Hierarchical Route Planning Based on Taxi GPS-Trajectories

Qingquan Li
State Key Laboratory of
Information Engineering in Surveying,
Mapping, and Remote Sensing,
Wuhan University,
Wuhan, Hubei, China
qqli@whu.edu.cn

Zhe Zeng
State Key Laboratory of
Information Engineering in Surveying,
Mapping, and Remote Sensing,
Wuhan University,
Wuhan, Hubei, China
zz0459@yahoo.com.cn

Bisheng Yang
State Key Laboratory of
Information Engineering in Surveying,
Mapping, and Remote Sensing,
Wuhan University,
Wuhan, Hubei, China

Tong Zhang
State Key Laboratory of
Information Engineering in Surveying,
Mapping, and Remote Sensing,
Wuhan University,
Wuhan, Hubei, China

***Abstract*—The popular methods of route planning, the planar and hierarchical algorithms, are mainly based on the Dijkstra's algorithm. The shortest path in machine computation is always different from the optimal one in a human cognition because travel experience information of human is neglected. In this paper,hierarchical travel experience information of road network is given by statistical analysis on a large amount of taxi GPS trajectories. According to taxi trajectories, the roads can be divided into frequent roads, secondary frequent roads and seldom roads. These trajectories can well reflect a hierarchical cognition of road network because considering taxi driver's cognition of the road network. With the hierarchical model of road network, the corresponding hierarchical route planning can be performed based on the experienced driver's cognition. The experiments show that the route results according to the experiential information approach to optimal results more than that according to static information of road network.**

***Index Terms*—Route planning,hierarchy,road network**

I. Introduction

The popular methods of route planning, the planar and hierarchical algorithms, are mainly based on the Dijkstra's algorithm. The Dijkstra's algorithm principally calculates a shortest path between source node and target node with the weight of edges in road network. The length or travel time of road is commonly used as weight of the edges. The shortest path in geometrical measure is computed from regarding length as weight of the edge while the fastest path in temporal measure from regarding travel time as weight. However, the shortest path does not mean it is optimal and the fastest path is not dynamic optimal because it is based on static travel time. The shortest path in machine computation is always different from the optimal one in a human cognition because travel experience information of human is neglected.

The extensively applied algorithms of planar route planning are mainly Dijkstra's algorithm and A* algorithm. The former is the fundamental solution for route planning in vehicle navigation system. And to improve its performance, the various priority queues(including binary heap, fibonacci heap, bucket and etc) had been implemented into Dijkstra's algorithms. On the other hand, A* algorithm ,a heuristic method, that incorporate information from the problem domain into a formal mathematical theory of graph searching was proposed by Hart [1]. However, the planar route algorithms are not so effective when the scale of road network increases to a certain extent. Therefore, the hierarchical strategies have been proposed.

The conventional approaches of constructing the hierarchy of road network are based on the following idea. The road network is primarily considered as an original graph. The graph is divided into several partitions. Then, the higher level sub-graph is formed by all border nodes and shortest paths between these partitions. The two steps are recursively performed not until that the hierarchy of road network can be completely formed. For example, this idea can be found in [2], [3]. The HEPV model [3] was proposed by Ning Jing and Yun-Wu Huang. To achieve an effective fragmentation, they developed a partitioning algorithm called spatial partitioning which clusters graph links into partitions based on spatial proximity [4]. Spatial partitioning takes advantage of ITS map characteristics such as the grid-like (near-planar) patterns, and the relatively short distance for the majority of links. In HiTi graph model [2], arbitrary shaped boundaries (e.g., political regional boundaries) partition a road map into a set of Component Road Maps (CROM). The CROM can be defined to contain a set of CROMs, thus creating a multilevel hierarchy. The HiTi graph is a graph whose nodes are the boundary nodes of the CROMs and edges are the path view and cut connections of CROMs.

Based on the idea of tailoring the route searching space, the latest hierarchical algorithms [5] separate large road network to multi-level graph. Highway Hierarchy [6] method, recently proposed by Sander and Schultes, classify nodes and edges fully automatically in a preprocessing step in such a way that all shortest paths are preserved. By this means, the algorithm can be performed with exactness and high speed.

In this research, hierarchical travel experience information of road network is given by statistical analysis on a large amount of taxi GPS trajectories. According to taxi trajectories, the roads can be divided into frequent roads, secondary frequent roads and seldom roads. These trajectories can well reflect a hierarchical cognition of road network because considering taxi driver's cognition of the road network. With the hierarchical model of road network, the corresponding hierarchical route planning can be performed based on the experienced driver's cognition. Consequently, the route results according to the experiential information should approach to optimal results more than that according to static information of road network.

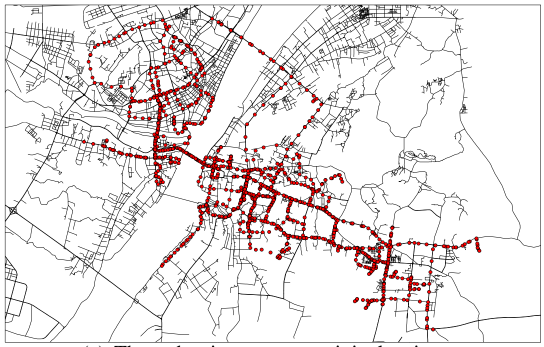
II. Experiential Hierarchy of road network

A taxi driver is a person acquainted with road network of a city. An optimal route can be obtained by statistical analysis with taxi GPS trajectories because these trajectories can reveal the taxi routes chosen by driver's experience. The experiential hierarchical network can be constructed by these taxi GPS trajectories. These points were firstly matched into road network and then the frequency of road travelled by taxi can be counted so that a hierarchy of road network was constructed by classification of the frequency.

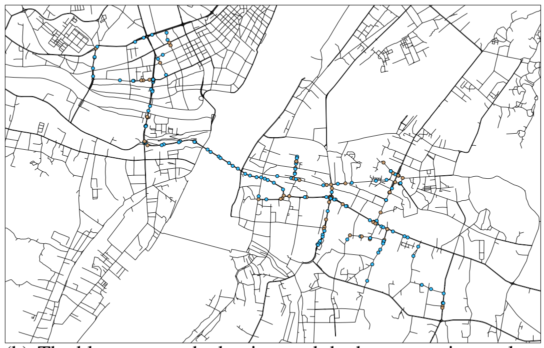
A. Taxi GPS Trajectories

Given a directed graph $G = \{V, E\}$, V is a set of intersections of roads and E is a set of road between two neighboring junctions. Let $T =< p_1, p_2, ..., p_k >$ be a GPS trajectory. A trajectory point p_i is matched into edge e_j of road network.The whole points of a trajectory are matched into $MT =< e_1, ..., e_j, ...e_l > (e_j \in E, 0 \leq l \leq k)$. Let $\Gamma = \{T\}$ be a set of the whole trajectories of one day and $\Psi = \{MT\}$ be a set of the results that all trajectories of Γ are matched into edges set E of road network G. The map-matched edge e_j and e_{j+1} isn't always neighbor in road network while the interval between the corresponding trajectory point p_i and p_{i+1} is regular. In this case, it is necessary to interpolate trajectory points between p_i and p_{i+1} in order that the corresponding matched edges can recover a route without losing any road travelled by the taxi. Therefore, the map matching process of taxi trajectories mainly consists of three steps.

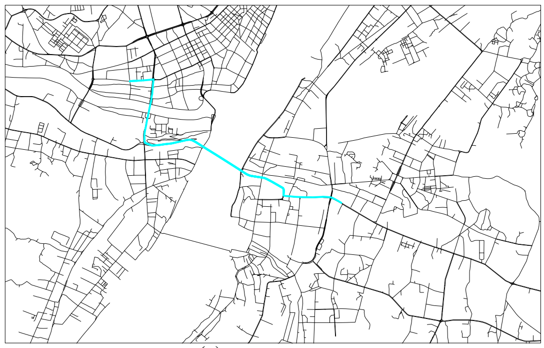
1) GPS trajectory points of taxi are matched into road.
2) Some points should be interpolated when two sequential roads into which trajectory points are matched are disconnected.
3) Then we recover a route which consists of a sequence of roads into which trajectory points are matched.



(a) The red points present original trajectory



(b) The blue are matched points and the brown are interpolated points between 7:00 AM and 9:00 AM



(c) recovery route

Fig. 1. A taxi trajectory of Wuhan city in one day

In figure 1(a), the red points present a taxi GPS trajectory of Wuhan city in one day. These trajectory points are matched into road and some points are interpolated into them according to connectivity between two sequential trajectory points. The blue points are matched points and the brow ones are interpolated points in figure 1(b). In figure 1(c), the blue route is a recovery one from roads into which the trajectory points are matched.

B. The road frequencies travelled by taxi

Let the function $g(e_i)$ be the travelled frequency of the edge e_i in Ψ by all trajectories in Γ. Let $Fn(g(e_i))$ be cumulative distribution function of $g(e_i)$. Let $0 < q_2 < q_1 < 1, Fn(g_1) = q_1, Fn(g_2) = q_2$. According to the value of $g(e_i)$ all roads are classified into three levels. The edge set in which their $g(e_i)$ values are above g_1 of all values is considered as the frequent roads, that between g_1 and g_2 is considered as the secondary frequent roads, and that below g_2 is considered as the seldom roads. The figure 2 is a cumulative distribution of road frequency by trajectories of 1400 taxis of Wuhan city in one day.

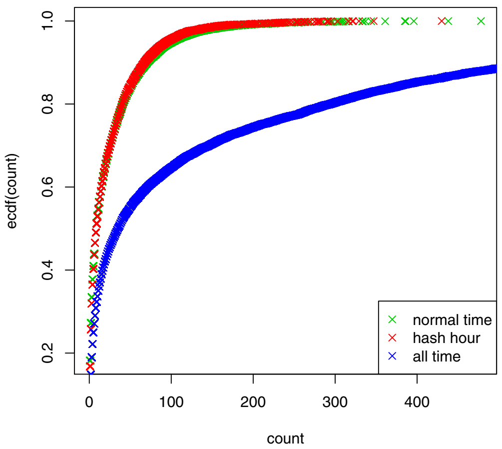


Fig. 2. trajectory frequency of road

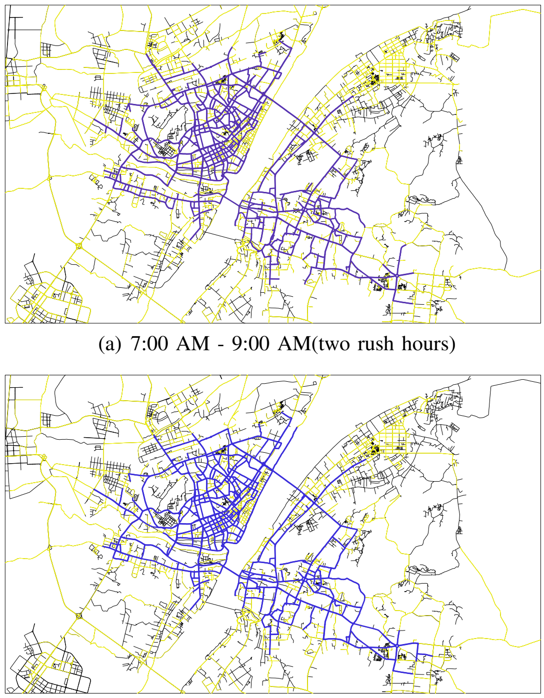
C. Experiential Hierarchy of road network

All roads are divided into three classes: frequent road,secondary frequent road and seldom road, according to statistical character with the frequencies of the routes recovered by taxis GPS trajectories. An experiential hierarchy of road network is constructed in terms of classification by frequency with which taxi routes travelled, as well as connectivity of every level is kept. The two steps that construct experiential hierarchical graph of road network is described in the following.

Step 1 *Hierarchy.* For each edge $e \in E$, we compare $g(e)$ with g_1 and g_2. The edge set E of road network is grouped into three subset, $E_f = \{e|g(e) \geq g_1\}$, $E_s = \{e|g_1 > g(e) \geq g_2\}$ and $E_n = \{e|g_2 > g(e) \geq 0\}$, where $E = E_f \cup E_s \cup E_n$. The three subset of E construct a hierarchy of road network. The hierarchical graph $H = \{G, H^0, H^1, H^2\}$ is a partition of graph $G = \{V, E\}$ according to frequency $g(e)$ of road. The bottom road network is $H^0 = G$. The level H^1 is constructed according to the edge set $E_f \cup E_s$, where $H^1 = \{V^1, E^1\}$, $E^1 = E_f \cup E_s$, $N^1 = \{v|(u_1, v) \in E^1 \cup (v, u_2) \in E^1\}$. So it is ensured that edge $e^1 \in E^1$ and two vertexes of it are also in H^1. In the same way, the level H^2 is constructed by E_f, where $H^2 = \{V^2, E^2\}$,$E^2 = E_f$,$N^2 = \{v|(v, u_1) \in E^2 \cup (u_2, v) \in E^2\}$. The sub-graph $H^i = \{V^i, E^i\}$ is the level i of graph G and it satisfies $G = H^0 \supset ... \supset H^i \supset ... \supset H^r$ for $0 \leq i \leq r$(in this case, $r = 3$). Let $v \in V^i$. And let $D_i(v)$ is defined as the degree of v in H^i.Then $V^i_{upper} = \{v \mid v \in V^i \cap V^{i+1}, D_i(v) \geq D_{i+1}(v)\}$. These nodes in V^i_{upper} are called as exit or entry nodes. They are not only connected to nodes in current level but also higher adjacent level. We record the node association with upper level $i+1$ in H^i. The association between level i and level $i+1$ is established by the node association in order that the hierarchical route planning can be done in the hierarchical road network.

Step 2 *Connectivity.* Let C^i be a maximal strongly connected subgraph of level H^i constructed by step 1. The vertexes and edges that don't belong to C^i in level H^i should be deleted. The C^2 is maximal SCC of top level H^2. The deletion of E^2/E_{C^2} and V^2/V_{C^2} from H^2 is firstly executed. The C^2 is a new H^2. Then the same operation is done on H^1. After deletion of these vertexes and edges that don't belong to C^i, the node association between neighboring levels should be upgraded because the change of edge set and vertex set had happened.

From the two steps, a hierarchical graph of road network is constructed, which is based on the experience of taxis. And a hierarchical route planning can be done on the hierarchical road network.



(a) 7:00 AM - 9:00 AM(two rush hours)

(b) 10:00 AM - 12:00 AM (two normal hours)

Fig. 3. Expierential hierarchy of WUHAN road network

Figure 3 shows hierarchical road network derived from GPS trajectories of 1400 taxis of Wuhan city. Figure 3(a) is hierarchy according to trajectories of 7:00 AM to 9:00 AM while figure 3(b) is one of 10:00 AM to 12:AM. In the two figures, the blue road present frequent roads, the yellow secondary frequent roads and the black seldom roads.

III. Hierarchical route planning with taxi frequency

According to above mentioned method, the experiential hierarchical road network can be consequently constructed by the taxi GPS trajectories. The connectivity of network is ensured so that the route planning can be performed by the following steps based on the hierarchy.

Let s be source point and t target point,

1) The forward search starts from s and the backward search from t in level H^i till they find nearest entry node v_s to upper level H^{i+1} and exit node v_t from H^{i+1}.
2) The bidirectional search subsequently continue from v'_s and v'_t, which are the corresponding nodes of v_s and v_t, regarded as a new pair of source and target nodes in upper level H^{i+1} .
3) If the forward search meets the backward at a node v_{meet}, the process of searching will be terminated.
4) If the search find entry or exit node of upper level of H^{i+1}, step 1 and step 2 will be recurred not until reaching the highest level.

The planned route HP is a sequence of paths $\langle P^0_s, ..., P^i_s, ..., P^k_{s,t}, ..., P^j_t, ..., P^0_t \rangle$ where

- $P^0_s = \langle s^0_s, ..., s^0_t \rangle$, $s^0_s, ..., s^0_t \in V^0$,
- $P^i_s = \langle s^i_s, ..., s^i_t \rangle$, $s^i_s, ..., s^i_t \in V^i$,
- $P^j_t = \langle t^j_s, ..., t^j_t \rangle$, $t^j_s, ..., t^j_t \in V^j$,
- $P^k_{s,t} = \langle s^k_s, ..., s^k_t \rangle = \langle t^k_s, ..., t^k_t \rangle$,
- $P^0_t = \langle t^0_s, ..., t^0_t \rangle$,
- $0 \leq i, j \leq k$.

The planned route HP should be connected path, so it must be $s^i_t = s^{i+1}_s$ and $t^{j+1}_t = t^j_s$ for $0 \leq i, j \leq k$. The nodes of hierarchical path given from forward searching are denoted by $s^i_{s,t}$ for $0 \leq i \leq r$. The nodes of hierarchical path given from backward searching are denoted by $t^j_{s,t}$ for $0 \leq j \leq r$. The route is planned on the principle that the frequent roads have a high priority of being chosen because it is based on the experiential hierarchy of taxi drivers according to their GPS trajectories.

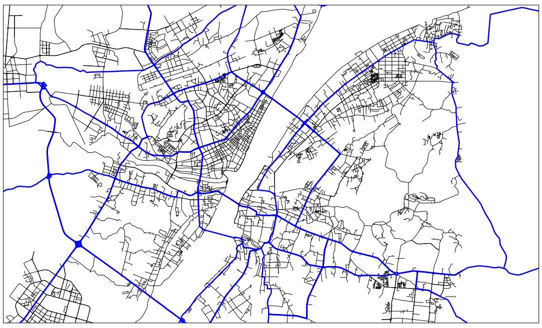
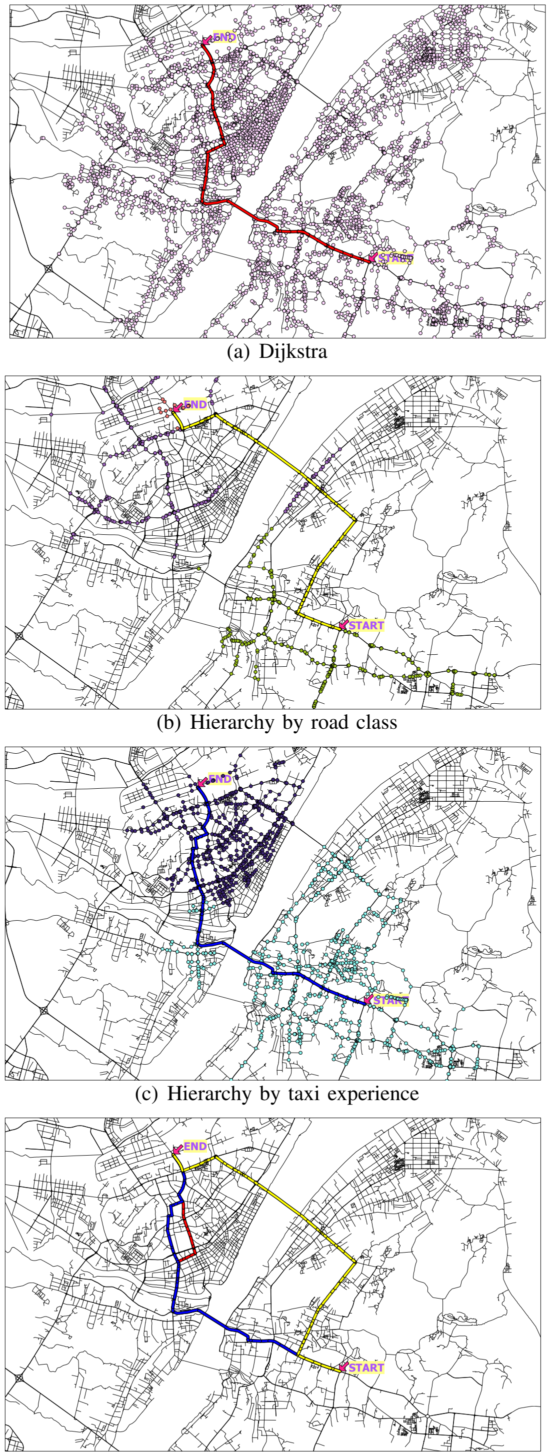


Fig. 4. hierarchy from road class



(a) Dijkstra

(b) Hierarchy by road class

(c) Hierarchy by taxi experience

(d) all

Fig. 5. route planning by three methods(first pair of table I)

IV. Experimental results and Conclusion

The experiments are implemented on a 32-bit PC with 512 MB main memory, using mono Intel Pentium 4 processor clocked at 2.4 GHz, running Windows XP (sp2). The programs are compiled by Visual C++ 6.0. We implement the hierarchical route planning algorithm on the road network of Wuhan which consists of 8446 nodes and 11148 edges. The GPS trajectories of 1400 taxis of Wuhan city in one day are chosen as an experiential data. From these trajectories, a hierarchical

TABLE I
EXPERIMENT RESULTS

NO	RP method	path length(m)	travelled time(s)	expanded node number	computing time(us)
1	flat(Dijkstra)	19096	2995	4347	35582
	Hierachical	20370	3140	757	5887
	Exp Hierarchical	19127	2852	1548	14900
2	flat(Dijkstra)	19096	2995	4347	36016
	Hierachical	20370	3140	757	5996
	Exp Hierarchical	19127	2852	1548	15200
3	flat(Dijkstra)	15187	1878	3801	30572
	Hierachical	15612	1958	488	3906
	Exp Hierarchical	15372	1813	1162	11202
4	flat(Dijkstra)	15254	1883	3820	35252
	Hierachical	16572	2210	536	3736
	Exp Hierarchical	15512	1871	1213	11880
5	flat(Dijkstra)	16585	2571	3715	30320
	Hierachical	18061	2789	620	4610
	Exp Hierarchical	17979	2507	1565	14236

road network based on taxi experience is constructed and time interval of these trajectories is chosen between 7:00 AM and 9:00 AM (two rush hours in morning). We perform three programs which respectively are Dijkstra algorithm based on original graph, hierarchical route planning algorithm based on hierarchical graph by road class and hierarchical route planning based on hierarchical graph by taxi experience. The weight of each edge is the length of corresponding road. Table I shows the results given by the three algorithms mentioned above, which are performed by random choice of five source and target pairs.The travelled time of route is estimated by speed of road taxi travelled according to GPS trajectories.

The first source target pair of table I is illustrated in figure 5. In figure 5(a), the red is a route planned by Dijkstra algorithm running on flat graph of road network and the colored nodes are expanded nodes in the process of route searching. In figure 5(b), the yellow is a route planned by hierarchical route planning algorithm based on hierarchical graph by road class and the colored nodes are expanded nodes in the process of route searching. In figure 5(c), the blue is a route planned by hierarchical route planning algorithm based on hierarchical graph by taxi experience and the colored nodes are expanded nodes in the process of route searching. The routes planned by the three methods are showed in figure 5(d).

The planning method by taxi experience give priority to frequently travelled roads by taxi. Therefore, the planning method by experience choose these frequent roads prior to others. However, the method by road class firstly finds optimal route in high class roads and the one by Dijkstra' algorithm finds optimal route in the whole road network. The results from figure 5 and the first row of table I indicate the path planned by taxi experience is an optimal one in the measure of travelled time. From the table I, it shows that the method by Dijkstra's algorithm computes the shortest path in the length of network distance but its travelled time isn't optimal and that the method by taxi experience can give a route, which is shorter in distance than the one according to road class as well as in travelled time than Dijkstra's algorithm.

Consequently, based on experiential hierarchical road network by taxi GPS trajectories, a hierarchical route planning is proposed according to experience of the taxi driver. In this paper, we give the detailed steps to construct experiential hierarchical road network by trajectories and to perform the corresponding route planning method. The experimental results from three different methods are given, compared and analyzed. An optimal route can obtained from the method we proposed because it performed based on the principle that the roads taxi frequently travelled have a priority to be chosen in the process of finding path.

ACKNOWLEDGMENT

Work described in this paper was jointly supported by the projects from NSFC (No. 4083050, No. 60872132, No.40871185), 863 program (No. 2007AA12Z41), and project from Ministry of Education of China (No. 108085).

REFERENCES

[1] P.E. Hart, N.J. Nilsson, and B. Raphael. A formal basis for the heuristic determination of minimum cost paths. *Systems Science and Cybernetics, IEEE Transactions on*, 4(2):100–107, July 1968.

[2] Sungwon Jung and Sakti Pramanik. An efficient path computation model for hierarchically structured topographical road maps. *IEEE Transactions on Knowledge and Data Engineering*, 14(5):1029–1046, 2002.

[3] Ning Jing, Yun-Wu Huang, and Elke A. Rundensteiner. Hierarchical encoded path views for path query processing: An optimal model and its performance evaluation. *IEEE Trans. Knowledge and Data Engineering*, 10(3), 1998.

[4] Yun-Wu Huang, Ning Jing, and Elke A. Rundensteiner. Effective graph clustering for path queries in digital map databases. In *CIKM*, pages 215–222, 1996.

[5] Martin Holzer, Frank Schulz, and Dorothea Wagner. Engineering multilevel overlay graphs for shortest-path queries. *J. Exp. Algorithmics*, 13:2.5–2.26, 2009.

[6] Peter Sanders and Dominik Schultes. Engineering highway hierarchies. In *ESA'06: Proceedings of the 14th conference on Annual European Symposium*, pages 804–816, London, UK, 2006. Springer-Verlag.