Modeling simplicity in drivers' route choice behavior

Martí Montesinos
Nicola Ortelli
Evanthia Kazagli
Leonardo Bellocchi
Nikolas Geroliminis

Conference paper STRC 2017

STRC | **17th Swiss Transport Research Conference**
Monte Verità / Ascona, May 17-19 2017

Modeling simplicity in drivers' route choice behavior

Martí Montesinos, Nicola Ortelli, , Evanthia Kazagli, Leonardo Bellocchi, Nikolas Geroliminis
LUTS & TRANSP-OR
Ecole Polytechnique Fédérale de Lausanne
Station 18
CH-1015 Lausanne
phone:
fax:
{marti.montesinosferrer,nicola.ortelli,evanthia.kazagli,leonardo.bellocchi,nikolas.geroliminis}@epfl.ch

May 2017

Abstract

The shortest path approach is commonly used in traffic simulators and route choice models, as well as route guidance systems, to represent drivers' behavior and, accordingly, to provide guidance back to them. The decisions of drivers though, are not merely based on such efficiency criteria. The simplicity of a way-finding, or a navigation, task is arguably an important factor. State-of-the art definitions of path simplicity are mainly based on geometric criteria. This paper presents a definition for the simplicity of a path that incorporates the type of road and driver's perception. The definition of simplicity is further exploited to develop a criterion allowing analysts to quantify the trade-offs between the shortest and the simplest path. An efficient algorithm is proposed to compute the simplest paths between any pair of nodes in the network. It extends the definition of simplicity proposed by Viala et al. (2013), by considering the road type additionally. Real data is used to test and demonstrate the validity of the approach. The impact of simplicity on route choice is identified in drivers observed choices. The chosen alternatives often lie in between the shortest and the simplest path.

Keywords

simplicity – route choice – networks – algorithms

1. Introduction

If one observes the trajectories of commuters in an urban network, what can be easily seen is that the chosen paths are not generally the most trivial or shortest possible. This is due to the fact that path choice is influenced by many factors, such as trip attributes and travelers' characteristics.

However, for many applications, such as traffic simulators, route choice models or route guidance systems, the shortest path – in terms of distance or time – approach is widely used. This approach may lead to simulated paths which do not describe accurately the behavior of the drivers, especially when travel distance is used instead of travel time – due to limited data availability, for example. The simplicity of a navigation task is arguably another important factor. In this study, the aim is to investigate if drivers are more inclined to choose the shortest or the simplest path and assess the issues of modelling the decisions of drivers merely based on the former efficiency criterion. According to the results, the shortest path used in traffic assignment problem could be corrected with an error term due to simplicity, in order to model better the chosen routes.

In order to do so, a definition of simplicity is proposed. It is inspired, as foundations, by the research done in Viana et al. (2013), in which simplicity is based on the number of turns. It is common practice to merely use geometric criteria in state-of-the art definitions of simplicity. However, this is due to the fact that these studies are not specific for road networks and focus on the spatial organization of planar graphs – e.g. biological systems such as leaves or dragonfly wings. On the other hand, Duckham and Kulik (2003) uses a simplicity definition from a cognitive point of view of a human navigator: in terms of how easy it is to explain, understand, memorize, or execute the instructions for a route. Thus, our paper complements the state-of-the art simplicity definitions, in order to assess driver's decision making involved in each alternative, from driver's origin to the desired destination.

Motivated on the findings of Venigalla et al. (2016), the type of road is added in our simplicity definition. In Venigalla et al. (2016) is shown that the proportion of primary roads in real paths was higher than the proportion in Shortest Distance Paths or Shortest Time Paths for the same O-D pairs.

In our work, simplicity is defined according to the number of changes instead of turns, in order to be easily differentiated from formulations such as the one in Viana et al. (2013), which are only geometrical. Our research is centered on driver's perception and, thus, the count of changes is path-based. On the contrary, in Viana et al. (2013), if a movement from a link to another one constitutes a turn or not is precomputed according to the network characteristics and does not depend on the current path a driver is following.

An efficient algorithm is proposed to compute the simplest paths between each pair of nodes in the network. Applying the logic of a conventional shortest path algorithm such as Dijkstra's algorithm (Dijkstra, 1959) leads to a much higher complexity of the simplest path algorithm. The reason is that due to the particularities of simplicity definition – e.g. the simplicity cost between two nodes depends on the previous visited node – Dijkstra's algorithm key properties cannot be exploited. Thus, using the medial graph to transform the network into a mapped graph becomes crucial in order to apply a conventional shortest path algorithm. For large networks other route planning algorithms may be more appropriate. Some speedup techniques such as Bidirectional Search can be applied directly to our transformed formulation. The more advanced ones (e.g. Transit Node Routing) require some heavy preprocessing but in exchange the query times are reduced dramatically. Sanders and Schultes (2007) reviews the progress of speed-up techniques for routing in road networks, describing and developing methods that are up to one million times faster than Dijkstra's algorithm.

Moreover, the definition of simplicity is further exploited to answer the following question: how large should be the improvement in simplicity of a route to compensate the additional travel time or length with respect to the shortest alternative? The intuition that route choice involves a trade-off between simplicity and length – or duration – is supported by the findings of Venigalla et al. (2016). It is observed that drivers are willing to travel longer distances or spend longer time on paths that have fewer turning movements.

Finally, the capabilities of our methodology to evaluate the simplest paths are shown in an artificial network, which includes the main variables of our simplicity definition. The simplest path algorithm has also been tested with two real datasets consisting of GPS trajectories, from the cities of Shenzhen, China and Borlänge, Sweden.

2. Methodology

2.1 Simplicity definition

The project is inspired by the research done in Viana et al. (2013). We started with a purely geometric definition of the simplicity, based on the number of turns. It was then modified to incorporate other factors that influence drivers' choices. Motivated by the findings of Venigalla et al. (2016), the type of road is added in our simplicity definition – not only from a hierarchical point of view, but also according to the higher understandability of the path (Duckham and Kulik, 2003). This way, the definition of a turn is made more flexible and drivers' perception is taken into account. As the type of road is considered, we will refer to turns using the term *changes* instead.

Hence, we propose the following definition of the simplest path: amongst all different paths with specific origin and destination, the simplest path is the one that needs the smallest number of changes, from the driver's point of view.

A change is incurred when the *natural extension* of the current link is not followed, either because of a change of direction or road type. It is assumed that the natural extension of a link is the one presenting the smallest deviation in terms of angles, amongst the adjacent links of the same type.

If one or more paths present the same number of changes, the shortest alternative amongst them is chosen as the simplest path. The difference of travel time would be a more behaviorally realistic criterion; however, if the aim is to put the focus on the spatial structure of the network, the length criterion is useful instead.

Additionally, the different types of change may be weighted differently, as they are not all equally inconvenient – e.g. a left turn usually causes more inconvenience than a right turn. However, this nuance is not considered for the sake of simplicity.

2.1.1 Interpretation

A way to interpret this definition is that according to the itinerary chosen by the driver, she will have to abandon the natural succession of links she is following a certain number of times, in order to reach her destination. The more times it occurs, the higher the complexity of the path. Following the simplest path implies taking as few decisions as possible, because each of those increases both the chances of making a mistake and the global inconvenience of the path.

In other words, the simplest path is the one that requires less attention. This can be related to the idea of complexity of the instructions describing a path, which is introduced by Duckham and Kulik (2003).

2.1.2 Modeling

Venigalla et al. (2016) investigate the effect of road classes in route choice, in the case of familiar networks. As expected, the result is that people prefer major roads, as they allow to travel at higher speed, leading to smaller travel times. However, this research was done only for local drivers, for which a good knowledge of the network could be assumed.

As this work intends to propose a more general methodology that considers drivers that are familiar with the network as well as occasional drivers, the road class has to be understood as the physical attributes shared by all roads of the same type. In other words, if a user travels along an arterial road and arrives at an intersection where she must choose between an arterial road or freeway (via an on-ramp), she will perceive the arterial one as the natural extension of her path – it will usually be the same street, by name. Therefore, the user has a previous intuition without considering the angles. In order to take the freeway, she is taking an active decision to exit her current path.

In order to consider driver's perception, the aim is to model the fact that two similar angles cannot be distinguished. In the case where this happens with two paths of the same type, both are equally penalized – i.e. both options add one change to the count –, as both options imply an active decision.

A realistic way of deciding when two angles are indistinguishable is to assume that drivers' perception depends on the magnitude of the angles they compare. That is, the bigger the angle, the more difficult to differentiate them. In many cases, like this one, human eyes resolution is modelled in relative terms.

2.2 Simplest path algorithm

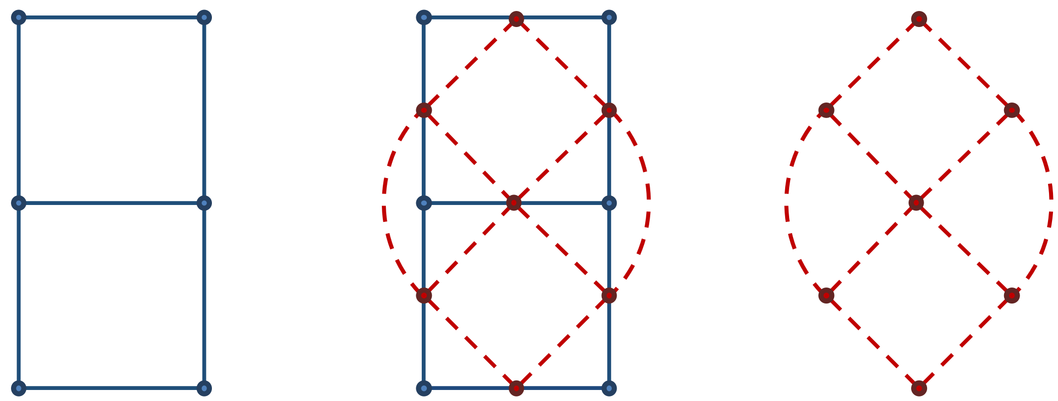
A naïve way of obtaining the simplest path from an origin to a destination would be to generate all the combinations of links that connect the two points and then compute their simplicity – i.e. number of changes. This method is very compute-intensive, not allowing to work with considerably large networks. For this reason, a new algorithm is proposed, inspired from the one presented in Duckham and Kulik (2003).

2.2.1 Medial graph and weighting

The main idea behind the above-mentioned article is that the developed definition of simplicity – as well as the one of this document – associates a weight with each pair of connected links, rather than each single link. This represents the essential difference between shortest and simplest paths algorithms, and motivates the use of the *medial graph* when working with the latter.

The medial graph *M(G)* of a plane graph *G* represents the adjacencies between the links of *G*. For each link of *G*, *M(G)* provides a node and for each pair of links of *G*, *M(G)* provides a link. An example is shown in Figure 1.

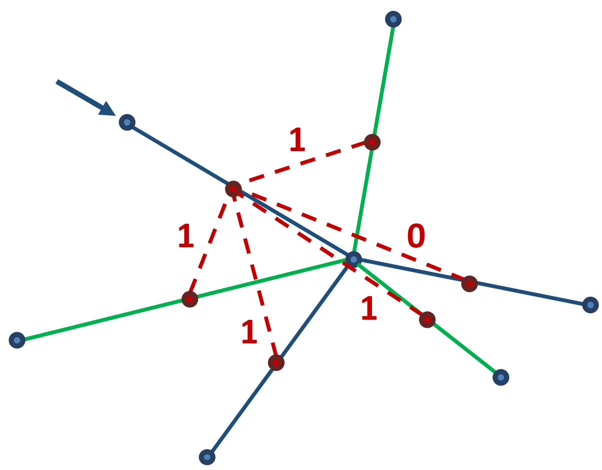
Figure 1: Example of a planar graph (blue) and its medial graph (red). All links of both graphs are bidirectional, for the sake of simplicity.



Once the medial graph is obtained only the weighting of its links remains to be done, before applying Dijkstra's algorithm or any other shortest path algorithm.

As mentioned above, the weighting is based on the concept of natural extension: a pair of adjacent links in the original graph – or a single link of the medial graph – has a weight of zero only if the second link is the natural extension of the first. All other pairs that include the same first link plus another adjacent link which therefore is not the natural extension, have a weight – or *penalty* – equal to one. An example of such weighting is presented in Figure 2.

Figure 2: Example of the weighting of an intersection that includes roads of two different types (blue and green). Among the links of the medial graph (red), only the one including the natural extension has its penalty set to zero.



Additionally, the length of the second link of each pair is added to the corresponding penalty, using an order of magnitude small enough to guarantee that the length of any considered path in the network will never reach a value of one. This addition allows to order any number of "equally simple" paths depending on their respective length, as explained in the previous section.

The final step of this method is to apply any algorithm that is able to find the shortest path between two nodes on the medial graph. In fact, given the weighting of the links, the shortest path in the medial graph corresponds to the simplest path in the original graph.

2.2.2 Generating more solutions

As the aim of this paper is to compare observed paths with their corresponding simplest paths, we are interested in generating more than one solution with the simplest path algorithm.

Many different algorithms solving this issue already exist, such as Yen's algorithm (Yen, 1971), which extends Dijkstra's algorithm in order to compute the k-shortest paths between two nodes of a graph. However, such extension consequently worsens the complexity of the algorithm – it increases from $O(N^2)$ to $O(kN^3)$, where N is the number of nodes in the considered graph –, which makes it unusable in the context of this research. For this reason, a simple heuristic method is developed and used instead.

The method is the following: once the shortest path in the medial graph is obtained, the links that compose the path are "unadvised" one at a time, by increasing their penalty to a sufficiently larger value. This value must be chosen in such way that the unadvised link will be used only it is the only possibility; in other words, the penalty of such link must be larger than the global penalty of any other considered path for the same O-D pair. The selected shortest path algorithm is then applied again on the modified network. For each modification, it generates an additional simple path that is necessarily different from the simplest.

The impact of such method on the overall complexity of the algorithm is rather small, in comparison to Yen's modification: it increases from $O(N^2)$ to $O(mN^2)$, where m is the number of links in the medial graph that compose the shortest path – obtained without any penalty modification. In other words, the selected shortest path algorithm is applied m times instead of only once.

The only issue with such method is that it generates a certain number of repetitions and there is no way of knowing in advance the number of unique solutions left after removing all repetitions. This makes the method highly unpredictable. Further work could consist in finding a way to solve this issue.

2.3 Trade-off between shortest and simplest paths

Drivers do not always choose the fastest or shortest route. There are other factors influencing the choice. For example, if the fastest route is very complex, an occasional user may choose a simpler option because she is not familiar with the network, or because the little time savings do not compensate the additional decisions she will have to take. On the other hand, some simple paths may be too long to be attractive and therefore should not be considered as valid alternatives for the driver.

The following condition – we will refer to it as the *π-criterion* – is proposed in order to discard the simplest paths which are not reasonable, in comparison with the shortest path connecting the same O-D pair:

$$\frac{\pi}{2} < A \cdot \frac{L_{simplest}}{L_{shortest}} - B \cdot \left(b \cdot \frac{1}{L_{shortest}} + \frac{C_{shortest}}{C_{simplest}} \right), \tag{1}$$

where $L_{simplest}$ and $L_{shortest}$ are the lengths of the simplest and the shortest path respectively and $C_{simplest}$ and $C_{shortest}$ are the number of changes of the two alternatives. A, B and b are three parameters to be calibrated.

This condition expresses the trade-off between simplicity and length of the simplest path, in comparison to the shortest one: the bigger the improvement in simplicity, the higher tolerance of the driver for a longer trip. Moreover, as the shortest and the simplest paths are compared relatively, an extra term – weighted by the bonus term b – is added to take into account that the trade-off equilibrium changes with the trip length. In other words, drivers are more likely to accept consequent increases of length in the case of short trips, as long as they reduce complexity.

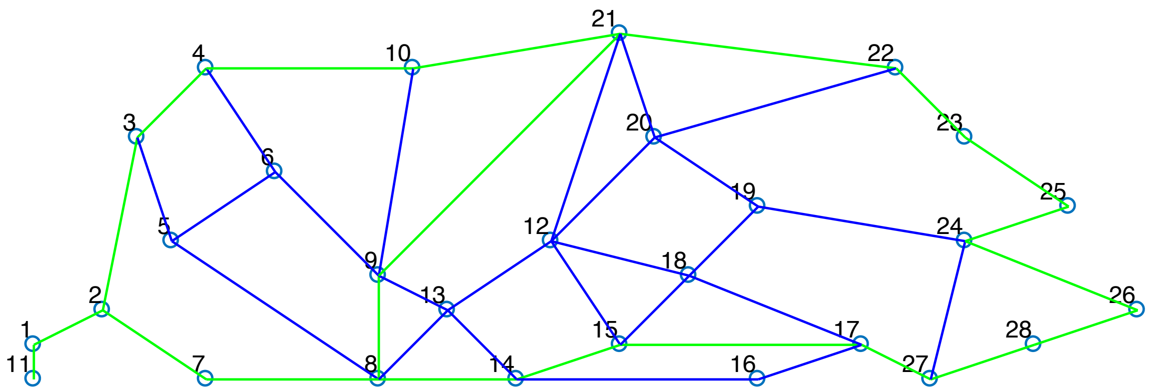
Additionally, both the *changes ratio* and the *length ratio* must also be weighted, which explains the existence of A and B.

3. Playground

3.1 Artificial network

The algorithm is initially tested in an artificial network, shown in Figure 3. This artificial network has been defined in order to reproduce the common feeling of simplicity on road networks. It includes bifurcations with similar angles to test the sensitivity of drivers' perception.

Figure 3: Representation of the artificial network used to test the simplest path algorithm. It includes two types of roads: the primary roads (green) prevail in the periphery, while secondary roads (blue) are located in the center. Peripheral roads lead to simple but long paths, while the center is denser in terms of nodes and links.



In this specific case, the difference between shortest and simplest paths can be shown as follows: the shortest path from node 11 to 22 passes connects nodes 11-1-2-7-8-13-12-20-22. The number of changes of such path is at least equal to two – penalties at node 2 because the angles are too similar and at node 8 because of the road type change – but might be higher, depending on the accuracy of angle perception in the center of the network, especially at node 12.

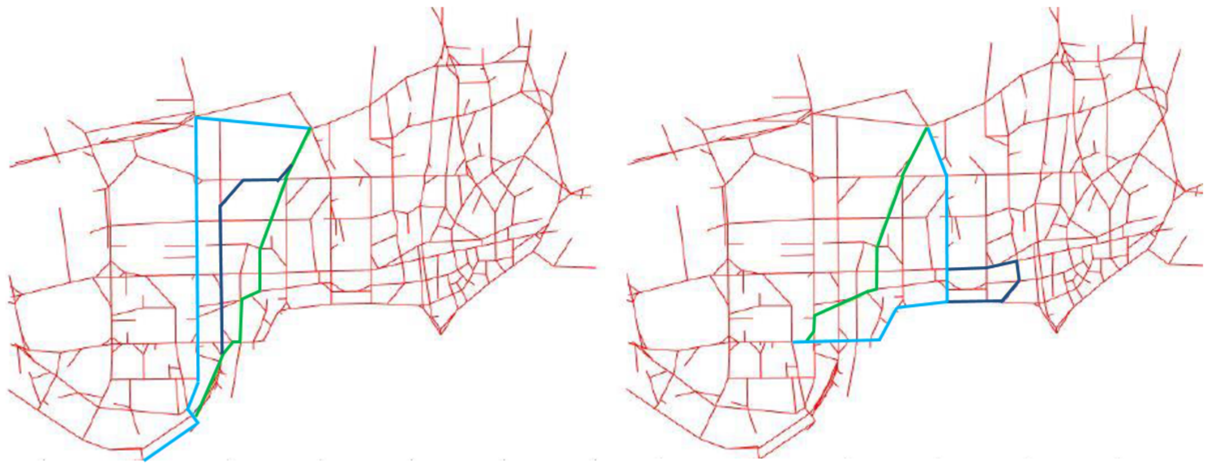
In comparison, the simplest path computed by the algorithm between the two same nodes is the following: 11-1-2-3-4-10-21-22. It is slightly longer than the previous, but presents only one change, at node 2, due to the same reasons as before.

3.2 Shenzhen case study

This case study exploits a GPS dataset from the city of Shenzhen, in China. From taxi GPS points, real paths are reconstructed using map-matching techniques. A simplified network is used, consisting of 635 nodes and 1523 links.

The influence of simplicity on route choice is clearly observed in Figure 4(a). The real path is geographically in between the shortest path and the simplest path. However, in Figure 4(b), a counter-intuitive set of shortest, real and simplest paths is shown.

Figure 4: Plot of shortest (green), real (blue) and simplest (cyan) paths in the simplified network of Shenzhen. (a) SI of the real path is 0.32 and its average speed is 27.5 km/h. (b) SI of the real path is 4.38 and average speed is 18.7 km/h.



The analysis of the data reveals that more than 30% of the real paths are *outlier candidates*. We use two informal criteria to identify outliers: numbers of changes and simplicity index (SI). A real path for a specific O-D pair is considered an outlier candidate if it has a higher number of changes than the corresponding shortest path. The SI is used in a similar fashion: a SI higher than one indicates an outlier candidate. The SI is defined as follows:

$$Simplicity\ index = \frac{\Delta real}{\Delta simplest} \tag{2}$$

where *Δreal* is the length difference between real and shortest paths and *Δsimplest* the difference between simplest and shortest paths. A trip with a SI close to zero means that the driver prefers the shortest path, meanwhile a SI close to one means that the driver appreciates the simplicity.

Several reasons may cause these outliers. They are still under discussion:

- Effect of congestion: the shortest path may be far from being the fastest one, under conditions of heterogeneous congestion. However, no data were available in order to compute the fastest paths, as the only speed data were from taxis.
- Directions of the links: wrong direction of the links may cause that the shortest path in reality may be longer and more complicated than shortest path computed for this study.
- Other factors that influence the decision-maker and that are not taken into account in this formulation, such as unobserved factors, irrationality, etc.

3.3 Borlänge case study

This case study uses a GPS dataset from the city of Borlänge, in Sweden. The general network consists of 3077 nodes and 7288 unidirectional links. Additionally, 239 observed paths distributed throughout the network are available.

Nonetheless, the dataset presents two major issues. The first one is related to the road types of the links, which are actually missing from the available data. They must be extracted from another source – *OpenStreetMap* in this case – and "map-matched" with the available network. As a perfect accuracy is never guaranteed, this first issue jeopardizes our results. The second issue is related to the roundabouts segmentation. Those are not modeled as unique nodes in the general network of Borlänge, which causes the algorithm to miscalculate the number of changes of a path passing through any of them. Depending on the way they are modeled, roundabouts may prioritize a counter-intuitive path or penalize the natural extension of the previous link. One could argue that roundabouts always require more attention from the driver, but the behavior of such model is too unpredictable to be left unchanged.

For each of the 239 O-D pairs, the simplest path algorithm and Dijkstra's algorithm are applied. We are therefore provided with the shortest path and a certain number of simplest paths, in addition to each observed path.

As a matter of example, such results are shown in Table 1 and Figure 5 for a single observed path, without considering road type in the algorithm. Table 1 gathers the characteristics of the observed path and the six corresponding simplest paths, while Figure 5 illustrates said paths in Borlänge network.

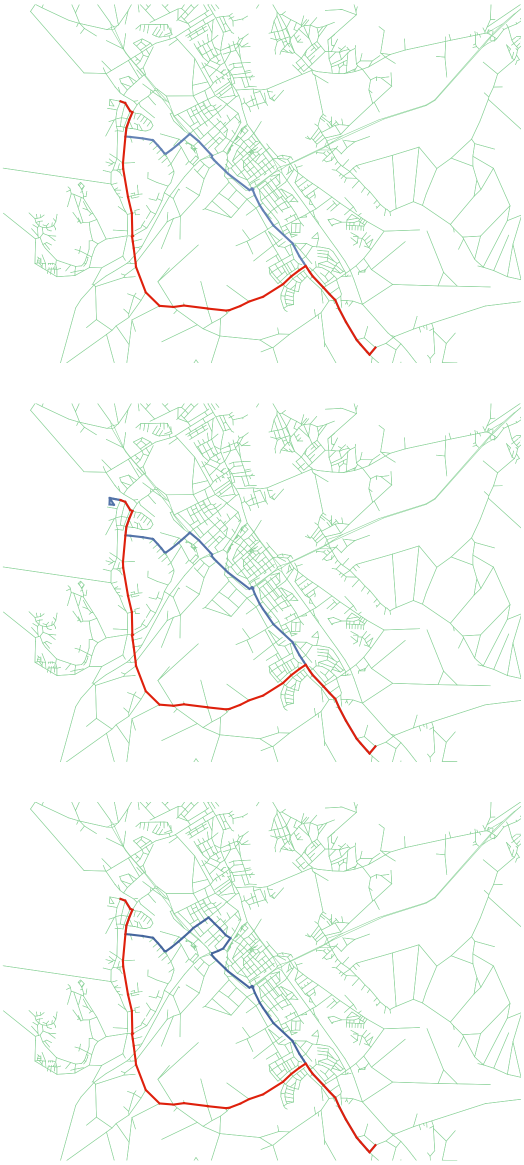
Table 1: Characteristics of the observed paths and the simplest paths between the same origin and destination nodes.

Type of path	Number of turns	Number of links
Observed	7	45
Simplest #1	5	57
Simplest #2	5	62
Simplest #3	5	69
Simplest #4	5	66
Simplest #5	5	75
Simplest #6	6	56

Figure 5 perfectly illustrate the two major issues of the dataset discussed above. The observed path, represented in red, clearly follows a unique road for most of its length. However, as the different links that compose this road have the same type as the ones around it, the algorithm counts a much higher number of turns than it should be. Without any road type distinction, the algorithm is unable to detect the natural succession of links that compose a curved road, as any link connected to such road with a small angle of incidence will disrupt the process.

One should also note that in this case, the driver is “more intelligent” than the simplest path algorithm. The driver knows that crossing the city center takes more time than using a circumvallation, despite the direct path being shorter in terms of length.

Figure 5: Plot of the observed path (red) in comparison to the six simplest paths (blue) presented in Table 1.



4. Conclusions and future research

In this paper, we investigate the impact of simplicity on route choice, for road networks. For this purpose, a definition of simplicity that focuses on drivers' perception is introduced. One of the main contributions of this work to state-of-art simplicity approaches is considering road type. As analysis is still ongoing, we do not provide any final results. However, from the preliminary results of Borlänge case study, the importance of road type in driver's perception of simplicity is pointed out. Moreover, the results obtained in Shenzhen case study show, under some circumstances, that the alternatives chosen by the drivers often lie in between the shortest and the simplest path. These promising results need to be validated, once the issues of the datasets described in the case studies will be overcome.

On the other hand, the simplicity definition may be extended including other variables to model in a more realistic way drivers' decision-making. In future work, we aim to exploit our abstract formulation of the simplest path algorithm to investigate whether drivers tend to choose routes that are easy to follow based on recognizable or well-known elements along the path. As an example, Kazagli et al. (2016) introduces the concept of *mental representation items* (MRI), which could be used in the context of our work in order to improve the algorithm.

5. Appendices

5.1 Simplest path algorithm: summary of the main steps

INPUTS: coordinates of the *N* nodes in a connected and directed graph *G*; starting and ending nodes for each of the *L* links in the form of the connectivity matrix of *G*; origin and destination nodes of the simplest path, *O* and *D*.

STEP 1: computation of the length and absolute angle of each link. Addition of two "ghost nodes" S_1 and S_2 and two "ghost links", connecting S_1 to *O* and *D* to S_2 respectively. They will be used later, when creating the medial graph of *G*.

STEP 2: computation of the link-adjacency matrix – link *B* is adjacent to link *A* if the ending node of *A* is the starting point of *B*. The link-adjacency matrix of graph *G* is the connectivity matrix of the medial graph of *G*, *M(G)*.

STEP 3: computation of the penalties matrix, which assigns a penalty – 0 or 1 – to each of the *L'* links in *M(G)*. The penalty is based on the concept of natural extension and on the length of the *L* links in *G*.

STEP 4: application of Dijkstra's algorithm on the medial graph. The origin and destination nodes in *M(G)* are *O'* and *D'*, chosen to be the ones that correspond to the ghost links of *G*. The obtained shortest path between *O'* and *D'* corresponds to the simplest path between *O* and *D* in the original graph *G*. Without the ghost nodes S_1 and S_2 defined earlier, nodes *O* and *D* are not "translated" unequivocally in the medial graph.

STEP 5: generation of other solutions, by sequentially "unadvising" – increasing the penalty of – the links that compose the shortest path from *O'* to *D'*. Dijkstra's algorithm is applied again on each modified version of *M(G)*, which generates each time an additional simple path that is necessarily different from the one obtained in STEP 4.

OUTPUTS: a certain number of simplest paths between *O* and *D*; the number of changes associated with each of these paths.

6. References

Dijsktra, E. W. (1959). A Note on Two Problems in Connexion with Graphs, *Numerische Mathematik 1*, 269–271.

Duckham, M. and Kulik, L. (2003). "Simplest" Paths: Automated Route Selection for Navigation, COSIT 2003, Lecture Notes in Computer Science, Vol. 2825, pp. 169–185, 2003.

Kazagli, E., Bierlaire, M., and Flötteröd, G. (2016). Revisiting the route choice problem: A modeling framework based on mental representations. Journal of Choice Modelling, 19, 1–23.

Sanders, P., and Schultes, D. (2007). Engineering fast route planning algorithms. Proceedings of the 6th International Conference on Experimental Algorithms, 23–36.

Sanders, P., and Schultes, D. (2012). Engineering highway hierarchies. Journal of Experimental Algorithmics, 17(1), 1.1.

Venigalla, M., Zhou, X., and Zhu, S. (2016). *Journal of Urban Planning and Development*, Vol. 143.

Viana, M. P., Strano, E., Bordin, P. and Barthelemy, M. (2013). The simplicity of planar networks, *Scientific Reports* **3**, Article number: 3495.

Yen, J. Y. (1971). Finding the K Shortest Loopless Paths in a Network, *Management Science*, Vol. 17, No. 11, Theory Series, pp. 712-716.