Improving Vehicle Speed Estimates Using Street Network Centrality

Anita Graser, Maximilian Leodolter, Hannes Koller, Norbert Bründle
AIT Austrian Institute of Technology, Vienna, Austria

ARTICLE HISTORY
Compiled March 1, 2017

ABSTRACT
This paper describes a novel approach to improve prediction models which estimate vehicle speeds and their diurnal variation for road network links in urban street networks using only static map attributes. The presented approach takes into account previously neglected spatial information by integrating network centrality measures for closeness (indicating how central a link is) and betweenness (indicating how important a road link is) into the prediction model. The model is calibrated with a real-world dataset of 100 million individual speed measurements from a fleet of 3,500 taxi probe vehicles in Vienna, Austria. Given that centrality can be derived directly from readily available street network data, the experimental results demonstrate that integrating centrality measures considerably improves the predictions without the need for adding a supplementary data source. Improvements for vehicle speed estimates are particularly prevalent on important street network links in the city center as well as rural streets in the periphery.

KEYWORDS
mobility; network centrality; traffic speed; travel time; prediction models

1. Introduction

Information about vehicle speeds and thus expected travel times in a street network is fundamental for many mobility applications and for applications evaluating effects of reachability. Speed estimates along street network links are, for example, crucial for vehicle routing applications and route travel time estimation, especially in congested networks Musolino, Polimeni, Rindone, and Vitetta (2013). Given that travel speeds in urban areas typically vary throughout the day – with usually lower speeds during peak travel hours – diurnal variation of travel speed should be captured and represented in a time series for every street network link in order to provide accurate information. Such time series of vehicle speed data may be captured with stationary traffic detectors such as loop detectors or from a fleet of probe vehicles (often taxis) reporting their position Jones, Geng, Nikovski, and Hirata (2013); Simroth and Zhie (2011). Since installation and/or maintenance costs for sensors are usually high, by far not all links of a street network can be captured. Similarly, vehicle speed information from probe vehicles may be very sparse on non-arterial streets, or completely absent when information of floating car data (FCD) is not exploited in an area. The main motivation for the development of travel time prediction models therefore is in estimating travel speeds.
and their diurnal variations for streets (within the same area or in different areas) where no sensor or FCD sources are available.

A common strategy for compensating lacking vehicle speed information for a street network link is to resort to estimated free-flow link speed, which is usually inferred from static parameters of the street network such as street category or legal speed limit and spacing between signalized intersections Dixon, Wu, Sarasua, and Daniel (1999); Moses and Mtoi (2013); Transportation Research Board (2010). Inferring information from such static information inevitably leads to single free-flow link speeds, rather than time series. A model estimating diurnal travel speed variations with 15-minutes granularity for the street network of the city of Vienna, as well as surrounding rural areas (depicted in Figure 1) is presented in Leodolter, Koller, and Straub (2015). This model estimates vehicle speeds on network links based on data from links sharing the same street category and speed limit. Grouping all street links based on such simple features, however, ignores spatial information, such as differences between street network links in urban city centers and more peripheral streets in the forested foothills in the west or flat agriculturally-used plains in the east. This hampers accurate predictions, as shown in Figure 2, which illustrates the spatial distribution of the prediction error of Leodolter et al. (2015) in the analysis area. This visualization shows that errors are not distributed randomly in space. Instead, the model tends to perform better on minor streets in the city center and on highways, while centers of towns outside of Vienna and rural streets connecting towns and villages show particularly bad results with mean absolute percentage error (MAPE) exceeding 40%.

Figure 1. Analysis area (red box) encompassing Vienna and surrounding areas; map © OpenStreetMap contributors

By combining street geographical factors and environmental factors (e.g. number of crosswalks in an area) with traffic flow properties from traffic detectors, Oh, Kim, and Hong (2015) predict traffic speed, traffic volume and density for a given time. In this paper, we propose an alternative approach which does not rely on information about environmental factors, but is able to exploit spatial information from the network itself via network centrality measures. In general, network centrality indicates which nodes of a network occupy critical positions Freeman (1979). In geography and geographic information science, centrality measures such as closeness, betweenness, straightness, and information centrality have been applied to urban street networks in
order to study city structure Crucitti, Latora, and Porta (2006) or to explain effects such as land use intensity Wang, Antipova, and Porta (2011) and retail and service activity Porta et al. (2009). Similarly, PageRank and geographically modified PageRank algorithms (Distance-Decay PageRank and Geographical PageRank) have been used by Jiang (2009a) and Chin and Wen (2015), respectively, to explain the concentration of human movement in street networks. In the context of motorized traffic, betweenness centrality already served as an indicator to predict traffic flows. For example, Jiang (2009b) uses taxi FCD to show that street hierarchies derived from street length, connectivity, and betweenness are a good indicator for traffic flow. Similarly, Huang, Zhu, Ye, Guo, and Wang (2015) find that daily and hourly traffic flow conforms well with street hierarchies derived from degree and betweenness centrality. On a different note, Kazerani and Winter (2009) discuss the integration of origin-destination matrices into betweenness centrality computations to address the non-uniform distribution of travel demand. Similarly, Puzis et al. (2013) present a betweenness-driven traffic assignment model which takes into account travel demand and model travel times and Gao, Wang, Gao, and Liu (2013) combine betweenness with travel demand data and geographical constraints to predict traffic flow.

While these existing research contributions provide travel demand predictions using street networks centrality, to the best of our knowledge, there currently exists no work using centrality measures to predict travel times and their diurnal variation. Our paper closes this gap by extending the work of Leodolter et al. (2015) with closeness and
betweenness centrality. Our approach is motivated by the following two hypotheses: The first hypothesis is that closeness centrality helps improve our model by providing a means to distinguish between central urban and peripheral rural streets which are expected to exhibit different vehicle speed patterns. The second hypothesis is that betweenness centrality provides a means to identify important links in the street network which are likely to attract a lot of traffic and therefore show vehicle speed patterns which vary from those found on less important links.

Section 2 provides a formal description of different network centrality measures evaluated in this study and discusses their suitability for the use case of predicting vehicle speeds. Section 3 then introduces the prediction model before Section 4 presents experimental results evaluating the prediction quality of our approach. Finally, Section 5 discusses potential future expansions and avenues for further research.

2. Street Network Centrality

In general, network centrality indicates which nodes of a network occupy critical positions Freeman (1979). This helps identify key network elements and detect structural differences. Networks can be represented geometrically, using the geometric locations of junctions and lengths of street segments, or topologically without locations and distances. For this work, a geometric representation of the street network was chosen since geometric link lengths and corresponding travel times provide highly relevant information for the computation of weighted centrality values and thus for our vehicle speed estimations. In the following, closeness and betweenness centrality are presented in detail.

Given a network \( \mathcal{N}(N, L) \) composed of sets of nodes \( N \) and links \( L \), closeness centrality measures how close a node is to all other nodes in the network. More formally, the closeness of a node \( n \) in the network \( \mathcal{N}(N, L) \) is defined as

\[
C_{\text{node}}(n) = \frac{1}{\sum_{m \in N \setminus \{n\}} d(n, m)},
\]

where \( d(n, m) \) is the sum of the weights of the links of the shortest path connecting the nodes. Link weights can be for example the link length in meters, or travel time.

Betweenness centrality is based on the concept that a node or link is important or central if it is traversed by a large number of shortest paths connecting all pairs of nodes in the network. Betweenness of a node \( n \in N \) is defined as

\[
B_{\text{node}}(n) = \sum_{m \neq o \in N \setminus \{n\}} \frac{\sigma_{mo}(n)}{\sigma_{mo}},
\]

where \( \sigma_{mo} \) is the number of all shortest paths connecting the nodes \( m \) and \( o \), and \( \sigma_{mo}(n) \) are those traversing node \( n \). It is worth noting that the betweenness computation for a node \( n \) excludes routes starting or ending in \( n \).

While centrality is typically computed for the nodes of a network, estimation of vehicle speeds requires centrality values for links, rather than nodes. We use link closeness to distinguish between streets in the network center and streets at the network periphery, whereas link betweenness is a potential indicator for street network links which are likely to attract much traffic due to their important connecting role.
within the network. One approach to compute link centrality values found in the literature is to compute centrality for the dual graph representation of the street network Gao et al. (2013); Huang et al. (2015); Porta, Crucitti, and Latora (2006a, 2006b). In the dual graph of a street network, streets are represented as nodes and intersections are turned into links. Evaluations presented by Porta et al. (2006b) show that this dual graph approach leads to more “abstract” results, since computations are based on topological distances instead of metric distances and they recommend to use the primal graph representation (with streets as links and intersections as nodes) instead. Therefore, we propose to continue to use the primal graph and to compute closeness values for network links as the average of the closeness values of the corresponding start and end nodes. This node-based closeness for a link \( l \) connecting nodes \( n \) and \( m \) is defined as

\[
\tilde{C}_{\text{link}}(l_{n,m}) = \frac{C_{\text{node}}(n) + C_{\text{node}}(m)}{2}.
\] (3)

In Wang et al. (2011), the same approach is used to compute link betweenness values by averaging node betweenness. This node-based betweenness value for a link \( l \) connecting nodes \( n \) and \( m \) is defined as

\[
\tilde{B}_{\text{link}}(l_{n,m}) = \frac{B_{\text{node}}(n) + B_{\text{node}}(m)}{2}.
\] (4)

While the closeness values of a link’s start and end nodes are a good representative of how close the link is to the rest of the network, the betweenness values of its start and end nodes can differ significantly (for example because the start node is at an important intersection with a high betweenness while the end node is at an intersection with low betweenness) and the average of these values is not necessarily a good representative of the link’s importance within the network. Therefore, an alternative approach is to define the link-based betweenness of a link \( l \) as the number of shortest path routes traversing \( l \) between all nodes, relative to all shortest paths, or more formally

\[
B_{\text{link}}(l) = \sum_{n \neq m \in N} \frac{\sigma_{nm}(l)}{\sigma_{nm}}.
\] (5)

Figure 3 illustrates the difference between \( \tilde{B}_{\text{link}} \) (4) and \( B_{\text{link}} \) (5) using an example network where all link weights are 1 except for the link from node \( c \) to node \( h \) in the center of the network which has a weight of 2. Due to its higher weight, the link \( l_{c,h} \) is avoided by the shortest path routing. This results in a correspondingly low link-based betweenness value, \( B_{\text{link}}(l_{c,h}) = 1 \). The node-based betweenness of \( l_{c,h} \), \( \tilde{B}_{\text{link}}(l_{c,h}) = 8 \), is much higher since both \( c \) and \( h \) have high betweenness values \( B_{\text{node}}(c) = B_{\text{node}}(h) = 8 \).

Figure 4 shows the effects of the different betweenness formulas (4 and 5) on a real-world street network: important streets (drawn using wider lines) bias node-based betweenness values of adjacent, less important streets. This effect can be seen clearly in Figure 4(a) where links connecting to major streets are shown as much more important than in Figure 4(b). We therefore propose to use link-based betweenness instead, since it captures the importance of the link itself rather than the importance of its start and end node, which may differ considerably.

A common issue with calculating centrality measurements for real world networks...
Figure 3. Link-based (a) and node-based (b) betweenness for a network where all links have a weight of 1, except for $l_{c,h}$ which has a weight of 2. Numerical values show the corresponding betweenness values, and node labels in (b) show $B^{\text{node}}$.

Figure 4. Real-world examples of node-based (a) and link-based (b) betweenness in the city of Vienna.
is the border effect or edge effect, that is the distortion that lowers centrality values near the border of the network. These distortions are significant for centrality values, particularly in fragmented networks. One solution to overcome this effect is to use local centrality measures rather than global centrality measures which we have discussed so far. Besides addressing the edge effect, local centrality measures also reveal network properties on a local scale which are not captured by global measures. Porta et al. (2009)

For local centrality measures, the calculations are restricted to the neighborhood of the respective node or link (which is a subset of the whole network) instead of using the whole network. The neighborhood is determined by a cut-off parameter which represents the maximum path length (or generalized path cost) which are considered for the calculations.

Figure 5 illustrates the differences between local and global betweenness in an example network where all link weights equal 1 and the local cut-off parameter in this example is 1.5. Given this cut-off value, only connections with a maximum length of two links are considered for the local calculations. For this reason, the central links $l_{a,f}$, $l_{f,g}$, $l_{g,h}$, $l_{a,h}$ all have the same local betweenness value while the global betweenness is higher for $l_{a,f}$ and $l_{f,g}$. According to (5) the local betweenness value for $l_{a,f}$ is calculated as follows: the connections a-e and a-f contribute both a value of 1, because in both cases there exists only one shortest path, and it traverses $l_{a,f}$; further b-f and a-g contribute 0.5, since two shortest paths exist for these connections. Other connections in this example network either don’t traverse $l_{a,f}$, or their path length exceeds the maximum path length at this cut-off parameter.

In our case, link weights are defined as travel time based on link length in meters and speed limit. Figures 6(a)-6(d) provide an overview of global and local closeness and betweenness centrality calculated for the Vienna city region using street network data from OpenStreetMap (OSM) and a local centrality cut-off value of 2 minutes travel time. In the first row, Figures 6(a)-6(b) show central links – with lower closeness values – in darker and peripheral links in lighter shades. Compared to global closeness, the local closeness measure provides more details about the smaller town centers surrounding Vienna (such as Wolkersdorf in the northeast and Mödling in the south). In the second row, Figures 6(c)-6(d) show more important links – with higher betweenness values – in darker and less important links in lighter shades. The distribution of local betweenness values shows more similarities with the closeness measures.
than global betweenness does. This is due to the fact that there are more routes within the local cut-off limit in dense urban areas than in open areas between towns. Moreover, arterial rural street links have a lower local betweenness than global, because the network within the cut-off area includes fewer routes traversing these links.

3. Regression Model for Vehicle Speed Estimation

The Ordinary Least Square (OLS) linear regression model for estimating vehicle speeds in 15-minutes intervals presented in Leodolter et al. (2015) (from here on referred to as “base model”) is extended to include closeness and betweenness centrality measures as well as their product. This extension addresses shortcomings of the base model approach which ignores spatial network aspects and only uses street category, legal speed limit, and time of day as explanatory variables. In order to obtain least square estimates for the model coefficients, the vehicle speed data is regressed onto individual street links and their attributes, and as extension in this work additionally on betweenness and closeness of the respective links.

Preliminary correlation analysis of speed records, speed limits, and calculated centrality values showed diverging results for different types of streets. Therefore, speed limit and centrality coefficients are modeled separately for four different street cate-
Table 1. Mapping of OSM highway tag values to street categories and number of corresponding network links.

<table>
<thead>
<tr>
<th>Street category</th>
<th>OSM highway tag</th>
<th>no. links</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>motorway, motorway_link, trunk, trunk_link</td>
<td>1765</td>
</tr>
<tr>
<td>2</td>
<td>primary, primary_link</td>
<td>5433</td>
</tr>
<tr>
<td>3</td>
<td>secondary, secondary_link, tertiary, tertiary_link, unclassified</td>
<td>11617</td>
</tr>
<tr>
<td>4</td>
<td>road, residential, living_street, track, service</td>
<td>20700</td>
</tr>
</tbody>
</table>

gories (for details about separate modeling see equation (9)) which are defined based on the input data OSM highway tags as shown in Table 1. In order to facilitate the interpretation of the prediction model, the centrality values are normalized and mapped to a value range between 0 and 1. (Original centrality values in this network of about 99,000 links take values of up to $5 \cdot 10^8$)

In general, a regression model investigates a supposed functional relation between a describing, or dependent variable $X$, and an independent variable $Y$.

$$Y = f(X) + \epsilon.$$ \hspace{1cm} (6)

To learn this functional relation, we need to observe $X$ and $Y$. Once this relation is understood and quantified, we can calculate an estimator for a missing or sparsely observed $Y$ in situations where $X$ is known. Thus, we can calculate speed estimates for street links, where only few, or no measurements were recorded.

In this work, as in Leodolter et al. (2015), a multiple least square linear regression model is applied to model the relation between the vector of FCD speed records $Y = (y_1 \ldots y_I) \in \mathbb{R}^I$ and the describing design matrix $X \in \mathbb{R}^{I \times J}$ by

$$Y = X \beta + \epsilon.$$ \hspace{1cm} (7)

The main objective of an OLS regression model is to estimate the vector $\beta \in \mathbb{R}^J$, which describes the functional relation between $X$ and $Y$. The novelty of this work, compared to Leodolter et al. (2015), is to add closeness and betweenness variables in the model’s design matrix $X$, that consists of the variables (represented as columns in the matrix): speed limit ($s$), daytime ($t$), street category ($\gamma$), betweenness ($b$), closeness ($c$) and the product of betweenness and closeness ($bc$), for modeling the interaction of the centralities (vehicle speeds differ for important links in the city center and outskirts of the city). Each row $i$ of $X$ describes the situation in which one single $y_i$, $i \in (1 \ldots I)$ was measured by the means of the independent variables. The term $\epsilon$ is the error term, modeled as an independent and identically distributed (i.i.d) Gauss distributed random variable, describing the unobserved part of $Y$. Daytime and street category are categorical variables and are therefore represented as design variables, taking values of 0 or 1 (indicating the 15 minute interval or the category, respectively).

The well known formula for the OLS estimation of $\beta$ is

$$\hat{\beta} = (X^T X)^{-1} X^T Y.$$ \hspace{1cm} (8)

This estimated $\hat{\beta}$ is the stacked vector of all the $\beta_t$ (one coefficient for each 15 minute interval) and $\beta_{t, \gamma,x}$, where $x \in \{b, c, bc, s\}$ and $\gamma \in \{1, 2, 3, 4\}$.

Finally, the speed estimate $\hat{y} \in \mathbb{R}$, for a given time interval $t$ and street category $\gamma$, can be calculated as a linear function of the speed limit, closeness, and betweenness depending on the time and street category specific coefficients:
\[ \hat{y}_{t,\gamma}(s, b, c) = \beta_t + \beta_{\gamma,s} \cdot s + \beta_{\gamma,b} \cdot b + \beta_{\gamma,c} \cdot c + \beta_{\gamma,bc} \cdot b \cdot c. \] (9)

The centrality values included in the regression model are \( \check{C}_{\text{link}} \) and \( B_{\text{link}} \) (see equations 3 and 5). Different model approaches with global and local centralities – with varying cut-off parameters – have been applied and will be discussed in the following Section 4.

The coefficients \( \beta_{\gamma,x} \) in Equation (9) are estimated separately for each street category \( \gamma \), given that preliminary analysis showed that the effects of different centrality values vary by street category. For example, Figure 7 shows the different distribution of speed estimates over the range of closeness and betweenness values for categories 3 and 4 with the same speed limit: while, in both cases, speeds are higher on peripheral streets (with closeness values near 0), the trends over betweenness point in different directions.

![Figure 7.](image)

This section presents the results which we achieved for the region of Vienna, Austria. We first introduce the data used to calibrate and evaluate the model and discuss the weaknesses of the base model which are addressed by our network centrality expansion. Then, we provide the results of the model with global centrality measures and discuss the achieved improvements and identified shortcomings. Finally, we provide the results of the model with local centrality measures and compare it to previous results.

The coefficients of the regression model were calibrated according to (9) using a large, real-life FCD dataset with about 100 million individual speed measurements for the city of Vienna which has an OSM street network size of about 99,000 directed links. This dataset represents one year’s worth of FCD collected on workdays (Monday to Friday) by about 3,500 taxis (for details see Leodolter et al. (2015)) which were
preprocessed using the FCD software FLEET (Fleet Logistics Service Enhancement with EGNOS and Galileo Satellite) Toplak, Koller, Dragaschnig, Bauer, and Asamer (2010). This includes the following steps: First, the raw GPS measurements are projected onto the street network graph using map matching techniques described in Koller, Widhalm, Dragaschnig, and Graser (2015). Then, vehicle speed measurements are obtained by analyzing the routes between consecutive projected positions. Finally, outliers are removed using a number of heuristics, such as, rejecting unrealistically high speeds. For model calibration, the data is aggregated into a design matrix $X$ with about $2.7 \cdot 10^6$ rows, by collecting all measurements of a link $l$ for the same time interval $t$.

As a first result, Figure 8 shows the diurnal distribution of all FCD records (boxplots) and resulting speed estimates for three different combinations of centrality values on links of street category 3 and speed limit 50 kph: unimportant rural streets are shown in red ($B, C \in [0, 1/3]$), streets with average centrality and betweenness in green ($B, C \in [1/3, 2/3]$), and important central streets in blue ($B, C \in [2/3, 1]$). The clear offset between the resulting estimates indicates that centrality enables us to distinguish between streets with different observed speed distributions.

![Boxplots of speed records and model estimates](image)

**Figure 8.** Diurnal variation of speed estimates on workdays for three different local betweenness and closeness values for one street category and speed limit.

To evaluate the performance of different model formulations, we chose the mean absolute percentage error (MAPE) which is defined as

$$MAPE = \frac{1}{I} \sum_{i \in I} \left| \frac{y_i - \hat{y}_i}{y_i} \right|,$$

based on disjunct learning and testing data subsets. For this purpose, the coefficient vector $\hat{\beta}$, which is used to calculate $\hat{y}$, is estimated based on a subset of the data $Y_j$ and $X_{j,k}$, such that $j \in J$ and $I \cap J = \emptyset$. Both, $I$ and $J$ are chosen randomly and of equal size ($10^5$), and repeated 1000 times. These 1000 MAPEs are averaged to one MAPE, which is then used in further analyses.

We calculate centrality using the *igraph* package of the popular statistical computing language R Csardi and Nepusz (2006). This package provides functionality to compute global centrality measures as well as local measures where computation is constrained using a cut-off parameter.

The base model without centrality results in a MAPE of 23.6% for the whole network.
Table 2. Comparison of model errors where relative MAPE changes show difference between base and global, global and local, and local and speed limit, respectively.

<table>
<thead>
<tr>
<th></th>
<th>Base model</th>
<th>Global cent.</th>
<th>Local cent.</th>
<th>Speed limit</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Whole network</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>23.6%</td>
<td>21.6%</td>
<td>20.4%</td>
<td>56.2%</td>
</tr>
<tr>
<td>MAPE change relative to base</td>
<td>-8.6%</td>
<td>-13.9%</td>
<td>+137.9%</td>
<td></td>
</tr>
<tr>
<td>MAPE change relative to global</td>
<td>-5.7%</td>
<td>+160.4%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE change relative to local</td>
<td></td>
<td>+176.2%</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>City center</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE</td>
<td>25.1%</td>
<td>20.4%</td>
<td>18.0%</td>
<td>71.3%</td>
</tr>
<tr>
<td>MAPE change relative to base</td>
<td>-18.8%</td>
<td>-28.0%</td>
<td>+184.1%</td>
<td></td>
</tr>
<tr>
<td>MAPE change relative to global</td>
<td>-11.4%</td>
<td>+249.9%</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MAPE change relative to local</td>
<td></td>
<td>+294.8%</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

or 25.1% for the central city districts. Figure 2 shows the spatial distribution of the base model MAPE values within the street network. The prevailing patterns are: good results on motorways and inner city residential streets, large errors (over 20%) on rural streets which are not motorways, similarly large errors on busy streets in the Vienna city center, and finally, very big errors (often 50% and more) in towns outside Vienna. Table 2 provides a first overview of the observed errors and error changes for different models which we discuss in detail next. For reference, the last column shows how our model performs compared to the common fallback solution using only speed limits to estimate vehicle speeds.

The expansion of the model to include global centrality values reduces the error by a relative 8.6% down to a MAPE of 21.6% for the whole network, and reduces the error by a relative 18.8% down to 20.4% for the central city districts. The spatial distribution of errors in Figure 9 shows improvements especially for important network links in the city center – such as arterials and bridges – as well as rural streets in the periphery. However, for some links the extended model with global centrality performs worse. This particularly affects links in towns outside of Vienna.

To address the issue of bad results for links in towns outside of Vienna, we propose to use local centrality measures which capture network properties on a local scale. Figure 10 shows the spatial distribution of improvements gained by switching from global centrality measures to their local equivalents. This comparison shows that it is possible to reduce the errors on links in towns outside of Vienna as well as achieve further improvements in the Vienna city center.

Including local centrality in the model further reduces the error by about 5.7% down to a MAPE of 20.4% for the whole network (and by 11.4% down to 18% for the central city districts) compared to the model using global centrality. This corresponds to an improvement of 13.9% compared to the base model. Figure 11 shows a comparison of the errors of the base model (Figure 11(a)) and the model with local centrality (Figure 11(b)) in the center of Vienna, where errors were reduced by 28%.

For the local centrality, we also evaluated different cut-off settings ranging between 0.5 and 10 minutes travel time. The blue lines in Figure 12 show the errors using different cut-off settings, with the smallest error being observed for a cut-off of 2 minutes. If shorter or longer cut-off values are chosen, the observed error approaches the error of the model with global centrality. Figure 12 furthermore shows how the model error relates to the characteristics of the data sample used to train the model. All errors reported so far correspond to a model trained with a cut-off parameter of 2 minutes and all available data – even if there was only a single speed record available for a certain link and time (minimum measurement count, \(mmc = 1\)). By requiring
Figure 9. MAPE difference between the base model and the model with global centrality: green links show reduced errors and pink links show increased errors.

better statistics, the model error can be reduced from 20.4% to 15.8% (for mmc = 3) or 14.2% (for mmc = 30). For practical applications, it is of course necessary to find a satisfactory balance between data availability and statistical robustness.

5. Conclusion and Outlook

We showed that adding centrality measures to our model – which predicts vehicle speeds for street network links depending on the time of the day using only static street network data as explanatory variables – considerably improves the model results without having to add new data sources since centrality can be derived directly from readily available street network data. Our expansion using street network centrality was motivated by two hypotheses: The first hypothesis was that closeness centrality would help distinguish between central urban and peripheral rural streets. The second hypothesis was that betweenness centrality would help identify important links in the street network. Our results confirm these hypotheses since improvements could be achieved especially for important network links in the city center as well as rural streets in the periphery. The proposed expansion improves the vehicle speed model introduced in Leodolter et al. (2015), which was developed to calculate speed estimates for street links, where only few, or no measurements where recorded. Compared to common fallback methods, such as speed limit as speed estimator, which performs much worse
with a MAPE value of 56.2%, our proposed model yields a MAPE of only 20.4%. We furthermore evaluated the use of both global and local centrality measures for this application and our results show that local centrality is a better predictor for speed and should therefore be preferred. In particular, local centrality helped reduce the prediction errors on links in smaller towns surrounding the city of Vienna.

It is worth noting that there are numerous conceivable expansions and alterations to the proposed model. In particular, the use of additional data sources, such as landuse or administrative data, could potentially improve model results. For example, landuse data could be used to determine whether a link is inside or outside of a village or town. This information could enhance or even replace closeness centrality in our model. The disadvantages of adding further data sources should not be ignored though: each data source adds a new source of uncertainty (for example, about the consistency of the landuse classification) and reduces transferability of the developed model to different areas where certain data sources might not be available in the same quality. One of the key advantages of our model formulation is its minimal data requirements, which makes it applicable worldwide.

For the future, one of our goals is to evaluate whether the integration of additional scores, such as geographically modified PageRank or different network centrality measures – particularly straightness – would further improve the model. Furthermore, we are planning to investigate whether models like the one presented in this paper can be transferred from one city or region to a similar one. This would enable us to predict...
travel times in networks were speed measurements are not available at all or at least not in sufficient quantity.

Acknowledgement(s)

The authors would like to thank the anonymous reviewers for their invaluable feedback, which helped improve the initial manuscript, as well as Taxi 31300 for providing taxi FCD in Vienna. An earlier abstract version of this paper was published in the proceedings of the 1st ICA European Symposium on Cartography under the title “Towards Better Urban Travel Time Estimates Using Street Network Centrality”.

Funding

This work was partially funded by the Joint Programming Initiative Urban Europe through the Austrian Ministry for Transport, Innovation and Technology (bmvit) under grant number 847350 (project e4share).

References


Figure 12. MAPE of the different models for varying minimum values of measurement counts per link and time interval (from 1 to 30 measurements) and different local centrality cut-off values (from 0.5 to 10 minutes).


