Understanding, Leveraging and Improving Human Navigation on the Web

Philipp Singer
Supervised by: Prof. Markus Strohmaier (University of Koblenz and GESIS Cologne)
Graz University of Technology
Inffeldgasse 13
Graz, Austria
philipp.singer@tugraz.at

ABSTRACT

Navigating websites represents a fundamental activity of users on the Web. Modeling this activity, i.e., understanding how predictable human navigation is and whether regularities can be detected has been of interest to researchers for nearly two decades. This is crucial for improving the Web experience of users by e.g., enhancing interfaces or information network structures.

This thesis envisions to shedding light on human navigational patterns by trying to understand, leverage and improve human navigation on the Web. One main goal of this thesis is the construction of a versatile framework for modeling human navigational data with the use of Markov chains and for detecting the appropriate Markov chain order by using several advanced inference methods. It allows us to investigate memory and structure in human navigation patterns. Furthermore, we are interested in detecting whether pragmatic human navigational data can be leveraged by e.g., being useful for the task of calculating semantic relatedness between concepts. Finally, we want to find ways of enhancing human navigation models. Concretely, we plan on incorporating prior knowledge about the semantic relatedness between concepts to our Markov chain models as it is known that humans navigate the Web intuitively instead of randomly. Our experiments should be conducted on a variety of distinct navigational data including both goal oriented and free form navigation scenarios. We not only look at navigational paths over websites, but also abstract away to navigational paths over topics in order to get insights into cognitive patterns.

Categories and Subject Descriptors

H.1.2 [**Models and Principles**]: User/Machine Systems—*Human information processing*; H.5.4 [**Information Interfaces and Presentation**]: Hypertext/Hypermedia—*Navigation*

Keywords

Markov chain; memory; navigation; pragmatics; semantics; semantic relatedness; Web

WWW'14 Companion, April 7–11, 2014, Seoul, Korea.
ACM 978-1-4503-2745-9/14/04.
http://dx.doi.org/10.1145/2567948.2567956 .

1. INTRODUCTION AND PROBLEMS

Human navigation in networks has been playing a crucial role for both navigators as well as researchers for decades. Initiated by one of the most well-known experiments on human navigation by Stanley Milgram [20], researchers got particularly interested in understanding and leveraging human navigation patterns as well as improving models of human navigation. In Stanley Milgram's so called "small-world" experiments randomly chosen persons received the task of sending a letter to a target person in Boston. The experiment was built upon the limitations that the persons do not know the target in person, but send the letter through their local social network. Even though of this limitation, the results indicated that the target person could be reached with six steps on average which formed the well-known term "six degrees of separation".

With the appearance and steady rise of the Web in the 90s the interest on human navigation has been extended to information networks on the Web. When navigating a set of pages on the Web, users typically need to tap into their intuitions about real-world concepts and the perceived relationships between them in order to progress towards their set of targeted pages. It is also known that humans tend to find intuitive paths instead of necessarily short paths [32] and that they navigate the Web rationally instead of randomly [8]. Early work by Huberman et al. [15] already showed that such human navigation exhibits strong regularities in the distributions of user page visits on a website. This study was followed by many others who tried to thoroughly investigate human navigation models (e.g., [12]). Just recently, Wang and Huberman [31] confirmed these observations by showing that sequences of online activities exhibit regularities. This also applies to human activities in general and may be reasoned by inherent regularities of human behavior in general [28].

Problems. There has been a vast amount of previous works that tried to model human navigation on the Web. One of the most frequently used model for this task is the Markov chain model [23], where webpages are represented as states and hyperlinks as probabilities of navigating from one page to another. Predominantly, the Markov chain model has been memoryless in a wide range of works (e.g., Google's PageRank [5]) indicating that the next state only depends on the current state of a user's Web trail. Contrary, above mentioned regularities in human navigation may indicate that such a memoryless model for explaining human navigation on the Web is not enough. This leads to the lack of a concrete **understanding** of human navigation on the Web.

Furthermore, the mostly intuitive navigational paths may also be **leveraged** in order to improve different tasks on the Web. One potential application is the improvement of the computation of semantic relatedness between concepts which represents a fundamental challenge on our way to a semantically-enabled web. While a large amount of methods for this task exists, they mostly only capture semantics from a limited set of people (e.g., those users that create content on the Web), while they neglect *pragmatics* (i.e., how the Web is used). These pragmatic patterns could be abstracted by human navigational trails which consequently, may help us to improve the task of computing semantic relatedness.

The final problem that should be tackled in this PhD thesis is the **improvement** of human navigation on the Web. E.g., this is necessary in order to improve the users' Web experience by improving user interfaces or information network structures [3]. Similarly to the previous problem, where we mentioned that human navigation might improve the computation of semantic relatedness between concepts, we may also be able to apply the other way around and use semantic relatedness scores between concepts – e.g.,websites – for enhancing models of human navigation. As humans tend to navigate intuitively – navigation is also guided by *information scent* [8] – they might prefer navigating semantically similar pages instead of very distinct ones.

Contributions. Concretely, the – partly expected – main contributions of this work are threefold: (i) We develop a framework for better understanding human navigation on the Web and shed light on the memory and structure in human navigation patterns by using Markov chain models. (ii) We leverage human navigational paths through the Web for enhancing the computation of semantic relatedness scores between concepts. This is done by introducing a novel and universal method. (iii) We enhance human navigation models by incorporating prior knowledge about semantic relatedness scores between concepts.

2. RELATED WORK

In this section we roughly capture the related work for our three problems we want to tackle in this PhD thesis: (i) understanding, (ii) leveraging and (iii) improving human navigation on the Web.

Understanding human navigation. Exceeding the work on detecting regularities in human navigation that have been mentioned in Section 1 of this article, there exists a large array of previous work that tries to directly model and understand human navigation on the Web. As previously mentioned, one of the most frequently applied model for this task is the stochastic Markov chain model [4, 12, 19, 23, 26, 33]. Specific configurations of model parameters – such as transition probabilities or model orders – have been used to reflect different assumptions about navigation behavior. The principle that human navigation might exhibit longer memory patterns than the first order Markov chain captures – which is indicated by the so-called *Markovian assumption* – has been investigated in the past (see e.g., [3, 23]). However, higher order Markov chains have been often disputed because the gain of a higher order model did not compensate for the additional complexity introduced by the model [23]. Therefore, it was a common practice to focus on a first order model since it was a reasonable but extremely simple approximation of user navigation behavior (e.g., [7, 24, 26, 33]). This discussion was previously picked up again by Chierichetti et al. [10] who showed that higher order Markov chains might be more appropriate. However, their study does not account for the higher complexity of such models and the possible lack of statistically significant gains of these models which should be an elemental part of the framework presented in this thesis.

Leveraging human navigation. West et al. [32] have been the first to leverage human navigational paths for the task of computing semantic relatedness scores between concepts. Their work demonstrates the great potential of this approach, but is limited in some ways: (i) the semantic relatedness between two concepts can only be calculated if they at least co-occur once in a navigational path, (ii) the work is limited to a small set of concepts and paths and (iii) their work was evaluated by some human judges only. Our proposed method in this work should overcome these problems.

Improving human navigation. In [22] and [6] the authors showed that semantics affect how user search visual interfaces on websites. Blackmon et al. [2] also used semantic relatedness scores between link or header texts of websites and navigational user goals to evaluate usability of Web platforms. Similar incorporation of semantics for the task of detecting web usability issues can e.g., be found in [1, 9]. Kaur and Hornof [18] applied semantic relatedness methodologies (WordNet, LSA and PMI-IR) in order to predict the link that people would select on a web page. We want to pick up on these observations and want to directly improve human navigation models by incorporating information about semantic relatedness between concept (e.g., websites) as prior knowledge.

3. PROPOSED APPROACH AND METHODOLOGY

In this section we present our proposed approaches and methodological concepts for tackling the research problems of (i) understanding, (ii) leveraging and (iii) improving human navigation on the Web which we postulated in Section 1. We also emphasize specific concepts for evaluating our studies and at the end of this section we also describe the datasets we have already used or which we are planning to use throughout the course of this PhD thesis.

3.1 Understanding human navigation

As already described previously in this article, we want to model human navigation on the Web by using the well-known stochastic Markov chain models; websites are described as states and the transition between these states as probabilities of navigating from one website to another. Formally, a finite and discrete (in time and space) Markov Chain can be seen as a stochastic process that contains of a sequence of random variables – e.g., $X_1, X_2, \ldots, X_n$. One of the most well-known hypotheses about Markov Chains is the so-called *Markovian assumption* that postulates that the next state only depends on the current state and not on a sequence of preceding ones. Such a first-order Markov chain holds if:

$$P(X_{n+1} = x_{n+1} | X_1 = x_1, X_2 = x_2, ..., X_n = x_n) = P(X_{n+1} = x_{n+1} | X_n = x_n) \quad (1)$$

We assume *time-homogeneity* for all our Markov chains and for simplification we will refer to data as a sequence $D = (x_1, x_2, ..., x_n)$ with states from a finite set S. Furthermore, as we are also interested in higher order Markov chains, we can state that in a k-th order Markov chain the next state depends on k previous ones. One advantage of such representation is that we can easily convert higher order Markov chains to first-order Markov chains by modeling all possible sequences of length k as states and adjusting the probabilities accordingly. Hence, we can focus on defining the concepts for first-order chains solely, as this applies for higher ones as well.

A Markov chain model is usually represented via a stochastic transition matrix P with elements $p_{ij} = p(x_j | x_i)$ where it holds that for all i:

$$\sum_j p_{ij} = 1 \quad (2)$$

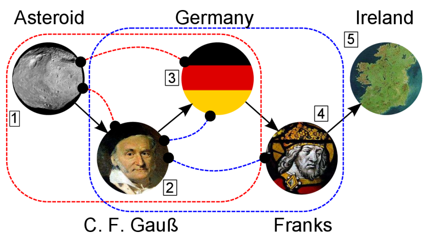

Figure 1: Demonstration of sliding windows

As we are interested about the appropriate Markov chain order – i.e., we want to identify memory patterns of humans navigating the Web – we need to establish methods for such a detection. Intuitively, we could think of picking the order with the highest likelihood. However, this is not enough as higher orders will always produce higher or at least similar likelihoods than lower order Markov chain models. The reason for this is that the number of parameters needed increases exponentially for higher orders and hence, this may result in overfitting[1] [21] – we need to note that lower order models are also always nested within higher order models.

In order to enhance existing work, we want to make use of a large variety of distinct methods that try to balance the goodness of fit for a given Markov chain order model with the actual number of parameters needed [17] – i.e., statistical significant improvements. We want to focus on two main methodological concepts: (1) Maximum Likelihood Estimation and (2) Bayesian Inference[2]. Both concepts give us several tools for detecting the best Markov chain order model in our framework: likelihood ratio test [30], Akaike information criterion (AIC) [30] and Bayesian information criterion (BIC) [17] for the maximum likelihood estimation and the Bayesian concept of using the Bayes rule of calculating $P(M_k|D)$ where M_k represents a model with specific order k and D represents the data [29]. For a robustness analysis of our model selection strategies we also use cross-validation for a prediction scenario (e.g., the next click of a user should be predicted) [10, 21].

3.2 Leveraging human navigation

As mentioned in Section 1 we want to leverage human navigational paths for calculating semantic relatedness between concepts users navigate over – e.g., Wikipedia pages. Schuetze and Pedersen [25] introduced the method for calculating semantic relatedness using lexical co-occurrence information between words – or in our case Web concepts. The basic idea is to represent each concept as a vector capturing the co-occurrence count to all other concepts in a multi-dimensional space. We adopt this concept and apply it to human navigational paths. Concretely, we use *second-order co-occurrence* information within paths and calculate the *cosine similarity* between co-occurrence vectors, given two concepts. To capture relatedness of two concepts in a corpus of human navigation paths, we use *sliding windows* of a variable size k over the path sequence. Thereby, we follow the natural assumption that the distance between two concepts in navigational paths is crucial for calculating precise semantic relatedness scores.

Figure 1 illustrates how we calculate the co-occurrence between concepts available in a path with a sample window size of $k = 3$. Circles represent articles (e.g., Wikipedia articles), rounded rectangles represent a window, the solid arrows represent the path taken and the dashed lines with dotted ends each represent a (symmetric) co-occurrence between two concepts. The final co-occurrence information is then projected to a large matrix where each row represents the co-occurrence vector of a specific concept, which is then used for the cosine similarity calculation.

To evaluate semantic relatedness, we compare our results to a golden standard dataset, specifically the *WordSimilarity-353* dataset [13]. We first map the Web concepts we look at the corresponding word pairs of the dataset and then calculate the semantic relatedness scores for each pair using our method and dataset and finally calculate the accuracy using the *Spearman rank correlation*. Further evaluation techniques like using the output for several NLP tasks like word sense disambiguation, recommendation or text segmentation are as well possible.

3.3 Improving human navigation

We also plan the reversal of Section 3.2 by improving human navigational models with prior knowledge about the semantic relatedness between concepts users navigate over. We again use Markov chain models which we described in Section 3.1 for modeling human navigation. Concretely, we want to apply only the Bayesian inference approach in this case as it allows us to intuitively incorporate prior knowledge to the model – priors allow to model our prior believe about the parameters before we see the data [29]. However, we first need to calculate the semantic relatedness between concepts. The methodology used will depend on the type of concepts we look at. There exists an array of well-performing methods for calculating semantic relatedness between concepts that we want to apply for tackling this research problem. Some well-known methods are: explicit semantic analysis (ESA) [14], latent semantic analysis (LSA) [11], WordNet based methods like Jiang-Conrath distance [16] and many more. To evaluate whether this prior incorporation of semantic relatedness scores between concepts can improve the Markov chain modelling process, we resort to the Bayesian model selection approach introduced in Section 3.1.

3.4 Datasets

For this thesis we want to study several human navigational datasets. First of all, we have access to human navigational paths over Wikipedia pages collected from TheWikiGame[3]. In this game users have to navigate over Wikipedia pages in order to reach a predefined Wikipedia concept starting from a given start page. Further datasets which we have already preprocessed are: (i) Wikispeedia[4] which is a similar game to the Wikigame and which has already been investigated in previous work by West et al. [32], (ii) navigational data from MSNBC[5] and (iii) human navigational paths over the collaborative tagging system BibSonomy[6]. Furthermore, we are in the process of acquiring click trails from users at the University of Indiana[7]. For our experiments we also abstract away from navigational paths over Web pages to navigational paths over topics. To give an example, we take the human navigational paths over Wikipedia pages from our Wikigame dataset and replace each page in all paths with a corresponding Wikipedia category. Hence, we reduce the state space by a large margin which also allows us to get insights into the cognitive navigational behavior of humans.

[1]This is a similar problem as the famous quote by John von Neumann states: "With four parameters I can fit an elephant, and with five I can make him wiggle his trunk."

[2]Hence, we cover the ideas of both big statistical parties of frequentists and Bayesians.

[3]`http://thewikigame.com/`

[4]`http://wikispeedia.net/`

[5]`http://msnbc.com/`

[6]`http://www.bibsonomy.org/`

[7]`http://cnets.indiana.edu/groups/nan/webtraffic/click-dataset`

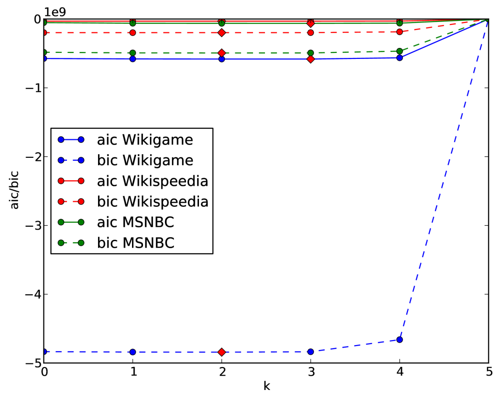

Figure 2: AIC and BIC values of varying order k for various navigational datasets. The lowest values depict the most appropriate models and are illustrated by diamond symbols.

4. PRELIMINARY RESULTS

In this section we present preliminary results that we have already achieved during the course of this PhD thesis. We focus on the problems of understanding and leveraging human navigation on the Web as we have not yet applied our ideas about improving human navigation models.

4.1 Understanding human navigation

We have constructed a framework that implements the Markov chain model fitting process for both the maximum likelihood and Bayesian variants. Furthermore, the framework is capable of determining the appropriate Markov chain order by applying several model estimation methods described in Section 3.1. Our current version of this framework can be found online[8].

We applied our framework to both goal-oriented navigational data (Wikigame and Wikispeedia) and free-form navigational data (MSNBC). We could confirm what we know from theory: It is indeed difficult to make plausible statements about the appropriate order of a Markov chain having insufficient data but a vast amount of states which results in too complex models. This suggests that the memoryless model is a plausible model for human navigation on a page level. Contrary, by representing pages by their corresponding topical category (see Section 3.4) we could identify that navigation on a cognitive level is not memoryless. Explanatory, we illustrate the AIC and BIC values for all topical datasets in Figure 2. We can clearly see that an order of two and respectively three best explain the observed data, independent whether the navigation is goal-oriented or free-form. Our other model selection techniques confirm these observations. Further structural investigations showed that users tend to stay in the same topic while navigating. However, this is much more frequent for our free form navigational dataset (MSNBC) as compared to both of the goal-oriented datasets (Wikigame and Wikispeedia).

4.2 Leveraging human navigation

For tackling the problem of leveraging human navigational paths for the task of calculating semantic relatedness scores between concepts we have applied our methodological concepts described in Section 3.2 to our Wikigame dataset. By evaluating the scores against the WordSimilarity-353 gold standard we could show that

[8] `https://github.com/psinger/PathTools/`

Table 1: Semantic relatedness accuracy

Window size	None	2	3	4	5
Accuracy	0.644	0.636	0.706	0.713	0.686

human navigational paths indeed may provide a viable source for calculating semantic relatedness between concepts in information networks. We found that our method can produce very precise semantic relatedness scores, but as hypothesized the window size is important. We could achieve the most accurate results by using a window size of 3 or 4 (see Table 1). Finally, we also found that not all navigational paths are equally useful. By sampling subsets of paths based on distinct characteristics (e.g., only using paths with low indegree nodes) we could outperform the complete corpus available. Preliminary work regarding this topic has already been published in poster format [27].

5. CONCLUSIONS AND FUTURE WORK

Human navigation on the Web represents a fundamental activity of Web users. This PhD thesis focuses on three core aspects: understanding, leveraging and improving human navigation on the Web. Each of these three investigations could stand on its own, but this thesis should combine them in a dynamic way. In this article we introduced preliminary steps in creating a Markov chain framework that allows us to get detailed insights into memory and structure in human navigational patterns. Furthermore, we already applied this framework to a small set of navigational datasets and our preliminary results indicate that human navigation exhibits memory patterns at least on a cognitive level. Our results also suggest that human navigational paths are perfectly suited for deriving common sense knowledge for the task of calculating semantic relatedness between concepts.

Currently, one limitation of our work is that we partly investigated navigational data that has been produced by a game and hence, this may bias our obtained results. However, the data represents an abstraction of real user navigation in information networks. In future, we want to focus more heavily on real user navigation by extending our experiments to such data which we have mostly already successfully preprocessed. Especially, we want to thoroughly investigate our navigational dataset from the collaborative tagging platform BibSonomy which should especially useful for the task of leveraging navigational paths. We also plan on enhancing our Markov chain framework and also take a look at navigational data from different domains (e.g., ontologies). Finally, we want to heavily focus on improving navigational models by incorporating prior knowledge about the semantic relatedness between concepts as proposed in this article.

Acknowledgements. This work is in part funded by the FWF Austrian Science Fund Grant I677.

6. REFERENCES

[1] M. H. Blackmon, M. Kitajima, and P. G. Polson. Repairing usability problems identified by the cognitive walkthrough for the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 497–504, New York, NY, USA, 2003. ACM.

[2] M. H. Blackmon, P. G. Polson, M. Kitajima, and C. Lewis. Cognitive walkthrough for the web. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '02, pages 463–470, New York, NY, USA, 2002. ACM.

[3] J. Borges and M. Levene. Data mining of user navigation patterns. In *Revised Papers from the International Workshop on Web Usage Analysis and User Profiling*, WEBKDD '99, pages 92–111, London, UK, UK, 2000. Springer-Verlag.

[4] J. Borges and M. Levene. Evaluating variable-length markov chain models for analysis of user web navigation sessions. *IEEE Trans. on Knowl. and Data Eng.*, 19(4):441–452, Apr. 2007.

[5] S. Brin and L. Page. The anatomy of a large-scale hypertextual web search engine. In *Proceedings of the seventh international conference on World Wide Web 7*, WWW7, pages 107–117, Amsterdam, The Netherlands, The Netherlands, 1998. Elsevier Science Publishers B. V.

[6] D. P. Brumby and A. Howes. Good enough but i'll just check: Web-page search as attentional refocusing. In *ICCM*, pages 46–51, 2004.

[7] I. Cadez, D. Heckerman, C. Meek, P. Smyth, and S. White. Model-based clustering and visualization of navigation patterns on a web site. *Data Min. Knowl. Discov.*, 7(4):399–424, Oct 2003.

[8] 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*, CHI '01, pages 490–497, New York, NY, USA, 2001. ACM.

[9] E. H. Chi, A. Rosien, G. Supattanasiri, A. Williams, C. Royer, C. Chow, E. Robles, B. Dalal, J. Chen, and S. Cousins. The bloodhound project: automating discovery of web usability issues using the infoscentπ simulator. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pages 505–512, New York, NY, USA, 2003. ACM.

[10] F. Chierichetti, R. Kumar, P. Raghavan, and T. Sarlos. Are web users really markovian? In *Proceedings of the 21st international conference on World Wide Web*, WWW '12, pages 609–618, New York, NY, USA, 2012. ACM.

[11] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, and R. Harshman. Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407, 1990.

[12] M. Deshpande and G. Karypis. Selective markov models for predicting web page accesses. *ACM Trans. Internet Technol.*, 4(2):163–184, May 2004.

[13] L. Finkelstein, E. Gabrilovich, Y. Matias, E. Rivlin, Z. Solan, G. Wolfman, and E. Ruppin. Placing Search in Context: The Concept Revisited. *ACM Transactions on Information Systems*, 20(1):116–131, Jan. 2002.

[14] E. Gabrilovich and S. Markovitch. Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis. In *Proceedings of The Twentieth International Joint Conference for Artificial Intelligence*, IJCAI '07, pages 1606–1611, San Francisco, CA, USA, 2007. Morgan Kaufmann Publishers Inc.

[15] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, and R. M. Lukose. Strong regularities in world wide web surfing. *Science*, 280(5360):95–97, Mar 1998.

[16] J. Jiang and D. Conrath. Semantic similarity based on corpus statistics and lexical taxonomy. In *Proc. of the Int'l. Conf. on Research in Computational Linguistics*, pages 19–33, 1997.

[17] R. W. Katz. On some criteria for estimating the order of a markov chain. *Technometrics*, 23(3):243–249, Aug 1981.

[18] I. Kaur and A. J. Hornof. 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*, CHI '05, pages 51–60, New York, NY, USA, 2005. ACM.

[19] R. Lempel and S. Moran. The stochastic approach for link-structure analysis (salsa) and the tkc effect. *Comput. Netw.*, 33(1-6):387–401, June 2000.

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

[21] K. P. Murphy. The encyclopedia of cognitive science, 2002.

[22] B. J. Pierce, S. R. Parkinson, and N. Sisson. Effects of semantic similarity, omission probability and number of alternatives in computer menu search. *International Journal of Man-Machine Studies*, 37(5):653–677, 1992.

[23] P. L. T. Pirolli and J. E. Pitkow. Distributions of surfers' paths through the world wide web: Empirical characterizations. *World Wide Web*, 2(1-2):29–45, Jan 1999.

[24] R. R. Sarukkai. Link prediction and path analysis using markov chains. In *Proceedings of the 9th international World Wide Web conference on Computer networks : the international journal of computer and telecommunications netowrking*, pages 377–386, Amsterdam, The Netherlands, The Netherlands, 2000. North-Holland Publishing Co.

[25] H. Schuetze and J. O. Pedersen. A Cooccurrence-based Thesaurus and two Applications to Information Retrieval. *Information Processing and Management*, 33(3):307–318, 1997.

[26] R. Sen and M. Hansen. Predicting a web user's next access based on log data. *Journal of Computational Graphics and Statistics*, 12(1):143–155, 2003.

[27] P. Singer, T. Niebler, M. Strohmaier, and A. Hotho. Computing Semantic Relatedness from Human Navigational Paths on Wikipedia. In *Proceedings of the 22nd International Conference on World Wide Web*, WWW '13, New York, NY, USA, 2013. ACM.

[28] C. Song, Z. Qu, N. Blumm, and A.-L. Barabási. Limits of predictability in human mobility. *Science*, 327(5968):1018–1021, 2010.

[29] C. C. Strelioff, J. P. Crutchfield, and A. W. Hübler. Inferring markov chains: Bayesian estimation, model comparison, entropy rate, and out-of-class modeling. *Phys. Rev. E*, 76:011106, Jul 2007.

[30] H. Tong. Determination of the order of a markov chain by akaike's information criterion. *J. Appl. Prob.*, 12:488–497, 1975.

[31] C. Wang and B. A. Huberman. How random are online social interactions? *Scientific reports*, 2, 2012.

[32] R. West, J. Pineau, and D. Precup. Wikispeedia: An Online Game for Inferring Semantic Distances between Concepts. In *Proceedings of the 21st International Joint Conference on Artifical Intelligence*, IJCAI '09, pages 1598–1603, San Francisco, CA, USA, 2009. Morgan Kaufmann Publishers Inc.

[33] I. Zukerman, D. W. Albrecht, and A. E. Nicholson. Predicting users' requests on the www. In *Proceedings of the Seventh International Conference on User Modeling*, pages 275–284, Secaucus, NJ, USA, 1999. Springer-Verlag New York, Inc.