9th International Symposium on Transportation Data & Modelling (ISTDM2023)
Ispra, June 19-22 2023

Generation of Aggregated Road Network by Vehicle Trajectory Data

Hengyi Zhong[a,*1], Toru Seo[a], Wataru Nakanishi[b], Shohei Yasuda[c], Yasuo Asakura[a], Takamasa Iryo[d]

[a] *Department of Civil and Environmental Engineering, Tokyo Institute of Technology*

[b] *Institute of Science and Engineering, Kanazawa University*

[c] *Faculty of Engineering, The University of Tokyo*

[d] *Department of Human-Social Information Sciences, Tohoku University*

Transportation network analysis commonly uses network data which are properly aggregated so that essential features of the network are directly investigated. However, most of the existing road network data are originally prepared for car navigation purposes which are too detailed for network analysis. This paper proposes a method to generate a properly aggregated network by using only vehicle trajectory data. The method consists of the following steps: node detection, link generation, and network reorganization. The characteristics of the proposed method was validated by using actual large-scale probe vehicle data in Tokyo. The results suggests that the method is useful to automatically generate proper network data for transportation analysis in various scales.

Keywords: road network generation, traffic network analysis, vehicle trajectories

1. Introduction

Properly aggregated road networks are usually welcomed in traffic network analysis. In those road networks, for example, one intersection area should be represented as only one node. One characteristic of those networks compared to existing road networks is that they may have a significantly smaller number of nodes and links. If there are too much links and nodes even if in a simple network, the calculation time and errors may both increase. To solve this problem, Boeing (2017) introduced a simplification method for OpenStreetMap road network by considering topology of the network. Yasuda et al. (2019) used ETC2.0 data, a set of vehicle trajectory data collected in Japan, to simplify National Digital Road Map (DRM) data by the recorded traffic volume of probe vehicles.

Global Navigation Satellite System (GNSS) trajectories can be considered as a useful data source in geographic data analysis nowadays, and it also can be used for generating road network data. Tang et al. (2019) tried to recognize the structure of intersections by using local features. Huang et al. (2018) introduced graph theory to generate complex road networks. Meanwhile, to generate a part of aggregated road network directly from GNSS trajectories, Zhong et al. (2022) applied spatial scan statistics on different distributions of trajectory dot data to determine intersection areas.

It is efficient to use such trajectory-based road network data in traffic network analysis when there is an absence of proper road network data. However, such a network should contain not only correct geographical positions, but also values which can be used in traffic network analysis. Also, the network should be aggregated, and networks with different aggregate levels can be generated automatically by only adjusting few parameters of the generation method, so that they can be applicated at different scales suitably.

In this paper, we developed a method of generating aggregated road network which is suitable for traffic network analysis directly from vehicle trajectory data. In this method, nodes and links are detected at an aggregated form in advance, and different networks can be generated by considering the traffic volume of probe vehicles. Also, the relationship between the road network generated by proposed method and the real road network was validated.

2. Methodology

2.1 Network generation

([1]) Corresponding author (zhong.h.aa@m.titech.ac.jp)

The methodology of network generation developed in this study consists of three steps: node detection, link generation, and network reorganization.

A method of detecting nodes by using spatial scan statistics proposed by Zhong et al. (2022) is applied in the node detection step. The area represents a node can be defined as the intersection area where vehicles can select their routes. Considering turning behaviors during route selections, an intersection area can be defined as an area where data dots with large turning angle accumulates.

Once node areas are determined, for each dot of a trajectory, the node or link which it is related to can be determined by checking the node pair $(p_{\mathrm{f}}, p_{\mathrm{b}})$, where p_{f} denotes the last node vehicle passed through before or at that dot and p_b denotes the first node vehicle passed through after or at that dot. We can separate the vehicle trajectory data into each node or link by using the node pair (p_f, p_b) of dots. Dots with the same node pair $(p_{\mathrm{f}}, p_{\mathrm{b}})$ can be considered as the set of fragmental trajectories $D(p_i)$ or $D(e_{ij})$ belong to the node p_i or link e_{ij} of that node pair. If $D(e_{ij}) \neq \emptyset$, then there may exist the link e_{ij}. For each $D(e_{ij})$, traffic characteristic values such as traffic volume q_{ij} of probe vehicles and link length l_{ij} for corresponding link can be easily calculated.

To reorganize the generated road network with different aggregation levels, major links and major nodes are defined by link daily probe traffic volume q (veh/day) and the traffic volume threshold τ_q (veh/day). A major link is defined as a link with large probe traffic volume, i.e., $q_{ij} > \tau_q$. A major node is defined as a node connected with three or more major links which performs not as an overpass, or a node performs as a dead-end connected with a major link.

Road networks generated by proposed method are called Probe Derived (PD) networks in this paper. Also, we call a PD network which major links and major nodes are defined by τ_q as a PD-τ_q network.

2.2 Validation methods

The validation of validity and usefulness of PD road networks are considered from following four aspects: graphical characteristics, correctness of nodes and links, stability of network generation, and the results of shortest path search. Comparison between PD road networks and the real road network is applied for each aspect. We used ETC2.0 data as the real observed trajectory data to generate PD road networks. This vehicle trajectory data was obtained in Tokyo, Japan for three months (July 2021-September 2021). OpenStreetMap (OSM) road network, which was obtained as in September 2022 by using the osmnx module by Boeing (2019), is used as the real road network data.

First, graphical characteristics such as number of nodes and links and link lengths show the level of aggregation in PD networks. In a PD network, which contains only major nodes and major links as well as one intersection appears as only one node, it is expected that it has less nodes and links comparing to existing real road networks. Also, links in PD networks are expected to have a larger length, as small intersections which split roads to several short portions are not expected to exist.

Secondly, the correctness of node means if there is a real intersection geographically or not at where the node exists in PD network. As a node in PD network can also be represented as node areas, it can be checked that whether real road network nodes exist inside the node areas for each node. Meanwhile, the correctness of links can be checked by link lengths from a viewpoint of link performance. The road section between two adjacent intersections always be the shortest path between those two intersections. A link is said to be able to represent real road section if the length of the corresponding shortest path on real road network is the same as the link length on PD network.

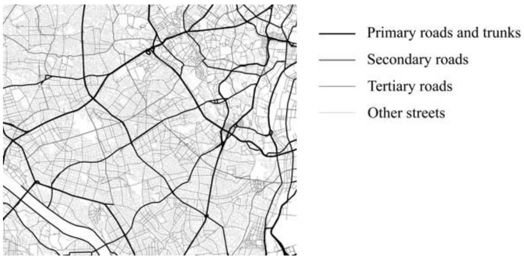
Thirdly, if a PD network can represent a real road network uniquely, networks of the same area should be generated in a stable manner even if data is obtained from different dates. That is, the similarity between those two PD networks is expected to be high. The similarity can be evaluated by the Jaccard coefficient. Let E_A and E_B be the sets of links of network A and B respectively. The Jaccard coefficient defined as the size of the intersection divided by the size of the union of E_A and E_B.

Finally, the usefulness of the PD network can be validated by applying shortest path search on the PD network and comparing the results on a real road network. We select origin and destination points randomly from the set of nodes in each PD network, and calculate the shortest path of length by the Dijkstra method. Similarly, on the real road network, the same origin and destination points as in the PD network are used. If the shortest path lengths in the PD network match the corresponding lengths in the real road network, there is a possibility that PD networks are useful for traffic network analysis.

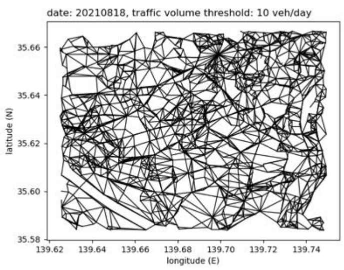
3. Results

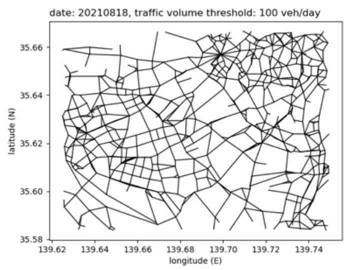
Different PD networks (Figure 1b-1d) perform as different aggregate levels for the real road network (Figure 1a). For PD-10 network, there remains many links which can be considered as aggregations of narrow roads. For PD-100, the shape of the network gets grid-liked, which is more similar to the real road network at normal. Those PD networks with low aggregate level can be used in traffic analysis focus on a part of the road network. As the aggregate level of probe traffic volume gets higher, there only remains links with high traffic volume such as national roads and expressways in the PD-500 network. This kind of PD networks with high aggregate level are suitable for large-scale analysis such as traffic assignment inside a metropolitan area.

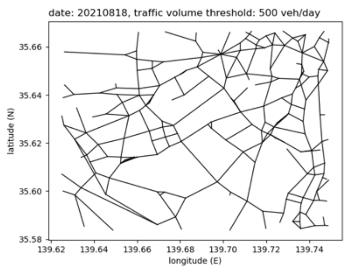
Figure 1. Different PD networks and the OSM network



a) OSM network







b) PD-10 **c)** PD-100 **d)** PD-500

Source: a): OpenStreetMap; b-d): Own elaborations.

Table 1. Graphical characteristics of PD and OSM networks.

	PD networks			OSM network
	PD-10	PD-100	PD-500	
Number of links	3,136	1,047	319	45,682
Number of nodes	1,025	579	208	30,473
Average link length (m)	417.04	414.87	655.90	58.85

Source: Own elaborations.

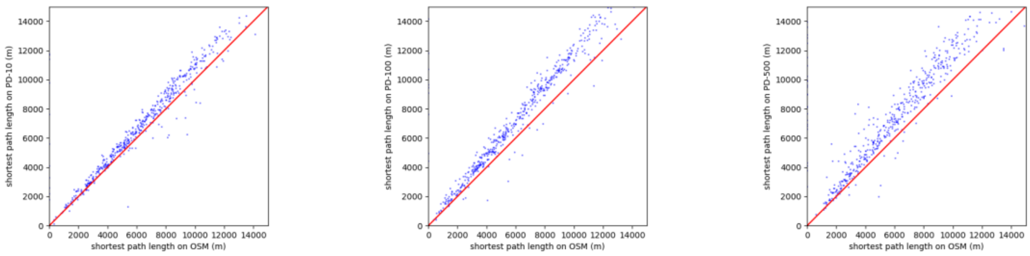
The graphical characteristics, which includes number of nodes, number of links and link lengths, show that PD networks can be considered as finely aggregated road networks at a viewpoint of nodes and links. The graphical characteristics of PD networks and OSM network are shown in Table 1. There is a large reduction on nodes and links comparing to the OSM network, and the average link length can be considered as the same scale as the length of the road between major intersections in general.

The correctness of nodes and links are confirmed in all generated PD networks. 981 of 1025 (95.7%) nodes in the PD-10 network correspond to intersection nodes of OSM network. And all nodes in PD-100 and PD-500 networks are confirmed to be able to represent real intersections. For most of the links, the lengths on PD networks are close to the lengths of corresponding shortest path on the OSM network. Though the lengths on PD networks tend to be slightly smaller especially on links with curves. This is because link lengths are calculated from the distance of straight line between two data dots, which is smaller than the real travel distance due to the long sampling interval of the ETC2.0 data.

For the stability of network generation, the Jaccard coefficients between two PD networks with same aggregate level can be 0.5-0.6, which means 70-75% of the links are strictly matched when the data is only from one day and obtained from both weekdays or both holidays. The similarity can be a little bit lower otherwise due to the lower traffic volume on holidays than weekdays. A usage of multiple-days data can mitigate this instability. When we use a combined data from more than 5 days, the Jaccard coefficients can be about 0.6-0.8, which means 80-90% of the links are strictly matched.

For the results of the shortest path search (Figure 2), most of the lengths of shortest paths on all PD networks are larger than them on OSM network. The results on PD-10, which includes links represent for narrow roads, however, are closer to the results on OSM network than PD-100 and PD-500. This is mainly because the searched shortest paths on OSM network including narrow roads which makes a shortcut. PD networks with high aggregate level perform more likely as aggregated networks with only major nodes and links.

Figure 2. Relationship of shortest path lengths between PD and OSM networks



a) PD-10 **b)** PD-100 **c)** PD-500

Source: Own elaborations.

4. Conclusion

In this study, we developed a method for generating aggregated road network data from only vehicle trajectory data. As a result of the validations, it was confirmed that PD networks generated by the proposed method are appropriately generated and aggregated. By adjusting the traffic volume threshold, networks with different aggregate levels can be easily generated, which can be applicated in traffic analysis with different scales. A future work of this study is to develop a general method to extract more information. A quantitative method to compare two spatial networks is also expected to be developed for verification.

Acknowledgements

The ETC2.0 data for this study is provided by Tokyo Metropolitan Government through their project called "offering of several datasets on mobility and transportation in Tokyo 2020 Games".

References

Boeing, G., 'Osmnx: New methods for acquiring, constructing, analyzing, and visualizing complex street networks', *Computers, Environment and Urban Systems*, No 65, 2017, pp. 126-139.

Huang, J., Deng, M., Tang, J., Hu, S., Liu, H., Wariyo, S. and He, J., 'Automatic generation of road maps from low quality gps trajectory data via structure learning', In *IEEE Access*, Vol. 6, 2018, pp. 71965–71975.

Tang, J., Deng, M., Huang, J. and Hu, S., 'A novel method for road intersection construction from vehicle trajectory data', In *IEEE Access*, Vol. 7, 2019, pp. 95065–95074.

Yasuda, S., Iryo, T., Sakai, K. and Fukushima, K., 'Data-oriented network aggregation for large-scale network analysis using probe-vehicle trajectories', In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019, pp. 1677–1682.

Zhong, H., Nakanishi, W., Yasuda, S. and Iryo, T., 'Extraction of major intersections by only vehicle trajectory data', *Theory and Applications of GIS*, Vol. 30, No 2, 2022, pp.7-18.