A probabilistic model for road selection in mobile maps

Thomas C. van Dijk* and Jan-Henrik Haunert

Chair for Computer Science I, University of Würzburg,
Am Hubland, D-97074 Würzburg,
{thomas.van.dijk,jan.haunert}@uni-wuerzburg.de

Abstract. Mobile devices provide an interesting context for map drawing. This paper presents a novel road-selection algorithm based on PageRank, the algorithm famously used by Google to rank web pages by importance. Underlying the PageRank calculation is a probabilistic model of user behavior. We provide suitable generalizations of this model to road networks. Our implementation of the proposed algorithms handles a sizable map in approximately a quarter of a second on a desktop PC. Therefore, our methods should be feasible on modern mobile devices.

Keywords: mobile maps, generalization, wayfinding, PageRank

1 Introduction

Maps on smartphones and other mobile devices are widely used for wayfinding, orientation, and exploration tasks. Most maps displayed by commercial software, however, are not well adapted for such applications. Often, they present too much unnecessary detail and thus distract users from comprehending the essential information quickly. Though the selection of a good subset of features from a geographic data set is a classical problem in map generalization, further research is needed, particularly to better address application-specific needs of users.

Several methods have been developed that generate maps of road networks for certain tasks, for example, you-are-here maps [11], which allow a user to localize himself, route maps [1], which allow a user to navigate from his current location to a specified destination, and destination maps [7], which allow users coming from different directions to navigate to a common destination. Two techniques that have been proven to be rather generally applicable are the grouping of sequences of almost collinear road segments into strokes [12] and the selection of roads based on betweenness centrality [4]. Generally, however, algorithms are missing that are configurable for multiple applications.

In this paper, we address a *range* of applications that we define by its two extreme cases, namely pure exploration and pure wayfinding. In a pure exploration application, the user needs detailed information about his surrounding but does not need to reach a certain destination. Therefore, he might favor the

* Corresponding author

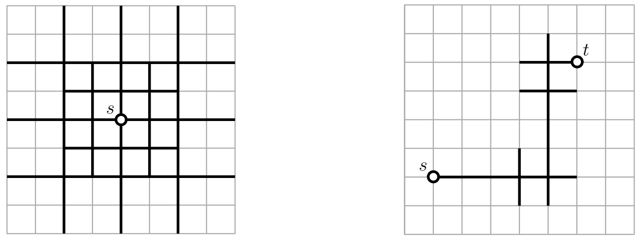

Fig. 1. Two maps that show different subsets of edges (black) of a road network (gray).

left map in Fig. 1, which displays the region around his location s in relatively high detail. With pure wayfinding, on the other hand, we mean that the user is solely interested in getting to a certain destination t and thus only needs a map similar to the right in Fig. 1. This map visualizes a single precomputed route from node s to node t plus some information that the user needs to not miss turns. Algorithms for maps similar to those in Fig. 1 have been presented before, for example, by Yamamoto et al. [16] and Agrawala and Stolte [1], respectively. We argue, however, that these maps might be of limited use if, for example, the user is interested in sights that lie close to his route or wants to have some freedom in selecting his route. In these cases, the user needs to be provided with a road network map that contains a sufficiently large choice of routes and relatively highly detailed information about his surroundings. We therefore present an algorithm for road selection that is configurable with respect to how freely the user moves through the road network. Note that this depends on the mode of transportation employed by the user, since, for example, pedestrians are usually less constrained in choosing their routes than cars. In this paper, our primary motivation is pedestrian navigation.

In this paper, we propose a technique related to PageRank, the algorithm famously used by Google to rank web pages by importance [9]. As a way to rank items in a network and a way to filter information, we consider how to apply it to road networks. In fact, PageRank has earlier been used by Jiang [3] to predict human movement in road networks, but has not been applied to road selection or map generalization before. PageRank has a probabilistic interpretation. For 'normal' PageRank this is the *random surfer* model, where the score of a page corresponds to the probabilistic behavior of a hypothetical user. In his paper, Jiang shows that his version of PageRank corresponds well to observed traffic amounts. This suggests that his (implied) model of traveler behavior is reasonable. It also suggests that specifically this underlying probabilistic model can be an interesting object of study. This provides proper interpretation for our results. We base our edge selection on the probabilities in a model of a 'random traveler:' if the user is likely to traverse a certain edge, then we should display it.

Map generalization is the problem of deriving a map of a smaller scale from a given map. Classical tasks in map generalization are simplification, aggregation, displacement, and feature selection, which has been studied at least since the 1960s when Töpfer and Pillewizer [13] investigated existing maps to derive guidelines about the number of features a map of a certain scale should contain.

For a long time, research on map generalization focused on static topographic maps. This changed with the advent of web mapping and mobile cartography, which challenged researchers to find solutions that also work on small screens and with limited data bandwidths [5] but also offers new possibilities of tailoring a map to a user's task, for example by highlighting regions of interest [17]. Several methods have been suggested to produce wayfinding maps, which visualize a single precomputed route a user has to follow, plus some context information. For such maps, geometric correctness is rather unimportant, thus they may be highly distorted and schematized to improve readability [1, 10]. In this paper, however, we concentrate on the selection of roads.

Schmid et al. [11], who mainly aim at supporting self-localization by providing the user with appropriate context, also propose a method for road selection, which, similar to the method of Jiang and Claramunt [4], is based on betweenness centrality.

The rest of the paper is structured as follows. In Sect. 2 we review some definitions, in particular PageRank. In Sect. 3 we develop the random traveler model. We finish in Sect. 4 with some experimental results using an implementation of our proposed algorithms.

2 Preliminaries

In this paper we use the concept of a *line graph* or *linear dual graph* [15]. We use a directed version.

Definition 1 (Line graph $\mathcal{L}(G)$). *Let $G = (V, A)$ be a directed graph. Then G's line graph $\mathcal{L}(G) = (A, B)$ is a directed graph on the arcs of G, where B is defined as follows. For all $a, b, c \in V$, we have $((a, b), (b, c)) \in B$ if and only if both $(a, b) \in A$ and $(b, c) \in A$.*

We interpret the input road network as a (possibly directed) graph. We replace any undirected edges with two arcs, one going each way. Let $G = (V, A)$ be the resulting directed graph and let $n = |V|$ and $m = |A|$. Then we take the line graph $L = \mathcal{L}(G)$. At this point it would be possible to drop some arcs from L, for example to model turn restrictions [15]. We have not done so, since our data sets did not contain turn restrictions for pedestrians.

We run our ranking algorithm on the line graph L. (For correctness of the algorithm it is not required that L is a line graph.) Since the input graph G is a road network, it is reasonable to assume that it has bounded degree. Then the line graph $\mathcal{L}(G)$ has $\mathcal{O}(\,n\,)$ nodes and arcs.

Random walks and PageRank

PageRank is a network ranking algorithm that is famous for its application in web search. It is typically formulated as an eigenvector calculation and as such is closely related to eigenvector centrality. It is well known from literature that PageRank can be interpreted using the *random surfer model* [9]. The high-level interpretation is that important pages are those that a user is likely to view. In this paper we apply a modified random 'surfer' model to road networks: important roads for a user are those that he is likely to traverse. This involves introducing appropriate generalizations to model user behavior and to handle road networks.

First we recall the random surfer model for PageRank. Consider a hypothetical user viewing a web page. At every time step this user follows a uniformly random link on the page he is viewing. The chosen outgoing link does not depend on the incoming link that the user arrived from, so in this model the location of the user has the Markov property. Then the probability distribution of the user location as time goes to infinity can be calculated using standard methods.

The above process will get stuck on any pages that have no outgoing links. As time goes to infinity, the hypothetical user would end up only in such nodes, which clearly does not correspond to actual user behaviour. Therefore, the model includes the possibility that the user jumps to a uniformly random page. This is called *damping* and happens, with some constant probability $1-d$, at every time step. Otherwise (with probability d) an outgoing link is followed as before. This is still a Markov process. The parameter d is called the damping factor.

Let W be the directed graph of the n pages and their links. Let M be the *link matrix*, which is the adjacency matrix of W where the columns are each scaled to sum to 1. That is, if j links to i, then $M_{ij} = 1/outdeg(v_j)$ and otherwise $M_{ij} = 0$. Let $\boldsymbol{\Delta}$ be a vector with all entries $1/n$. Then by definition, in a steady state the PageRank vector $\boldsymbol{r}$ satisfies

$$\boldsymbol{r} \;=\; dM\boldsymbol{r} + (1-d)\boldsymbol{\Delta}. \tag{1}$$

Such $\boldsymbol{r}$ exists and a numerical solution can be found efficiently, for example by iterative methods (see for example [8]).

3 Random Traveler Model

We now introduce the *random traveler model*. Like PageRank, it concerns random walks. Conceptually it applies to the line graph of a road network instead of to a web graph. Concretely, it differs from the normal random surfer model in three ways. First, we use non-uniform transition probabilities: to model user behavior, not every outgoing arc in the line graph is equally likely. Secondly, we use non-uniform damping. This allows us to focus the road selection on the user. Lastly, we handle non-uniform transition times: in contrast to links in the world-wide web, the time it takes to traverse a road segment is relevant.

Non-uniform transition probability Let G and $L = \mathcal{L}(G)$ be graphs as defined before. Note that a node in L corresponds to a direction on a road segment in G. We consider the user task of navigating from a node s to a node

t. A path in L corresponds to a route in the road network. Every arc on such a path corresponds to transitioning from one road segment to another. In this way, every arc represents a *routing decision*. We want to model that some routing decisions are more likely than others. This can easily be achieved by modifying M in Equation (1): each non-zero entry in M corresponds to an arc in L and non-uniform transition probabilities can be put in the matrix instead of the uniform ones.

Then the question is of course: what transition probabilities do we use? In Jiang's Weighted PageRank [3], a method for traffic prediction, the transition probabilities are proportional to indegrees (though not specifically in a line graph). As our application is different, we pick different probabilities.

At first we tentatively consider that each arc is equally likely; we assign to every arc odds 1. Then we will increase the odds of certain arcs and afterward normalize to get probability distributions. We propose two factors for determining these odds. First, note that the user is trying to navigate to the node t. We argue that the user is more likely to navigate onto road segments that lie on the shortest path tree towards t, since that is a good decision for getting to t. Therefore, a bonus of β_{SP} is added to the odds of such arcs, where β_{SP} is a tweakable parameter. These arcs can be found using a standard shortest path tree algorithm (for example Dijkstra's algorithm [2]). The set of arcs does not depend on s and therefore only needs to be recomputed when t changes.

Secondly, we propose a bonus for arcs that represent going approximately straight on. This corresponds for example to the *non-turning* concept of Klippel et al. [6]. Taking such non-turning transitions means routes with low turn complexity; people prefer such routes [14]. Additionally, we argue that the user might 'miss a turn' and mistakenly go straight ahead. Therefore, we give a bonus of $\beta_{\|}$ to arcs that correspond to a turn of less than $\alpha_{\|}$. This depends neither on s nor on t so it needs to be computed only once per road network.

With a higher value for this non-turning bonus $\beta_{\|}$, we model a dislike of turns. Note that we could instead have put a cost for turns into the shortest-path calculations. (See for example Winter [15].) That results in prescribing to the user what the exact trade-off is between turn minimization and distance minimization. By putting this trade-off into our random traveler model we allow for a more nuanced calculation that is likely to find multiple reasonable paths. We think this better reflects the nature of pedestrians.

In Sect. 4 we explore how the values for these bonus parameters influence subsequent edge selection. Throughout the paper we use $\alpha_{\|} = 30°$. For now, we mention that on our data, values like $\beta_{\mathrm{SP}} = 5$ and $\beta_{\|} = 5$ give reasonable results.

Non-uniform damping Our second modification to PageRank is non-uniform damping. With a motivation different from ours, a modified damping vector is known in web-search literature as a *personalization vector*. The damping in PageRank corresponds to the hypothetical user jumping to a random web page. Instead, we set the damping to always return to node s. In terms of Equation (1), this is accomplished by setting $\boldsymbol{\Delta}$ to 1 at s and 0 elsewhere.

To see why this makes sense, consider a random walk starting from s, influenced by damping. As argued before, the ranking vector $\boldsymbol{r}$ that we calculate is the probability distribution of the location of the traveler as time goes to infinity. Because of our damping vector, at every step of the process there is a probability $(1 - d)$ that the walk returns to s. Consider splitting the walk at these jumps so that each sub-walk starts at s. Sub-walks of length k occur with exponential distribution in k: at every time step, there is a constant probability that the sub-walk terminates. Then $\boldsymbol{r}$ is also the probability distribution of the location of the traveler, taking a random walk of exponentially distributed length, starting at s. This is a reasonable distribution for how far the user's interest deviates from his current location.

The interpretation of the calculated scores is then: starting from s, where is the traveler likely to go soon? This way we can use a real-world property to pick the damping factor d, a parameter for which it is generally not clear how to pick a value. A sub-walk of k steps occurs by taking k normal steps, each at probability d, and then a damping step, at probability $(1 - d)$. The relation between the expected length $\mathcal{E}$ and d is then $\mathcal{E} = \sum_{k=1}^{\infty} k \cdot (1-d) \cdot d^k$ which results in $d = \frac{\mathcal{E}}{\mathcal{E}+1}$.

Non-uniform transition time As third and final modification to PageRank, we handle non-uniform transition time. A node in L corresponds to (a direction on) a road segment, so spending time traversing a road segment corresponds to spending time in a node. The amount of time does not depend on where the user came from or where the user is going. The ranking vector $\boldsymbol{r}$ as used up to now is the proportion of events where the user goes to a certain node. To get the probability over time instead of over events, simply weigh by the time spent at each node and renormalize.

4 Experiments

We have implemented the ranking algorithm proposed in this paper. To get an edge selection from the calculated scores, we select the minimum number of edges that together sum to score at least 90%. According to the random traveler model, this set of edges contains wherever the user is likely to go soon.

In summary, the runtime on reasonably large data sets is good and certainly feasible on a mobile device. In this section, we give details of this claim. All experiments have been run on a desktop PC with an Intel® Core™ i5-2400 CPU at 3.10Ghz. Though that is certainly more powerful than a current smartphone, it is not so by orders of magnitude.

Our examples are run on a road network of the German city of Würzburg and its surroundings, consisting of 3786 nodes and 4987 road segments. This network is a crop of a much larger OpenStreetMap data set of Lower Franconia (approximately 10^5 nodes) that is available online[1] in the ESRI shapefile format.

[1] `download.geofabrik.de/osm/`

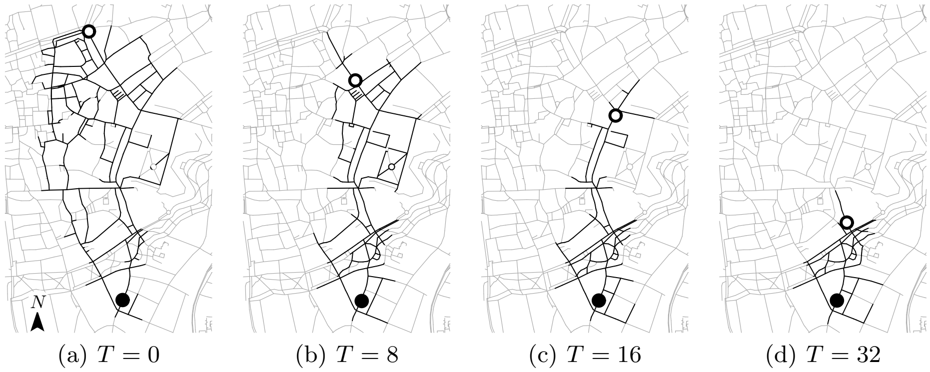

Fig. 2. Several edge selections while a user travels a shortest path to a destination. The value T indicates the number of road segments traveled. Current user location is marked with ○, destination is marked with ●.

This is generally a reasonable approach for our method: first generously clip an area of a larger map and then run our algorithm on the cropped map. By the nature of the random traveler model, we will not select edges very far from the initial user location. Since our algorithm is fast and space efficient, a very wide crop can be used so as not to influence the results. The figures in this paper have been clipped further to show only the area where edges are actually selected.

In Fig. 2 we show the edge selections at several times for a user traveling a shortest path toward a destination. The value T indicates the number of road segments traveled. The focus area shifts to match the current user location. Our algorithms are fast enough to do these calculations in realtime on a smartphone.

Note that in the first selection, possibilities for various distinct routes are presented. As time goes on, the selection becomes more concentrated. Also note that more detail is added near the destination as the user approaches. These things follow from the random traveler model.

To calculate the scores in our random-traveler model, we use the typical iterative method, where nodes distribute score to their neighbors. It should be possible to use more sophisticated methods as known from PageRank literature, but we have not investigated this: a simple implementation already gives good runtime on reasonable data sets.

As termination criterion for our calculations we have used a threshold of 10^{-6}: we continue as long as the sum of score changes in an iteration is more than this value. Then it is reasonable to assume convergence. In an experiment with 1000 random source/destination pairs, this procedure took an average of 102ms (719 iterations).

All experiments use damping factor $d = 0.98$. As calculated in Sect. 3, this means the expected length of paths considered is 49 transitions, which is reasonable for queries in this network. As mentioned before, all our experiments use $\alpha_{\parallel} = 30°$. Figures 3 and 4 show the effect of the shortest-path bonus β_{SP} and the

(a) $\beta_{\mathrm{SP}} = 0$ (b) $\beta_{\mathrm{SP}} = 1$ (c) $\beta_{\mathrm{SP}} = 5$

Fig. 3. Effect of the shortest-path-bonus parameter β_{SP}. Notice that a higher value directs the edge selection toward wayfinding versus exploration. All of these pictures have $\beta_{\|} = 0$. User location is marked with ○, destination is marked with ●.

(a) $\beta_{\|} = 0$ (b) $\beta_{\|} = 5$ (c) $\beta_{\|} = 15$

Fig. 4. Effect of the non-turning-bonus parameter $\beta_{\|}$. Notice that a higher value leads to the selection of longer non-turning stretches of road. Here more edges are selected if $\beta_{\|}$ is higher: the random traveler distributes over a wider area. User location is marked with ○.

non-turning bonus $\beta_{\|}$. Notice how β_{SP} allows a gradual change from exploration to wayfinding. Some discussion is included in the captions.

5 Conclusion

We have presented a novel algorithm for edge selection. Partially inspired by PageRank, it is based on a 'random traveler model' and has a solid foundation in probability theory. Using an implementation we have shown that the proposed algorithm runs in realtime.

As regards future work, we have currently only concerned ourselves with road selection. It could be interesting to apply schematization to the resulting selection. Additionally, we note that it may be interesting to use data mining to train a traveler model based on real-world data.

References

1. M. Agrawala and C. Stolte. Rendering effective route maps: improving usability through generalization. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques*, SIGGRAPH '01, pages 241–249. ACM, 2001.
2. E. W. Dijkstra. A note on two problems in connexion with graphs. *Numerische Mathematik*, 1:269–27, 1959.
3. B. Jiang. Ranking spaces for predicting human movement in an urban environment. *International Journal of Geographical Information Science*, 23(7):823–837, 2009.
4. B. Jiang and C. Claramunt. A structural approach to the model generalization of an urban street network. *Geoinformatica*, 8(2):157–171, 2004.
5. C. B. Jones and J. M. Ware. Map generalization in the Web age. *International Journal of Geographical Information Science*, 19(8):859–870, 2005.
6. A. Klippel, H. Tappe, L. Kulik, and P. U. Lee. Wayfinding choremes – a language for modeling conceptual route knowledge. *Journal of Visual Languages and Computing*, 16(4):311–329, 2005.
7. J. Kopf, M. Agrawala, D. Bargeron, D. Salesin, and M. Cohen. Automatic generation of destination maps. *ACM Transactions on Graphics*, 29(6):158:1–158:12, 2010.
8. A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings*, chapter 4. Princeton University Press, 2006.
9. L. Page, S. Brin, R. Motwani, and T. Winograd. The pagerank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford University, 1999.
10. F. Schmid. Knowledge-based wayfinding maps for small display cartography. *Journal of Location Based Services*, 2(1):51–83, 2008.
11. F. Schmid, C. Kuntzsch, S. Winter, A. Kazerani, and B. Preisig. Situated local and global orientation in mobile you-are-here maps. In *Proceedings of the 12th international conference on Human Computer Interaction with mobile devices and services*, MobileHCI '10, pages 83–92. ACM, 2010.
12. R. C. Thomson and R. Brooks. Exploiting perceptual grouping for map analysis, understanding and generalization: The case of road and river networks. In *Selected Papers from the Fourth International Workshop on Graphics Recognition Algorithms and Applications*, GREC '01, pages 148–157. Springer-Verlag, 2002.
13. F. Töpfer and W. Pillewizer. The principles of selection. *The Cartographic Journal*, 3(1):10–16, 1966.
14. A. Turner. The role of angularity in route choice. In *Spatial Information Theory*, volume 5756 of *Lecture Notes in Computer Science*, pages 489–504. Springer-Verlag, 2009.
15. S. Winter. Modeling costs of turns in route planning. *GeoInformatica*, 6:345–361, 2002.
16. D. Yamamoto, S. Ozeki, and N. Takahashi. Focus+glue+context: an improved fisheye approach for web map services. In *Proceedings of the 17th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '09, pages 101–110. ACM, 2009.
17. A. Zipf and K.-F. Richter. Using focus maps to ease map reading: Developing smart applications for mobile devices. *Künstliche Intelligenz (KI)*, 2(4):35–37, 2002.