The Application of ETC2.0 Probe Data and Network Clustering in Investigating Urban Mobility Patterns During an Extreme Snow Event

Tran Vinh Ha [1] & Mikiharu Arimura [2]

[1,2] *Division of Sustainable and Environmental Engineering, Muroran Institute of Technology, Muroran, 050-8585, Hokkaido, Japan.*

Abstract

Implementing the Vehicle-to-Infrastructure with the upgraded electronic toll collection systems (ETC2.0) has transformed Japan's transportation infrastructure by elevating it into one of the foremost Intelligent Transport Systems. Despite the wealth of data provided by ETC2.0, its application in studying urban mobility patterns under extreme weather conditions remains limited. This study examines the impact of heavy snowfall on the mobility network in Sapporo, Japan, using ETC2.0 probe data. By comparing mobility patterns on selected heavy snow and normal days in February 2022, the study identifies significant changes in network structure and community distribution. Findings reveal that heavy snowfall causes fragmentation of mobility networks, with notable shifts in community locations and node centrality. The study underscores the importance of maintaining connectivity to industrial and commercial areas during extreme weather events and highlights the need for further research into the relationship between community structures and travel behavior.

Keywords ETC2.0, Network clustering, Urban mobility, Heavy snow, Sapporo

1 Introduction

Since its inception in 1989, electronic toll collection (ETC) has emerged as one of the most successful Intelligent Transport Systems (ITS) [1]. The primary benefits of ETC include the elimination of congestion, enhanced toll gate capacity, and a consequent reduction in pollution. In Japan, ETC began operations in March 2001 and has since expanded nationwide. As reported by the Ministry of Land, Infrastructure, Transport, and Tourism (MLIT), approximately 94.0% of vehicles in Japan were using the ETC service by October 2022 [2].

To further promote Intelligent Transport Systems, the Japanese government launched a Vehicle-to-Infrastructure (V2I) project in 2011, marking the first V2I system globally. In 2014, the country introduced an enhanced version of electronic toll collection, ETC2.0, as a component of the V2I project. This new system includes onboard units (OBUs), roadside units (RSUs), and the Vehicle Information and Communication System (VICS) [3]. The OBUs installed in vehicles gather probe data, such as travel and behavior history, in a privacy-preserving format and communicate with the RSUs on the roads. VICS then aggregates the collected data and provides information to drivers through the RSUs. This information may include traffic and weather conditions, accidents, or roadwork, aiding drivers in detouring and safe driving [4].

Given the wealth of information provided by ETC2.0 probe data, transport researchers have explored various topics using this data source. For instance, Katoh et al. [5] and Maki et al. [6] used ETC2.0 probe data to estimate traffic volume and identify bottlenecks. In the context of travel behavior, Matsushita and Hayashi [7] and Sekine et al. [8] analyzed driving maneuvers and traffic congestion. Similarly, Kaneko et al. [9] and Goto et al. [10] developed a route choice model based on vehicle trajectories derived from ETC2.0 data. Additionally, Homma et al. [11] introduced inundation monitoring by comparing ETC2.0 probe data with numerical flood simulations.

Despite ETC2.0 being a reliable and promising data source, its application in studying urban mobility patterns remains limited. This shortcoming raises the question of how to exploit the ETC2.0 data source effectively to solve the current transport problems. One of these problems is the impact of disasters or extreme weather on human mobility.

Japan is one of the most vulnerable countries to natural disasters, including earthquakes, tsunamis, and storms [12]. In addition to these disasters, heavy snow is a significant extreme weather condition that impacts human activities. Figure 1 shows regions with heavy and extremely heavy snowfall, as defined by the National Spatial Planning and Regional Policy Bureau of Japan [13]. The former category represents areas where industrial development is hindered by snow, affecting economic growth. Additionally, heavy snowfall in these areas impedes the improvement of residents' living standards. The latter category includes regions with particularly high snowfall, leading to significant disruptions in daily life due to prolonged automobile traffic suspensions.

According to the Road Bureau of the Ministry of Land, Infrastructure, Transport, and Tourism of Japan, about a quarter of Japan's population lives in cold and snowy regions, which have the highest population density

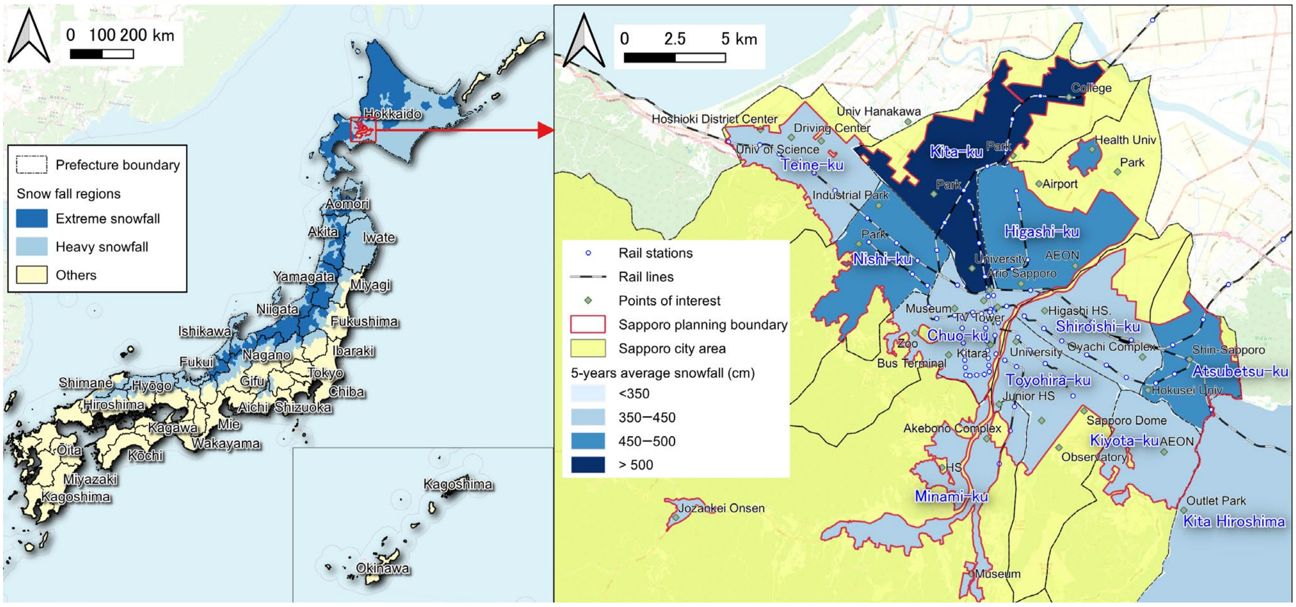


Figure 1. Snowy regions of Japan (left side) and recent 5-year average snow depth in Sapporo city (right side) (Reproduced from [18] with the authors' permission).

compared to other snowy countries [14]. This region includes several large cities, such as Sapporo, the capital of Hokkaido prefecture. Recently, Sapporo experienced two heavy snowfalls in February 2022. The first occurred from the 6th to the 7th, and the second from the 21st to the 23rd. During these periods, the transportation system was severely disrupted. Traffic was stagnant due to obstructed streets, poor visibility, and vanished road facilities. Additionally, all flights were canceled, and up to 90% of trains were suspended [15–17].

Although more than two years have passed since the events, we have found only one study that examined their effects on city mobility, produced by [18]. In this study, the authors successfully revealed the quantitative changes in traffic volume over time and by location, which partly reflect the negative impacts of the whiteout. However, the transformation of city mobility as a whole network caused by the extreme event is still not fully conveyed.

To fill the gap remaining in [18], the current study aims to address the following questions:

(1) Are mobility networks fragmented, and how do they change during heavy snow days?
(2) In each network, which locations play an important role in the cluster as well as the whole network?

The rest of the paper is organized into four sections. Section 2 details the methodology for data collection and analysis. Section 3 outlines the results obtained from the analysis. Sections 4 and 5 cover the discussions and limitations, respectively.

2 Data and Methodology

2.1 ETC 2.0 data set

In this study, we used the same data set utilized in the research of [18]. However, instead of examining all days in February 2022, we selected only two pairs of days to compare the mobility patterns (Figure 2). The first pair includes 2022-02-06 (day1 – Sun*) and 2022-02-27 (day4 - Sun), representing the weekend and holiday. The second pair consists of 2022-02-15 (day3 - Tue) and 2022-02-22 (day3 - Tue*), representing the weekdays.

The data set includes the number of trips generated by the hour. Additionally, each trip contains the origin and destination (by 500x500 m mesh), trip duration, and length. We aggregated trip volume for each day and used only trips that have both origin and destination within the city boundaries (the intra trips defined by [18]).

2.1 Network clustering technique

To address the study's purposes, we proposed using the network clustering technique to examine the mobility patterns on the selected days.

Network clustering, also known as community detection, involves identifying clusters of nodes within a network that exhibit stronger connections to each other than to nodes outside their group. This technique is valuable for revealing the hidden organization within intricate networks. Its applications extend across diverse fields such as e-commerce, social network, economics, etc., [19] and also in transportation [20–23].

In the current study, we treated the mobility pattern on a given day as a network, where mesh centers act as nodes or vertices, and movements between these nodes represent links or edges. Additionally, we assigned the number of trips connecting two nodes as the weight of each link. All

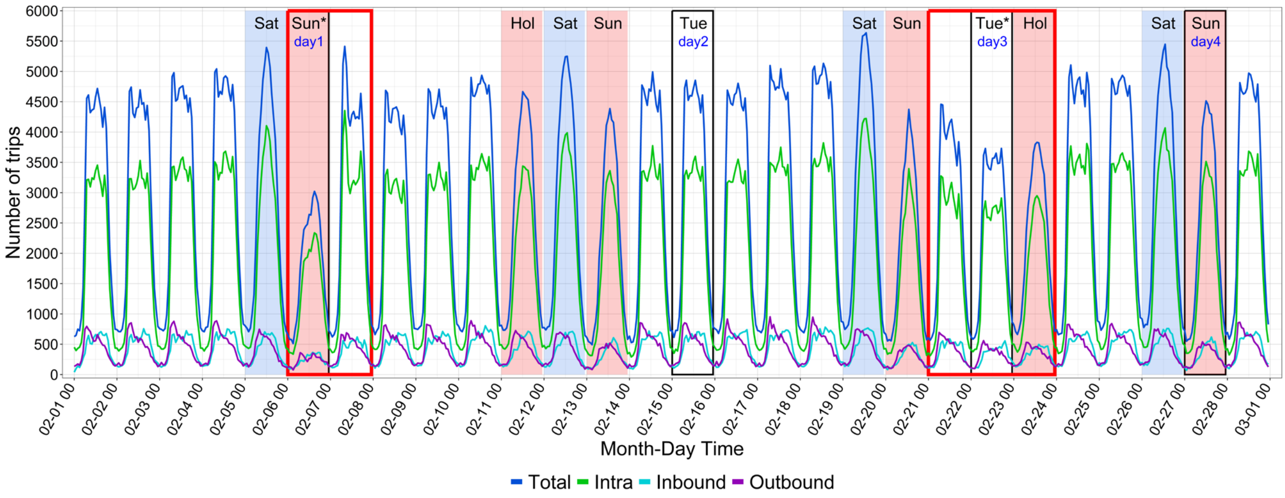


Figure 2. Generated trips variation in February 2022 in Sapporo. (Reproduced from [18] with the authors' permission; Light blue shade: Saturday; Light pink shade: Sunday/Holiday; Red box: Heavy snow days)

links are directed based on the movement direction from the origin node to the destination node.

Among the various network clustering techniques, we applied the Infomap algorithm developed by [24, 25] due to its ability to handle directed, weighted networks as well as medium-sized networks [22, 23]. The Infomap algorithm is based on the map equation (1) [25].

$$L(M) = q_{\curvearrowright}\mathrm{H}(\mathcal{Q}) + \sum_{i=1}^{m} p_{\circlearrowright}^{i}\, H(\mathcal{P}^{i}) \qquad (1)$$

where L is the expected description length of a random walk on a network, M is the partition of the network into modules, m is the number of modules, $q_{\curvearrowright}$ is the probability that the random walk switches modules on any given step, $\mathrm{H}(\mathcal{Q})$ is the entropy of the module names, $p_{\circlearrowright}^{i}$ is represents the probability of moving from one community to another, and $(\mathcal{P}^{i})$ is the entropy of the within-module movements.

After detecting network communities, we computed the network modularity value using Equation (2) proposed by [26] to test whether the network is well-structured. According to [26], the modularity value of about 0.3 indicated the community structure is significant.

$$Q = \frac{1}{2m}\sum_{i,j}\left[A_{ij} - \frac{k_i^{out}k_j^{in}}{2m}\right]\delta(c_i, c_j) \qquad (2)$$

where Q is the modularity, A_{ij} is the weight of the edge from node i to node j, k_i^{out} and k_j^{in} are the sum of the weights of the edges leaving node i and entering node j, respectively, m is the sum of the weights of all edges in the network, $\delta(c_i, c_j)$ is the Kronecker delta, which is 1 if nodes i and j are in the same community, and 0 otherwise. c_i, c_j are the communities to which nodes i and j belong.

To clarify the differences between networks, we computed the adjusted Rand index (ARI) proposed by L. Hubert and P. Arabie [27], which is developed based on the Rand index (RI) introduced by Rand [28]. Equations (3) and (4) summarize the ARI computation as described by [29].

$$ARI = \frac{RI - Expected\_RI}{\max(RI) - Expected\_RI} \qquad (3)$$

$$RI = \frac{a + d}{n(n+1)/2} \qquad (4)$$

where *Expected_RI* is the expected value of the Rand index, a and d are the number of pairs of elements that are in the same and different subsets in networks, respectively, and n is the number of elements in the network.

In theory, the ARI ranges from -1 to 1. A value of +1 indicates a perfect match between the two networks, while a value of 0 implies high disagreement between the two networks.

To gain a deeper understanding of the network structure, we continued to visualize the network communities. We first extracted the main communities that contribute at least 1% of the total trips in a day.

Along with extracting the main communities, we estimated the role of nodes in the networks by computing the vertex betweenness centrality value *Cb(v)* using Equation (5). A higher value of *Cb(v)* indicates a higher importance of that node, implying it is a bottleneck.

$$Cb(v) = \sum_{s\neq t\neq v}\frac{\sigma_{st}(v)}{\sigma_{st}} \qquad (5)$$

where σ_{st} is the total number of lowest weight paths from vertex s to vertex t, $\sigma_{st}(v)$ is the number of those paths that pass-through vertex v and n is the number of vertexes.

3 Results

3.1 Network clustering and comparison

Table 1 shows the values of network modularity and ARI for each pair of chosen days. As expressed in the table, all modularity values are higher than 0.3 (ranging from 0.40 to 0.45), indicating that the structures of the four networks are significant.

Regarding network comparison, the ARI scores do not exceed 0.6, implying a medium level of agreement between networks. Notably, the Sunday pair (day1 and day4) has a lower ARI value than the Tuesday pair (day2 and day3), at 0.45 and 0.57, respectively.

The lowest ARI is 0.37, which accounts for the "day1 and day3" pair (heavy snow on Sunday and Tuesday). Meanwhile, the ARI value for heavy snow Tuesday and normal Sunday (day2 and day4) is equal to that of the Tuesday pair (day2 and day3). For heavy snow Sunday and normal Tuesday (day1 and day2), the ARI is comparable to the Sunday pair (day1 and day4).

Table 1. Network modularity (diagonal values) and ARI across network pairs (upper diagonal values)

	day1 (Sun*)	day2 (Tue)	day3 (Tue*)	day4 (Sun)
day1 (Sun*)	0.45	0.44	0.37	0.45
day2 (Tue)		0.40	0.57	0.57
day3 (Tue*)			0.40	0.50
day4 (Sun)				0.44

3.2 Network structure

Figures 3 and 4 illustrate the network structure for the four days. In these figures, the community indexes were sorted by the total number of nodes. For example, the first community (e.g., 02-06_1) is the largest cluster in the network for February 6, 2022. Note that the largest community does not necessarily have the highest number of trips compared to others.

As shown in the figures, the number of large communities on Sundays is higher than on Tuesdays, with 9 and 11 communities compared to 7 and 6 communities, respectively. Additionally, large communities mainly appear on the outskirts of the city center. Meanwhile, although the downtown community (Chuo district) does not contain many nodes, it has a relatively high trip volume.

In Figure 1, the largest community on a normal Sunday is located on the northwest side of the city (Tenei and Nishi districts). However, on a heavy snow Sunday, the largest community shifts to the northeast side of the city (Higashi and Shiroishi districts). Additionally, the second-largest community on a normal Sunday (on the north side of the city) splits into three smaller communities on the heavy snow Sunday. Similarly, the two communities on the west side (in Teine, Nishi, and Chuo districts) rearrange into three communities. In contrast, the two communities in the Toyohira district combine into one community on the heavy snow Sunday.

In the case of the Tuesday pair, the distribution of large communities is similar except on the east side of the city. On a normal Tuesday, there are two communities located in the Higashi, Shiroishi, and Atsubetsu districts. These communities merge into one and become the largest community on the heavy snow Tuesday. The second and third largest communities on both Tuesdays are on the north and northwest sides of the city. The remaining communities are also in the same locations and have the same shapes, but they change in order of size.

Regarding node betweenness, all networks present high betweenness values in the Higashi Aeon and Oyachi areas. We also found an agreement between networks when the medium betweenness value appeared in the community in the Minami district.

High centrality indexes are also found in the Nishi district and around Hokkaido University. However, these nodes become less important during heavy snow on Sundays. Interestingly, Sapporo Station did not have a high betweenness value, even though it is a major transportation hub of the city.

4 Discussions

As expected, the mobility network is clustered into communities across the city area. These communities are situated in specific areas and rarely overlap with each other. This might imply a relationship between working and living locations. When people choose their accommodation, it is ideal to control the commuting time to work, school, and other facilities. Although travel time depends on various factors, such as the distance of trips, mode of transport, trip purpose, and built-up environment, we suggest further investigation on this matter to clarify the relationship between moving communities and these terms.

Interestingly, we found that the mobility network on weekends has more clusters than that on weekdays. This finding might indicate changes in travel behavior. For instance, commuting to work trips would decrease, and leisure trips would increase on weekends. Additionally, people seem to make shorter trips on weekends compared to working days, as presented in [18].

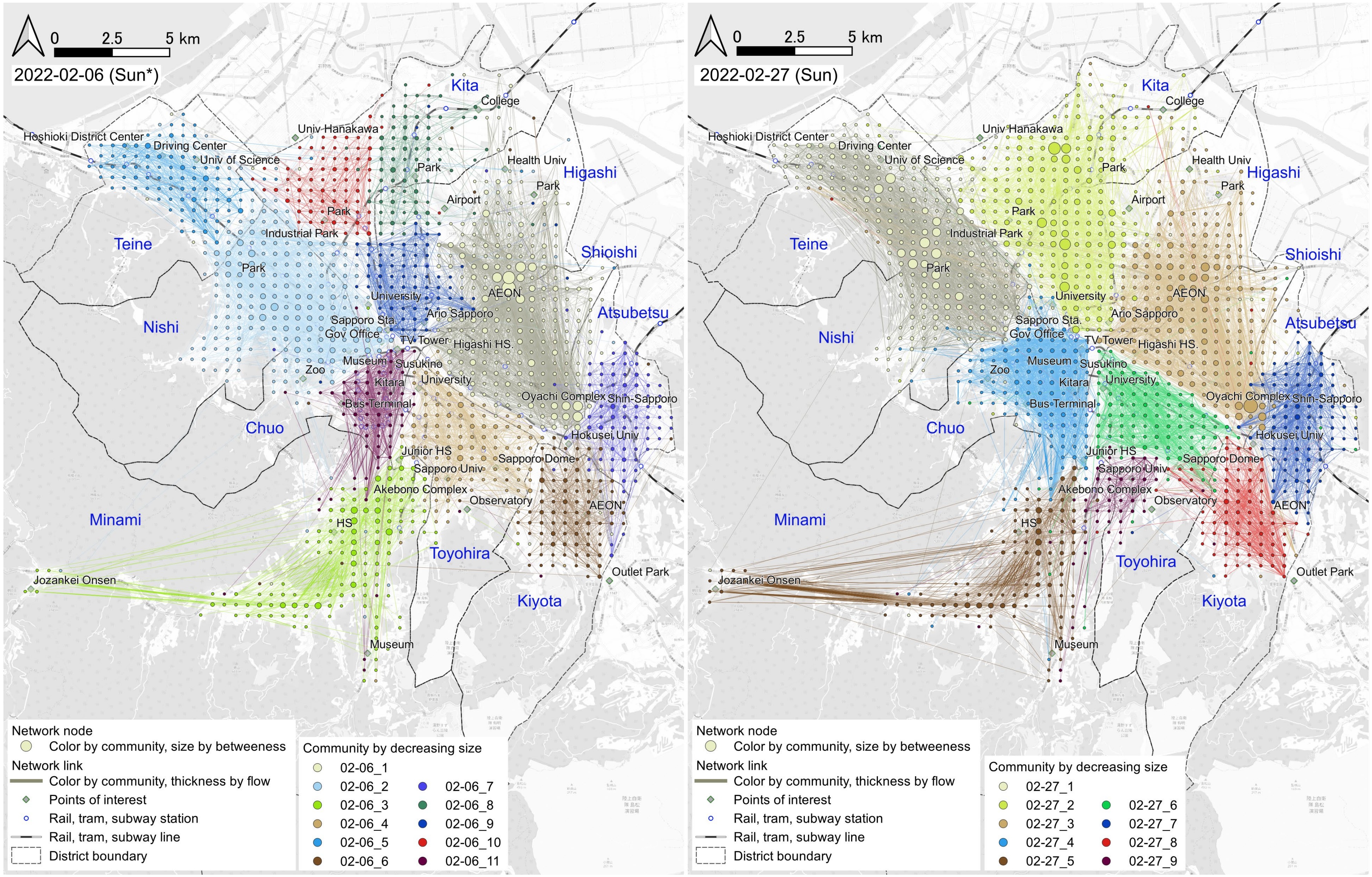


Figure 3. Network communities for the weekend (Left side: Heavy snow on Sunday; Right side: Normal Sunday)

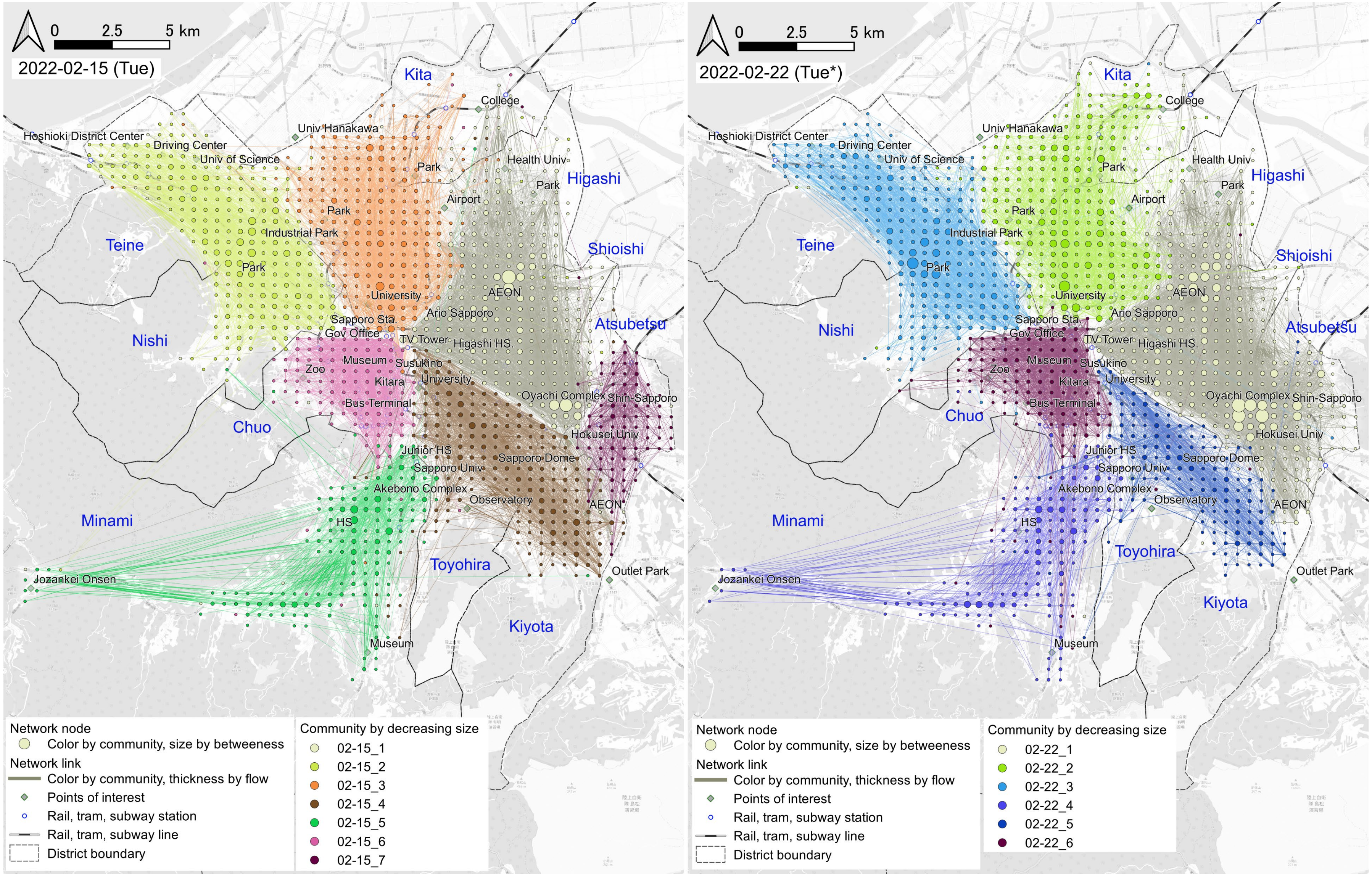


Figure 4. Network communities for weekday (Left side: Normal Tuesday; Right side: Heavy snow on Tuesday)

Notably, the results indicate that the first heavy snow event had a stronger impact on city mobility than the second event. One reason for this is that the snowfall during the first event was heavier and occurred over a shorter period (from the 6th to the 7th of February). At that time, city commuters might not have been prepared for the sudden change in weather conditions, making it difficult for them to make their daily trips. As a result, network mobility was fragmented due to the whiteout. However, awareness about heavy snow seemed to increase after the first event. The city government proactively deployed snow removal patrols before the second heavy snow occurred. Additionally, citizens became more attentive to the weather conditions in the upcoming days. By observing weather forecasts, they could adjust their trip plans, thereby minimizing the impact of the heavy snow.

Lastly, the current study highlights the importance of industrial and commercial areas on all days. These locations act as bottlenecks in the network due to their function as workplaces and logistics hubs. Since they serve the daily demands of the city, a large volume of trips connect to them even on weekends or during extreme weather events. Thus, we suggest that it would be good practice to maintain the connection between these locations and other areas of the city in the case of heavy snow or extreme weather events.

5 Conclusions and Limitations

The current study examined the urban mobility under the heavy snow condition using ETC2.0 probe data and network clustering method. The results demonstrates that heavy snowfall significantly impacts urban mobility networks by causing fragmentation and reorganization of community structures. Key findings include the higher number of large communities on Sundays compared to Tuesdays and the noticeable shifts in community locations during heavy snow days. Locations with high centrality, particularly in industrial and commercial areas, play a crucial role in maintaining network connectivity. The results highlight the need for effective planning and preparedness to minimize disruptions caused by heavy snowfall. Additionally, the study suggests further investigation into the relationship between community structures and travel behavior to enhance urban planning and resilience strategies.

Although ETC2.0 probe data contain rich information, they may cause biases related to the specific population using ETC2.0-equipped vehicles, potentially excluding insights from other modes of transportation. Additionally, aggregating trips by day may not fully reveal the dynamic attributes of the network, particularly during rush hours. We propose that further work should address these shortcomings to enhance the application of ETC2.0 data.

List of Abbreviations

ARI: Adjust Rand Index
ETC: Electronic Toll Collection
ITS: Intelligent Transport Systems
MLIT: Ministry of Land, Infrastructure, Transport, and Tourism
OBUs: Onboard Units
RI: Rand index
RSUs: Roadside Units
V2I: Vehicle-to-Infrastructure
VICS: Vehicle Information and Communication System

Acknowledgements

The authors greatly thank the Regional Road Economy Strategy Study Group, Hokkaido Regional Study Group, and Ministry of Land, Infrastructure, Transport and Tourism, Hokkaido Regional Development Bureau Construction Department Road Planning Division for providing the data for this study.

This research conducted with the support from the National Institute for Land and Infrastructure Management under the Ministry of Land, Infrastructure, Transport and Tourism that contributes to improving the quality of road policy, “R&D on the winter road traffic management system using digital twin”.

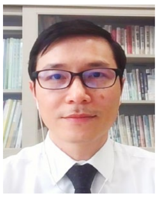
Statements and Declarations

Conflicts of interest: The authors declare that they have no conflict of interest.

References

1. Toriumi, A.: Electronic Toll Collection. In: International Encyclopedia of Transportation. pp. 60–67. Elsevier (2021)
2. MLIT: ETC2.0, https://www.mlit.go.jp/road/ITS/j-html/etc2/index.html
3. NILIM: ETC2.0 Probe Data, http://www.nilim.go.jp/lab/qcg/research/etc-2.0_en.html
4. MLIT: Intelligent Transport Systems, http://www.mlit.go.jp/road/road_e/p1_its.html
5. Katoh, S., Hashimoto, H., Setoshita, S., Matsuda, N.: A study on the effectiveness of utilizing ETC 2.0 probe data to identify bottleneck and its influence range. J. Japan Soc. Civ. Eng. Ser. F3 (Civil Eng. Informatics). 73, I_242-I_250 (2017).
6. Maki, Y., Goto, A., Ooshima, K., Ikeda, Y.: A study on the estimation of capacity at expressway bottlenecks using the ETC 2.0 probe data. Proc. Civ. Eng. Plan. Res. Conf. 58, (2018)
7. Matsushita, H., Hayashi, T.: Quantification of Abrupt Driving Maneuver Utilizing ETC 2.0 Probe Data: A Case Study in Japan. 8th Int. Conf. Logist. Informatics Serv. Sci. LISS 2018 - Proceeding. 1–6 (2018).
8. Sekine, Y., Yamamoto, T., Hayashi, Y., Ohtsubo, Y.: An analysis of the relationship between traffic congestion on

main roads and the traffic environment in surrounding areas (An evaluation of residential road safety from analysis of traffic accident and ETC 2.0). Trans. JSME (in Japanese). 87, 20-00392-20–00392 (2021).
9. Kaneko, N., Oka, H., Chikaraishi, M., Becker, H., Fukuda, D.: Route Choice Analysis in the Tokyo Metropolitan Area Using a Link-based Recursive Logit Model Featuring Link Awareness. Transp. Res. Procedia. 34, 251–258 (2018).
10. Goto, A., Itsubo, S., Maki, Y., Ikeda, Y.: Estimation of route choice models on the tokyo metropolitan expressway network using the ETC2.0 probe data. Proc. Civ. Eng. Plan. Res. Conf. 58, 1–8 (2018)
11. Homma, S., Ota, K., Yaginuma, H., Onomura, S., Kataoka, T., Nihei, Y.: Case study for monitoring inundation process during floods using ETC 2.0 probe data. J. Japan Soc. Civ. Eng. Ser. B1 (Hydraulic Eng. 76, I_721-I_726 (2020).
12. Dilley, M., Chen, R.S., Deichmann, U., Lerner-Lam, A., Arnold, M., Agwe, J., Buys, P., Kjekstad, O., Lyon, B., Yetman, G.: Natural disaster hotspots: A global risk analysis. (2005)
13. NSPRPB: Heavy snow type code, https://nlftp.mlit.go.jp/ksj/gml/codelist/HeavySnowTypeCode.html
14. MLIT: Roads in Japan. Road Adminstration Japan. 1–39 (2021)
15. Kyodo: Heavy snow batters northern Japan, disrupting rail and air traffic, https://www.japantimes.co.jp/news/2022/02/21/national/snow-transport-northern-japan/
16. Tandler, A.: Sapporo is Snowed In, and it's All Natural, https://japan-forward.com/sapporo-is-snowed-in-and-its-all-natural/
17. Taniguchi, T.: Sapporo sees record 60 cm of snow in 24 hours; trains brought to halt, https://mainichi.jp/english/articles/20220207/p2a/00m/0na/005000c
18. Ha, T.V., Sato, Y., Van, H.T.K., Arimura, M.: A descriptive analysis of heavy snow effects on travel demand patterns using ETC2 . 0 probe data , the case study of Sapporo city , Hokkaido , Japan . Proc. East. Asia Soc. Transp. Stud. 14, (2023)
19. Javed, M.A., Younis, M.S., Latif, S., Qadir, J., Baig, A.: Community detection in networks: A multidisciplinary review. J. Netw. Comput. Appl. 108, 87–111 (2018).
20. Dantsuji, T., Sugishita, K., Fukuda, D.: Understanding changes in travel patterns during the COVID-19 outbreak in the three major metropolitan areas of Japan. Transp. Res. Part A Policy Pract. 175, 103762 (2023).
21. Dantsuji, T., Hirabayashi, S., Ge, Q., Fukuda, D.: Cross Comparison of Spatial Partitioning Methods for an Urban Transportation Network. Int. J. Intell. Transp. Syst. Res. 18, 412–421 (2020).
22. Xie, C., Yu, D., Zheng, X., Wang, Z., Jiang, Z.: Revealing spatiotemporal travel demand and community structure characteristics with taxi trip data: A case study of New York City. PLoS One. 16, 1–21 (2021).
23. Yildirimoglu, M., Kim, J.: Identification of communities in urban mobility networks using multi-layer graphs of network traffic. Transp. Res. Procedia. 27, 1034–1041 (2017).
24. Rosvall, M., Axelsson, D., Bergstrom, C.T.: The map equation. Eur. Phys. J. Spec. Top. 178, 13–23 (2009).
25. Rosvall, M., Bergstrom, C.T.: Maps of random walks on complex networks reveal community structure. Proc. Natl. Acad. Sci. U. S. A. 105, 1118–1123 (2008).
26. Clauset, A., Newman, M.E.J., Moore, C.: Finding community structure in very large networks. Phys. Rev. E - Stat. Physics, Plasmas, Fluids, Relat. Interdiscip. Top. 70, 6 (2004).
27. Hubert, L., Arabie, P.: Comparing partitions. J. Classif. 2, 193–218 (1985).
28. Rand, W.M.: Objective Criteria for the Evaluation of Clustering Methods. J. Am. Stat. Assoc. 66, 846–850 (1971).
29. Qannari, E.M., Courcoux, P., Faye, P.: Significance test of the adjusted Rand index. Application to the free sorting task. Food Qual. Prefer. 32, 93–97 (2014).



Tran Vinh Ha completed his Ph.D. and currently works as a postdoctoral research fellow at Muroran Institute of Technology. His research interests include transportation, ITS, data mining, disaster prevention, urban planning, and infrastructure management.

Mikiharu Arimura completed his Ph.D. in the Department of Civil & Architectural Engineering at Muroran Institute of Technology in 2000. He is currently a professor in the College of Design and Manufacturing Technology at Muroran Institute of Technology. His research interests include civil engineering, urban planning, transportation planning, big data analysis, and ITS.