayed image. We use an implementation that does not revisit previously explored nodes. In case no unvisited item is present, the simulation backtracks.

A number of information seeking models have been established in the literature. To investigate the general suitability of recommendation algorithms to navigation based on different approaches, we evaluate navigation scenarios based on three of these models: point-to-point navigation [16], berrypicking [2], and information foraging [28]. For all scenarios, the start and target nodes in the network are determined independently of the network structure, i.e., regardless of whether the recommendation algorithm actually enabled a path between them. This allows us to fairly compare all recommendation algorithms and shows how well they support navigability. In what follows, we describe the three navigation scenarios in more detail (cf. Figure 1).

Point-To-Point Navigation. Point-to-point navigation [16] represents the task of finding a single target item in a recommendation network and models the navigational behavior of users with a specific item in mind that they cannot explicitly describe. For example, a user could try to find a science-fiction movie with a specific motif or to rediscover something on tip of their tongue. As such, this scenario covers point (ii) ("retrieval of items that cannot be explicitly described") of Toms's ways of information retrieval [32]. We then simulate navigation starting at the start node of a pair and with the objective of reaching the target node. As start-target pairs we sample pairs of nodes proportionally to how often they were corated by users in the corresponding rating dataset.

Navigation via Berrypicking. Berrypicking is an information seeking model which regards information seeking as a dynamic process where the information need is evolving and can be satisfied by multiple pieces of information in a *bit-at-a-time retrieval*—an analogy to picking berries on bushes [2]. Berrypicking can be thought of as covering points (ii) ("retrieval of items that cannot be explicitly described") and (iii) ("serendipitous discovery") of Toms's ways of information retrieval [32].

We model this scenario based on clusters, which we obtain with k-means based on the rating vectors. We randomly pick a first cluster and then draw one of the top four closest clusters based on Euclidian distance randomly. We then repeat this to find two more clusters. Starting from a randomly chosen node in the first cluster, the objective of the scenario is then to reach any node from the second cluster, followed by any node from the third and then the fourth cluster. In this way, the scenario models the evolving stages of berrypicking, where users inspect an item and adapt their information needs based on it.

Navigation via Information Foraging. Information foraging [28] is an information seeking theory inspired by optimal foraging theory in nature, where organisms have adopted strategies maximizing energy intake. For instance, when foraging on a patch of food, an animal must decide when to move on to the next patch (e.g., when finding apples on a tree is becoming too tedious). Some of the same mechanisms have identified for human information seeking behavior, where humans try to maximize information gain. Information can be modeled as occurring in patches, and information seekers as guided by *information scent* [9]. In a scenario based on information foraging, we model the scenario of depleting a patch of information. We assume that the objective is to retrieve nodes in a patch—guided by information scent in terms of the search heuristic. We take information foraging to model points (ii) and (iii) (”retrieval of items that cannot be explicitly described”) and ”serendipitous discovery”) of Toms’s ways of information retrieval [32].

Baselines. We also evaluate two baseline solutions: An *optimal solution* makes use of the shortest possible paths for a scenario (that users with perfect knowledge of the network could take). A *random solution* performs a random walk with no background knowledge at all.

4 Experimental Setup

We use three datasets for this paper:

- **MovieLens** is a film recommender systems maintained by GroupLens Research at the University of Minnesota. For this work, we use their dataset consisting of one million ratings from $6,000$ users on $4,000$ movies.
- **BookCrossing** is a book exchange platform. For this work, we use a 2005 crawl of the website [35]. We use only the explicit ratings, combine ratings for duplicate books and use ratings from users with ≥ 20 ratings on ≥ 5 books. This leaves us with roughly $50,000$ ratings by $1,088$ users on $3,637$ books.
- **IMDb** is a database of movies and TV shows. We use a 2015 crawl of the website [20], from which we use ratings for items published in 2013 and 2014 and condense them in the same way as for the BookCrossing dataset, resulting in 2.3M ratings for $6,690$ titles by $37,216$ users.

We calculate recommendations in the following way: For a given set of items I and a recommendation algorithm R, we use R to compute the pairwise similarities for all pairs of items $(i,j) \in I$. For each item $i \in I$, we then define the set of the top-N most similar items to i as $L_{i,N}$. We then create a directed top-N recommendation network $G(V,N,E)$, where $V = I$, N is the number of recommendations available for each item and $E = \{(i,j) \,|\, i \in I, j \in L_{i,N}\}$. This method leads to recommendation networks with constant outdegree and varying indegree—representing a typical setting.

For simplicity's sake, we investigate recommendation algorithms based on non-personalized recommendations. The similarities these recommendations are based on, however, are directly taken from the similarities used in the personalized recommendation algorithms. They therefore represent the recommendation networks as an unregistered or newly registered user would see them. For most websites, the vast majority of visitors does not contribute or register—this is known as the *90-9-1 Rule* (90% lurkers, 9% intermittent contributers and 1% heavy contributers) [26, 27]. However, our method is general and also applicable to personalized recommendation algorithms, which we exemplarily demonstrate in Section 6.

We use the following four recommendation algorithms in this work:

Association Rules (AR). Association rules are based on the market-basket model, where, in this case, we put all items rated by the same user into a basket and regard ratings as binary (i.e., rated/not rated). For every ordered pair of items (i,j), we then rank all items by how much more likely an item is to be consumed after a given item was consumed (similiar to the Apriori algorithm [1]). Specifically, we compute the fraction of co-ratings of i and j over the total ratings of i (i.e., the fraction users who rated both i and j, out of those who rated i). Let U_i be the set of users who rated item i. We can then compute this as as $(|U_i \cap U_j|)/(|U_i|)$. To compensate for the popularity of j, we then divide by the fraction of users who did not rate i but still rated j. Let $\overline{U}_i$ be the set of users who did not rate item i. We can then divide by $(|\overline{U}_i \cap U_j|)/(|\overline{U}_i|)$ to counter the effect of highly popular items that are likely to be co-rated with every item, but would not be very useful as a recommendation. We then take the top-N items most likely to be co-rated with it.

Collaborative Filtering (CF). For a given user u and an unrated item i, item-based collaborative filtering predicts the rating of u for i from a small number of other items that u previously rated. These other items are commonly selected as the ones maximizing the centered cosine similarity to i. The rating prediction is then computed as the weighted sum of their ratings, weighted by their similarity. To obtain unpersonalized recommendations, we compute the centered cosine similarity of an item i to all other items j in the dataset and use the top-N.

Interpolation Weights (IW). Interpolation weights are computed in a similar way to item-based collaborative filtering. However, instead of using a predefined similarity measure (such as the centered cosine similarity) to weight the contributions of other ratings, *interpolation weights* representing the relations between pairs of items are learned from the data. We use gradient descent to learn item-based interpolation weights by minimizing the root-mean square error for predictions on a test set [3] and then use the resulting weights as the similarity measure to obtain the top-N most similar items to an item.

Matrix Factorization (MF). Matrix factorization describes both items and users of a recommender system by affinities to a number of latent factors [18]. To find these factors, this algorithm factorizes the rating matrix U into two matrices as $U = Q^T P$ that represent the associations of users and items with the latent factors. We learn these matrices by minimizing the root-mean-square prediction error on a test set with gradient descent. After this minimization, we represent each item by the vector of its association with the latent factors and compute the centered cosine similarity between the latent factors for all pairs of items to obtain the top-N most similar items.

As the heuristic for decentralized search, we use the TF-IDF cosine similarity of brief textual descriptions of titles (namely title and plot summary of IMDb for the movies and the summary provided by GoodReads for the books). At each step, the simulation uses this heuristic to select the link leading to the item that has the highest TF-IDF cosine similarity to the navigation goal. We use a heuristic independent of ratings to decouple it from the recommendations used to generate the networks. For sake of brevity, we only report the results for a deterministic greedy search with 50 steps. However, we also evaluated all simulations for 10 and 25 steps as well as with an ε-greedy approach [13] and found that, while the total success rates decreased, the relative differences between the approaches did not change.

We evaluate a total of $1,200$ navigation simulations per scenario. For the clusters, we only use those consisting of 4–30 nodes to balance the difficulty. As a difference to the point-to-point navigation scenario, the target of the navigation for the berrypicking and information foraging scenario is not represented by a single node but by the centroid of the target cluster. The TF-IDF cosine similarity of a potential link target l is therefore represented by the average of the similarity between l and all items in the target cluster.

5 Results

Point-To-Point Navigation. The first row of Figure 2 displays the success rate (i.e., the fraction of successful simulations) for point-to-point navigation. Since the number of steps per simulation (50) is larger than the distances between all start-target pairs in the recommendation networks, the optimal solutions (shown in gray bars) correspond to all start-target pairs between which a path of any length existed. The optimal solution is therefore a measure of how well a recommendation algorithm theoretically supports this navigation scenario. The second baseline approach is a random walk, which shows the success rates achievable by an uninformed random process and serves to demonstrate that the simulations based on greedy search are able to exploit the link selection heuristic to reach navigation goals. The simulation for point-to-point navigation with greedy search for $N = 5$ recommendations leads to an average success rate of 6.86%. This indicates that users would be able to retrieve only a very small share of items in the recommender systems by focused point-to-point navigation. For $N = 20$ recommendations, the success rates increase substantially (average of 24.4%). Recommendations generated by interpolation weights lead to the best success rates (42–48%).

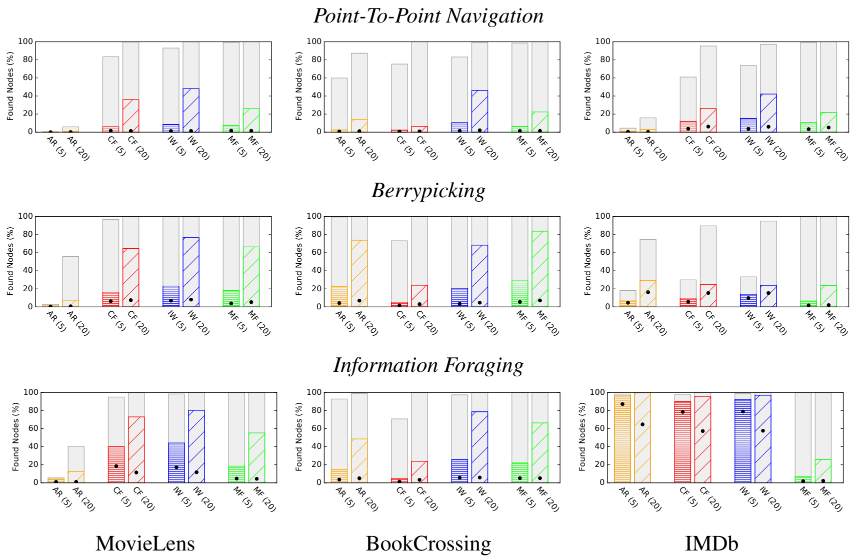


Fig. 2: **Success Ratios for the navigation simulations.** The bars depict the average percentage of found targets. Baseline success rates are depicted as gray bars (optimal solutions) and black dots (random walk solutions). Success rates are computed as the average number of found targets. Recommendation networks generated by interpolation weights (IW) generally performed best.

Navigation via Berrypicking. For five recommendations, the success rates for the case of genre-based clusters are 14.5% on average. With 20 recommendations, this increases to to 47%. Since the targets consists of three clusters, a success rate of 33% indicates that an average of one cluster was found.

The success rates for the IMDb dataset are substantially lower than for the other two datasets. A more detailed analysis shows that the networks for IMDb are clustered more strongly than those of the other two datasets. For a dynamic information seeking scenario such as berrypicking, this means that the simulation of adapting information needs was not very well supported for IMDb. Overall, reommendations generated by matrix factorization and interpolation weights fared best.

Navigation via Information Foraging. A priori, it is not clear if retrieving multiple items from the same cluster represents an easier task than retrieving them from different clusters, as a cluster of items does not necessarily mean that items are located in proximity in the recommendation network. However, the resulting success rates show that items from the same clusters in the network are easier to retrieve: five recommendations lead to a success rate of 38.3%, and twenty recommendations to 63.1%. This indicates that the recommendation algorithms are able to use the characteristics in the ratings to support both genre-based and rating-based clustering.

The success rates again measures the number of found items in a cluster. The results for this scenario show that the success rates for the baselines, namely the random walks and the optimal solutions are consistently very high. This also indicates that the network structures reflect the clustering very well. Whereas for berrypicking, the simulations on the IMDb dataset perform poorly, the contrary is the case for information foraging, where the success rates range up to 99%. This again confirms the strong clustering in these networks, that lead to densely interconnected regions among similar items and facilitate retrieval of items in the same cluster. Recommendations generated by the interpolation weights algorithm generally fare best.

6 Personalized Recommendations

We now demonstrate the general suitability of our method to personalized recommendation approaches and report initial results for a sample configuration of parameters. The key difference for personalized recommendations is that a separate recommendation network emerges for every user based on their rated items. For this illustration, we follow the approach of Amazon.com, as detailed by Linden, Smith and York in 2003 [23], which consists of two steps: First, a set of similar items is determined for each item. Second, the items with the highest predicted rating among this set are recommended. We study two variants of this:

- **Pure.** We first compute a candidate set of similar items for an item—these are simply the non-personalized recommendations. Then we select the N items from this set that have the highest predicted rating for the specific user.
- **Mixed.** We again compute the set of similar items, but only use the $N/2$ recommendations with the highest predictions and add the $N/2$ top non-personalized recommendations (without introducing duplicates).

For both algorithms, we allow the recommendation of items that the user had already rated (which is yet another parameter to tune). We note that for this setting, the differences between the personalized networks for users decrease. When not allowing this, the resulting recommendation networks show a decrease in navigability the more items a user has already rated. For sake of space, we only report results for a restricted set of parameters. The results for the other combinations of parameters were similar, but we leave it to future work to examine them in more details.

Figure 3 shows the evaluation for recommendations generated by interpolation weights and matrix factorization for the user with the median number of ratings in the BookCrossing dataset. The outcome is generally similar to non-personalized networks. The pure algorithm leads to notably higher success rates for the optimal solution, but not for the simulation results themselves. This indicates that while the mixed algorithm leads to a better connectivity in the networks, this was not necessarily the case for navigability. This in turn suggests that the recommendations generated by this algorithm did not capture the intuitions used in the navigation

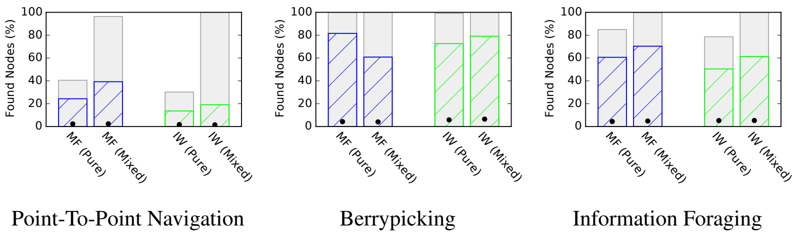


Fig. 3: **Navigational Success Rates for Personalized Recommendations.** All simulations were evaluated for BookCrossing, 20 recommendations and personalized for the user with the median number of ratings in the dataset. The results show that while the mixed recommendations enable a better optimal solution, the recommendations did not reflect the intuitions of the navigation simulations very well.

simulations very well. In future work, the evaluation method proposed in this paper could be used to develop a more effective personalized recommendation selections.

7 Discussion

We have presented a novel evaluation method that expands the repertoire of recommendation evaluation measures with a technique to assess navigability. The proposed method evaluates the navigation dynamics of recommendation networks by simulating three different navigation models, namely point-to-point navigation, navigation via berrypicking and navigation via information foraging. We believe that applying this method can broaden our understanding of recommendation algorithms and lead to a more complete characterization of their properties.

To demonstrate the feasibility of our method, we applied it to three exemplary datasets and highlighted differences in navigability for four different, non-personalized, recommendation algorithms. For five recommendations per item, we find that the recommendation algorithms we investigate considerably limit the navigability. However, we find that it can be improved by raising the number of recommendations. For the three navigation scenarios we investigate we find that the explorative scenarios inspired by berrypicking and information foraging lead to the best retrieval performance, while the scenario based on point-to-point navigation was less well supported. While increasing the number of recommendations represents a simple solution, a large number of recommendations could potentially clutter the interface and overwhelm users [5]. This shows that there is still a substantial potential to improve recommendation algorithms to better support navigation dynamics. As for the recommendation algorithms, we find that the recommendations generated by interpolation weights and matrix factorization performed best overall. However, more work is necessary to confirm these findings.

The selection of algorithms and datasets was naturally arbitrary, but they serve the purpose of illustrating the evaluation and therefore do not limit our main contribution of presenting a novel evaluation method. We have shown the suitability of our method for non-personalized recommendation algorithms and thereby effectively inspected recommendation networks for users who are either new to the system or simply browsing without being registered, and have also illustrated the applicability of our method to personalized recommendations.

The navigation models applied in this method are well-established in the research community and cover a wide range of typical user interaction scenarios with information systems in general, and recommender systems in particular. Greedy decentralized search, the basis for our navigation scenarios based on these models, has been used in previous work to analyze navigation dynamics in networks [13, 14] and has been found to produce comparable results to human navigation patterns [21, 33]. The navigation models we used do, however, have limitations and were deliberately kept simple, as the focus of our work was not on the information seeking models and their validity but on the properties of the recommendation algorithms. However, this does not limit our work, as our evaluation method does not depend on this particular model, which can easily be adapted or exchanged in future work. Possible enhancements to the navigation models could include a teleportation element (as in PageRank) modeling jumps between items without recommendations.

In summary, our work extends common evaluation measures of recommendation algorithms towards a path-based evaluation. Just as the evaluation of recommender systems has been shifting from accuracy-based measures towards diversification, coverage and time-dependent evaluations, we believe that our method helps to push the frontier of recommendation algorithms towards producing recommendations that make it easier for users to discover and explore items.

References

1. R. Agrawal and R. Srikant. Fast algorithms for mining association rules in large databases. In *VLDB'94*, 1994.
2. M. J. Bates. The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424, 1989.
3. R. M. Bell and Y. Koren. Improved neighborhood-based collaborative filtering. In *KDD Cup and Workshop at the KDD'07*, 2007.
4. R. Boim, T. Milo, and S. Novgorodov. Diversification and refinement in collaborative filtering recommender. In *CIKM'11*, 2011.
5. D. Bollen, B. P. Knijnenburg, M. C. Willemsen, and M. Graus. Understanding choice overload in recommender systems. In *RecSys'04*. ACM, 2010.
6. P. Cano, O. Celma, M. Koppenberger, and J. M. Buldú. Topology of music recommendation networks. *Chaos*, 16(1), 2006.
7. P. Castells, N. J. Hurley, and S. Vargas. Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 881–918. Springer, 2015.
8. Ò. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *RecSys'08*, 2008.
9. E. H. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions on the web. In *CHI'01*, 2001.

10. J. Davidson, B. Liebald, J. Liu, P. Nandy, T. V. Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, and D. Sampath. The youtube video recommendation system. In *RecSys'10*, 2010.
11. M. Ge, C. Delgado-Battenfeld, and D. Jannach. Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proc. of the 4th ACM Conference on Recommender Systems*, 2010.
12. A. Gunawardana and G. Shani. Evaluating recommender systems. In *Recommender Systems Handbook*, pages 265–308. Springer, 2015.
13. D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. Models of human navigation in information networks based on decentralized search. In *HT'13*, 2013.
14. D. Helic, M. Strohmaier, C. Trattner, M. Muhr, and K. Lerman. Pragmatic evaluation of folksonomies. In *WWW'11*, 2011.
15. J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, 2004.
16. J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.
17. J. M. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *TOC'00*, 2000.
18. Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009.
19. D. Lamprecht, D. Dimitrov, D. Helic, and M. Strohmaier. Evaluating and improving navigability of wikipedia: A comparative study of eight language editions. In *OpenSym'16*, 2016.
20. D. Lamprecht, F. Geigl, T. Karas, S. Walk, D. Helic, and M. Strohmaier. Improving recommender system navigability through diversification: A case study of IMDb. In *IKNOW'15*, 2015.
21. D. Lamprecht, M. Strohmaier, D. Helic, C. Nyulas, T. Tudorache, N. F. Noy, and M. A. Musen. Using ontologies to model human navigation behavior in information networks: A study based on wikipedia. *Semantic Web*, 6(4):403–422, 2015.
22. K. Lerman and L. Jones. Social Browsing on Flickr. In *ICWSM'07*, 2007.
23. G. Linden, B. Smith, and J. York. Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80, 2003.
24. S. Milgram. The small world problem. *Psychology Today*, 1(2):60–67, 1967.
25. T. T. Nguyen, P.-M. Hui, M. F. Harper, L. Terveen, and J. A. Konstan. Exploring the filter bubble: the effect of using recommender systems on content diversity. In *WWW'14*, 2014.
26. J. Nielsen. The 90-9-1 rule for participation inequality in social media and online communities, 2006. `www.nngroup.com/articles/participation-inequality`.
27. B. Nonnecke and J. Preece. Lurker demographics: Counting the silent. In *CHI'00*, 2000.
28. P. Pirolli. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, 2007.
29. K. Seyerlehner, P. Knees, D. Schnitzer, and G. Widmer. Browsing music recommendation networks. In *ISMIR'09*, 2009.
30. J. Su, A. Sharma, and S. Goel. The effect of recommendations on network structure. In *WWW'16*, 2016.
31. J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *CHI'04*, 2004.
32. E. G. Toms. Serendipitous information retrieval. In *DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries*, 2000.
33. C. Trattner, P. Singer, D. Helic, and M. Strohmaier. Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks. In *IKNOW'12*, 2012.
34. D. J. Watts, P. S. Dodds, and M. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.
35. C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *WWW'05*, 2005.

3.4 Improving Recommender System Navigability through Diversification: A Case Study of IMDb

This article tackles the second and third research questions, investigating how navigation can be modeled and subsequently improved for information systems. The article focuses on the two recommender systems of the Internet Movie Database (IMDb), the largest online database of movies, TV shows and the like.

IMDb employs two recommender systems, one of which is based on ratings, and the other based on content features. The article studies these in two steps: the first step evaluates the reachability in the recommendation networks in terms of components, path lengths and a bow-tie analysis. The results for this step show that the reachability provided by the recommender systems is fairly poor, with as little as 10% of items in the largest strongly connected component and path lengths between items as long as 140 hops. As a method to improve this, the article presents the introduction of diversified recommendations to replace a few of the displayed recommendations. This leads to a substantial improvement in reachability.

The second step applies an approach based on decentralized search to model navigation dynamics on the recommendation networks (cf. Article 3 [Lamprecht et al., 2015a]). The approach simulates the retrieval of top-rated items in the networks by an automated method. Like for the analysis of reachability, the results for the simulations show that the networks are poorly suited to retrieve items. This can again be improved by introducing diversified recommendations, which leads to an up to four-fold increase in success rates.

This article presents a comprehensive evaluation approach for both reachability and navigability in recommendation networks and demonstrates it on two real-world examples of recommendation networks from IMDb. As an important finding regarding the improvement of navigability, the article shows the important role that the introduction of a few diversified items can play in making networks more navigable.

Improving recommender system navigability through diversification: A case study of IMDb

Daniel Lamprecht
KTI, Graz University of Technology
Graz, Austria
daniel.lamprecht@tugraz.at

Florian Geigl
KTI, Graz University of Technology
Graz, Austria
florian.geigl@tugraz.at

Tomas Karas
KTI, Graz University of Technology
Graz, Austria
karas@student.tugraz.at

Simon Walk
IICM, Graz University of Technology
Graz, Austria
simon.walk@tugraz.at

Denis Helic
KTI, Graz University of Technology
Graz, Austria
dhelic@tugraz.at

Markus Strohmaier
GESIS and University of Koblenz-Landau
Cologne, Germany
strohmaier@uni-koblenz.de

ABSTRACT

The Internet Movie Database (IMDb) is the world's largest collection of facts about movies and features large-scale recommendation systems connecting hundreds of thousands of items. In the past, the principal evaluation criterion for such recommender systems has been the rating accuracy prediction for recommendations within the immediate one-hop-neighborhood. Apart from a few isolated studies, the evaluation methodology for recommender systems has so far lacked approaches that quantify and measure the exposure to novel content while navigating a recommender system. As such, little is known about the support for navigation and browsing as methods to explore, browse and discover novel items within these systems. In this article, we study the navigability of IMDb's recommender systems over multiple hops. To this end, we analyze the recommendation networks of IMDb with a two-level approach: First, we study reachability in terms of components, path lengths and a bow-tie analysis. Second, we simulate practical browsing scenarios based on greedy decentralized search. Our results show that the IMDb recommendation networks are not very well-suited for navigation scenarios. To mitigate this, we apply a method for diversifying recommendations by specifically selecting recommendations which improve connectivity but do not compromise relevance. We demonstrate that this leads to improved reachability and navigability in both recommender systems. Our work underlines the importance of navigability and reachability as evaluation dimension of a large movie recommender system and shows up ways to increase navigational diversity.

i-KNOW '15, October 21-23, 2015, Graz, Austria

DOI: http://dx.doi.org/10.1145/2809563.2809603

CCS Concepts

•**Human-centered computing** → **Web-based interaction;** *Collaborative filtering;* •**Information systems** → **Recommender systems;**

Keywords

Recommender Systems, IMDb, Navigation, Diversification

1. INTRODUCTION

Recommender systems support users in filtering information and selecting items among huge numbers of possible options. By connecting users with appropriate, relevant, or novel items, recommender systems also help to reduce information overload by filtering out unwanted items and reducing cognitive load on users [9, 10, 20]. By establishing connections between items, recommender systems enable users to browse and peruse a system. Users enjoy browsing a recommender system without the intention of making a purchase [9], which is especially relevant on systems where users immediately consume items (such as on YouTube [5]). Finally, recommendations are also important in the discovery of novel content [17].

In the past, the majority of research and development on recommender systems has focused on improving rating prediction accuracy. Spurred by the Netflix Prize challenge[1], where the evaluation criterion was the root mean squared error (RMSE) calculated on the rating predictions, researchers have found substantial improvements in terms of computing rating predictions [13].

So far, comparatively little attention has been paid to supporting, evaluating, or improving navigation and exploration properties of recommender systems. As a consequence, we still do not know much about how these scenarios are supported in state-of-the-art recommender systems. Learning more about the conditions of navigability in recommender systems is vital for researchers and practitioners who want to gain insight into how well these systems support navigation.

In this paper, we set out to analyze such properties in a real-world recommender system. To this end, we apply

[1]http://www.netflixprize.com

a recently presented network-theoretic framework [15] that proposes a two-level approach:

1. The first step investigates the *reachability* of recommendation networks (i.e., the networks formed by items as nodes connected by recommendations as links) by analyzing the topological characteristics in terms of components, clustering, path lengths and partitions. This analysis quantifies what parts of the network are connected via links and how many hops it takes to reach them.

2. The second step analyzes the results of these findings in a more practical way by simulating browsing scenarios on these networks. This provides us with insight into how well these networks fare in real-world navigation scenarios.

We apply this approach to investigate the case of IMDb, the largest movie database in the world. In particular we are interesting in answering the following research questions:

RQ 1 How well do the recommendations of IMDb support reachability and navigability?

RQ 2 How can reachability and navigability on IMDb be improved?

RQ 3 What are differences between collaborative filtering and content-based recommendations in terms of reachability and navigability?

In order to answer these questions, we analyze the two types of recommender systems present on IMDb in their entirety (see Figure 1 for an example of an IMDb page). Our results show that the recommendation networks on IMDb are split into a large number of disconnected components with large distances within components. As a result, the current state of IMDb recommendations does not support any kind of exploration scenario very well. As a remedy, we introduce recommendation diversification to better distribute the recommendations among items and show that two diversification approaches are able to substantially improve navigability.

2. RELATED WORK

The study of human navigation in networks was strongly influenced by Milgrams and Travers [19, 24], who performed a series of experiments on navigation in social networks. They found that even within very large social entities, such as the entire United States, humans were able to find connections to others through a very small number of intermediaries. This coined the term of *Six Degrees of Separation*. The notion of an efficiently navigable network was later formalized by Watts and Strogatz, who described high clustering and short path lengths as characteristics of highly navigable small-world networks [25, 26]. Kleinberg identified further properties that rendered networks efficiently navigable with decentralized search algorithms [11, 12]. The navigation model of greedy decentralized search was later used to analyze human navigation dynamics in networks [8, 16, 23].

West and Leskovec [27] studied human goal-oriented navigation in the information network of Wikipedia and found that humans took only a few clicks longer than the shortest possible paths. However, in contrast to the shortest paths,

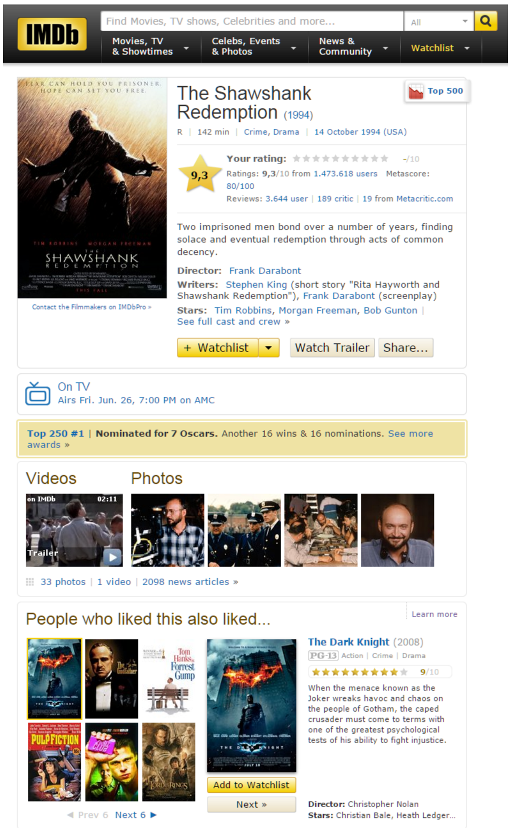


Figure 1: **IMDb page.** Example of an IMDb movie page, displaying facts, a voting score, links to videos and photos and collaborative filtering recommendations.

the resulting click trails exhibited a characteristic zoom-out phase (leading to more general concepts), followed by a phase of homing in to the target based on similarity.

Human navigation on recommender systems can occur in a range of uses cases. Recommendation browsing helps in discovering novel content [17], and the same has been found for search [28], where some users prefer navigation to search even when they know the target [22]. Generally, recommender systems help in learning and decision making [18, 20]. Users are more likely to follow links on movie recommendation sites than on factual websites such as Wikipedia [6]. On YouTube, recommendations fulfill the need for *unarticulated want* [5] and form a vital part of the user experience by connecting items.

A few studies have already investigated navigability on recommender systems. Music recommender systems were found to show heavy-tail degree distributions as well as small-world properties [2]. Several variations of IMDb recommendation networks have been found to exhibit long-tail degree distributions [7]. Celma and Herrera [3] found that collaborative filtering led to popularity bias and that a trade-off existed between accuracy and other evaluation metrics.

A simple method to improve navigability by selecting recommendations based on reachability was proposed by Seyerlehner et al [21]. We improve on this by taking the relevancy of recommendations as wells as their directionality into account.

3. MATERIALS AND METHODS

3.1 Data Sets

The Internet Movie Database (IMDb) is a database of facts about movies and television shows. The website started out as a hobby project on Usenet and has since grown to be the largest movie website worldwide[2] . The website presents facts and details about titles (movies, TV shows, short films and so forth), such as plot, cast, trailers and reviews, as well as information about actors and actresses, directors and crew. As of January 2015, the database contained facts about 3.1 million titles[3].

Users on IMDb can contribute and edit facts, although changes are moderated before being entered into the database. Users can also rate movies, write reviews and participate in messaging forums.

IMDb offers two different recommender systems:

Collaborative Filtering Recommendations (CF).

IMDb uses non-personalized rating-based recommendations for its titles, listed as *People who liked this also liked...* on title pages. The interface shows a total of 12 CF recommendations, from which 6 are immediately visible (see Figure 1 for an example of this interface).

Content-based Recommendations (CB).

Up until a site redesign in 2010, IMDb used non-personalized content-based recommendations [4]. These recommendations were computed from a proprietary combination of facts such as title, keywords, genre and user votes. This interface including the recommendations is still available through a change in the user preferences. In the interface, 5 recommendations are visible initially, and a total of 10 are available by following a link.

The presence of two parallel recommendation engines enabled us to directly compare the navigability within two real-world recommender systems side-by-side. To obtain the data on the recommender systems, we performed an exhaustive search over the IMDb title IDs by enumerating the space of 10 million possible values. During our crawl in January 2015, we were able to obtain the entire database of about 3.1 million titles in this way. We then extracted facts, such as release date, plot, storyline and average rating, as well as all available recommendations of both types. In total, we obtained 785, 019 nodes with content-based recommendations and 168, 560 nodes with collaborative-filtering recommendations.

As the basis for the diversification approaches, we also inspected the reviews for each title and downloaded all ratings assigned as part of a review. After that, we visited the profile pages of all users who had written at least one review and additionally downloaded all of their ratings they had assigned without an associated review, if they were publicly available. To avoid problems with sparse data, we only used data from films with at least three ratings and users who had rated at least three titles. By combining the profile ratings with the reviews ratings, we obtained a total of 25, 290, 692 million ratings from 149, 240 users for 168, 078 titles.

[2]http://http://www.imdb.com/pressroom

[3]http://http://www.imdb.com/stats

[4]http://www.imdb.com/help/show_leaf?history

3.2 Recommendation Networks

We constructed unpersonalized top-N recommendation networks from the recommendations we obtained from IMDb. In each of these networks, the items were represented as nodes and recommendations formed directed edges. We constructed a total of four different networks: Two for collaborative filtering, with 6 and 12 recommendations per node (denoted as CF (6) and CF (12)), and two for content-based recommendations with 5 and 10 recommendations per node (denoted as CB (5) and CB (10)). The number of recommendations was therefore the same as in the user interfaces.

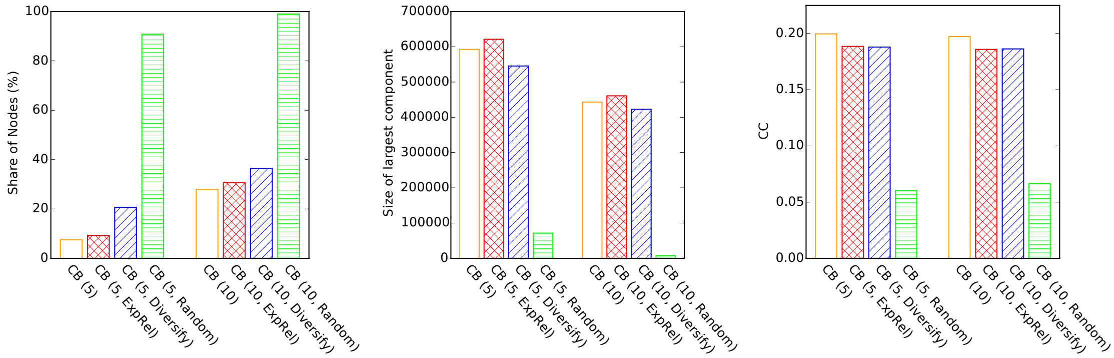
For the collaborative filtering networks, a fraction of nodes did not have any outgoing recommendations and were thus reachable via recommendations but then constituted a dead end. These nodes made up 11% of the CF (6) and 21% of the CF (12) network.

3.3 Diversification

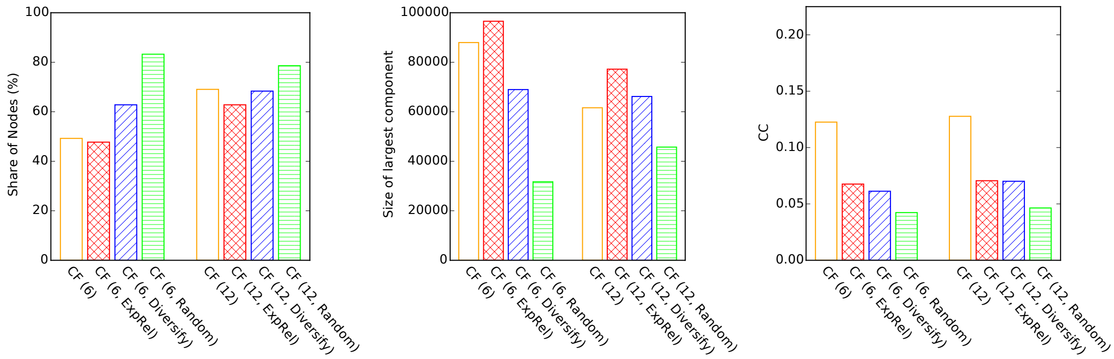
To improve navigability, we introduced diversity into the networks. User satisfaction with diversity for collaborative filtering has been found to peak between 30-40% diversity [29]. Based on this, we replaced recommendations as follows: For the immediately visible recommendations (5 for CB and 6 for CF), we replaced two recommendations. For the total recommendation list (10 for CB and 12 for CF) we replaced 4 recommendations. We use the following three approaches for diversification:

- **Random Recommendations.** The introduction of random links generally leads to well-connected networks with a small diameter. As such, introducing random recommendations effectively constituted an upper bound on the possible improvement through diversification.
- **Diversify.** Ziegler et al [29] proposed a method called *Diversify*. To apply it, we first build the recommendation network with the desired number of non-diversified recommendations (e.g., 4 for CF (6)). *Diversify* then introduces diverse recommendations for each node as the ones minimizing the similarity to the recommendations already present. We compute similarities between items by comparing their rating vectors.
- **Expanded Relevance (ExpRel)** Küçüktunç et al. [14] proposed a method to take the location of recommendations in the network into account. We use a simplified version thereof: We first build the recommendation network $G = (V, E)$ with of the desired number of non-diversified recommendations (e.g., 4 for CF (6)). Based on this, for each node $n \in V$ we compute $\Gamma(n, 2)$, the set nodes reachable in the two-hop neighborhood of each node. We then rank potential diverse recommendations $d \in V$ based on the number of nodes in $\Gamma(d, 2) - \Gamma(n, 2)$, the number of additional nodes the recommendation would add.

For the random recommendations, we added the diverse recommendations to all nodes that had existing outgoing recommendations in the graph. For *Diversify* and *ExpRel*, we first computed the cosine similarities between all pairs of items for which co-ratings were present in our dataset. We then selected all items for which at least 100 similarities to other nodes could be computed and used the top 50 most similar nodes to select diversified recommendations from. This left us with 145, 504 nodes for the CB and 118, 691 for the CF networks.



(a) Size of largest component (CB) (b) Number of components (CB) (c) Clustering coefficient (CB)



(d) Size of largest component (CF) (e) Number of components (CF) (f) Clustering coefficient (CF)

Figure 2: **Topology Analysis.** The figures show the sizes of the largest component, the numbers of components and the clustering coefficients. The unmodified recommendation networks, as present on IMDb, exhibit a comparatively small largest components and a high number of disconnected components. Diversification approaches change this and result in a larger component, while reducing clustering.

4. REACHABILITY

As the first part of our analysis, we study reachability of recommendation networks and analyze what parts of the graph are connected by paths of arbitrary lengths. This represents the basis for further analyses of efficient reachability and partition reachability, which permit us to gain more detailed insight into navigational dynamics.

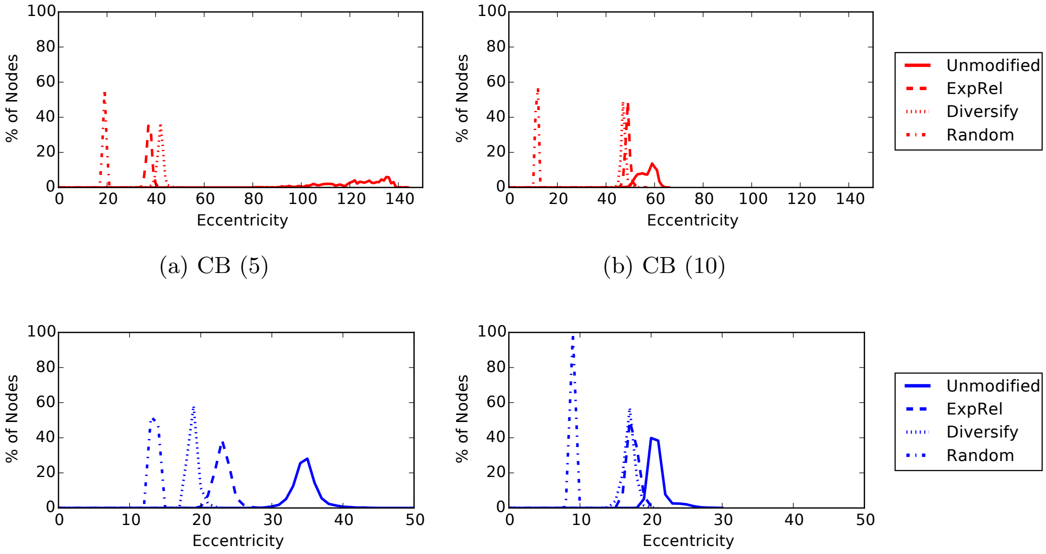
4.1 Effective Reachability

As the first step, we investigate the fundamental problem of whether a connection between pairs of nodes exists at all.

Strongly connected components.

The largest component enables users to explore all of its items by following recommendation links and is a direct measure for the fraction of the network reachable via navigation. In addition to the largest component, the number of components present in the network shows the division into separate parts that are not interconnected by recommendation links. Figure 2 shows that in their unmodified versions, content-based recommendations led to substantially smaller largest components than collaborative filtering recommendations. This confirms results from a previous study which found collaborative filtering to lead to larger components [15]. Possible contributing factors are the higher number of recommendations for collaborative filtering (6 and 12 versus 5 and 10 for content-based recommendations) as well as the higher total number of nodes in the content-based network. Diversification approaches were able to increase the size of the largest component substantially. The random diversification demonstrated the maximally achievable improvement, as random graphs are among the graphs with the highest possible reachability.

In terms of numbers of components, the results show that there exists a large number of disconnected components within the recommendation network. For the collaborative filtering recommendations, a major contributor to this is the fact that recommendations pointing away from items existed only for 79% of the nodes in CF (6) and 89% of nodes in CF (12). Those nodes therefore each formed a separate component with only one node in the directed graph. This problem was not present for the content-based recommendation network. However, the large number of disconnected components clearly hinders navigation in both networks. Diversification again mitigated this issue.



(a) CB (5) (b) CB (10)

(c) CF (6) (d) CF (12)

Figure 3: **Eccentricity Analysis.** This figure shows a sampled eccentricity distribution of both unmodified and diversified recommendation networks for a sample of 15% of the nodes in the largest strongly connected component (chosen uniformly at random). Eccentricity measures the longest shortest path from a node to any other node in the same component. With distances of up to 140 hops, the unmodified networks as present on IMDb do not lend themselves to navigation very well. Eccentricities can be reduced by introducing diversified recommendations.

Clustering Coefficient.
The clustering coefficient measures the fraction of neighbors that have a connection among themselves. High clustering implies more predictable browsing (with a large overlap of recommendations between related nodes) while low clustering increases the chance of being able to break out of the local context and follow a diverse or novel recommendation. Generally, high clustering with a few diverse links best supports navigation [26]. We define the clustering coefficient for recommendation networks as

$$C = \frac{1}{|V|} \sum_{i \in V} \frac{|\{(j,k) \in E | j,k \in \Gamma(i,1)\}|}{|\Gamma(i,1)| \left(|\Gamma(i,1)| - 1\right)}, \tag{1}$$

where $\Gamma(i,1)$ is the set of nodes reachable from i in one hop. The results show that the content-based networks exhibit higher clustering coefficients than the collaborative filtering networks. This indicates that content-based recommendations led to more redundancy in the resulting network. Together with the component sizes, it becomes apparent that a trade-off exists between reachability (i.e., the size of the largest component) and navigation predictability (i.e., higher clustering, which leads to better predictability of the area of a network a recommendation leads to).

4.2 Efficient Reachability

As the second step, we study the actual distance between pairs of nodes (given that there exists a path that connects them). This allows us to further investigate how well these networks support navigability and browsing. The probability that a user follows a link instead of typing in another URL or using the search function is around 65% [6] in movie recommender systems. This indicates that path lengths need to be short to properly support browsing scenarios.

To assess the difficulty of navigation, we evaluate eccentricity distribution on the largest strongly connected component. The eccentricity of a node measures the longest shortest distance between the node and any other node of the same component, therefore allowing us to learn about distances in the recommendation network. For a node $i \in SCC(G)$,

$$\mathrm{ecc}(i) = \max_{j \in \mathrm{SCC}(G)} d(i,j), \tag{2}$$

where $\mathrm{SCC}(G)$ is the largest strongly connected component in G and $d(i,j)$ is the geodesic distance between i and j. To evaluate eccentricity, we sampled the values for 15% of the nodes in the largest strongly connected component (between $8,000$-$112,000$ nodes, chosen uniformly at random).

For the content-based network, eccentricities were comparatively large (cf. Figure 3), with distances reaching up to 140 hops. The collaborative filtering networks exhibited lower eccentricities, rendering them better suited for browsing. Diversification measures lowered eccentricities for both networks.

4.3 Partition Reachability

As the third step, we study reachability based on the bow-tie model. The bow-tie model is a partitioning of a graph into three major components: IN, SCC and OUT, as well as a few additional ones, with the disconnected nodes collected in *OTHER* [1] (see Figure 5 for details). A bow-tie analysis allows us to learn more about the navigational structures in recommendation networks beyond the largest strongly connected component.

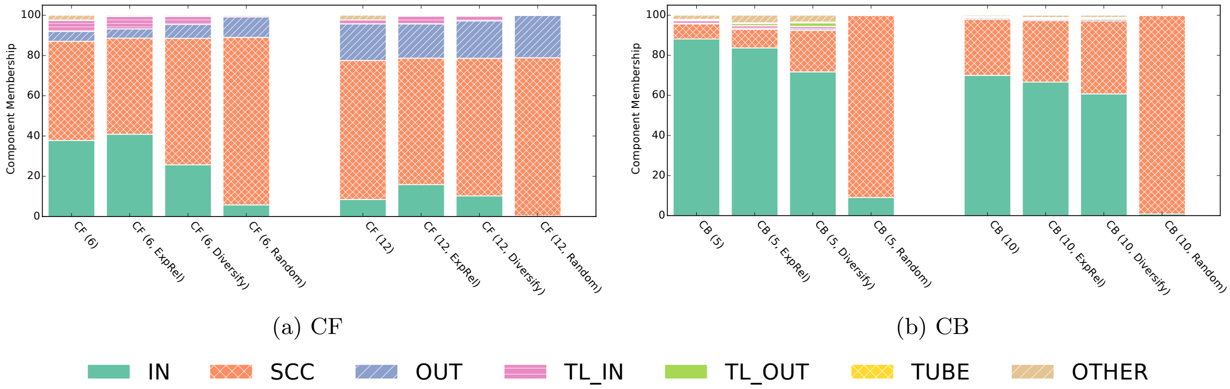


Figure 4: **Bow-Tie Analysis.** The figure shows the partition of the recommendation networks based on the bow-tie model (cf. Figure 5). The recommendation networks of IMDb consisted mainly of nodes in *IN*, *SCC* and *OUT*, implying that they were not in completely disconnected components. Diversification led to a larger share of nodes in the *SCC*.

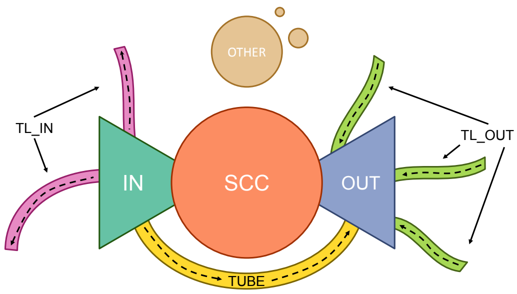


Figure 5: **Bow Tie Model.** The bow tie model [1] is a partitioning of a graph into a strongly connected component or core (*SCC*) as well as *OUT* which is reachable from it and *IN* which is able to reach it. Nodes in *TUBE* are on a detour from *IN* to *OUT*. The *TENDRILS (TL_IN, TL_OUT)* contain nodes pointing away from *IN* or pointing to *OUT*. Remaining nodes are collected in *OTHER*.

The bow-tie analysis confirmed that the size of the largest strongly connected component increased with diversification (cf. Figure4). Moreover, it shows that most other nodes were in the *IN* and *OUT* components. From a navigation perspective, this is desirable as it implies that these nodes are either able to reach the largest component or are reachable from it. When following recommendations from a node contained in *IN*, it is likely that a user will be able to reach the *SCC*. Figure 6 depicts the changes in component membership from the unmodified network to a diversified one. Increasing the size of the *SCC* via diversification implies that some of the recommendations from items previously in *IN* now point to nodes in the *SCC* and therefore become themselves a part of it. Note that the number of nodes in *OTHER* components slightly increases due to the fact that diversifying recommendations removes some of the recommendations to sink nodes (that do not have any outgoing recommendations). Navigationwise, this implies that the number of dead-ends encountered by users browsing the recommendation network decreases.

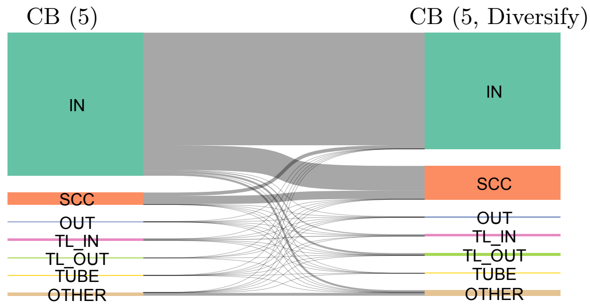


Figure 6: **Bow-Tie Membership Change Analysis for CB (5) to CB (5, Diversify).** Nodes were mostly part of *IN* in the unmodified recommendation network. Diversification moved items from *IN* to *SCC*.

5. NAVIGABILITY

As the first part of our analysis, we studied the reachability of recommendation networks. In the second part, we are now interested in how well the networks fare in terms of actual browsing scenarios. To this end, we simulate browsing in the networks and evaluate the results.

5.1 Start and target nodes

We evaluate browsing scenarios inspired by the desire to find a few movies relevant to certain genres. To this end, we take the genres (e.g., *Action*) as well as the genre combinations (e.g., *Action, Comedy*) as listed on IMDb[5] for a total of 93 target genres. For each of these genres, we compute the 25 top-rated items with at least 1,000 ratings from our rating dataset and take them as target sets. We restrict our analysis to the largest strongly connected components (cf. Figure 2) and sample 100 start nodes for every target set, leaving us with 9,300 start-target missions to simulate.

5.2 Link selection strategy

Our simulation approach was based on *greedy decentralized search*, a method to analyze navigation dynamics in net-

[5] http://www.imdb.com/genre/

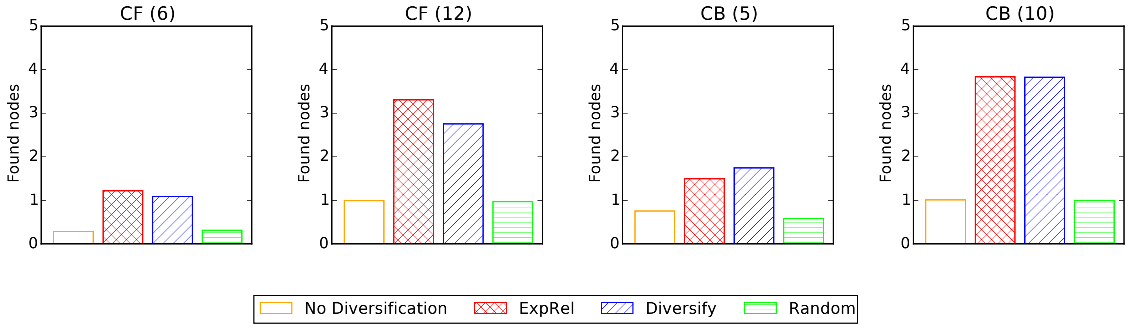


Figure 7: **Nodes found in navigation simulations.** The navigation scenarios were not very well-supported in the unmodified networks, where the simulations found between 0.2 and 1 node per run on average. With diversification approaches, the number of nodes found in the simulation of exploratory navigation scenarios increased by $100 - 300\%$.

works [8, 16, 23]. The simulation started at a target node and at each step greedily selected the outgoing recommendation with the highest similarity to the target set. The simulation kept track of visited nodes and explored each node only once. In case of a dead-end (no outgoing recommendations at all or no unvisited outgoing recommendations), the simulation backtracked to the previously visited node. We simulated each mission for a total of 50 clicks.

As the background knowledge to inform link selection, we computed the TF-IDF similarities between items by making use of the words contained in the title, plot and storyline descriptions. The value for a potential recommendation link was computed as the similarity between the current item and the average vector of the 25 target nodes. This is similar to the concept of information scent [4], where a link is thought to emanate a certain smell based on its usefulness with respect to the target.

We believe that these are plausible assumptions for users who have some idea where a recommendation could lead based on the information present with the recommendation in the interface (i.e., title and image).

5.3 Results

Figure 7 shows the results for the simulations. Overall, the navigation scenarios were not very well-supported in the unmodified networks, where the simulations found between 0.2 and 1 node per run on average. The diversification approaches were able to improve the outcomes compared to the unmodified recommendation networks substantially: both *ExpRel* and *Diversify* improved the number of found target nodes by $100 - 300\%$, thus strongly improving navigability in these networks.

Random diversification, however, did not lead to better results than the unmodified networks. Even though the injection of random links led to large components (cf. Figure 2), the resulting lower clustering meant that the similarity information was of little use in informing a navigation process.

6. DISCUSSION AND CONCLUSIONS

In this paper, we analyzed two recommendation networks from IMDb by applying a two-level evaluation approach for recommender systems to study reachability and navigability. In the following, we discuss the findings in the context of our research questions.

RQ 1 *How well do the recommendations of IMDb support reachability and navigability?*

The results of our analysis and our simulations show that with the unmodified recommendations present on IMDb, navigating the network (if at all possible) represents a very hard task for users. Within our navigation simulations, it was possible to retrieve only about one out of 25 target node within 50 steps, even though the target nodes were chosen to be the most popular items in terms of ratings and had a high number of votes.

RQ 2 *How can reachability and navigability on IMDb be improved?*

Applying two simple diversification measures led to improvement of reachability and navigability for both recommendation networks. The number of items the simulations was able to retrieve saw an up to threefold increase, thus making it more realistic for users to be able to gain useful knowledge from exploratory browsing in the network.

RQ 3 *What are differences between collaborative filtering and content-based recommendations in terms of reachability and navigability?*

The collaborative filtering recommendations (the approach currently in use by IMDb) led to a larger strongly connected component, resulting in a larger reachable share of the network than the network for content-based recommendations. However, in terms of the simulated navigation scenarios, content-based recommendation networks fared slightly better. This suggests that content-based recommendations make it easier to reach more popular nodes. Content-based recommendations led to networks with higher clustering—thus making link selection in them more predictable. Another possible explanation for this is that the information used for generating these recommendations overlapped with the background knowledge used in the simulations.

In the diversification measures we applied, we made the assumption that users would prefer between $30 - 40\%$ diversification, based on a study by Ziegler et al. [29]. In future work, it would be interesting to conduct a usability study investigating the results of applying diversification on a live system and testing different fractions of diversified recommendations.

The results of this paper suggest that navigating and browsing recommendations are currently not very-well supported on IMDb. Our work shows a possible way of improving this via diversification measures.

7. ACKNOWLEDGMENTS

This research was supported by a grant from the Austrian Science Fund (FWF) [P24866].

8. REFERENCES

[1] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1):309—-320, 2000.

[2] P. Cano, O. Celma, M. Koppenberger, and J. M. Buldu. Topology of music recommendation networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 16(1):013107, 2006.

[3] Ò. Celma and P. Herrera. A new approach to evaluating novel recommendations. In *Proceedings of the 2nd ACM Conference on Recommender Systems*, pages 179–186. ACM, 2008.

[4] E. H. Chi, P. Pirolli, K. Chen, and J. Pitkow. Using information scent to model user information needs and actions and the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 490–497. ACM, 2001.

[5] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. Van Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, et al. The YouTube Video Recommendation System. In *Proceedings of the fourth ACM Conference on Recommender Systems*, pages 293–296. ACM, 2010.

[6] D. F. Gleich, P. G. Constantine, A. D. Flaxman, and A. Gunawardana. Tracking the random surfer: Empirically measured teleportation parameters in pagerank. In *Proceedings of the 19th International Conference on World Wide Web*, pages 381–390. ACM, 2010.

[7] J. Grujić. Movies recommendation networks as bipartite graphs. In *International Conference on Computational Science*, pages 576–583. Springer, 2008.

[8] D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. Models of human navigation in information networks based on decentralized search. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 89–98. ACM, 2013.

[9] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, 2004.

[10] D. Jannach, M. Zanker, A. Felfernig, and G. Friedrich. *Recommender Systems: An Introduction*. Cambridge University Press, 2010.

[11] J. Kleinberg. The small-world phenomenon: An algorithmic perspective. In *Proceedings of the thirty-second annual ACM Symposium on Theory of Computing*, pages 163–170. ACM, 2000.

[12] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.

[13] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *Computer*, (8):30–37, 2009.

[14] O. Küçüktunç, E. Saule, K. Kaya, and Ü. V. Çatalyürek. Diversified recommendation on graphs: Pitfalls, measures, and algorithms. In *Proceedings of the 22nd International Conference on World Wide Web*, pages 715–726, 2013.

[15] D. Lamprecht, M. Strohmaier, and D. Helic. Improving Reachability and Navigability in Recommender Systems. *arXiv preprint arXiv:1507.08120*, 2015.

[16] D. Lamprecht, M. Strohmaier, D. Helic, C. Nyulas, T. Tudorache, N. F. Noy, and M. A. Musen. Using ontologies to model human navigation behavior in information networks: A study based on wikipedia. *Semantic Web*, 6(4):403–422, 2015.

[17] K. Lerman and L. Jones. Social Browsing on Flickr. In *Proceedings of 1st International Conference on Weblogs and Social Media*, 2007.

[18] G. Marchionini. Exploratory search: from finding to understanding. *Communications of the ACM*, 49(4):41–46, 2006.

[19] S. Milgram. The small-world problem. *Psychology Today*, 1:60–67, 1967.

[20] P. Resnick and H. R. Varian. Recommender systems. *Communications of the ACM*, 40(3):56–58, 1997.

[21] K. Seyerlehner, P. Knees, D. Schnitzer, and G. Widmer. Browsing music recommendation networks. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, 2009.

[22] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: a study of orienteering behavior in directed search. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 415–422. ACM, 2004.

[23] C. Trattner, P. Singer, D. Helic, and M. Strohmaier. Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, page 14. ACM, 2012.

[24] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32, 1969.

[25] D. J. Watts. Networks, dynamics, and the small-world phenomenon. *American Journal of Sociology*, 105:493–527, 1999.

[26] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[27] R. West and J. Leskovec. Human wayfinding in information networks. In *Proceedings of the 21st International conference on World Wide Web*, pages 619–628. ACM, 2012.

[28] R. W. White and J. Huang. Assessing the scenic route: measuring the value of search trails in web logs. In *Proceedings of the 33rd International ACM SIGIR conference on Research and development in information retrieval*, pages 587–594. ACM, 2010.

[29] C.-N. Ziegler, S. M. McNee, J. A. Konstan, and G. Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pages 22–32. ACM, 2005.

3.5 Evaluating and Improving Navigability of Wikipedia: A Comparative Study of Eight Language Editions

This article addresses the third research question, which is how navigability can be improved, and does so for the case of Wikipedia. Wikipedia supports a wide variety of use cases, such as quick look up of facts, researching a topic or discovering new articles. For many of these use cases, navigating between articles is of vital importance. Specifically, this article assesses the state of navigability on the eight largest Wikipedia language editions and investigates methods to improve it.

The article shows that when the entirety of all links between Wikipedia articles is considered, the resulting article networks are very well connected and allow mutual reachability of almost all articles. However, research on the influence of page organization (cf. Article 1 [Lamprecht et al., 2016b]) has shown that the position of a link exerts a substantial influence on the likelihood that users will select it. In fact, most of the attention of Wikipedia users is directed at the top of articles.

For this reason, this article investigates Wikipedia navigation with a restricted user view port and analyzes gradually more restricted areas of articles. The results show that even a comparatively loose restriction to the links occurring in the entire lead section drastically limits the set of mutually reachable articles to roughly a third.

To make Wikipedia more navigable even with restricted view ports, the article presents an algorithm to recommend links from the remainder of the article to move towards the top sections. The recommended links are therefore all semantically relevant to the article, as they have been previously introduced by an editor. The findings demonstrate the feasibility of this method and open up new possibilities to improve navigability in real-world systems by restructuring content.

Evaluating and Improving Navigability of Wikipedia: A Comparative Study of Eight Language Editions

Daniel Lamprecht
KTI, Graz University of Technology
Graz, Austria
daniel.lamprecht@tugraz.at

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

Denis Helic
KTI, Graz University of Technology
Graz, Austria
dhelic@tugraz.at

Markus Strohmaier
GESIS - Leibniz Institute for the Social Sciences
and University of Koblenz-Landau
Cologne and Koblenz, Germany
markus.strohmaier@gesis.org

ABSTRACT

Wikipedia supports its users to reach a wide variety of goals: looking up facts, researching a topic, making an edit or simply browsing to pass time. Some of these goals, such as the lookup of facts, can be effectively supported by search functions. However, for other use cases, such as researching an unfamiliar topic, user need to rely on the links to connect articles. In this paper, we investigate the state of navigability in the article networks of eight language versions of Wikipedia. We find that, when taking all links of articles into account, all language versions enable mutual reachability for almost all articles. However, previous research has shown that visitors of Wikipedia focus most of their attention on the areas located close to the top. We therefore investigate different restricted navigational views that users could have when looking at articles. We find that restricting the view of articles strongly limits the navigability of the resulting networks and impedes navigation. Based on this analysis we then propose a link recommendation method to augment the link network to improve navigability in the network. Our approach selects links from a less restricted view of the article and proposes to move these links into more visible sections. The recommended links are hence semantically valid to the article. Our results are relevant for researchers interested in the navigability of Wikipedia and open up new avenues for link recommendations in Wikipedia editing.

CCS Concepts

•Information systems → Web searching and information discovery; •Human-centered computing → Hypertext / hypermedia; Wikis;

OpenSym '16 August 17–19, 2016, Berlin, Germany
© 2016 ACM. ISBN 123-4567-24-567/08/06...$15.00
DOI: 10.475/123_4

Keywords

Wikipedia, Navigability, Reachability, Bow Tie Model, Link Recommendations

1. INTRODUCTION

The multiple language editions of Wikipedia serve around 16 billion views per month as of 2016, with the English Wikipedia accounting for almost half of them [1]. Visitors to the free encyclopedia pursue a wide range of goals: looking up specific facts, learning about a topic of interest, making an edit, or simply browsing to pass time. Generally, the goals of users in retrieving information on the Web can be classified into three different ways [34]:

1. lookup of a specific object or fact,
2. search for something that cannot be explicitly described but will be recognized once retrieved, and
3. serendipitous (accidental) discovery.

Wikipedia supports these three ways of information retrieval by different means. The lookup of specific facts (or articles) is generally satisfiable with the Wikipedia-internal search engine or an external Web search engine. The other two ways, however, are not as well-supported by search engines. Users therefore need to rely on the hyperlinks that join the vast number of separate pieces of knowledge on Wikipedia in order to reach their goals. There exists several types of links to support users in connecting articles: links in the running text, links in tabular summaries such as infoboxes and links to groupings of articles such as categories or portals.

For the English Wikipedia, out of 3, 279, 134, 602 visits to articles collected in a clickstream in February 2015 [43], 30% were referred to from other Wikipedia articles, and 70% from external sources. This suggests that external links to Wikipedia, such as search engines, play an important role in referring visitors to the encyclopedia. However, it also shows that 30% of all clicks relied on links within Wikipedia and this implies that navigation plays a vital role for users.

On the Web, some users have been found to prefer the incremental process of navigation to direct retrieval, even

[1] http://stats.wikimedia.org/EN/TablesPageViewsMonthlyCombined.htm

Figure 1: **Navigational Views of a Wikipedia Article.** We investigate five different navigational views of Wikipedia articles: the unrestricted view of all links, the view of the links in the lead (shown with a blue border), the first lead paragraph (shown with a red border), the infobox (shown with a green border) and the very first link in the article (shown with brown background). These views enable us to understand the effects of limiting the number of links on the navigability of the article network.

when knowing what they are looking for [33]. Users also enjoy browsing without specific objectives, for example in recommender systems [12] or on entertainment websites such as YouTube [5]. Navigating articles via links is thus a vital component of Wikipedia even in the presence of powerful search engines. This is also manifested in the detailed description of the rules for placement and linking within Wikipedia's manual of style.

Problem. Recent research has shown the majority of the attention of Wikipedia users is focused on the areas located near the top, namely the lead section and the infobox [7, 19]. While previous work has already provided us with insights into the network structure of Wikipedia [4, 8], little is known about the effects of viewport restriction to, for example, the lead section. Understanding the impact of viewport restriction on the navigability of Wikipedia's article network would allow us to better understand and support the needs of users that only read parts of articles. We therefore investigate the following research questions.

RQ1. What is the state of navigability across multiple language versions of Wikipedia?

RQ2. How can this analysis be exploited to suggest improvements to navigability?

Approach. To assess navigability we make use of elements of a framework to evaluate navigability introduced in the context of recommender systems [18, 20]. We use this framework to study the navigability of eight language versions of Wikipedia based on the bow tie model, which partitions a network based on reachability criteria. five different navigational views constructed by taking into account a subset of all links in articles: (i) the unrestricted view of all links, (ii) the view of the links in the lead section, (iii) the view of the first lead paragraph, (iv) the view of the infobox and, finally, (v) the view of the very first link. Figure 1) shows an example of the areas used for these views.

Based on the results of this analysis we then propose a method to improve navigability in Wikipedia by recommending specific links from articles to move closer towards the more visible regions of an article or otherwise emphasize. We demonstrate our method by applying it to two Wikipedia versions and showing the effects.

Contributions. Our main contribution is a comparative study of navigability in terms of components and mutual reachability for the eight largest Wikipedia language versions. We show that for a restricted views of articles, only a comparatively small share of the articles in Wikipedia is reachable by following links. Based on this analysis, we then propose a method to improve the reachability in the article networks by suggesting specific links that could be emphasized, such as moving them to sections that users pay more attention to.

2. RELATED WORK

Navigation in networks. Stanley Milgram's small world experiments [25, 36] investigated decentralized search in the social network of the United States. Participants of the experiments were provided with a short description of a target person and were asked to forward the letter to a first-name acquaintances with the goal of reaching the target person. The striking result of the experiment was that participants were able to find very short paths to the target person across the social network of the entire United States.

As a result, the enabling property for efficient wayfinding in a network was named small world property. Many networks that emerge in nature are in facts small world networks. Watts and Strogatz proposed a generative model for these networks based on rewiring of a ring lattice [39]. These models were subsequently extended based on networks with nodes organized in two-dimensional grid lattices and hierarchies [15, 16, 38]. Jon Kleinberg identified the properties that made these networks efficiently navigable with decentralized search algorithms.

The model of decentralized search was applied to model human navigation in information networks such as folksonomies and a subset of Wikipedia [11, 10, 21, 35].

For Wikipedia, human navigation was extensively studied based on wikigames. In these navigational games, players are challenged to get to a target article by following links in the text and not using the search function. Based on log files from these games, researchers have studied goal-directed navigation with explicitly specified target articles. Players are, in general, very efficient in finding targets on Wikipedia and use high-degree hubs as landmarks in navigation [40]. The resulting paths have also been found useful to compute semantic relatedness of articles [30].

Wikipedia network analysis. The article network of

Table 1: **Datasets.** We used the eight Wikipedia language versions with the highest number of edits for this work.

Language	Articles	Edits
English	5,072,214	812,170,986
German	1,905,450	156,204,159
French	1,721,902	125,368,222
Spanish	1,232,123	94,262,365
Russian	1,287,687	88,544,149
Italian	1,251,650	83,993,233
Japanese	1,001,180	59,504,158
Dutch	1,854,708	46,955,102

Wikipedia has been found to grow by preferential attachment, with most of the articles are contained in a (mutually reachable) strongly connected component [4]. In comparison to general Web pages, Wikipedia has been found to be more densely linked and contain a larger strongly connected component [14]. Authors contributing to different areas in the network have been found to support knowledge integration [8].

Ibrahim, Danforth and Dodds have studied the structures and cycles emerging in the network of the first links of the English Wikipedia [13]. The same first-link network has also been used to automatically categorize articles in Wikipedia based on its *is-a* descriptions and shown to lead to categories with high precision when compared to human-curated categories [2].

Wikipedia link suggestions. The problem of improving the link structure of an information system has been studied in the context of social and information networks cast as a link formation (creation) [6, 22, 23, 32] and link removal problem [17, 29].

For Wikipedia, a number of approaches have been proposed: Extracting potential links from text based on keyword extraction and word sense disambiguation [24], machine learning [26], factorization of the adjacency matrix [42], based on clustering documents and linking related cluster documents [1] and based on mining navigational traces [41, 28]. In addition, there have been several approaches to suggest crosslingual links [27, 31, 37, 44].

3. MATERIALS AND METHODS

3.1 Datasets

For this work, we investigate the navigational properties of the top eight language editions of Wikipedia based on the number of edits. While there exist language versions containing a larger number of articles than the ones we investigate (e.g., the Swedish Wikipedia contains almost three million articles and the Cebuano Wikipedia contains almost two million), these versions have comparatively few edits and a large share of the articles are bot-created stubs. We hence restrict our analysis to Wikipedia language versions where the majority of edits is done by human editors. Table 1 shows the languages, numbers of articles and numbers of edits for the eight versions.

For all language versions, we used the articles in the version present at February 3, 2016, based on the Wikipedia IDs in the dump from that date. We obtained the HTML pages for the articles corresponding to all IDs from the Wikipedia

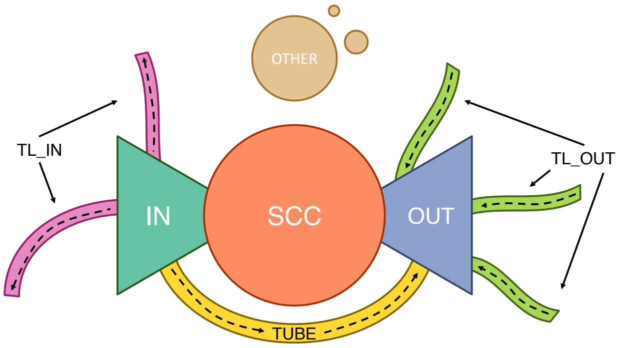


Figure 2: **Bow Tie Model.** For a directed network, the bow tie model [3] defines the largest strongly connected component (*SCC*) as the largest set of nodes which are all mutually reachable. Directly adjacent to the *SCC* is the *IN* component, in which all nodes can reach the *SCC*. Note that the set of nodes in *IN* in general does not form a strongly connected component but consists of multiple components that all share that they have a connection to the *SCC*. Nodes in *OUT* are reachable from *SCC* but not the other way around. In addition to these three main components, the bow tie model defines disconnected components (*OTHER*), *TUBES* (which connect *IN* to *OUT*) and tendrils leading away from *IN* or into *OUT*.

API [2]. This had two distinct advantages over the XML dump containing Wiki Markup: First, it allowed us to view all templates in their resolved forms, which otherwise would be a very cumbersome to achieve. Second, it permitted us to resolve redirects by using the information contained in the API response, rather than relying on the (incomplete) redirect list part of the official Wikipedia dumps. In addition to the articles themselves, we used the page view count information for the entire month of January 2016 [3].

3.2 Navigational Views

After having downloaded all articles, we parsed the HTML files and extracted all links. We then constructed the article networks for all Wikipedia language versions we investigated. To this end, we regard all articles as nodes of a network and insert all links between pages as directed edges. We restrict the link analysis to those links occurring in the article itself (i.e., links in the text, in tables, divs, etc. within the content part) and exclude the remaining links, such as links to categories, links in the menu on the left side or links to external websites.

Previous research has shown that users of Wikipedia dedicate a large proportion of their attention to the top of articles (namely to the lead section and the infobox [7, 19, 28]). To better understand the implications of these behaviors, we investigate five distinct *navigational views* of Wikipedia.

1. **Entire Article.** This view represents the links from the entire article, including those in the lead section and any tables (such as infoboxes). This view shows how users would navigate if they considered the links from the entire article.

[2] https://wikipedia.org/w/api.php

[3] https://dumps.wikimedia.org/other/pagecounts-raw/2016/2016-01

Figure 3: **Bow Tie Membership Change Analysis.** The figure shows the transitions from the unrestricted navigational view containing all articles to more restricted views. Colors and labels correspond to the ones used in 2. The leftmost view (entire article) contains the the second view (entire lead) and the membership transitions are shown between the states. For the unrestricted view of all links, the large majority of all articles are mutually reachable in the *SCC*. Restricting the navigational views to include fewer links decreases the size of the largest strongly connected component (*SCC*).

2. **Entire Lead.** The links in the lead section receive a large share of user attention [19]. As the table of contents after the lead is by default expanded on Wikipedia, this presents an obstacle to users and frequently requires scrolling to read the first section.

3. **First Lead Paragraph.** This view comprises of all links in the first paragraph of the lead. This is similar to the excerpt shown by a Google search result for a search term. While the excerpt does not highlight the links themselves, the information contained in it represents what users learn from it. If users are interested to learn more based on the excerpt, they might look into the Wikipedia articles for corresponding concepts.

4. **Infobox.** Infoboxes are tabular representation of the most important facts of an article and are present for 40% of articles on the English Wikipedia (and between 32 and 69% for the eight Wikipedia versions we investigated). Limiting the view to links contained in infoboxes represents users that look only at tabular information and the key facts of an article.

5. **First Link.** This view is restricted to the very first link that is not in parentheses, italics, or contained in a table. In a sense, the set of articles reachable this way represents the backbone concepts of a Wikipedia version.

Table 2: **Correlation Analysis.** This table shows the correlation of the *SCC* sizes for all eight Wikpedia language versions with the number of edits and the median outdegree. The correlation was computed with Spearman's ρ. The median outdegree strongly correlates with the *SCC* size when taking the entire articles into account. This implies that longer articles are correlated with a larger *SCC*. The number of edits correlates with the *SCC* size for the navigational view of the first lead paragraph, which suggests that a large number of edits introduces more navigable links into the lead section.

Navigational View	# edits	median outdegree
First Lead Par.	0.74	-
Entire Lead	-	-
Infobox	-	-
Entire Article	-	0.83

3.3 Bow Tie Analysis

To analyze the structure and connectivity of the article networks, we make use of the bow tie model. This model was proposed to study the structure of the Web graph [3] and took its name from the resemblance of a bow tie. Figure 2 shows the structure of the bow tie model. For a directed network, the model defines the largest strongly connected component (*SCC*) as the largest set of nodes which are all mutually reachable. Directly adjacent to the *SCC* is the *IN* component, in which all nodes can reach the *SCC*. Note that the set of nodes in *IN* in general does not form a strongly connected component but consists of multiple components that all share that they have a connection to the *SCC*. Nodes in *OUT* are reachable from *SCC* but not the other way around. In addition to these three main components, the bow tie model defines disconnected components (*OTHER*), *TUBES* (which connect *IN* to *OUT*) and tendrils leading away from *IN* or into *OUT*.

Navigability of a Wikipedia article network measures the extent to which articles can be reached by following links. An important metric for navigability is hence the size of the largest strongly connected component (*SCC*), which measures the number of mutually reachable articles. We use the bow tie model to study navigability in the following way: Firstly, we study the membership of articles to partitions and the sizes of these partitions (in particular the size of the *SCC*). Secondly, we make use of the flow information contained in the model: For example, for all articles in *IN*, there exists a path to a node in the *SCC*. This allows us to identify one-way reachability in the article network.

4. EVALUATING NAVIGABILITY

We evaluate the bow tie structure of the article networks based on the navigational views with a membership change analysis, shown in Figure 3. The sizes of the components represent the percentage of articles contained in them, and the transition of node membership between navigational views is shown with connections between the partitions.

4.1 Entire Article

When taking the entirety of the links of Wikipedia articles into account, the *SCC* covers the vast majority of the articles. For the English Wikipedia, the *SCC* contains 94% of all articles, and for the remainder of language versions the coverage is between $87-97\%$, with the only exception being the Dutch version, for which only 63% of articles are contained in the *SCC*. A small share of articles for each investigated Wikipedia furthermore belongs to the *IN* component. These are articles that have outgoing paths into the *SCC*, but cannot be reached from it. Frequently, these are very short articles: For example, for the English Wikipedia the articles in *IN* have a median 6 links, while those in the *SCC* have a median 42. This indicates that articles in the *SCC* are substantially longer than those in *IN*.

Assuming that visitors carefully explore all present links on an article, these results imply that they could reach almost all of the articles on the encyclopedia by navigation. The coverage of articles by the *SCC* has notable increased since the earlier days of Wikipedia: A study of several language versions of Wikipedia from 2004 found that the *SCC* covered between 72 and 89% of articles for the Italian, Spanish, French, German and English Wikipedias and 67% for the Portuguese Wikipedia [4]. In addition to the larger percentage of articles covered by the *SCC*, the sizes of all these Wikipedia language versions have vastly increased since then (e.g., the English Wikipedia grew from 340k to 5M articles).

4.2 Entire Lead, First Lead Paragraph, and Infobox

When restricting the navigational view to links occurring in the lead section, the sizes of the *SCC* decrease to 16% (Dutch) to 37% (Italian). This implies that for visitors not looking *below the fold* (which for Wikipedia mostly implies going further than the lead section and the table of contents), the share of mutually reachable articles in the network shrinks to $20-40\%$ of the *SCC* that would be available for all links in the articles.

For the links in the first lead paragraph, the *SCC* sizes range between 3% (Dutch) and 7% (English). This implies that Wikipedia becomes effectively unnavigable for this view, except for very few concepts. The transitions of articles between the partitions of the bow tie model reveal that most of the articles that are in the *SCC* for the unrestricted navigational view of all links become part of the *IN* component when looking at links from the lead and the first lead paragraph. From a navigational perspective, this means that while fewer articles are mutually reachable, those articles still lie on paths leading to the *SCC*. Visitors looking at the lead section of these articles can therefore still reach the the *SCC*. However, this navigation is necessarily one-way: Once in the *SCC*, the number of reachable articles is severely limited.

A possible explanation for this can be found in the guidelines for the lead section in Wikipedia. For example, for the English Wikipedia [4], these guidelines state that the lead section should summarize the most important aspects and provide links to more general articles. This likely restricts links from the less general articles to point only to the more general ones. This in turn leads to the former becoming part of *IN* and the latter becoming part of the *SCC*.

A similar observation can be made for the infobox view. The *SCC* sizes for the corresponding navigational view range between 4% (Dutch) and 19% (Japanese). Like for the first lead paragraph, these reduced component sizes result in networks that allow only for a small fraction of mutually reach-

[4] http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section

Table 3: **Cycles in the First-Link Networks of Wikipedias.** The first-link networks limit the navigational view to the first link not in a table, in parentheses or italics. The table shows the resulting cycles (i.e., the strongly connected components) in the Wikipedias. The *IN* component shows the percentage of articles which eventually lead to the cycle when repeatedly following first links. The percentage after the first listed article states the size of the *IN* component for this article, excluding incoming links via the cycle. The articles in the cycles belong to very general topics and show the effect of the first sentence in articles frequently making use of an *is-a* relation. The article on philosophy is central to four of the eight investigated language versions.

Language	*IN* Component	Article titles (translated)	Article titles (original)
English	97.0%	**Philosophy (92.1%)**, Existence, Ontology, Reality	Philosophy, Existence, Ontology, Reality
German	95.8%	**Philosophy (95.8%)**, World, Totality	Philosophie, Welt, Totalität
French	85.0%	**Philosophy (68.8%)**, Linguistics, Discipline (academia), Knowledge, Ancient Greek, Knowledge, Notion (philosophy), Greek language, Feature (linguistics), Isogloss, Centum and satem languages, Hellenic languages	Philosophie, Linguistique, Discipline (spécialité), Connaissance, Grec ancien, Savoir, Notion, Grec, Trait (linguistique), Isoglosse, Isoglosse centum-satem, Langues helléniques
Spanish	87.8%	**Psychology (87.7%)**, Profession, Activity, Specialization	Psicología, Profesión, Actividad, Especialización
Russian	73.7%	**Philosophy (58.9%)**, Mathematics, Science, Cognition, Object (philosophy), Set theory, Method (philosophy), Systematization, Objectivity (philosophy)	Философия, Математика, Наука, Познание, Объект (философия), Теория множеств, Метод, Систематизация, Объективность
Italian	73.2%	**Science (39.0%)**, Knowledge, Biology, Psychology, Psyche (psychology), Tissue (biology), Central nervous system, Brain, Nervous system, Awareness	Scienza, Conoscenza, Biologia, Psicologia, Psiche, Tessuto (biologia), Sistema nervoso centrale, Cervello, Sistema nervoso, Consapevolezza
Japanese	82.3%	**Person (82.3%)**, Interpersonal relationship	人間, 人間関係
Dutch	67.0%	**Knowledge (67.0%)**, Know-how	Kennis, Weten

able articles. Again, there is a large number of articles in *IN* that can at least reach the *SCC*. The explanation for this behavior is also similar as for the lead: since infoboxes state the key facts of articles, links to less general articles have a lower chance of being placed. There also exists a comparatively large fraction of articles in *OTHER*: These are the articles that do not possess an infobox and hence do not have any outlinks in this view.

In general, the difference in sizes of the *SCC* could be explained by the length of the article, as a longer text offers the possibility to include more links. To investigate this, we compute the correlation between the number of outlinks and the size of the *SCC* for all eight Wikipedia language version. Table 2 shows that there is indeed a correlation, but it only occurs for the unrestricted view of the links in the entire article. For the restricted views, the two features do not significantly correlate. A potential explanation for this is the restricted length of the lead section, the first paragraph and the infobox, which dampens the differences between long and short articles. We also find a correlation between the number of edits to a Wikipedia and the size of the *SCC* for the view of the first lead paragraph. This suggests that with an increasing number of edits, editors attach great importance to the links in the lead section, which as a result become more useful for navigational purposes.

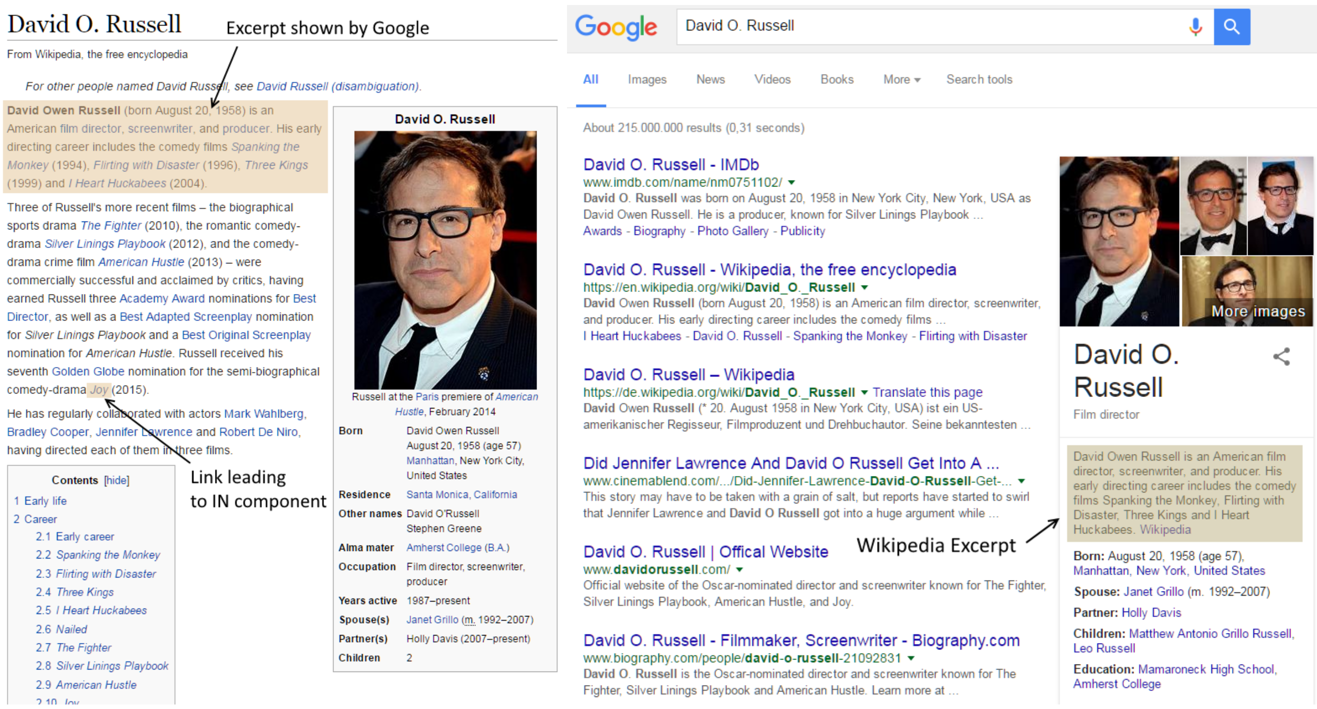
4.3 First Link

The manual of style for the English Wikipedia states that the first sentence should give an easy-to-understand introduction, define the title, and put it in context [5] (other languages have similar guidelines). For example, in Figure 1, the first sentence defines Reykjavík to be the capital of Iceland and places the first link on the word for the country. In fact, many of the first links in the English Wikipedia are *is-a* relations. The tree structure created by these first links has been shown to lead to a category hierarchy with high precision as compared to human-curated categories [2].

The navigational view of the first links is interesting from a theoretical perspective: According to popular belief, repeatedly following the first link in the English Wikipedia will eventually lead to the article on Philosophy [6]. To investigate the link structure of the first-link network, we look at its *SCC*. By definition, it consists of a cycle of articles (or a dead-end in the degenerate case). We then compute the number of articles in *IN*, which measures the number of articles from which navigation would end up at that cycle. For sake of clarity, we use the *SCC* with the largest corresponding *IN* component (and not the *SCC* containing the largest number of articles, which was different for several cases). Table 3 shows the results of this analysis. For all investigated Wikipedias, there is a single article in the cycle that accounts for more than half of the number of articles in its *IN* component, which confirms the funnels identified by Ibrahim et al. [13]. For the English Wikipedia we indeed find that the vast majority of articles (97%) lead to the cycle containing the philosophy article. This finding also holds true for a large majority of articles in the German, French and Russian Wikipedias. For the Spanish and Italian Wikipedias, the dominant cycle contains the article on psychology, while for Dutch, the dominant cycle consists

[5] https://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Lead_section#First_sentence

[6] https://en.wikipedia.org/wiki/Wikipedia:Getting_to_Philosophy



Wikipedia Article

Google Search Result

Figure 4: **Link Recommendation Example.** The English Wikipedia article on film director *David O. Russell* is in the largest strongly connected component (*SCC*) for the navigational view of the first lead paragraph. A Google search for the director shows a knowledge panel on the right side that shows the same first paragraph of the article as an excerpt. However, the first paragraph only includes information about *David O. Russell*'s early career. Moving the link to the film *Joy* to the first paragraph would not only better inform users of Google about the director but also add the article on *Joy*, which is in *IN*, to the *SCC* by closing a circle. Additionally, the article on *Joy Mangano*, which lies on a path from *Joy* to the *SCC* would also be added to the *SCC*. The two articles accounted for a total of 874, 423 pageviews in January 2016.

of the articles on knowledge and know-how. Interestingly, for the Japanese Wikipedia, the main cycle consists of the articles on person and interpersonal relationship. A possible explanation for this could be that this is an artifact of the importance of status of relationships in the Japanese language, which uses an extensive range of honorific suffixes for addressing conversational partners.

5. IMPROVING NAVIGABILITY

The previous section has shown that restricted navigational views exert a strong influence on navigability. For visitors who only look at the links in a part of the article, navigability of the resulting article network is substantially reduced due to reduced size of the set of articles in the *SCC*.

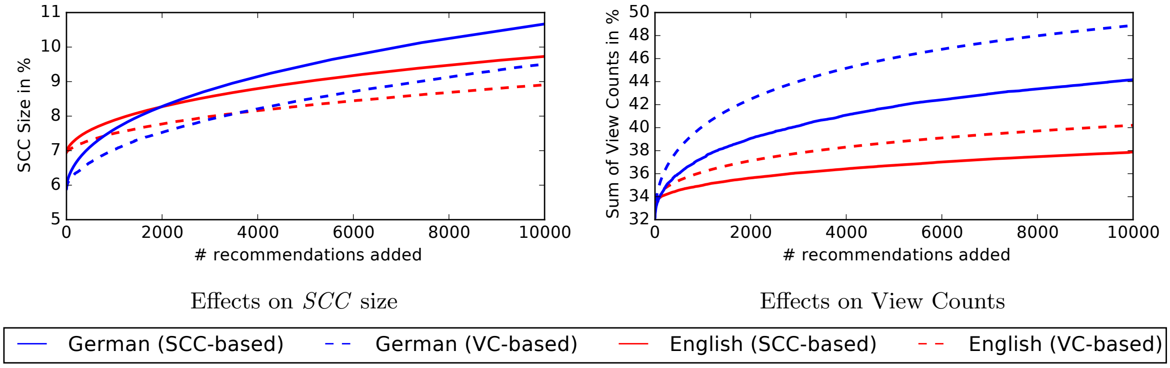
In this section, we propose an approach for improving navigability in an article network. Several methods already exist for augmenting the link structure of Wikipedia [24, 26, 41]. However, none of the approaches introduced so far considers restricted navigational views as criteria for navigability. The novel method we propose chooses additional links based on the bow tie structure of the network and takes the size of the *SCC* into account as well as exploiting the information about connectivity between the *IN*, *SCC*, and *OUT* components of the bow tie model.

Figure 4 shows an example of the application of our method. The main article shown is part of the *SCC*. The highlighted link points to an article in *IN*. Moving this link to the first lead paragraph would add it to the *SCC* by closing a circle to it, and also have the side effect of introducing a second article (that it links to) to the *SCC*. In what follows, we describe our method in more details.

5.1 Link Recommendation Approach

Our proposed approach of recommending links to increase navigability in the article networks makes use of the bow tie analysis. Generally, in order to add an article A to the mutually reachable set of articles of the *SCC*, two links are necessary: one link from A to an article of the *SCC* and a second link from an article in the *SCC* back to A. However, should one of the two links already be present, then adding the other type of link suffices. Moreover, there need not be a direct link from A to an *SCC* node: It is enough that there be a path from A that reaches an *SCC* node.

The information about the links and paths is contained in the bow tie model: articles in *IN* have a path leading to the *SCC*, and articles in *OUT* are reachable from it. As the *IN* component largely dominates the *OUT* component in the Wikipedia article networks, we will focus our attention on this component in what follows. However, the approach works for *OUT* in much the same way. In the following, we



Effects on *SCC* size — Effects on View Counts

Figure 5: **Effects of Recommendations on the *SCC* Size and the Sum of View Counts**. For demonstration purposes, we compute the effects of adding the top $10,000$ recommendations to the first paragraphs of the English and German Wikipedias. The y-axes show the fraction of articles in the *SCC* (left) and the fraction of the sum of view counts covered by articles in the *SCC*. The x-axes show the number of link recommendations added. The ranking of link candidates by *SCC* size (SCC-based) and by the sum of view counts introduced to the *SCC* (VC-based) shows a trade-off between these two effects.

will describe the steps necessary to compute the recommendations.

Computation of Link Candidates. The proposed method selects link recommendations from a given navigational view for inclusion in a more restricted view. For example, all links that are present in article in the entire lead but not in the first lead paragraph can potentially be recommended for inclusion in the first lead paragraph if that were to make the network more navigable as a result. The recommendations are therefore links to articles that have already proven to be semantically relevant for the article by the community of Wikipedia editors. The recommended links could then be made more visible to users. This could be accomplished in several ways—for example, links could be emphasized by displaying them in italics or bold face. Perhaps an easier way, however, would be to move a link into a more visible section of the article, if this makes sense in the context of that section. As such, our proposed method would lend itself well to a link recommender that offers suggestions to Wikipedia editors wishing to make an article more navigable.

Ranking of Link Candidates. The computation of link candidates results in a large number of potential links. Next, we propose two methods to rank the links.

1. **Ranking by number of articles added to the *SCC*.** Each link increases the number of articles added to the *SCC* by at least one article, but potentially several more. Specifically, consider an article that is located at the start of a path leading to the *SCC* via several hops. If that article receives a link from the *SCC*, all articles on the path become part of the *SCC* as well. A natural ranking method is therefore to rank link candidates by the number of articles that the link would add to the *SCC*.

2. **Ranking by sum of article view counts added to the *SCC*.** A second method to rank the link candidates is to take their importance in terms of view counts into account. We can hence rank link candidates by the sum of the view counts of all articles that a link adds to the *SCC*. This approach favors popular articles, which cannot be reached from the *SCC* without the added link and which can only be found via a search engine or a direct URL manipulation.

Introduction of Recommendations. Finally, the links can be moved or otherwise emphasized in the corresponding Wikipedia article. We propose that the computed information could be used as supplemental information for Wikipedia editors and shown alongside links in the edit view. In this way, the decision for what links to include would remain in the hands of the editors. In addition, the navigational effects of all other links could be made available to establish what effects the removal of a certain link would have.

5.2 Example of Link Recommendations

We now demonstrate the effects of making use of the recommended links. To this end, we compute both candidate ranking methods for the two largest Wikipedia language versions in our datasets (namely English and German) and incorporate the $10,000$ top-ranked links in the network. For the example, we assume a navigational view of the links in the first link paragraph. This is the view that users looking at results provided by the Google search engine showing excerpts of articles in the knowledge panel would have. We then select link candidates from the next-largest view, which are the links from the entire lead.

Figure 5 shows the results for the exemplary application of this method to the English and German Wikipedias. The results show that both ranking methods lead to increases for the corresponding metric. However, it also shows that a trade-off exists between increasing the size of the *SCC* and increasing the number of page views covered by the *SCC*. Both effects could be made available to Wikipedia editors, who could then make the editorial decisions whether to emphasize a link.

6. DISCUSSION AND CONCLUSION

In this article we have assessed the navigability of eight large Wikipedia language versions and suggested a method to recommend links.

RQ1. What is the state of navigability across multiple language versions of Wikipedia?

When taking all links in the articles into account, the vast majority of all articles are contained in the largest strongly connected component and are mutually reachable. However, if we look at Wikipedia with a more restricted navigational view, navigability is substantially reduced. When looking only at the links in the lead section, the fraction of mutually reachable nodes decreases to 16 to 37%. When further restricting the view to the first lead paragraph, which mirrors the excerpt that could be shown as supplementary information by an external search engine, navigability further decreases.

RQ2. How can this analysis be exploited to suggest improvements to navigability?

To improve navigability, we have proposed a link recommendation algorithm based on the bow tie analysis of the article network. The algorithm selects links for a navigational view from a less restricted view. The suggested links are therefore semantically relevant and could be introduced by editing the text and emphasizing them or moving them into a more visible region of the article. We have shown the effects of introducing link recommendations based on two examples. In a real-world setting, the decisions about link recommendations would be left to Wikipedia editors. Editors wishing to take navigability into account could be shown the additional information about the effects of specific links and receive suggestions to better connect articles.

Limitations and Future Work. The presented navigational views were selected based on evidence in previous studies suggesting that users dedicate a large portion of their attention to the area located close to the top. Due to the dynamic width of the Wikipedia in its Desktop view, the exact size of the viewport or the area above the fold is dynamic as well and no universally applicable method of establishing the number of links visible exists. The selected navigational views are therefore necessarily approximations to the true user viewports. However, the evaluation approach we presented is general and can be applied to any navigational view to analyze its effects on navigability. In future work, it could easily be adapted to test the effects of several specific screen resolutions.

The different language versions of Wikipedia generally have little overlap in their coverage of concepts [9]. In addition, every language edition of Wikipedia can establish its own policies and guidelines. Despite this, the eight language versions we investigated led to similar structure in terms of the bow tie model. While the navigational guidelines are likely to be influenced by the English language versions, it would be interesting to explore the differences and commonalities and their effects on navigability of the article networks in future work.

Finally, the use of the view counts to rank the importance of articles is a proxy measure that is subject to influence of crawlers, Wikipedia bots, traffic spikes due to external factors and periodic events such as holidays. For this work, we used the sum of all view counts within the most recent month before the page dump. While this is potentially subject to these limitations, it also carries with it the advantage of bringing to the attention articles that were popular but not reachable at that specific point in time.

We hope that our work stimulates discussion about the navigational effects of restricted views of Wikipedia articles and about methods to highlight the navigational effects of link editing to the Wikipedia community.

7. ACKNOWLEDGMENTS

This research was supported by a grant from the Austrian Science Fund (FWF) [P24866].

8. REFERENCES

[1] S. F. Adafre and M. de Rijke. Discovering missing links in wikipedia. In *Proceedings of the 3rd International Workshop on Link Discovery*, pages 90–97, 2005.

[2] D. Brezhnev, S. Trusheim, and V. Yendluri. All paths lead to philosophy. part of the Stanford Network Analysis Project, 2013.

[3] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener. Graph structure in the web. *Computer Networks*, 33(1):309—-320, 2000.

[4] A. Capocci, V. D. Servedio, F. Colaiori, L. S. Buriol, D. Donato, S. Leonardi, and G. Caldarelli. Preferential attachment in the growth of social networks: The internet encyclopedia wikipedia. *Physical Review E*, 74(3):03611:3–6, 2006.

[5] J. Davidson, B. Liebald, J. Liu, P. Nandy, T. V. Vleet, U. Gargi, S. Gupta, Y. He, M. Lambert, B. Livingston, and D. Sampath. The youtube video recommendation system. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 293–296, 2010.

[6] D. Davis, R. Lichtenwalter, and N. V. Chawla. Multi-relational link prediction in heterogeneous information networks. In *Proceedings of the 2011 International Conference on Advances in Social Networks Analysis and Mining*, 2011.

[7] D. Dimitrov, , P. Singer, F. Lemmerich, and M. Strohmaier. Visual positions of links and clicks on wikipedia. In *Proceedings of the 25th International Conference on World Wide Web*, 2016.

[8] I. Halatchliyski, J. Moskaliuk, J. Kimmerle, and U. Cress. Who integrates the networks of knowledge in wikipedia? In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, pages 1:1–1:10, 2010.

[9] B. Hecht and D. Gergle. The tower of babel meets web 2.0: User-generated content and its applications in a multilingual context. In *Proceedings of the SIGCHI Conference on Human Factors in Computing systems*, pages 291–300, 2010.

[10] D. Helic, M. Strohmaier, M. Granitzer, and R. Scherer. Models of human navigation in information networks based on decentralized search. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 89–98, 2013.

[11] 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*, pages 417–426, 2011.

[12] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53, 2004.

[13] M. Ibrahim, C. M. Danforth, and P. S. Dodds. Connecting every bit of knowledge: The structure of wikipedia's first link network. *arXiv pre-print*, 2016. arXiv:1605.00309 [cs.SI].

[14] J. Kamps and M. Koolen. Is wikipedia link structure different? In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 232–241, 2009.

[15] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.

[16] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14*, pages 431–438, 2001.

[17] J. Kunegis, J. Preusse, and F. Schwagereit. What is the added value of negative links in online social networks? In *Proceedings of the 22nd International Conference on World Wide Web*, pages 727–736, 2013.

[18] D. Lamprecht, F. Geigl, T. Karas, S. Walk, D. Helic, and M. Strohmaier. Improving recommender system navigability through diversification: A case study of IMDb. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pages 21:1–21:8, 2015.

[19] D. Lamprecht, K. Lerman, D. Helic, and M. Strohmaier. How the structure of wikipedia articles influences user navigation. *New Review of Hypermedia and Multimedia*, 2016. to appear.

[20] D. Lamprecht and D. H. M. Strohmaier. Improving reachability and navigability in recommender systems. *arXiv pre-print*, 2015. arXiv:1507.08120 [cs.IR].

[21] D. Lamprecht, M. Strohmaier, D. Helic, C. Nyulas, T. Tudorache, N. F. Noy, and M. A. Musen. Using ontologies to model human navigation behavior in information networks: A study based on wikipedia. *Semantic Web*, 6(4):403–422, 2015.

[22] J. Leskovec, D. Huttenlocher, and J. M. Kleinberg. Predicting positive and negative links in online social networks. In *Proceedings of the 19th International Conference on World Wide Web*, pages 641–650, 2010.

[23] D. Liben-Nowell and J. M. Kleinberg. The link-prediction problem for social networks. *Journal of the American Society for Information Science and Technology*, 58(7):1019–1031, 2007.

[24] R. Mihalcea and A. Csomai. Wikify!: Linking documents to encyclopedic knowledge. In *Proceedings of the 16th ACM Conference on Information and Knowledge Management*, pages 233–242, 2007.

[25] S. Milgram. The small world problem. *Psychology Today*, 1(2):60–67, 1967.

[26] D. Milne and I. H. Witten. Learning to link with wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 509–518, 2008.

[27] J.-H. Oh, D. Kawahara, K. Uchimoto, J. Kazama, and K. Torisawa. Enriching multilingual language resources by discovering missing cross-language links in wikipedia. In *Proceedings of the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*, pages 322–328, 2008.

[28] A. Paranjape, R. West, L. Zia, and J. Leskovec. Improving website hyperlink structure using server logs. In *Proceedings of the 9th International Conference on Web Search and Data Mining*, pages 615–624, 2016.

[29] J. Preusse, J. Kunegis, M. Thimm, and S. Sizov. Decline - models for decay of links in networks. *arXiv pre-print*, 2014. arXiv:1403.4415 [cs.SI].

[30] P. Singer, T. Niebler, M. Strohmaier, and A. Hotho. Computing semantic relatedness from human navigational paths: A case study on wikipedia. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 9(4):41–70, 2013.

[31] P. Sorg and P. Cimiano. Enriching the crosslingual link structure of wikipedia—a classification-based approach. In *Proceedings of the AAAI 2008 Workshop on Wikipedia and Artifical Intelligence*, pages 49–54, 2008.

[32] Y. Sun, J. Han, C. C. Aggarwal, and N. V. Chawla. When will it happen?—relationship prediction in heterogeneous information networks. In *Proceedings of the 5th ACM International Conference on Web Search and Data Mining*, pages 663–672, 2012.

[33] J. Teevan, C. Alvarado, M. S. Ackerman, and D. R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 415–422, 2004.

[34] E. G. Toms. Serendipitous information retrieval. In *Proceedings of the DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries*, 2000.

[35] C. Trattner, P. Singer, D. Helic, and M. Strohmaier. Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, pages 14:1–14:8, 2012.

[36] J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(1):425–443, 1969.

[37] Z. Wang, J. Li, Z. Wang, and J. Tang. Cross-lingual knowledge linking across wiki knowledge bases. In *Proceedings of the 21st International Conference on World Wide Web*, pages 459–468, 2012.

[38] D. J. Watts, P. S. Dodds, and M. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.

[39] D. J. Watts and S. H. Strogatz. Collective dynamics of small-world networks. *Nature*, 393(6684):440–442, 1998.

[40] R. West and J. Leskovec. Human wayfinding in information networks. In *Proceedings of the 21st International Conference on World Wide Web*, pages 619–628, 2012.

[41] R. West, A. Paranjape, and J. Leskovec. Mining missing hyperlinks from human navigation traces: A case study of wikipedia. In *Proceedings of the 24th*

International Conference on World Wide Web, pages 1242–1252, 2015.

[42] R. West, D. Precup, and J. Pineau. Completing wikipedia's hyperlink structure through dimensionality reduction. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 1097–1106, 2009.

[43] E. Wulczyn and D. Taraborelli. Wikipedia clickstream, 2015. http://dx.doi.org/10.6084/m9.figshare.1305770, accessed Februar 17, 2016.

[44] E. Wulczyn, R. West, L. Zia, and J. Leskovec. Growing wikipedia across languages via recommendation. In *Proceedings of the 25th International Conference on World Wide Web*, 2016.

4 Conclusions

The emergence of electronic devices and the ensuing digital revolution has greatly simplified information storage, and information production rates have skyrocketed as a result. A 2010 study estimated that the same amount of information from the beginning of humanity up until the present was then being produced in just two days [Siegler, 2010]. Information is rarely stored in isolation—in fact, most pieces of information depend on other pieces and reference them, thereby forming an information network. The ubiquity of digital information storage has made the ability to successfully retrieve information a vital everyday skill, and navigating information networks forms an essential part of this process.

This thesis has investigated *navigability*, the ability of a networked information system to support information retrieval by means of following links between nodes. The nature of link placements is becoming increasingly dynamic and adaptive, as is made evident by collaborative websites such as Wikipedia or recommender systems now in use at most large e-commerce sites. This thesis has analyzed some of the factors influencing navigation decisions of users and has highlighted the substantial role of page organization. Most centrally, this thesis has presented a coherent evaluation framework for navigability, whose main part is a model for navigation based on decentralized search. Finally, this thesis has produced suggestions for website operators who wish to improve navigability on their sites.

This chapter concludes this thesis and is structured as follows. Section 4.1 summarizes the research questions, results, and contributions. Section 4.2 discusses the implications of this thesis. Section 4.3 lists the limitations, and, finally, Section 4.4 discusses potential avenues for future work.

4.1 Results and Contributions

This section summarizes the answers to the research questions presented in Section 1.7.

RQ1: What factors influence the navigational choices of users?

To model and ultimately improve navigability in networked information systems, it is fundamental to first develop a thorough understanding of the link selection process in human navigation. Analyzing the influences on the navigation behavior of users allows to more accurately model navigation and to assess the weaknesses of navigability in a system. This in turn paves the way for the generation of effective suggestions to improve the retrieval of information via navigation, which can then be applied by website operators as necessary.

There exists a variety of potential influences on link selection in navigation. To answer the first research question, I have presented Article 1 [Lamprecht et al., 2016b] which investigates influencing factors in Wikipedia navigation. The article models a range of influences, representing the popularity and generality of link target articles, the semantic similarity to the target articles and the influence of page organization. The influences are modeled by a probabilistic click model that selects links with probability proportional to a given influencing factor, e.g., with respect to degree.

The findings of this article have shown that page organization, including the arrangement into sections and the positioning of links, is able to best explain the majority of all link selection cases. Additionally, for goal-directed Wikipedia navigation, the article has shown that the first step (i.e., the first click going away from the article a user starts from) can be best explained by a bias towards generality, as modeled for example by the in-degree of a target article. Likewise, the last step can be best explained by a bias towards the semantic similarity with the navigation target.

A better understanding of human navigation behavior is essential to making information more accessible to a broader audience. The analysis of influencing factors has provided important insights into human navigation behavior and has shown up the importance of page organization with respect to link selection behavior. These findings enable us to better characterize the navigation behavior of users based on structural influences and open up ways to address page organization to better support navigability on a website. The answers to the first research question therefore form the foundation for further analysis and serve to inspire ideas to improve navigability by reorganizing webpage content.

RQ2: How can navigation be modeled?

Information systems support information retrieval with a variety of techniques and interfaces. The availability of a method to automatically evaluate the effects of interface changes on navigability would greatly simplify adjusting and testing interfaces regarding the accessibility to information on a website.

To tackle this research question, I have presented two articles that introduce a method to model navigation. Article 2 [Lamprecht et al., 2016c] has introduced the general-purpose method to model and simulate navigation dynamics on networked information systems and demonstrated it on recommender systems. The method models navigation based on three information seeking scenarios that cover a variety of use cases, namely point-to-point search, berrypicking and information foraging. The underlying link selection algorithm is based on the established decentralized search algorithm. The method evaluates simulations of navigation on networks to reveal insights into the navigation dynamics and offer information about navigational success rates. The code to reproduce this method is available as an open-source software[1].

Article 3 [Lamprecht et al., 2015a] applies this method to the real-world recommender systems of the Internet Movie Database (IMDb). To this end, the article first evaluates topological characteristics in the recommendation

[1] https://github.com/lamda/RecNet

networks by investigating components, path lengths and by performing a bow-tie analysis. Then, the article applies decentralized search to simulate the navigation dynamics by evaluating navigation scenarios designed to retrieve a subset of popular IMDb titles.

The presented method is able to model navigation dynamics on arbitrary networked information systems and can supply website operators with a powerful tool to automatically evaluate the navigational ramifications of changes to the structure of an information network. By potentially avoiding the need for user studies, the method allows to quickly gather facts for decision-making. The findings show how the method can be applied to adjust recommender systems by selecting an appropriate recommendation algorithm or fine-tuning an existing one.

RQ3: How can navigability in information systems be improved?

The first two research questions have investigated the factors influencing navigation behavior, and the corresponding research articles have presented a method to model navigation in an automated manner based on decentralized search. The application of this method to evaluate networked information systems has shown up the state of navigability and exposed some of the properties affecting it. To answer the third research question, I have presented two articles that suggest improvements to navigability in networked information systems based on the findings of the first two research questions.

Article 3 [Lamprecht et al., 2015a] has investigated the navigability of the recommender systems of IMDb and has presented the diversification of an existing link structure as a way to increase navigability. Whereas the simplest method to increase navigability consists of the introduction of novel links into a network, an increasing number of links per page can be difficult for humans to process. This is made evident by the focus of users on just the links located at the top of Wikipedia articles and other websites, also known as a position bias. It is therefore often desirable to modify an existing link structure by changing some of the link targets while keeping the number of links constant. Article 3 has demonstrated this on recommender systems, where a selected few of the existing recommendations were swapped out for diversified ones. Thereby, recommendations were kept relevant and yet pointed to a link target with a different topic, located at a different area in the network. This approach has led to substantial improvements in the number of mutually reachable nodes in the network, and success rates of the evaluation based on decentralized search saw an up to threefold increase.

Based on the findings from Article 1 [Lamprecht et al., 2016b] with respect to the influence on link selection, Article 4 [Lamprecht et al., 2016a] has evaluated the influence of page organization on the navigability of the eight largest Wikipedia language versions. The findings have shown that overall, the article networks of Wikipedia are very well-connected. However, when restricting the view port to the sections that receive the majority of user attention, navigability substantially decreases and the number of mutually reachable articles plummets to around a third. To improve this, the article has presented a recommendation algorithm that suggests links to move from the body of the article into the more visible sections. The recommendations are able to increase the number of mutually reachable articles while keeping the number of links constant.

4.2 Implications of this Work

Being able to successfully retrieve information in complex and changing information networks has become a critical skill. To facilitate access to information, a better understanding of the navigation behavior of users is necessary. This can then be translated into models of navigation to yield valuable insights and advice. This thesis has shed some new light on these aspects, and I am confident that both future research and real-world applications can benefit from it. The following section lists the implications of this thesis in more details.

Influence of Page Organization on Navigability. Link selection during navigation of information networks is subject to a variety of influences, such as the similarity of a potential link to a navigation goal, the generality or abstractness of a link target, or the organization of the current page in terms of layout and link positions. In this thesis, I have shown that page organization exerts the strongest influence of all of the investigated factors. This has a range of implications for interface design. First of all, website operators may wish to keep all navigationally important links in the most visible sections of a webpage, and should keep in mind that links that are not immediately visible will potentially not affect navigability. From a different perspective, a nefarious agent may attempt to change the perception of, for example, a Wikipedia article not by changing any of its words but by instead tampering with the order of sections, thus substantially changing how the article is perceived. Finally, website operators could exploit this knowledge to actively direct users towards certain pages (e.g., for promotions) by moving certain links into more visible sections.

A Novel Evaluation Method for Navigability. Traditionally, a common method to expose the actual navigation dynamics on a network has been the evaluation of a user study probing the usability and effectiveness of the navigational aids. However, involving humans as test subjects is expensive, rendering it infeasible in all but a very limited number of cases. To avoid user studies, navigational models such as PageRank or information foraging have been used as a proxy. This thesis has presented a novel evaluation method for navigability based on decentralized search, which has previously been successfully applied to model human navigation characteristics in information networks. The presented method extends the arsenal of navigation models with a powerful tool to assess the state of navigability in arbitrary information networks. It models navigation based on given targets, uses an on an explicitly specified background knowledge, and is adaptable to a wide range of scenarios. By automatically simulating multiple navigation scenarios, the presented method can be used to efficiently compare the effects of changes to the network structure. This can reduce or altogether avoid the need for user studies and help to quickly narrow down the space of possible adaptations to make a site more accessible.

Diversified Recommendation Algorithms. This thesis has demonstrated the feasibility and utility of the presented evaluation method by applying it to several recommendation algorithms. The findings have shown that comparably simpler algorithms, such as text-book collaborative filtering, can lead to poorer navigability and are potentially subject to a popularity bias. More complex algorithms, such as matrix factorization, are not subject to these factors as much, and generally fared better. An alternative way to overcome issues with navigability consists of the introduction of diversified recommendations to a standard recommendation algorithm. By replacing a small fraction of the existing recommendations with still relevant but more diverse recommendations, navigability can be substantially increased. This approach is generally applicable to any kind of information network for which a better navigability is desired. To this end, the approach would first inspect the links present on each page, then compute a larger set of still relevant links, and finally introduce or replace a few links with links leading to more diverse areas of the network.

4.3 Limitations

The methods and analyses presented by this thesis have been subject to certain limitations that will be discussed in what follows.

Generality of Findings. The generality of the presented findings is naturally restricted to the evaluated datasets. For this thesis, this means that the analysis of the influencing factors in navigation is limited to Wikipedia. The generality of the findings for the evaluation method for navigability is likewise limited to recommender systems. However, the presented methods are comprehensive and general, and the presented research reports contain detailed information so that the methodology can be easily adapted for arbitrary information networks. To extend the validity of these findings towards further kinds of systems, more work is required to corroborate it.

Data Restrictions. The analysis of human navigation relies on the availability of detailed log data from information systems. However, due to limitations concerning privacy or the confidentiality of business processes, only a very limited number of datasets containing click log data is available publicly for research purposes. To study navigation behavior on Wikipedia, this thesis has made use of a clickstream of the English Wikipedia, provided by the Wikimedia foundation. This has the advantage of providing real click data from users of Wikipedia, at the disadvantage of not allowing to view separate click trails by individual users. To supplement this, this thesis has also used click trails from wikigames, which do include separate paths for users, at the disadvantage of having been collected as part of navigational games, therefore potentially introducing biases into the data. Likewise, hardly any click data is available to study recommender systems. For this reason, this thesis has studied recommendation algorithms, which can be used to compute recommendation networks from more easily available rating data. Additionally, the structure of real-world recommendation networks was obtained with the help of a web crawler.

Modeling of Human Behavior. Decentralized search, the search algorithm used as the foundation for the presented evaluation method, has been established as a method leading to characteristics comparable to human navigation behavior. However, the extension of decentralized search to recommender systems and further information seeking scenarios has not been subject to a comparative user study. As such, this thesis does not claim that the presented evaluation method can completely accurately model human navigation by default, as it is subject to many parameters such as the population of users on a site, their background knowledge, and their familiarity with the site's topic. However, the method can provide important and relevant insights into navigation dynamics, and can also be easily fine-tuned to fit these parameters and produce more accurate results tailored towards a specific system.

Restrictions to Desktop Views. The presented analysis has largely focused on devices with larger screens, such as desktop or laptop computers. This is especially the case for the Wikipedia clickstream, which does not include data for the mobile Wikipedia site. Likewise, for recommender systems, the assumptions about the number of links shown hold mainly for the desktop view of a website. However, all analyses can be repeated analogously for smaller devices.

4.4 Future Work

This final section shows up potential avenues for future work based on the findings of this thesis.

Extending the Analysis of Navigational Influences. The analysis of the factors influencing navigation (and in particular, link selection) represents a first foray into measuring and comparing navigational influences. The presented methodology is extensible and can be used to take the effects of further cognitive biases occurring on the Web into account. The study of the influence of page organization can be logically extended to compare competing sections on a webpage, such as redundant navigational structures on Wikipedia (namely infoboxes, category information and navboxes), frequently containing a large overlap of links. This could also be used to evaluate the effects of policies on the resulting Wikipedia articles. As another example, the policy of only converting the first occurrence of a concept into a link could be evaluated based on the number of times links are clicked if they do erroneously occur multiple times. Finally, repeating the analysis on click paths from real-world systems would give an indication about any biases presented in the data stemming from wikigames.

Refinement of the Evaluation Method. The presented evaluation method focuses on modeling the navigational parts of information seeking. By doing so, it allows to assess the state of navigability of a system. In future work, the method could be adjusted in two ways. Firstly, it could be tuned to the actual demands of a specific site. This would involve the comparison with click data (or a user study) to identify appropriate components for the background knowledge and the identification of any further modifications to best match user behavior in any desired criteria.

As an example, if users of a site commonly combine the use of a search function and navigation, a teleportation property could be built into the method, modeling jumps between disconnected nodes enabled by searching. Secondly, the method could generally be extended to model further information seeking scenarios. An example of this is the representation of users with varying degrees of information about a site, such as having memorized preferred paths or subpages, or using specific landing pages. More attention could also be dedicated to differences between user groups, such as novices and experts, or users on handheld and desktop devices, all potentially requiring different navigational structures.

Apply Findings to Real Systems. The analyses conducted by this thesis have led to several practical suggestions for real-wold information systems. As the logical next step, these suggestions could be implemented and rolled out to real-world systems and compared in A/B tests. The execution of these ideas in real-world scenarios will likely raise new questions and issues, and show up any side effects that the proposed changes might have. A rewarding first step could be the introduction of a Wikipedia bot or tool that shows up the navigational effects of changes introduced by editors. This would have the advantage of leaving the decisions of what links to introduce and where to place them in the hands of a human, while still offering suggestions to improve navigability on the encyclopedia. In a similar way, recommender system operators wishing to make their systems more navigable can easily introduce a few diversified recommendations and evaluate the effects.

Research on navigation in networks has progressed a long way since Milgram and Traver's small world experiments relying on forwarding letters. With this thesis, I hope to have shown up ways to encourage future research forays in this field, as well as to have provided valuable suggestions to make information networks more accessible.

List of Figures

List of Algorithms

Bibliography

Adafre, S. F. and de Rijke, M. (2005). Discovering missing links in wikipedia. In *Proceedings of the 3rd International Workshop on Link Discovery*, pages 90–97.

Adamic, L. and Adar, E. (2005). How to search a social network. *Social Networks*, 27(3):187–203.

Adamic, L. A., Lukose, R. M., Puniyani, A. R., and Huberman, B. A. (2001). Search in power-law networks. *Physical Review E*, 64(4):046135:1–046135:8.

Agrawal, R. and Srikant, R. (1994). Fast algorithms for mining association rules in large databases. In *Proceedings of the 20th International Conference on Very Large Data Bases*, pages 487–499.

Alexa (2017). Wikipedia.org traffic statistics. http://www.alexa.com/siteinfo/wikipedia.org, accessed April 26, 2017.

Andrews, K., Kappe, F., and Maurer, H. (1996). The hyper-g network information system. In *Journal of Universal Computer Science*, pages 206–220. Springer.

Backstrom, L., Boldi, P., Rosa, M., Ugander, J., and Vigna, S. (2012). Four degrees of separation. In *Proceedings of the 4th Annual ACM Web Science Conference*, pages 33–42. ACM.

Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5):407–424.

Bell, R. M. and Koren, Y. (2007). Improved neighborhood-based collaborative filtering. In *KDD Cup and Workshop at the 13th ACM SIGKDD*

International Conference on Knowledge Discovery and Data Mining, pages 7–14.

Berners-Lee, T. and Cailliau, R. (1990). Worldwideweb: Proposal for a hypertext project. www.w3.org/Proposal.html, accessed September 21, 2016.

Bestavros, A. (1995). Using speculation to reduce server load and service time on the www. In *Proceedings of the Fourth International Conference on Information and Knowledge Management*, pages 403–410. ACM.

Bilenko, M. and White, R. W. (2008). Mining the search trails of surfing crowds: Identifying relevant websites from user activity. In *Proceedings of the 17th International Conference on World Wide Web*, pages 51–60.

Blum, A., Chan, T.-H. H., and Rwebangira, M. R. (2006). A random-surfer web-graph model. In *Proceedings of the Meeting on Analytic Algorithmics and Combinatorics*, pages 238–246. Society for Industrial and Applied Mathematics.

Boguñá, M., Krioukov, D., and Claffy, K. C. (2009). Navigability of complex networks. *Nature Physics*, 5(1):74–80.

Boguñá, M., Papadopoulos, F., and Krioukov, D. (2010). Sustaining the internet with hyperbolic mapping. *Nature Communications*, 1:62.

Boim, R., Milo, T., and Novgorodov, S. (2011). Diversification and refinement in collaborative filtering recommender. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 739–744.

Bollen, D., Knijnenburg, B. P., Willemsen, M. C., and Graus, M. (2010). Understanding choice overload in recommender systems. In *Proc. of the fourth ACM Conference on Recommender Systems*, pages 63–70. ACM.

Brin, S. and Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the 7th International Conference on World Wide Web*.

Broder, A. (2002). A taxonomy of web search. In *ACM Sigir Forum*, pages 3–10. ACM.

Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopalan, S., Stata, R., Tomkins, A., and Wiener, J. (2000). Graph structure in the web. *Computer Networks*, 33(1):309—-320.

Budiu, R. (2014). Search is not enough: Synergy between navigation and search. www.nngroup.com/articles/search-not-enough, accessed September 20, 2016.

Buscher, G., Cutrell, E., and Morris, M. R. (2009). What do you see when you're surfing? using eye tracking to predict salient regions of web pages. In Dan R Olsen, J., Arthur, R. B., Hinckley, K., Morris, M. R., Hudson, S., and Greenberg, S., editors, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 21–30, New York, NY. ACM.

Bush, V. (1945). As we may think. *The Atlantic Monthly*, 176(1):101–108.

Cano, P., Celma, O., Koppenberger, M., and Buldú, J. M. (2006). Topology of music recommendation networks. *Chaos*, 16(1).

Capocci, A., Servedio, V. D., Colaiori, F., Buriol, L. S., Donato, D., Leonardi, S., and Caldarelli, G. (2006). Preferential attachment in the growth of social networks: The internet encyclopedia wikipedia. *Physical Review E*, 74(3):03611:3–6.

Carmody, S., Gross, W., Nelson, T. H., Rice, D., and Dam, A. V. (1969). A hypertext editing system for the/360. *Pertinent Concepts in Computer Graphics*, pages 291–330.

Castells, P., Hurley, N. J., and Vargas, S. (2015). Novelty and diversity in recommender systems. In *Recommender Systems Handbook*, pages 881–918. Springer.

Celma, Ò. and Herrera, P. (2008). A new approach to evaluating novel recommendations. In *Proceedings of the 2nd ACM Conference on Recommender systems*, pages 179–186.

Chi, E. H., Pirolli, P., Chen, K., and Pitkow, J. (2001). Using information scent to model user information needs and actions on the web. In

Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pages 490–497.

Chierichetti, F., Kumar, R., Raghavan, P., and Sarlos, T. (2012). Are web users really markovian? In *Proceedings of the 21st International Conference on World Wide Web*, pages 609–618. ACM.

Clarke, C. L., Agichtein, E., Dumais, S., and White, R. W. (2007). The influence of caption features on clickthrough patterns in web search. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 135–142. ACM.

Covington, P., Adams, J., and Sargin, E. (2016). Deep neural networks for youtube recommendations. In *Proceedings of the 10th ACM Conference on Recommender Systems*, pages 191–198. ACM.

Davidson, J., Liebald, B., Liu, J., Nandy, P., Vleet, T. V., Gargi, U., Gupta, S., He, Y., Lambert, M., Livingston, B., and Sampath, D. (2010). The youtube video recommendation system. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 293–296.

Dimitrov, D., , Singer, P., Lemmerich, F., and Strohmaier, M. (2016). Visual positions of links and clicks on wikipedia. In *Proceedings of the 25th International Conference on World Wide Web*.

Dimitrov, D., Singer, P., Helic, D., and Strohmaier, M. (2015). The role of structural information for designing navigational user interfaces. In *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, pages 59–68. ACM.

Dodds, P. S., Muhamad, R., and Watts, D. J. (2003). An experimental study of search in global social networks. *Science*, 301(5634):827–829.

Engelbart, D. C. (1988). A conceptual framework for the augmentation of man's intellect. In *Computer-Supported Cooperative Work*, pages 35–65. Morgan Kaufmann Publishers Inc.

Erickson, B. H. (1979). Some problems of inference from chain data. *Sociological Methodology*, 10(1):276–302.

Fountain, A. M., Hall, W., Heath, I., and Davis, H. C. (1990). Microcosm: An open model for hypermedia with dynamic linking. In *Proceedings of the European COnference on Hypertext*, pages 298–311.

Ge, M., Delgado-Battenfeld, C., and Jannach, D. (2010). Beyond accuracy: Evaluating recommender systems by coverage and serendipity. In *Proceedings of the 4th ACM Conference on Recommender Systems*, pages 257–260.

Geigl, F. and Helic, D. (2014). The role of homophily and popularity in informed decentralized search. In *Proceedings of the 2nd International Workshop on Dynamic Networks and Knowledge Discovery*.

Geigl, F., Lamprecht, D., Hofmann-Wellenhof, R., Walk, S., Strohmaier, M., and Helic, D. (2015). Random surfers on a web encyclopedia. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pages 5:1–5:8.

Gleich, D. F., Constantine, P. G., Flaxman, A. D., and Gunawardana, A. (2010). Tracking the random surfer: Empirically measured teleportation parameters in pagerank. In *Proceedings of the 19th International Conference on World Wide Web*, pages 381–390. ACM.

González-Caro, C. and Marcos, M.-C. (2011). Different users and intents: An eye-tracking analysis of web search. *Proceedings of the 4th International Conference on Web Search and Data Mining*, pages 9–12.

Grujić, J. (2008). Movies recommendation networks as bipartite graphs. In *International Conference on Computational Science*, pages 576–583. Springer.

Gunawardana, A. and Shani, G. (2015). Evaluating recommender systems. In *Recommender Systems Handbook*, pages 265–308. Springer.

Halasz, F. (1988). Reflections on notecards: Seven issues for the next generation of hypermedia systems. *Communications of the ACM*, 31(7):836–852.

Halasz, F. G. (2001). Reflections on "Seven Issues": Hypertext in the Era of the Web. *ACM J. Comput. Doc.*, 25(3):109–114.

Halasz, F. G., Moran, T. P., and Trigg, R. H. (1987). Notecards in a nutshell. In *Proceedings of the SIGCHI/GI Conference on Human Factors in Computing Systems and Graphics Interface*, pages 45–52. ACM.

Halatchliyski, I., Moskaliuk, J., Kimmerle, J., and Cress, U. (2010). Who integrates the networks of knowledge in wikipedia? In *Proceedings of the 6th International Symposium on Wikis and Open Collaboration*, pages 1:1–1:10.

Hearst, M. A. (2009). *Search User Interfaces*. Cambridge University Press.

Helic, D., Körner, C., Granitzer, M., Strohmaier, M., and Trattner, C. (2012). Navigational efficiency of broad vs. narrow folksonomies. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, pages 63–72. ACM.

Helic, D. and Strohmaier, M. (2011). Building directories for social tagging systems. In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, pages 525–534. ACM.

Helic, D., Strohmaier, M., Granitzer, M., and Scherer, R. (2013). Models of human navigation in information networks based on decentralized search. In *Proceedings of the 24th ACM Conference on Hypertext and Social Media*, pages 89–98.

Helic, D., Strohmaier, M., Trattner, C., Muhr, M., and Lerman, K. (2011). Pragmatic evaluation of folksonomies. In *Proceedings of the 20th International Conference on World Wide Web*, pages 417–426.

Helic, D., Trattner, C., Strohmaier, M., and Andrews, K. (2010). On the navigability of social tagging systems. In *Proceedings of the 2nd IEEE International Conference on Social Computing*, pages 161–168.

Herlocker, J. L., Konstan, J. A., Terveen, L. G., and Riedl, J. T. (2004). Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems*, 22(1):5–53.

Hertzum, M. and Frøkjær, E. (1996). Browsing and querying in online documentation: A study of user interfaces and the interaction process.

ACM Transactions on Computer-Human Interaction (TOCHI), 3(2):136–161.

Hotchkiss, G., Sherman, T., Tobin, R., Bates, C., and Brown, K. (2010). Search engine results: 2010. *Enquiro Search Solutions (Technical Report)*, pages 1–61.

Jannach, D., Zanker, M., Felfernig, A., and Friedrich, G. (2010). *Recommender Systems: an Introduction*. Cambridge University Press.

Jansen, B. J., Spink, A., and Saracevic, T. (2000). Real life, real users, and real needs: A study and analysis of user queries on the web. *Information Processing & Management*, 36(2):207–227.

Juvina, I. and van Oostendorp, H. (2008). Modeling semantic and structural knowledge in web navigation. *Discourse Processes*, 45(4-5):346–364.

Juvina, I., van Oostendorp, H., Karbor, P., and Pauw, B. (2005). Toward modeling contextual information in web navigation. In Bara, B. G., Barsalou, L., and Bucciarelli, M., editors, *Proceedings of the 27th Annual Meeting of the Cognitive Science Society*, pages 1078–1083, Mahwah, NJ. Lawrence Erlbaum Associates.

Kamps, J. and Koolen, M. (2009). Is wikipedia link structure different? In *Proceedings of the Second ACM International Conference on Web Search and Data Mining*, pages 232–241.

Kaur, I. and Hornof, A. J. (2005). A comparison of lsa, wordnet and pmi-ir for predicting user click behavior. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 51–60. ACM.

Kitajima, M., Blackmon, M. H., and Polson, P. G. (2000). *People and Computers XIV — Usability or Else!*, chapter A Comprehension-Based Model of Web Navigation and its Application to Web Usability Analysis, pages 357–374. Springer London.

Kitajima, M., Polson, P. G., and Blackmon, M. H. (2007). CoLiDeS and SNIF-ACT: Complementary models for searching and sensemaking on the web. In Bhavnani, S., Chi, E., and Kellogg, W., editors, *Human*

Computer Interaction Consortium 2007 Winter Workshop, Ann Arbor, MI. Human Computer Interaction Consortium.

Kleinberg, J. M. (2000a). Navigation in a small world. *Nature*, 406(6798):845.

Kleinberg, J. M. (2000b). The small-world phenomenon: An algorithmic perspective. In *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing*, pages 163–170.

Kleinberg, J. M. (2001). Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems 14*, pages 431–438.

Koren, Y., Bell, R., and Volinsky, C. (2009). Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37.

Krioukov, D., Papadopoulos, F., Kitsak, M., Vahdat, A., and Boguñá, M. (2010). Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106:1–036106:18.

Lamprecht, D., Dimitrov, D., Helic, D., and Strohmaier, M. (2016a). Evaluating and improving navigability of wikipedia: A comparative study of eight language editions. In *Proceedings of the 12th International Symposium on Open Collaboration*, pages 17:1–17:10.

Lamprecht, D., Geigl, F., Karas, T., Walk, S., Helic, D., and Strohmaier, M. (2015a). Improving recommender system navigability through diversification: A case study of IMDb. In *Proceedings of the 15th International Conference on Knowledge Management and Knowledge Technologies*, pages 21:1–21:8.

Lamprecht, D., Helic, D., and Strohmaier, M. (2015b). Quo vadis? on the effects of wikipedia's policies on navigation. In *Proceedings of the Workshop on Wikipedia, a Social Pedia at the Ninth International AAAI Conference on Web and Social Media*, pages 64–66.

Lamprecht, D., Lerman, K., Helic, D., and Strohmaier, M. (2016b). How the structure of wikipedia articles influences user navigation. *New Review of Hypermedia and Multimedia*, 23(1):29–50.

Lamprecht, D., Strohmaier, M., and Helic, D. (2016c). A method for evaluating the navigability of recommendation algorithms. In *Proceedings of the 5th International Workshop on Complex Networks and their Applications*, pages 247–259.

Lamprecht, D., Strohmaier, M., Helic, D., Nyulas, C., Tudorache, T., Noy, N. F., and Musen, M. A. (2015c). Using ontologies to model human navigation behavior in information networks: A study based on wikipedia. *Semantic Web*, 6(4):403–422.

Lattanzi, S., Panconesi, A., and Sivakumar, D. (2011). Milgram-routing in social networks. In *Proceedings of the 20th International Conference on World Wide Web*, pages 725–734. ACM.

Lerman, K. (2016). Information is not a virus, and other consequences of human cognitive limits. *arXiv pre-print*. arXiv:1605.02660 [cs.CY].

Lerman, K. and Hogg, T. (2014). Leveraging position bias to improve peer recommendation. *PLOS ONE*, 9(6):e98914.

Lerman, K. and Jones, L. (2007). Social Browsing on Flickr. In *Proceedings of the 1st International Conference on Weblogs and Social Media*.

Li, Z. and Tian, J. (2003). Testing the suitability of markov chains as web usage models. In *Proceedings of the 27th Annual International Conference on Computer Software and Applications*, pages 356–361. IEEE.

Linden, G., Smith, B., and York, J. (2003). Amazon.com recommendations: Item-to-item collaborative filtering. *IEEE Internet Computing*, 7(1):76–80.

Liu, Y., Wang, C., Zhou, K., Nie, J., Zhang, M., and Ma, S. (2014). From skimming to reading: A two-stage examination model for web search. In *Proceedings of the 23rd ACM International Conference on Conference on Information and Knowledge Management*, pages 849–858. ACM.

Marchionini, G. (2006). Exploratory search: From finding to understanding. *Communications of the ACM*, 49(4):41–46.

Maurer, H. (1996). *HyperWave—The Next Generation Web Solution*. Addison-Wesley Longman Publishing Co., Inc.

Menczer, F. (2002). Growing and navigating the small world web by local content. *Proceedings of the National Academy of Sciences*, 99(22):14014–14019.

Meusel, R., Vigna, S., Lehmberg, O., and Bizer, C. (2014). Graph structure in the wweb—revisited. In *Proceedings of the Companion Publication of the 23rd International Conference on World Wide Web*, pages 427–432.

Mihalcea, R. and Csomai, A. (2007). Wikify!: Linking documents to encyclopedic knowledge. In *Proceedings of the 16th ACM Conference on Information and Knowledge Management*, pages 233–242.

Milgram, S. (1967). The small world problem. *Psychology Today*, 1(2):60–67.

Milne, D. and Witten, I. H. (2008). Learning to link with wikipedia. In *Proceedings of the 17th ACM Conference on Information and Knowledge Management*, pages 509–518.

National Center for Supercomputing Applications (NCSA) (2017). Ncsa mosaic™. http://www.ncsa.illinois.edu/enabling/mosaic, accessed February 4, 2017.

Nelson, T. H. (1965). Complex information processing: A file structure for the complex, the changing and the indeterminate. In *Proceedings of the 1965 20th National Conference*, pages 84–100. ACM.

Nguyen, T. T., Hui, P.-M., Harper, M. F., Terveen, L., and Konstan, J. A. (2014). Exploring the filter bubble: the effect of using recommender systems on content diversity. In *Proceedings of the 23rd International Conference on World Wide Web*, pages 677–686.

Nielsen, J. (1988). Hypertext'87 trip report. https://www.nngroup.com/articles/trip-report-hypertext-87.

Nielsen, J. (2006). F-shaped pattern for reading web content. www.nngroup.com/articles/f-shaped-pattern-reading-web-content.

O'Day, V. L. and Jeffries, R. (1993). Orienteering in an information landscape: How information seekers get from here to there. In *Proceedings of the INTERACT'93 and CHI'93 Conference on Human Factors in Computing Systems*, pages 438–445. ACM.

Pan, B., Hembrooke, H., Joachims, T., Lorigo, L., Gay, G., and Granka, L. (2007). In google we trust: Users' decisions on rank, position, and relevance. *Journal of Computer-Mediated Communication*, 12(3):801–823.

Paranjape, A., West, R., Zia, L., and Leskovec, J. (2016). Improving website hyperlink structure using server logs. In *Proceedings of the 9th International Conference on Web Search and Data Mining*, pages 615–624.

Pirolli, P. (2007). *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press.

Pirolli, P. and Card, S. (1999). Information foraging. *Psychological Review*, 106(4):643.

Pirolli, P. and Fu, W.-T. (2003). *9th International Conference on User Modeling*, chapter SNIF-ACT: A Model of Information Foraging on the World Wide Web. Springer.

Resnick, P., Iacovou, N., Suchak, M., Bergstrom, P., and Riedl, J. (1994). Grouplens: An open architecture for collaborative filtering of netnews. In *Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work*, pages 175–186. ACM.

Resnick, P. and Varian, H. R. (1997). Recommender systems. *Communications of the ACM*, 40(3):56–58.

Rosvall, M. and Bergstrom, C. T. (2008). Maps of random walks on complex networks reveal community structure. *Proceedings of the National Academy of Sciences*, 105(4):1118–1123.

Salganik, M. J., Dodds, P. S., and Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856.

Salmerón, L., Kammerer, Y., and García-Carrión, P. (2013). Searching the web for conflicting topics: Page and user factors. *Computers in Human Behavior*, 29(6):2161–2171.

Scaria, A. T., Philip, R. M., West, R., and Leskovec, J. (2014). The last click: Why users give up information network navigation. In *Proceedings of the 7th ACM International Conference on Web Search and Data Mining*, pages 213–222. ACM.

Schnettler, S. (2009). A small world on feet of clay? a comparison of empirical small-world studies against best-practice criteria. *Social Networks*, 31(3):179–189.

Sen, R. and Hansen, M. H. (2003). Predicting web users' next access based on log data. *Journal of Computational and Graphical Statistics*, 12(1):143–155.

Serrano, M. A., Krioukov, D., and Boguñá, M. (2008). Self-similarity of complex networks and hidden metric spaces. *Physical Review Letters*, 100(7):078701:1–078701:4.

Seyerlehner, K., Flexer, A., and Widmer, G. (2009a). On the limitations of browsing top-N recommender systems. In *Proceedings of the 3rd ACM Conference on Recommender systems*, pages 321–324.

Seyerlehner, K., Knees, P., Schnitzer, D., and Widmer, G. (2009b). Browsing music recommendation networks. In *Proceedings of the 10th International Society for Music Information Retrieval Conference*, pages 129–134.

Shani, G. and Gunawardana, A. (2011). *Recommender Systems Handbook*, chapter Evaluating Recommendation Systems, pages 257–297. Springer.

Siegler, M. (2010). Eric schmidt: Every 2 days we create as much information as we did up to 2003. www.techcrunch.com/2010/08/04/schmidt-data, accessed October 26, 2016.

Singer, P., Helic, D., Taraghi, B., and Strohmaier, M. (2014). Detecting memory and structure in human navigation patterns using markov chain models of varying order. *PloS One*, 9(7):e102070:1–e102070:7.

Singer, P., Niebler, T., Strohmaier, M., and Hotho, A. (2013). Computing semantic relatedness from human navigational paths: A case study on wikipedia. *International Journal on Semantic Web and Information Systems (IJSWIS)*, 9(4):41–70.

Su, J., Sharma, A., and Goel, S. (2016). The effect of recommendations on network structure. In *Proceedings of the 25th International Conference on World Wide Web*.

Teevan, J., Alvarado, C., Ackerman, M. S., and Karger, D. R. (2004). The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 415–422.

Toms, E. G. (2000). Serendipitous information retrieval. In *Proceedings of the DELOS Workshop: Information Seeking, Searching and Querying in Digital Libraries*.

Trattner, C., Singer, P., Helic, D., and Strohmaier, M. (2012). Exploring the differences and similarities between hierarchical decentralized search and human navigation in information networks. In *Proceedings of the 12th International Conference on Knowledge Management and Knowledge Technologies*, pages 14:1–14:8.

Travers, J. and Milgram, S. (1969). An experimental study of the small world problem. *Sociometry*, 32(1):425–443.

Watts, D. J., Dodds, P. S., and Newman, M. (2002). Identity and search in social networks. *Science*, 296:1302–1305.

Watts, D. J. and Strogatz, S. H. (1998). Collective dynamics of small-world networks. *Nature*, 393(6684):440–442.

Weinreich, H., Obendorf, H., Herder, E., and Mayer, M. (2006). Off the beaten tracks: Exploring three aspects of web navigation. In *Proceedings of the 15th International Conference on World Wide Web*, pages 133–142. ACM.

West, R. and Leskovec, J. (2012a). Automatic versus human navigation in information networks. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*.

West, R. and Leskovec, J. (2012b). Human wayfinding in information networks. In *Proceedings of the 21st International Conference on World Wide Web*, pages 619–628.

West, R., Paranjape, A., and Leskovec, J. (2015). Mining missing hyperlinks from human navigation traces: A case study of wikipedia. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1242–1252.

West, R., Pineau, J., and Precup, D. (2009a). Wikispeedia: An online game for inferring semantic distances between concepts. In *Proceedings of the 21st International Joint Conference on Artificial Intelligence*, pages 1598–1603.

West, R., Precup, D., and Pineau, J. (2009b). Completing wikipedia's hyperlink structure through dimensionality reduction. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management*, pages 1097–1106.

White, R. W. and Huang, J. (2010). Assessing the scenic route: Measuring the value of search trails in web logs. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Inform ation Retrieval*, pages 587–594. ACM.

Wikipedia (2017a). Wikipedia: Manual of style. http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style.

Wikipedia (2017b). Wikipedia: Manual of style/linking. http://en.wikipedia.org/wiki/Wikipedia:Manual_of_Style/Linking.

Wikipedia (2017c). Wikipedia: Policies and guidelines. https://en.wikipedia.org/wiki/Wikipedia:Policies_and_guidelines.

Wikipedia (2017d). Wikipedia statistics. https://tools.wmflabs.org/wmcharts/wmchart0002.php.

Wulczyn, E. and Taraborelli, D. (2015). Wikipedia clickstream. http://dx.doi.org/10.6084/m9.figshare.1305770, accessed Februar 17, 2016.

Ziegler, C.-N., McNee, S. M., Konstan, J. A., and Lausen, G. (2005). Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web*, pages 22–32.