

Exploring Differences and Similarities between Hierarchical Decentralized Search and Human Navigation in Information Networks

Christoph Trattner
Knowledge Management
Institute
Graz University of Technology
Graz, Austria
ctrattner@icm.edu

Philipp Singer
Knowledge Management
Institute
Graz University of Technology
Graz, Austria
philipp.singer@tugraz.at

Denis Helic
Knowledge Management
Institute
Graz University of Technology
Graz, Austria
dhelic@tugraz.at

Markus Strohmaier
Knowledge Management
Institute
Graz University of Technology
Graz, Austria
markus.strohmaier@tugraz.at

ABSTRACT

Decentralized search in networks is an activity that is often performed in online tasks. It refers to situations where a user has no global knowledge of a network's topology, but only local knowledge. On Wikipedia for instance, humans typically have local knowledge of the links emanating from a given Wikipedia article, but no global knowledge of the entire Wikipedia graph. This makes the task of navigation to a target Wikipedia article from a given starting article an interesting problem for both humans and algorithms. As we know from previous studies, people can have very efficient decentralized search procedures that find shortest paths in many cases, using intuitions about a given network. These intuitions can be modeled as hierarchical background knowledge that people access to approximate a networks' topology. In this paper, we explore the differences and similarities between decentralized search that utilizes hierarchical background knowledge and actual human navigation in information networks. For that purpose we perform a large scale study on the Wikipedia information network with over 500,000 users and 1,500,000 click trails. As our results reveal, simulations based on decentralized search with hierarchies created directly from the link structure of the information network are more similar to human navigation behavior than simulations based on hierarchies that are created from network-external knowledge.

Categories and Subject Descriptors

H.5.4 [Information Interfaces and Presentation]: Hypertext/ Hypermedia–Navigation

General Terms

Human Factors

Keywords

Navigation, Decentralized Search, Simulations, Wikipedia

1. INTRODUCTION

In 1967, Milgram conducted his now famous small-world experiment [17], in which randomly selected people from Nebraska had to pass on a letter to a specific target person in Boston. The specific experimental setup required the participants to pass the letter in a decentralized manner, i.e. they were only allowed to pass the letter through their local social networks. Despite this restriction, the average chain length of those letters that reached the target person was only six - thus, giving rise to the hypothesis that the USA constituted a small-world.

One of the most interesting research questions raised by this experiment was to understand and characterize the algorithm that people use to efficiently find other distant people in social networks. To that end, among others, Kleinberg introduced the theory of *decentralized search* and provided a theoretical explanation of this human ability [13, 14, 15]. In a number of studies Kleinberg showed that social networks possess certain latent structural properties that humans are aware of and are able to utilize in their search for other people. This allows them to find short paths between two arbitrary network nodes efficiently even with only local knowledge of the network. Consequently, Kleinberg also examined the structure of such latent structural properties that he called background knowledge, and discovered that

social networks can be efficiently searched, i.e. in $\log(N)$, where N are the number of nodes in the network, if the nodes of the network can be organized into a hierarchy. This theoretical model is also known as Kleinberg’s hierarchical network model [15].

Based on these ideas, Lada Adamic [1] implemented a decentralized search algorithm that utilizes hierarchical background knowledge of a network and applied that algorithm in a number of experiments. Adamic showed that the algorithm performs well in simulating human-like search behavior in social network. Furthermore, she demonstrated that the performance of the simulator depends on the quality of the background knowledge of the network.

In our previous work [11, 9, 10], we applied a variant of Adamic’s algorithm for simulation of navigation in information networks. Navigation in information networks is a kind of decentralized search, as users at each particular step of their navigation are only aware of links emanating from the current document. Thus, this situation is intuitively very similar to decentralized search in social networks. For example, in [11] we developed a hierarchical decentralized search algorithm based on the ideas of Adamic that allows decentralized search in social tagging systems. By constructing tag hierarchies from the bipartite tag-resource network structures of a number of tagging systems and by using this background knowledge as input for our hierarchical decentralized search algorithm, we could show that tag hierarchies perform extremely well in searching social tagging systems. In subsequent work [21], we also demonstrated that the most semantically sound tag hierarchies are also those that perform well on navigational tasks. However, our previous experiments were based on intuitions how humans navigate and we have not yet compared our simulations (based on decentralized search) with real human navigation paths.

Human navigation has been investigated in some recent studies. In [23], West and Leskovec analyzed click trails from over 9,000 users navigating a subset of Wikipedia¹. An interesting finding of the study was that humans require only two more clicks on average (median 1 click) to navigate to articles in Wikipedia than the shortest possible paths. In subsequent work, [24] then developed machine learning agents navigate in Wikipedia significantly faster than human navigators.

The purpose of this paper however is not to develop the most efficient strategy to search information networks in a decentralized manner, instead we are interest in exploring the differences and similarities between hierarchical decentralized search (as introduced in previous work) and human search behavior in information networks. For that purpose, we examine more than 150,000 click trails of users navigating the complete English Wikipedia.

As our results reveal, simulations based on hierarchies created directly from the link structure of the information network are more similar to human navigational behavior than hierarchies created from network external knowledge.

Algorithm 1 Hierarchical Decentralized Search

```

1: INPUT: Network  $N$ , Hierarchy  $H$ , start-node  $s$ , target-
   node  $t$ 
2:  $c \leftarrow s$ 
3: while  $c \neq t$  do
4:    $o \leftarrow -1$ 
5:    $dist_{min} \leftarrow \infty$ 
6:   /*  $\Gamma(c)$  is a set of all neighbors of  $c$  */
7:   for each  $n \in \Gamma(c)$  do
8:      $dist \leftarrow h(n, H) + h(w, H) - 2h(n, w, H) - 1$ 
9:     if  $dist < dist_{min}$  then
10:       $dist_{min} \leftarrow dist$ 
11:       $o \leftarrow n$ 
12:     end if
13:   end for
14:    $c \leftarrow o$ 
15: end while

```

The remainder of the paper is structured as follows: In Section 2, we discuss related work. In Section 3 we shortly present our simulation model for user navigation in information network. In Section 4, we outline our experimental setup and in Section 5 we present our experimental results. In Section 6 we discuss the results and conclude our paper.

2. RELATED WORK

Related work in this area can be broadly divided into the following three areas: Web click-trail analysis, navigation in complex networks and hierarchy creation from networks.

2.1 Click-Trail Analysis

Click-trail analysis has been mainly performed to improve the Web search results of users. For instance, in [5, 20] the authors assessed the possibility to rank search results more efficiently by taking the users click-trails into account. In [2] a large scale study was conducted to investigate how often users revisit the same Web page. To the best of our knowledge, there is only one study that tries to understand how people navigate in information networks by analyzing a large click-trail log from the online game Wikispeedia. In [23] West and Leskovec performed a study of users navigating Wikipedia articles. In their work they found out that user navigation behavior is close to the short paths of the network. In subsequent work [24], the authors analyzed a number of decentralized search algorithms and benchmarked them against their human click corpus. The most interesting result was that even simple search strategies such as utilizing node degrees outperforms human information seeking.

2.2 Navigation in Networks

Research on navigation in complex networks was initiated by the famous small-world experiment conducted by Milgram [17]. Apart from the work on the algorithmic perspective of search in social networks that we mention in 1, a number of studies recently dealt with navigability of other types of complex networks. In [19], the authors extend the notion of Kleinberg’s background knowledge to the notion of *hidden metric spaces*. In such hidden metric spaces nodes are identified by their co-ordinates – distance between nodes is their geometric distance in a particular metric space. Navigation strategies in complex networks are then based on

¹<http://schools-wikipedia.org/>

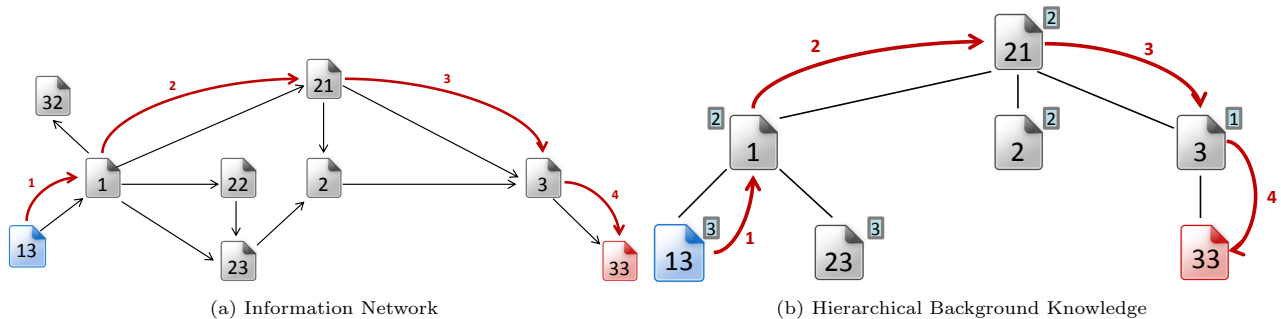


Figure 1: An example of decentralized search in an information network (a) using hierarchical background knowledge of this network (b). The information network links information for instance document pages (i.e., Wikipedia pages) with each other. The search begins at the blue node 13. The destination node is the red node 33. At each step, the search algorithm selects one of the current node’s adjacent nodes, which is the closest to the target node in the hierarchy. The numbers in boxes in (b) provide the distance between the current node and the destination node 33. At step one, node 13 has a single adjacent node 1, so search continues to 1. At step two, 1’s adjacent nodes include 21, 22, 23 and 32. The algorithm consults the hierarchy finding out that node 21 is the closest to the destination node. At step three, the algorithm has an option to move to nodes 2 or 3. The simulation selects node 3, since again, it has the smallest distance to the destination node. Finally, at step 4, the target node is successfully reached.

the distances between nodes – an agent always navigates to the node with the smallest distance to a particular destination node. An interesting research question is the structure of such hidden metric spaces that underlie observable networks. In [6], the authors introduce a model with the circle as a hidden metric space and show its effects on routing in the global airport network. In [16] the authors discuss hyperbolic geometry as a hidden metric space (which can be approximated by a node hierarchy) whereas in [7] the authors apply hyperbolic geometry as a model of the hidden metric space of the Internet and design a novel greedy Internet routing algorithm. In this work we will focus on Kleinberg’s hierarchical network model.

2.3 Extracting Hierarchies from Networks

Hierarchies that are extracted from networks play an important role in many of these network navigation models. Apart from the tag hierarchy induction algorithms based on bipartite networks such as e.g. [12, 3, 10], researchers also proposed hierarchy extraction algorithms for general networks. In [18] the authors discuss an algorithm for hierarchy construction in Wikipedia networks based on metrics for estimating hierarchy level of single nodes. Also, Clauset et al. [8] present a hierarchy induction algorithm based on prediction of hierarchical links. To extract hierarchical background knowledge as hidden metric space for our decentralized search algorithm, we rely on the hierarchy induction algorithms of [12, 18] in this paper.

3. HIERARCHICAL DECENTRALIZED SEARCH ALGORITHM

In order to simulate human information seeking behavior in information networks, we implemented a hierarchical search algorithm (see Algorithm 1) based on the ideas of Lada Adamic in the past. The algorithm takes as input a given network, start and target nodes and a hierarchical represen-

tation of the given network. To navigate from one node in the network to another, all adjacent nodes of the current node are examined and the distance to the target node is calculated over the input hierarchy. The simulation then selects as the next step the node with the minimal distance to the target (see Figure 1).

The pseudo code of our algorithm does not include the cancellation strategy – we cancel navigation if the simulation visits the same node again. As shown by [23] only a small fraction of users choose the same link again for navigating from one resource to another in an information network. For that purpose, we ignore back tracking. We also cancel search in the case we can not find a particular node of the network in hierarchy. When the distance function returns the same minimum distance for more than one adjacent node, we try to avoid the nodes that we already visited. To simplify the pseudo code in Algorithm 1, we omit this avoiding strategy from the code.

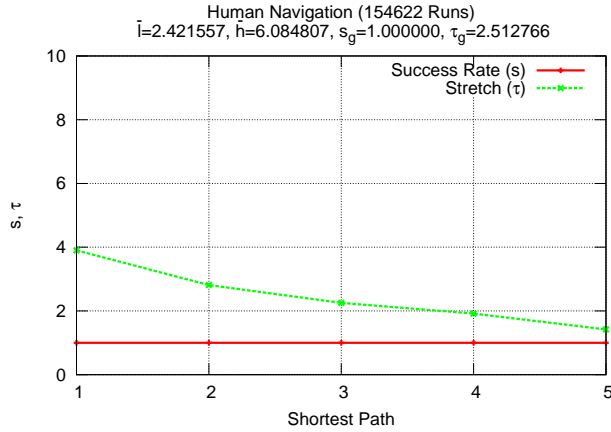
4. DATASETS

The following section provides a brief overview of the datasets used in our experiments.

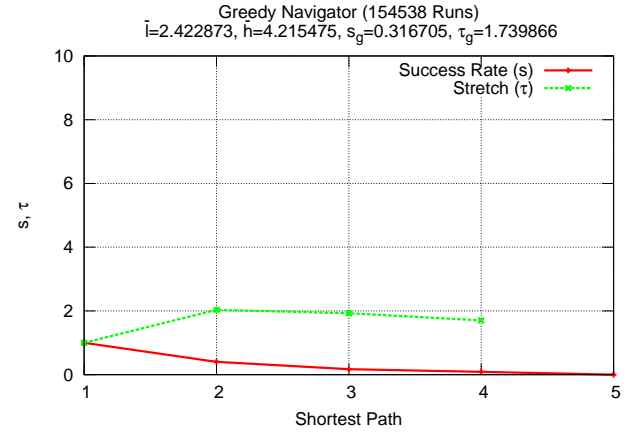
4.1 Wikipedia Click Dataset

In order to compare the behavior of the search algorithm with human navigation, we analyze a click dataset from the complete English Wikipedia. The dataset comes from the online platform the wikigame². There are two reasons for our decision on this dataset. First, there are no freely available datasets that include complete click paths from a specific start node to a specific target node. Typically, one has to apply heuristics to extract users, their sessions, and their click trails. In Wikigame, we have a complete sequence of clicks of different users participating in a game that requires

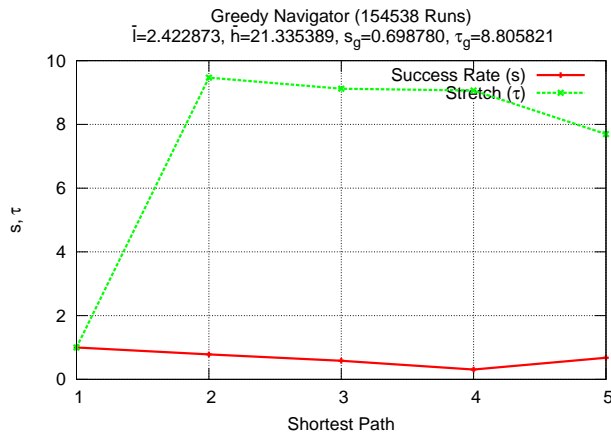
²<http://thewikigame.com/>



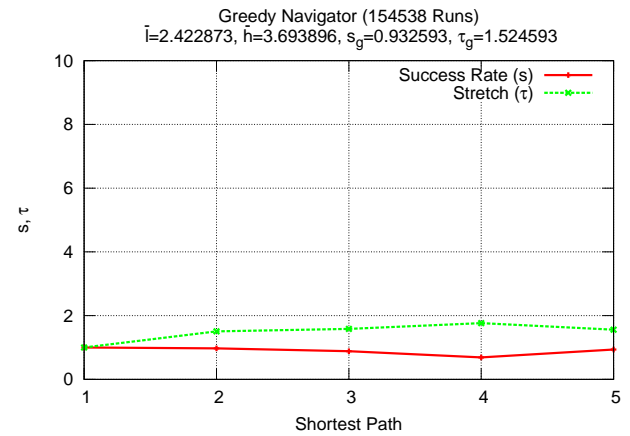
(a) Humans navigators



(b) Searcher (Simulation) w. Wiki. Cat. hierarchy as background knowledge



(c) Searcher (Simulation) w. Del. Tags hierarchy as background knowledge



(d) Searcher (Simulation) w. Wiki. Net. hierarchy as background knowledge

Figure 2: Results of Human navigators vs. simulator (=greedy navigator): Average shortest path \bar{l} , average hop count h , success rate s and stretch τ – global average values (s_g and τ_g) and distribution over shortest paths. As shown, the simulator with the hierarchy based on the Wikipedia link structure simulates humans best (highest success rate $s_g = 0.93$ and stretch that is close to the human navigators). The simulator with the Wikipedia category label hierarchy performs worst, success rate is only $s_g = 0.31$.

from the users to navigate from e.g. “*Wolfgang Amadeus Mozart*” to e.g. “*Arnold Schwarzenegger*”. In turn, other datasets do not include explicit (*start, target*) information. The second reason is basically the large scale of the dataset, with records of more than 500,000 users and 1,000,000 click trails. However, for the purposes of this study we analyze only a subset of this large scale dataset.

4.2 Wikipedia Network Dataset

Additionally to the dataset record of Wikigame click paths, our work is based on an information network dataset (= directed link-network dataset) of the English-Wikipedia from February 2012. We use this kind of dataset as basis for our simulations. All in all, the dataset includes around 10,000,000 articles and around 250,000,000 links.

4.3 Wikipedia Category Label Datasets

Since our decentralized search simulations are based on hierarchical background knowledge of the information network, the question arises how can we extract this kind of knowledge from our Wikipedia dataset. A simple idea is to use Wikipedia category labels for constructing a hierarchy representation of the network. The other idea is to use external meta-data information, such as social tags which were shown useful to classify information such as Web pages [26, 25]. In our case we used a dataset of Wikipedia category labels as well as a dataset of social tags from Delicious which only consists of annotated Wikipedia articles. Overall, the Wikipedia category label dataset includes around 2,300,000 category labels, 4,500,000 articles and 30,000,000 category label assignments. The Delicious tag dataset includes around 440,000 tags, 580,000 articles and 3,400,000 tag assignments.

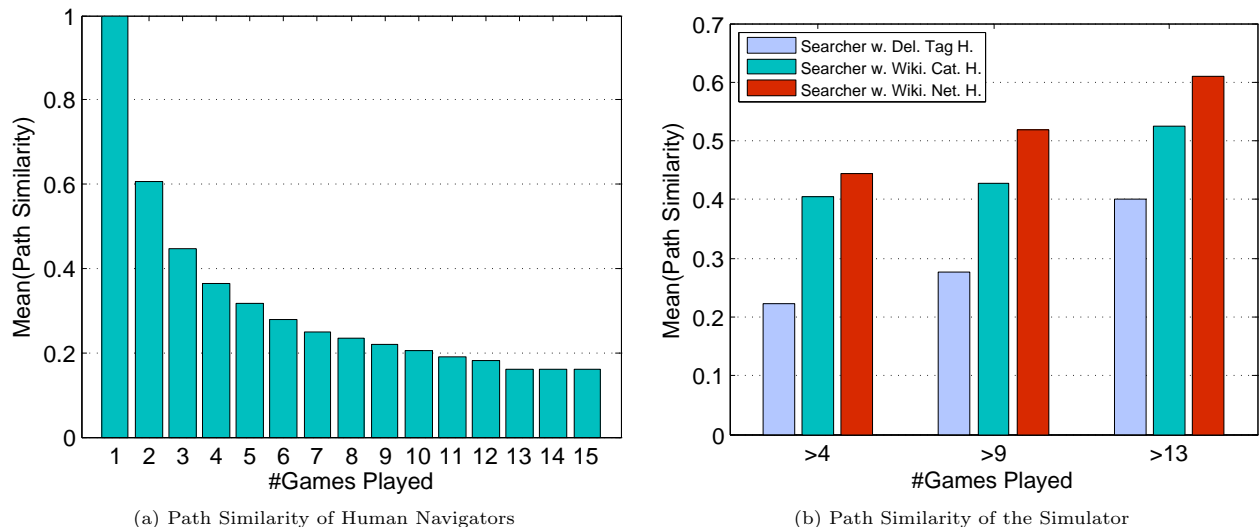


Figure 3: Path Similarity of Human Navigators (a) and Path Similarity of the simulator grounded on human click trails (b) (only successful paths). As shown, in Figure (a) path similarity drops significantly the more people play the same game. For games that are played more the 13 times, the path similarity drops down to 18% indicating that humans agree not very much in taking the same paths to reach their target nodes. In Figure (b), path similarity of the simulator with different background knowledge is shown. As presented in Figure (b), simulation paths based on the Wikipedia network hierarchy fit human navigator paths most.

5. METHODOLOGY

In the following section we present the methodology for creating our decentralized search simulations.

5.1 Click-Trail Selection

For the purpose of our study, we only considered games (= click trails) that were successfully accomplished. We also selected only those click trails where the start and target node were present in all of the hierarchies that we produced. At the end, we analyzed around 160,000 click trails.

5.2 Hierarchy Creation

Since our simulations depend on hierarchical knowledge of the information network, we created hierarchies that we can use for decentralized search in the Wikipedia information network. As mentioned before, there are several ways to extract hierarchies from information networks. In related work, we have shown that it is possible to create good hierarchies from tagging data by inducing graph based clustering algorithms that are based on the tag network’s tag-cooccurrence graph. Specifically, in [21] we revealed that the most semantically correct tag hierarchies also produce good decentralized search results, in terms of for example success rate and shortest paths.

5.2.1 Extracting Hierarchies from External Knowledge

In this work, we use two different types of hierarchy induction algorithms. The first approach we use is based on the ideas of [12]. In their work the authors introduce a generic algorithm for producing hierarchies from bipartite networks such as tag-to-resource networks. The algorithm

can be applied to arbitrary bipartite structures. The algorithm takes as input two parameters. The first is a ranked list of tags sorted by their centrality in the projected tag-to-tag network. This centrality ranking acts as a proxy to the generality ranking of tags. Benz et al. [4] showed that the centrality provides a viable approximation for term abstractness in tags. The second input parameter is the tag similarity matrix. The algorithm starts then by a single node hierarchy with the most general tag as the root node and then iterates through the centrality list. At each iteration step, the algorithm adds the current tag to the hierarchy as a child to its most similar tag. The centrality and similarity measure are exchangeable – in [12] the authors use closeness centrality and cosine similarity, whereas in [3] the authors select degree centrality and co-occurrence similarity measure. As both combinations perform similarly in supporting navigation [11], we select in this work the latter combination because of better computational properties. Furthermore, we adopted the algorithm of Benz et al. to produce a resource taxonomy instead of a tag taxonomy. We achieve that by simply switching our computations from the projected tag-to-tag network to the projected resource-to-resource network. This algorithm is then applied to generate a Wikipedia resource hierarchy on the basis of the Delicious tag dataset as well on the basis of the Wikipedia category label dataset.

5.2.2 Extracting Hierarchies from the Network

The second type of hierarchy we produce for our simulation is based on the ideas of [18]. The algorithm is based on the idea that each network possesses an inherent hierarchical structure that leads to the emergence of observable structural properties such as power-law degree distributions and high node clustering (cf. [8]). The algorithm aims then

to recognize and extract that hierarchical structure. Thus, the algorithm iterates through all links in the network and decides – using a simple criteria – if that link is of a hierarchical type, in which case it remains in the network, or if that link is of some other kind (e.g. a synonym link), in which case the link is removed from the network. To that end, the algorithm assigns to each node a so-called hierarchical score, which is a measure of the generality of a node. For each link the ratio between hierarchical scores of two incident to that link is calculated. The simple idea is that if that ratio is close to 1 then those two nodes are very close in their generality and they are situated in the same hierarchy level – thus, the link between those two nodes is not a hierarchical one and is therefore removed from the network. Similarly, if the hierarchical ratio for a link is close to 0 then those two nodes are very far away from each other in the hierarchy and the link is removed (e.g. an article on a very small town in the USA, say Paris, Texas, links to the article on the United States). Technically, the authors define two thresholds – high and low threshold – to decide on the links removal. Thus, a link is removed if the hierarchical ratio is greater than the high threshold or smaller than the low threshold. Another technical issues is the decision on how to calculate the hierarchical score. In their paper, the authors compare five different hierarchical scores ranging from global scores such as betweenness centrality to local scores such as ratio of in-degree and out-degree of a node. In our experiments we use a local score, defined as:

$$hs(n) = \frac{d_{in}(n)}{d_{out}(n)} \sqrt{d_{in}(n)}. \quad (1)$$

The term $\sqrt{d_{in}(n)}$ ensures that a node having e.g. 200 in-degree and 100 out-degree is rendered more general than a node having e.g. 2 in-degree and 1 out-degree. As thresholds we choose 0.6 and 0.2 for high and low thresholds respectively (cf. [18]).

5.3 Simulations

We simulate search on the Wikipedia link-network. To make results comparable, we run our simulations on the Wikipedia link-network using the same start, target node pairs and trails as present in the human click-trail dataset. Additionally, we use three different hierarchies as background knowledge for our simulations.

5.4 Comparison

To compare our simulations with human navigation, we define a number of measures. In the following list, we give a short overview of these measures and how they are calculated:

- **Success Rate:** As discussed before, we use in our analysis only success games (=click trails), i.e. the success rate of human navigators is 100%. Since we perform our simulations on the same search trails, we can identify with this measure to which extent the simulation differs from reaching the destination node in each step or on average. In our analysis we calculate the mean local s and global (=overall) success rate s_g .

- **Number of Hops:** Another interesting measure is the number of hops needed to reach the target node. We capture this on a global basis \bar{h} .
- **Stretch:** Stretch captures the ratio of the number of steps and the global shortest path. As shown in [23] humans are typically very efficient at finding shortest paths. On average, they find information in Wikipedia in not more than two more steps than the shortest possible path. Thus, with this measure we identify how good our simulation is in finding shortest paths in each step τ and on average overall τ_g compared to human navigators.
- **Path Similarity:** We calculate path similarity to determine the extent to which successful paths of our simulations differ from real user navigational trails. Since the user’s click paths in general show a high diversity by terms of similarity (see Figure 3(a)), we calculate path similarity as

$$\frac{CTL(humans)_{(a,b)} \cap CTL(sim)_{(a,b)}}{CTL(sim)_{(a,b)}} \quad (2)$$

where $CTL(humans)_{(a,b)}$ is the set of human click trails for the search pair (a, b) and where $CTL(sim)_{(a,b)}$ is the set of click trails for the same pair.

- **Degree:** Finally, we also investigate the median in- and out-degree values of the nodes visited by the simulator and the human navigator (we use the median in this case since the values are not normally distributed).

6. RESULTS

In Figure 2 we illustrate the first results of our comparison. As shown, the simulator utilizing the hierarchy based on the Wikipedia link structure generates the best results. We can observe the a highest success rate $s_g = 0.93$ of all other simulators. The worst performance $s_g = 0.31$ is achieved by the simulator with hierarchical background knowledge generated from the Wikipedia category labels. Interestingly, the success rate of the simulations based on the Delicious tag hierarchy is rather good, taking into account that the Delicious tag dataset covers five times less articles in Wikipedia. This leads to the situation that the Delicious hierarchy contains also five times fewer nodes than the hierarchy extracted from the Wikipedia category labels, which means that the simulation is more likely to fail the search, since a possible selected node of the simulation is not present in the hierarchy. However, as also shown in Figure 2 the average hop length is quite high $\bar{h} = 21.34$. This demonstrates that it is possible to navigate successfully through an information network even though the hierarchy is not complete. On the other hand we can see that hierarchies directly extracted from the information network is better suited as hierarchical background knowledge than hierarchies extracted from external knowledge.

In addition to the previous results, we illustrate in Figure 3(a) path similarity of human navigators that play the same game on the same start and target nodes. As shown, when more people play the same game, the average path similarity drops significantly. This indicates that humans have only little agreement on how they route through an

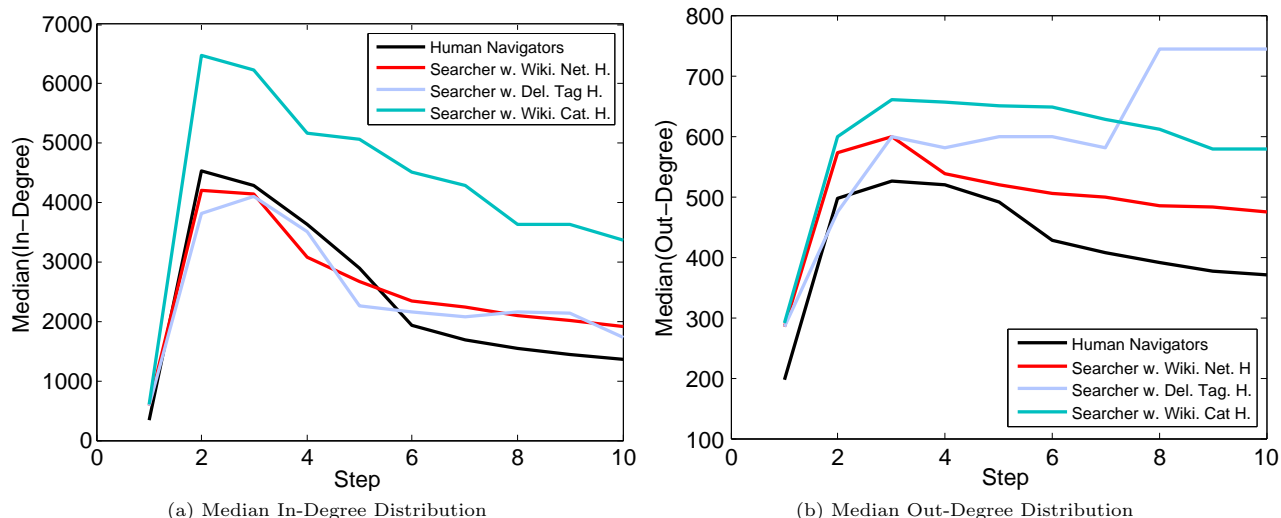


Figure 4: Median in and out-degree distributions for human navigators and simulations. As shown, again the hierarchical decentralized simulation utilizing the Wikipedia network hierarchy as background knowledge comes human search behavior closest. However, as shown in Figure (b), the simulator in general favors higher out-degree nodes than the human navigators.

information network. This could be explained by their familiarity with the target item or their search expertise level [22]. In Figure 3(b), we compare the similarity of the successful paths conducted by human navigators and the ones resulting from our simulator on different hierarchies. As the results reveal, again simulations based on the Wikipedia network hierarchy are most similar to human navigator paths.

Finally, Figure 4 shows the median in- and out-degree distributions for human navigators and simulations. As already observed in related work by West and Leskovec [23], humans follow certain patterns in their information seeking behavior. In particular, high degree nodes are typically used in the first steps of the search, while similar nodes are used in the end. Since degree is highly correlated to similarity [23], we focus in our analysis only on degree. As shown in Figure 4, human navigators as well simulators choose high degree nodes in the first step of their search. Again we can see that the hierarchical decentralized simulations utilizing the Wikipedia network hierarchy as background knowledge is most similar to human search behavior. The simulations based on the Wikipedia category label hierarchy differ the most – regarding the median in-degree – from the other distributions. This behavior might be an explanation for the bad performance of the simulation as shown in Figure 2.

7. CONCLUSIONS

In this work we explored the differences and similarities between hierarchical decentralized search – as introduced in previous work [11, 21] – and human navigational behavior in information networks. Based on a click dataset of over 150,000 click trails from the online platform the Wiki Game, we performed a number of experiments to gain insights into human search behavior and simulations with hierarchical decentralized search in information networks. Generating background knowledge from various sources, we could show

that hierarchies generated directly from the link-network structure perform best by terms of navigational efficiency. In contrast, hierarchies generated from external knowledge sources such as tags from Delicious or Wikipedia category labels, perform significantly worse.

8. ACKNOWLEDGMENTS

The authors wish to acknowledge Alex Clemesha for providing the Wiki Game dataset. This work is in part supported by funding from - BMVIT - the Federal Ministry for Transport, Innovation and Technology (grant no. 829590), by the FWF Austrian Science Fund Grant I677 and the Knowledge Center Graz.

9. REFERENCES

- [1] L. Adamic and E. Adar. How to search a social network. *Social Networks*, 27(3):187 – 203, 2005.
- [2] E. Adar, J. Teevan, and S. T. Dumais. Large scale analysis of web revisitation patterns. In *Proceedings of the twenty-sixth annual SIGCHI conference on Human factors in computing systems*, CHI '08, pages 1197–1206, New York, NY, USA, 2008. ACM.
- [3] D. Benz, A. Hotho, and G. Stumme. Semantics made by you and me: Self-emerging ontologies can capture the diversity of shared knowledge. In *Proc. of the 2nd Web Science Conference (WebSci10)*, Raleigh, NC, USA, 2010. Web Science Trust.
- [4] D. Benz, C. Körner, A. Hotho, G. Stumme, and M. Strohmaier. One tag to bind them all: Measuring term abstractness in social metadata. In *Proceedings of the 8th Extended Semantic Web Conference (ESWC 2011)*, Heraklion, Crete, May 2011.
- [5] M. Bilenko and R. W. White. Mining the search trails of surfing crowds: identifying relevant websites from user activity. In *Proceedings of the 17th international*

- conference on World Wide Web, WWW '08, pages 51–60, New York, NY, USA, 2008. ACM.
- [6] M. Boguñá, D. Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5:74–80, Jan. 2009.
- [7] M. Boguñá, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1:62, Sept. 2010.
- [8] A. Clauset, C. Moore, and M. E. J. Newman. Hierarchical structure and the prediction of missing links in networks. *Nature*, 453(7191):98–101, 2008.
- [9] D. Helic, C. Körner, M. Granitzer, M. Strohmaier, and C. Trattner. Navigational efficiency of broad vs. narrow folksonomies. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, HT '12, New York, NY, USA, 2012. ACM.
- [10] D. Helic and M. Strohmaier. Building directories for social tagging systems. In *Proceedings of the 20th ACM international conference on Information and knowledge management*, CIKM '11, pages 525–534, New York, NY, USA, 2011. ACM.
- [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*, WWW '11, pages 417–426, New York, NY, USA, 2011. ACM.
- [12] P. Heymann and H. Garcia-Molina. Collaborative creation of communal hierarchical taxonomies in social tagging systems. Technical Report 2006-10, Stanford InfoLab, April 2006.
- [13] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, STOC '00, pages 163–170, New York, NY, USA, 2000. ACM.
- [14] J. M. Kleinberg. Navigation in a small world. *Nature*, 406(6798):845, August 2000.
- [15] J. M. Kleinberg. Small-world phenomena and the dynamics of information. In *Advances in Neural Information Processing Systems (NIPS) 14*, page 2001, Cambridge, MA, USA, 2001. MIT Press.
- [16] D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat, and M. Boguñá. Hyperbolic geometry of complex networks. *Phys. Rev. E*, 82(3):036106, Sep 2010.
- [17] S. Milgram. The small world problem. *Psychology Today*, 1:60–67, 1967.
- [18] L. Muchnik, R. Itzhack, S. Solomon, and Y. Louzoun. Self-emergence of knowledge trees: Extraction of the wikipedia hierarchies. *Phys. Rev. E*, 76:016106, Jul 2007.
- [19] M. A. Serrano, D. Krioukov, and M. Boguñá. Self-similarity of complex networks and hidden metric spaces. *Phys. Rev. Lett.*, 100(7):078701, Feb 2008.
- [20] A. Singla, R. White, and J. Huang. Studying trailfinding algorithms for enhanced web search. In *Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval*, SIGIR '10, pages 443–450, New York, NY, USA, 2010. ACM.
- [21] M. Strohmaier, D. Helic, D. Benz, C. Körner, and R. Kern. Evaluation of folksonomy induction algorithms. *ACM Trans. Intell. Syst. Technol.*, 2012.
- [22] C. Trattner, Y.-L. Lin, D. Parra, W. Real, and P. Brusilovsky. Evaluating tag-based information access in image collections. In *Proceedings of the 23rd ACM Conference on Hypertext and Social Media*, HT '12, New York, NY, USA, 2012. ACM.
- [23] R. West and J. Leskovec. Human wayfinding in information networks. In *Proceedings of the 21th international conference companion on World wide web*, WWW'12, New York, NY, USA, 2012. ACM.
- [24] R. West and J. Leskovec. Human wayfinding in information networks. In *Proceedings of the 6th International AAAI Conference on Weblogs and Social Media*, ICWSM'12, Dublin, Ireland, 2012. AAAI.
- [25] A. Zubiaga, C. Körner, and M. Strohmaier. Tags vs shelves: from social tagging to social classification. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*, HT '11, pages 93–102, New York, NY, USA, 2011. ACM.
- [26] A. Zubiaga, R. Martínez, and V. Fresno. Getting the most out of social annotations for web page classification. In *Proceedings of the 9th ACM symposium on Document engineering*, DocEng '09, pages 74–83, New York, NY, USA, 2009. ACM.