Souknilanh Keola, “Measuring Connectivity Within and Among Cities in ASEAN,” BRC Research Report, Bangkok Research Center, JETRO Bangkok/IDE-JETRO, 2019

Chapter 2
Measurement of Inter- and Intra-city Connectivity Using Vehicle Probe Data

Hiroyuki Miyazaki[#]

Asian Institute of Technology, Thailand

Abstract: This chapter analyzes intra- and inter-city connectivity using the vehicle probe data for selected 48-hour slots in March and September in 2017 and 2018. I demonstrate the potential analyses by aggregating the probe data of commercial vehicles with overlay to geographical extents of the major cities identified by night-time light satellite image data. The cities could be classified into more vehicles in the daytime or night time, which were likely associated with drivers’ preference on traffic conditions by the time. Some cities indicated notable changes of driving speeds by the time, possibly owing to traffic condition with people’s commuting as well as transport infrastructure, such as highways. More than half of the vehicles were traveling only two cities within the 48-hour periods, which were possibly shuttle trips between two cities. Some cities in the large industrial areas and inland cities indicated high proportion of vehicles were travelling more than two cities, indicating contribution to connectivity among the city.

Keywords: Urban, Connectivity, Probe data

JEL Classification: O18, C55, C80

---

1. Introduction

Connectivity is a crucial aspect of the dynamics in socio-economic activities. For example, the movement of people is challenged in developing an application such as migratory flow estimation, dynamic urban population, and urban mobility patterns (Barbosa et al., 2018). This could be helpful for urban planning and management issues in urban areas, including the equality of service access. Moreover, in order to provide efficient and equitable essential services to the public without fighting the increasing congestion and population in metropolitan areas (Yang et al., 2018). Including hidden populations in urban areas such as Bangkok that create a difficult decision for the

[#] Autor: Hiroyuki Miyazaki; Address: Km.42 Phaholyothin Highway Klong Luang Pathumthani 12120 Thailand. Tel: +66-2-524-6394; Fax: +66-2-524-5597;
E-mail: heromiya@csis.u-tokyo.ac.jp.

government to provide the necessary infrastructure facilities and services for the transit system.

For this reason, geospatial data analysis can be used to understand and present the current status of urban activities. However, location information of human activity is not always available, particularly at a temporal resolution (Z. Liu et al., 2018). Integrating various data sources can be used to present transportation services in big cities that can support urban planning development and management.

Recently, several types of research about connectivity using mobility data have been studied in many big cities, such as New York (Qian & Ukkusuri, 2015), Boston (Malleson et al., 2018), Shanghai (X. Liu, Gong, Gong, & Liu, 2015), etc. Thereby, developing a novel approach or an application which can be useful and helpful to study the connectivity within cities. This chapter provides a review of past publications and a preliminary data analysis using vehicle probe data, which paves the way for future research.

2 Literature Reviews

In this section, we present reviews on data sources of probe data and analysis methods for probe data. The following subsections are with details of the reviews.

2.1 Types of Probe Data

2.1.1 Census Data and Surveys

Census data is collected by periodical national surveys in which householders are asked questions about the socio-demographic and economic status of the household's members. In terms of mobility there are questions related to the location of the workplace, or the place of current and previous residence. Collectively, this information can then be used to estimate commuting flows or internal migration flows within a country (Barbosa et al., 2018). Nowadays, national censuses are held in most countries, typically every ten years.

According to the National Statistical Office of Thailand, the population and housing census is collection of data of the population and their place of residence for everyone who currently resides in the country. It is equivalent to a photo “snap-shot” on the date of the census to show how many people on that date are living in Thailand, where are they living, how many are males/females, children, in the labor force, elderly, disabled, what is their level of education, literacy, employment status, occupation, what type of residence they live in, how hygienic is that residence, and how often do they move, among other issues.

2.1.2 Urban Geospatial Database

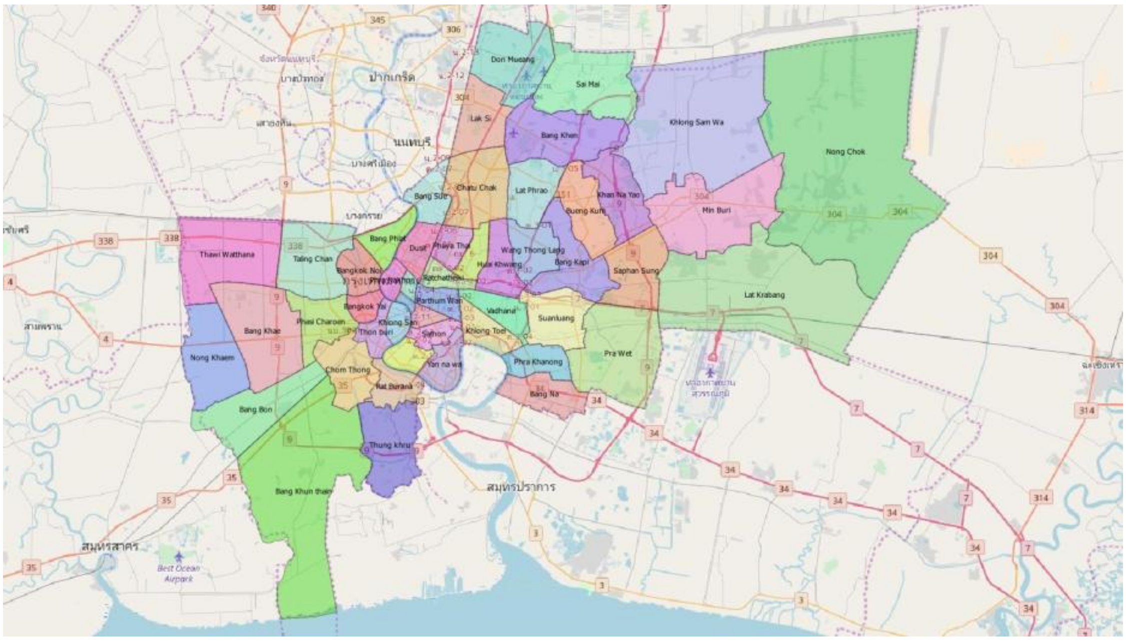
Several geospatial datasets for cities have been published and provided by governmental agencies and private companies. The datasets are typically related to location information or the coordinate system for latitude and longitude in the real world. Mainly, there are two types of GIS data, called spatial data and attribute data. Spatial data describes the absolute and relative location of geographic features and attribute data describes the characteristics of spatial features, which can be quantitative or qualitative. Usually, attribute data is often referred to as tabular data.

Regarding human mobility analysis, several geospatial urban data sets have been used as input data like the following:

- Land use data
- Access to bus and subway services
- Mobile base stations
- Boundary data
- Open Street Map

Example of metro stations in Washington D.C. (Z. Liu et al., 2018), the geolocation of the metro stations and bus stops in the entire metropolitan area were used to measure the access to other transport modes. This data was collected from the D.C. GIS Open Data Site and the General Transit Feed Specification (GTFS) website.

Figure 1. Example of Bangkok's boundaries overlaid on the Open Street Map



Source: Author based on Open Street Map.

2.1.3 Mobile Telephone Data

Mainly, there are data sources for mobile phone data collection in this review. The first data set covers calls Call Detail Record (CDR) data. Most cell phone companies maintain detailed records of customer information related to their calls and text messages. The information contained in these CDRs includes the time of the call or message as well as the company’s cell tower routing the communication (Barbosa et al., 2018). CDR data covers both phone calls and Short Message Service (SMS) exchanges, and always includes the following information: time stamp, caller ID, recipient ID, call duration, and the antenna's (cellular tower) code.

The second data set collects information through a mobile application for smartphone location data (Malleson et al., 2018), which contains location information from the installed application on the mobile phone. The application runs in the background of the phone’s operation, automatically keeping track of the user’s movements using the device’s motion co-processor to record the time and movement of the phone. Geographic information is assigned to such activities using the device’s geolocation service, including assisted GPS, tower-based positioning, and satellite-based GPS.

2.1.4 Vehicle Probe Data

The greatest level of accuracy on movement trajectory is provided by Global Positioning System (GPS) data, which can provide precise information on any location covered by at least four GPS satellites. GPS trackers are units which receive signals from GPS satellites and compute the device’s position at regular intervals. This technology allows researchers to trace the movement of an individual with a high degree of accuracy and temporal frequency, thus providing a rich source of data that can be analyzed and mapped directly to a person’s mobility pattern (Barbosa et al., 2018).

Taxi GPS data and bus route trajectories collected from GPS locations are used to map where people are in terms of location distribution and pattern, such as to explore travel patterns and a city's structure using taxi GPS data (X. Liu et al., 2015), taxi cruising behavior using the taxi's GPS trace (Zong, Wu, & Jia, 2018), urban human activity using taxi GPS data (Tang, Liu, Wang, & Wang, 2015), etc.

Figure 2. Example of GPS taxi data overlaid to an Open Street Map

Source: Author based on Probe Data and Open Street Map.

2.1.5 Comparison of Data Sources

Table 1 presents a comparison of data sources with advantages and disadvantages.

Table 1. Data Sources Comparisons

	Data source	**Advantage**	**Disadvantage**
Spatial Data	GIS Data: Point Feature Accessibility points Mobile base stations CDRs Smartphone location Taxi, bus GPS data GIS Data: Line Feature Bus trajectory Taxi trajectory GIS Data: Polygon Feature Land use data Boundary data\	Easy to visualize Large scale - Useful to detect patterns Precise location accuracy Overhead processing	Accuracy of data Competency of data Duplicated data sources Privacy issues for CDR data
	Voluntarily collected data - Open Street Map	Large coverage area	- Resolution of data
Non-Spatial Data	Census Data and Survey Population Households Employment	Individual details National coverage Multiple dimensions	Time-consuming to survey Less frequent Outdated

Source: Author.

2.2 Analysis Methods for Probe Data

Many types of research have been published related to human mobility, using various data sources such as census data, surveys, GIS data, as well as GPS data from taxis, buses, mobile phones, etc. In order to better understand and obtain innovative ideas, eleven research publications have been selected to review. For the selection of research papers step, a few keywords are used to identify the most relevant research papers related to urban mobility, such as urban structure, urban mobility, community detection, spatial variation, etc.

To identify and justify the movement of people, several methods and approaches are applied. In this section provide an example of the methodology that recent researchers have used to process outcomes which can clarify people's mobility patterns.

2.2.1 Regression Model

Ordinary Least Square (OLS) regression models are applied to relate taxi demand with land use characteristics and transit supplies. The case study is in the Washington D.C. metropolitan area (Yang et al., 2018).

2.2.2 Clustering Model

Given that a small spatial unit usually has the same land use, we can aggregate trips to obtain the spatial interaction among these small regions. The small units could be traffic analysis zones (TAZs), grids, or parcels segmented by major roads. By treating the units as nodes, and the movement flows as edges, we can construct spatially embedded networks and apply complex network methods to study their properties and structures further. Community detection methods can partition an entire network into tightly connected sub-network communities, and reveal that network’s clustering characteristics (Liu et al., 2015).

2.2.3 Origin-Destination Metrix

This review's study is taxi trips is classified into two parts based on their status: the first is picking up passengers from origin to destination, and the second is cruising on the road to find the next passenger. The overall distribution of the origins and destinations reflect the travel demand by the citizens (Tang et al., 2015). The results show that the origins and destinations of most trips are located in the center of the city as in the figure below.

2.2.4 Gravity Model

In 1946, George K. Zipf proposed an equation to calculate mobility flows inspired by Newton’s law of gravity (Barbosa et al., 2018). In his work, he highlighted the importance of the distance for human migration patterns where the magnitude T_{ij} of a migratory flow between two communities i and j can be approximated by:

$T_{ij} \propto P_i r_{ij} P_j$,

where P_i and P_j are the respective populations, and r_{ij} the distance between i and j. The basic assumptions of this model are that the number of trips leaving i is proportional to its population, the attraction of j is also proportional to P_j, and finally, that there is a cost effect in terms of the journey's distance.

According to Chen et al. (2018), the concept of a gravity model is applicable to find the relationship with health care services. The extracted hospital visits are estimated based on the gravity model (Hansen, 1959), so that the influence of hospital capacity and spatial impedance on health care activities can be obtained. In this study, we have reviewed taxi data to characterize the demand-supply relationship of the health care services in Guangzhou, China. The hospital visits during one week are analyzed using the gravity model. The estimated coefficients of the gravity model suggest a significant distance-decay effect for hospital visits.

2.2.5 Summary of the Methods

Table 2 presents the reviews of recent methods of analyzing the probe dataset.

Table 2. List of references to methods for analyzing probe data

Paper title	Year	Study area	Periods	Data	Methods	Results	Advantage	Remained issues
Spatial variation of the urban taxi ridership using GPS data	2015	New York City, USA	Dec 2008 to Jan 2010	Daily trips (about 500,000 trips) Demographic data Land use Accessibility to bus and subway	Ridership analysis using spatial variation Geographically Weight Regression (GWR)	Good explanatory accuracy. A significant impact of urban form on urban taxi ridership and strong spatial variability. Relationships between income and taxi trips Accessibility to subway promotes taxi ridership.	Providing an in-depth understanding of how explanatory variables vary locally. Eliminating potential bias from spatial heterogeneity Comprehensive understanding of the impact of independent variables.	The level of demographics and socioeconomics of a particular area matters, the functionality and the geographical location of the place are also crucial factors when determining the number of ridership.
Analysis of Washington, DC taxi demand using	2018	Washington D.C., USA	May 2015 to April 2016	Taxi ridership log All public vehicle for	Ordinary Least Square (OLS) regression models to relate	A strong link between demand for taxi, land use patterns, and	OLS models can be used for Investigating the correlation	Consistent with the model based on daily ridership, the correlation with

Paper title	Year	Study area	Periods	Data	Methods	Results	Advantage	Remained issues
GPS and land-use data				hire trips - Land use data Demographic data	taxi demand with land use characteristics, and transit supplies.	accessibility to other modes. The taxi mode is likely to complement metro trips but compete with bus trips, although both modes of travel are considered public transit.	between demand for taxi, land use patterns, and accessibility to other modes using detailed GPS and GIS information	the number of bus stops was still negative during midnight and early morning although the magnitude was much smaller. More research is needed to address the different impact of the metro and buses on taxi demand.
Revealing travel patterns and city structure with taxi trip data	2015	Shanghai, China	1 June 2009 to 4 June 2019 (4 days)	Taxi trip data about 6,600 taxis - Lands use data	Complex Network Science (weighted and directed network)	The community detection result mainly consists of large, spatially continuous regions indicating the influence of the distance decay effect on spatial interactions.	Collective intra-city trips extracted from emerging taxi GPS trajectory data can explore travel patterns and city structure. Taxi data are reasonable to represent intra-city spatial interactions and reveal city structure.	Communities with less than ten cells using 1 km cell are not displayed. Taxi trips are only able to represent a part of intraurban travel. Some travel, such as long-distance commuting, is hardly reflected by taxi trajectories. Further study can use more data sources such as a

Paper title	Year	Study area	Periods	Data	Methods	Results	Advantage	Remained issues
								private car, bus, metro trip, CRD, etc.
Mapping hourly dynamics of the urban population using trajectories reconstructed from mobile phone records	2017	Beijing, China	6 Dec 2015	CDR (about 600 million records) from 27 million users 20,790 mobile base stations	Mobility pattern using ROG (Radius of Gyration). Non-linear and self-learning BP neural networks. A hierarchical clustering algorithm	The time series of hourly population estimates can be used to group blocks based on similar patterns and then delineate urban functional areas	Hourly population density can be estimated by examining the time series of individual trajectories. The hourly population dynamic is also useful in studies delineating urban functional areas.	Information about human activity is not covered - The time series of hourly population estimates can be used to group blocks based on similar patterns and then delineate urban functional areas. Dynamic
Taxi Drivers’ Cruising Patterns Insights from Taxi GPS Traces	2018	Shenzhen, China	18 Nov 2011 to 26 Nov 2011 (9 days)	GPS traces from 2,536 taxis land use data	Probabilistic weighted distance ZINB (zero-inflated negative binomial) models	Drivers follow different patterns during various times of day relying on traffic conditions and land use in the morning and evening peak hours. High-earning drivers and roaming drivers	The cruising patterns revealed by the ZINB models align with the cruising patterns from the taxi survey data - external factors (land use, traffic conditions, and road grade) and internal factors (previous pick-up	In reality, there may be other internal factors that have potential effects on taxis’ cruising behavior, such as drivers’ cruising preferences or familiarity with street networks. Other external factors, such as

Paper title	Year	Study area	Periods	Data	Methods	Results	Advantage	Remained issues
						prefer to cruise in areas with high-density land use and optimal traffic conditions, whereas low-earning drivers and target drivers tend to cruise in areas with more previous pick-up points	experience) affect cruising behaviors, although external factors appear to have a stronger influence than internal ones	driving distance and waiting time, may also have an impact on cruising decisions.
Discovering Corridors from GPS Trajectories	2017	Dublin, Ireland	Not mentioned	5,000 buses' trajectories	Minimum description length (MDL) principle Corridor mining	The movements of several buses that have different origins and destinations, but they share a common movement behavior where is detected as corridors.	A large collection of bus trajectories to a few numbers of frequently accessed paths can be detected corridors where buses have shared common routes.	Using taxi trajectories may detect smaller corridors
Uncovering urban human mobility from	2015	Harbin, China	July 2012 to December	Taxi GPS data from	DBSCAN OD matrix	By extracting taxi trips from GPS data, travel	The distribution patterns of origins and destinations on	Travel distance, time and speed are used to explore

Paper title	Year	Study area	Periods	Data	Methods	Results	Advantage	Remained issues
large scale taxi GPS			2012	about 1,100 drivers		distance, time and average speed in occupied and unoccupied status can be used to investigate human mobility	weekday and weekend ban be analyzed by using characterize people travel movement from taxi GPS data	human mobility by extracting taxi trips from GPS trace data. But, at large scale of the city is lacking.
Mapping the spatial disparities in urban health care service using taxi trajectory data	2018	Guangzhou, China	5 May 2009 to 11 May 2009 (7 days)	GPS records from taxi about 172 million records	Gravity model	Taxi data is used to find the spatial disparities in health care service access. About 21.05% of the total population has high hospital accessibility, while the remaining 78.95% has relatively low hospital accessibility	Taxi data can be used to characterize the actual demand-supply of health care services. Improve the understanding of spatial inequalities in public service provision	Others public service provisions can be applied to study spatial inequalities using taxi data

Paper title	**Year**	**Study area**	**Periods**	**Data**	**Methods**	**Results**	**Advantage**	**Remained issues**
Agent-Based Modeling of Taxi Behavior Simulation with Probe Vehicle Data	2018	Bangkok, Thailand	1 June 2015 to 31 July 2015	Taxi Probe GPS data from about 10,000 taxis	Agent-based simulation model OD analysis	A significant level of similarity of different taxi behavior, such as trip generation; trip time; trip distance as well as trip occupancy, based on its distribution.	Grid-based behavior of taxi	The current agent-based model system utilizes an offline learning method.
The characteristics of asymmetric pedestrian behavior A preliminary study using passive smartphone location data	2018	Boston, USA	15 May 2014 to 1 May 2015	- Passive smartphone data from 6,424 users - Open Street map	- OD matrix (Transition matrix) - Hidden Markov Model	People change their route approximately 15% of the time. Although this varied little when observing trips made at the weekend or on a weekday, people taking journeys.	A greater understanding of the degree of asymmetry in route choice has the potential to lead to a better understanding of the factors that influence walkability and ultimately inform many urban modeling and design initiatives.	The subsequent highlighting of these travel behaviors may implicitly reveal factors about the built environment amenities, stores, and social conditions that may explain why certain route choices are made along one path versus another.

Source: Author.

3 Data Analysis Results

In this section, we demonstrate some preliminary analysis using probe data to measure intra- and inter-city connectivity.

3.1 Input Data

The probe data we use is a subset of the GPS logs of commercial vehicles, such as trucks and taxis, collected by Toyota Tsusho Nexty Electronics Thailand. The data was sampled according to the following criteria:

a) Appeared in predefined border areas or

b) appeared in more than two predefined cities

Between

a) 5th and 6th March 2017

b) 13th and 14th September2017

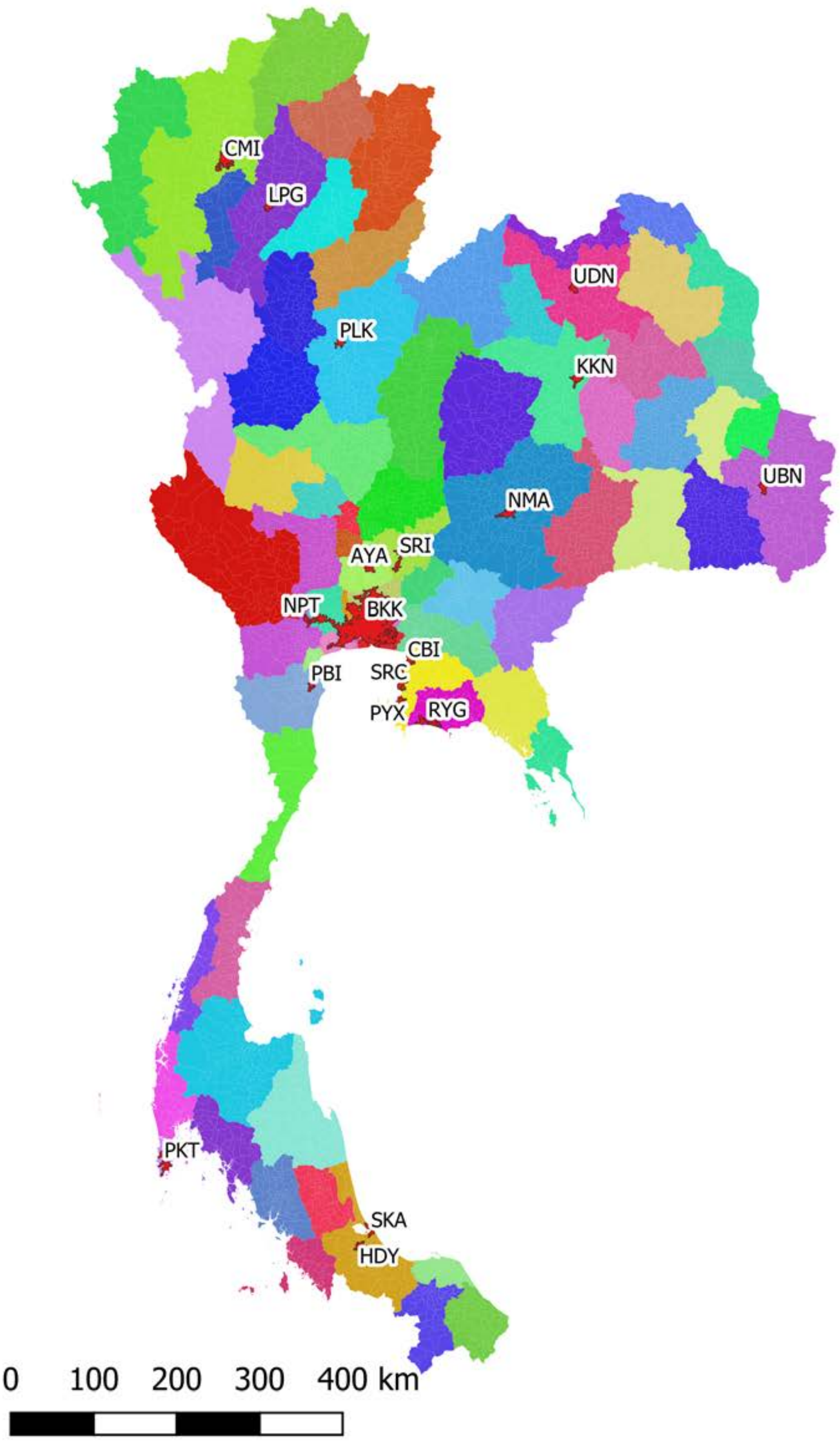
c) 4th and 5th March 2018

d) 12th and 13th September 2018

The number of records is 140,815,982 in total.

The probe data was analyzed through spatial aggregation of various areas of the cities identified from the night-time light satellite images by the author (Figure 3).

Figure 3. Location of the target cities



Source: Author based on Shapefile prepared in Chapter 1.

3.2 Intra-City Analysis

We defined intra-city connectivity with the number of vehicles within a city as an indicator of the capacity of the city, and average driving speed as an indicator of the flow within a city. As both indicators vary notably by time, we calculate the indicators in hours. The greatest number of vehicles is likely indicating the capacity of a city, whereas the driving speed is measured during the daytime, especially at the time with the highest number of vehicles. The indicators are useful to compare the intra-city connectivity of the cities.

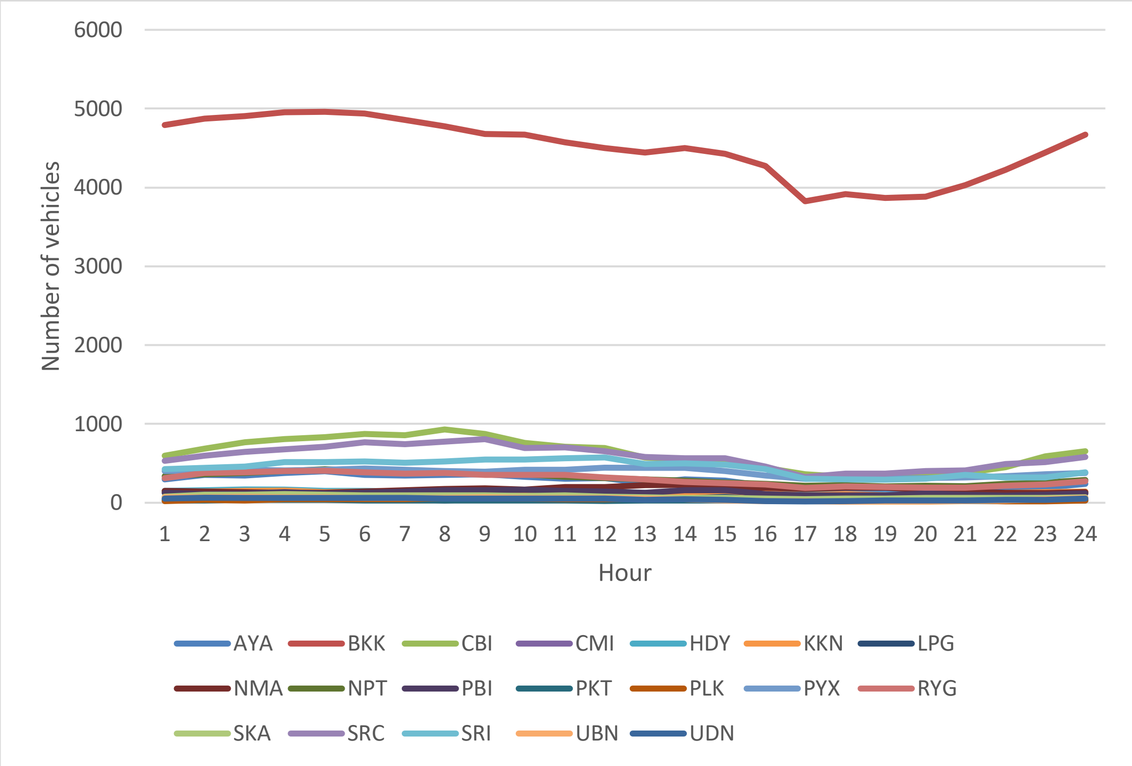
3.2.1 Number of Vehicles per Hour

The number of vehicles is calculated through summarizing the number of unique vehicles contained within the city's boundaries by the hour (i.e., 0:00-1:00, 1:00-2:00, … 22:00-23:00, 23:00-00:00; Figure 4 - Figure 7 for all cities; Figure 8 - Figure 11 for the cities without BKK). The results indicate that cities are likely classified into two types: more vehicles in the daytime, or more vehicles in the night time.

a) More vehicles in the daytime: KKN, NMA, PBI, PKT, PLK, PYX, RYG, SRC, SRI, UDN
b) More vehicles in the night time: AYA, BKK, CBI, CMI, HDY, LPG, NPT, SKA, UBN

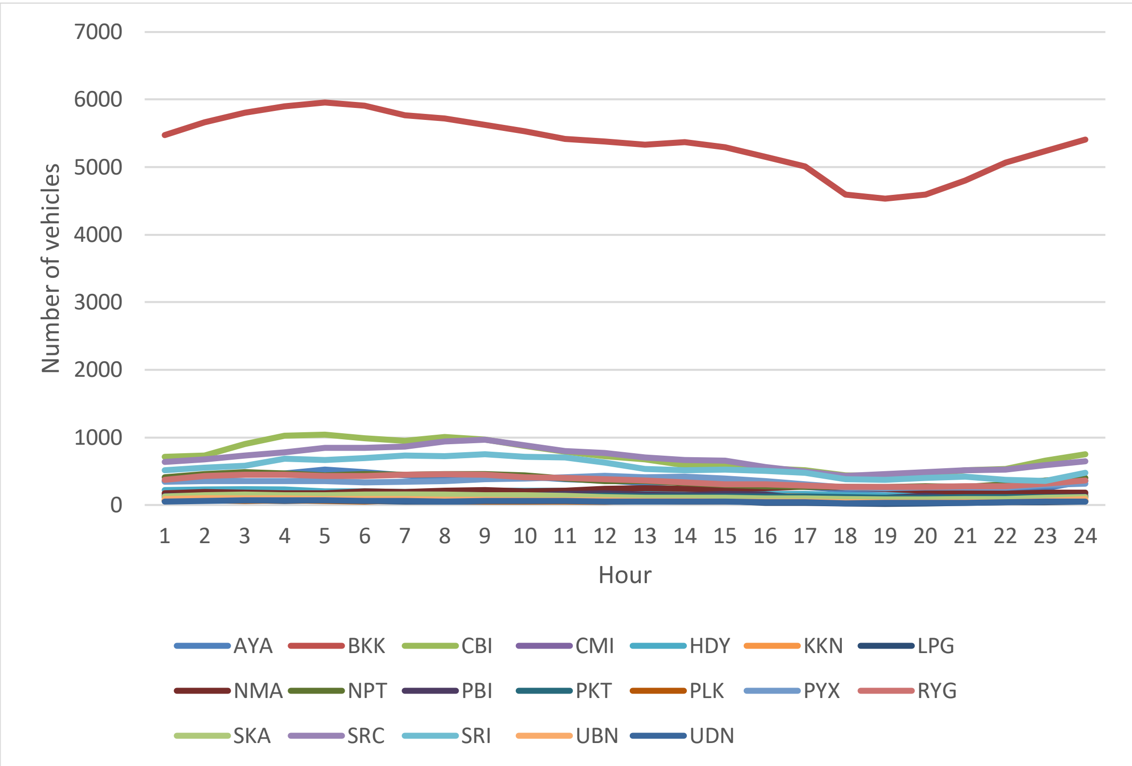
Because the most populated cities, such as BKK and CMI, have more vehicles in the night time, commercial vehicles seem to avoid the heavy traffic in the daytime and mainly drive in the cities in the night time. This is only for trucks, not taxis. Taxis could be counted as the baseline number of vehicles in the city. In contrast, cities with fewer vehicles in the daytime likely have little heavy traffic in the daytime. Therefore, truck drivers prefer to drive there in the daytime.

Figure 4. Number of vehicles per hour for 5-6 March 2017



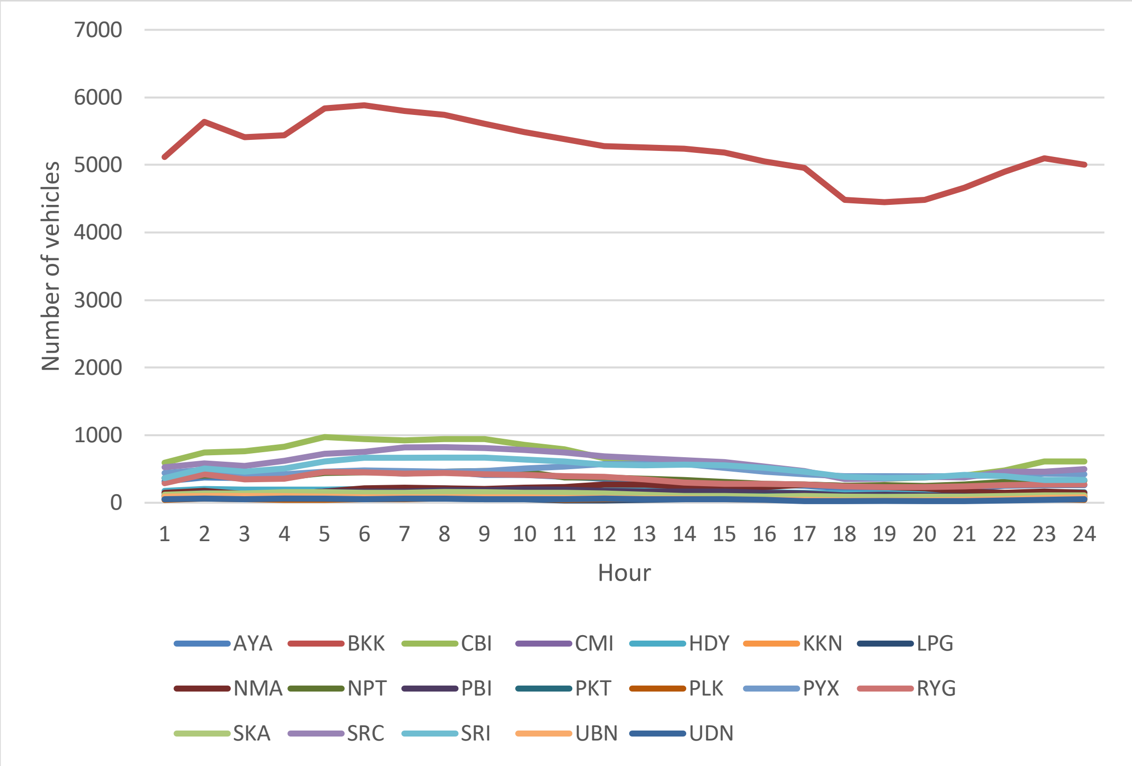
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 5. Number of vehicles per hour for 13-14 September 2017



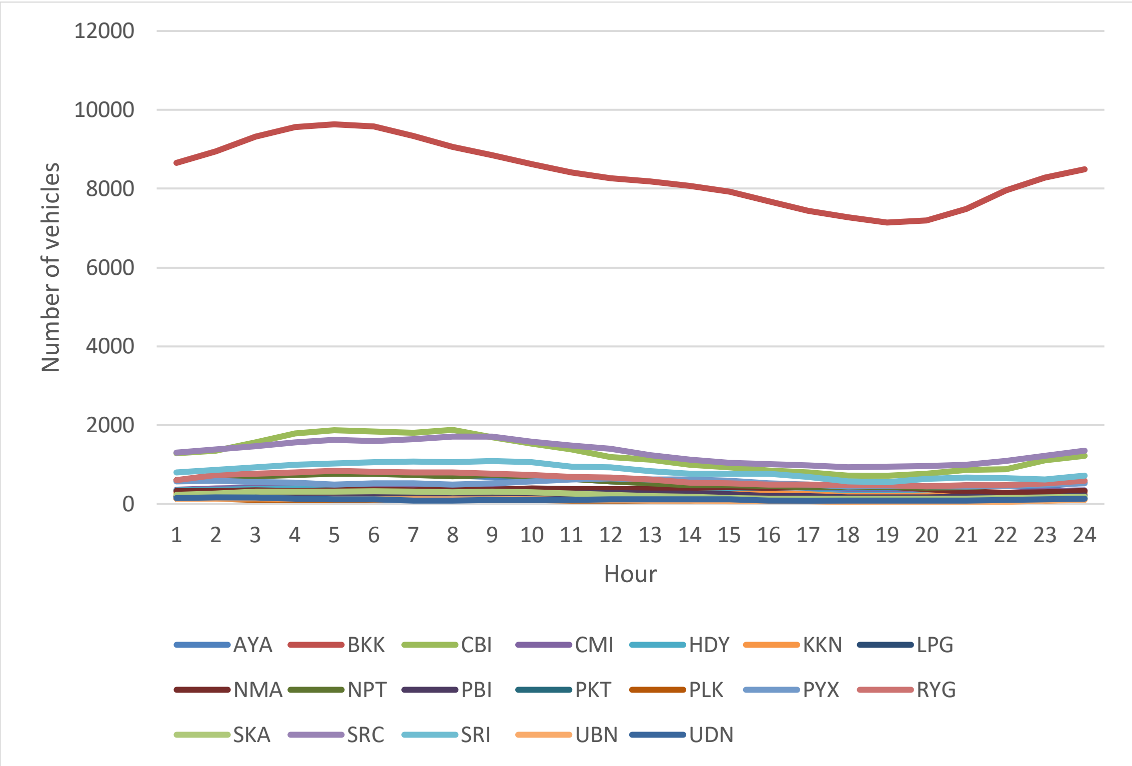
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 6. Number of vehicles per hour for 4-5 March 2018



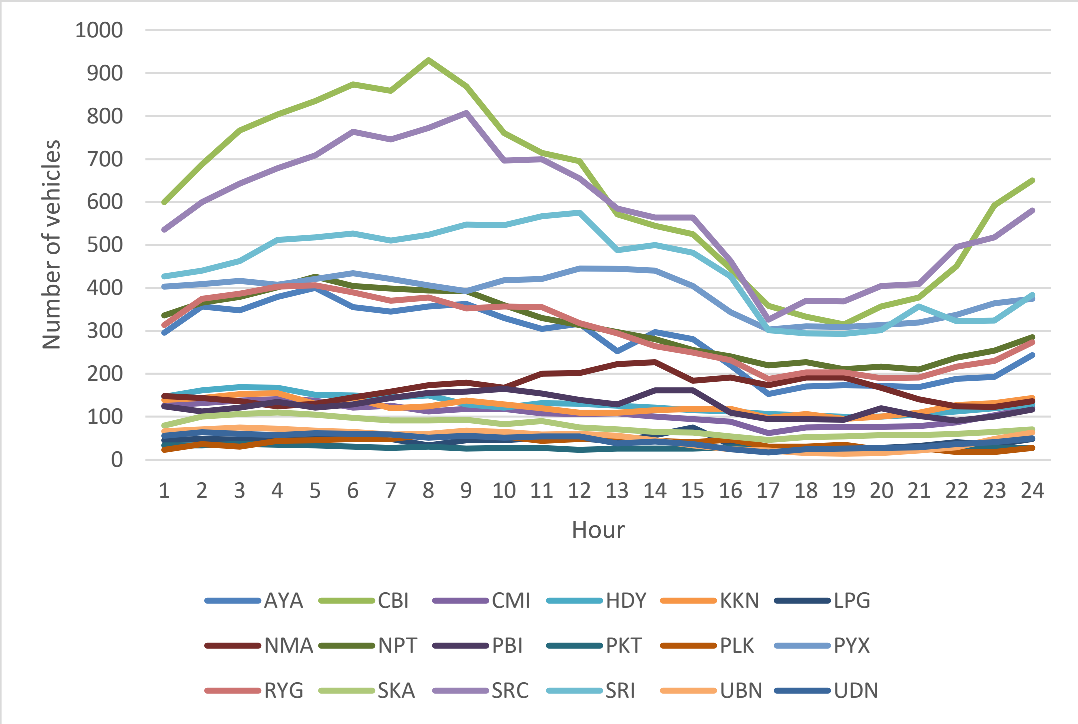
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 7. Number of vehicles per hour for 12-13 September 2018



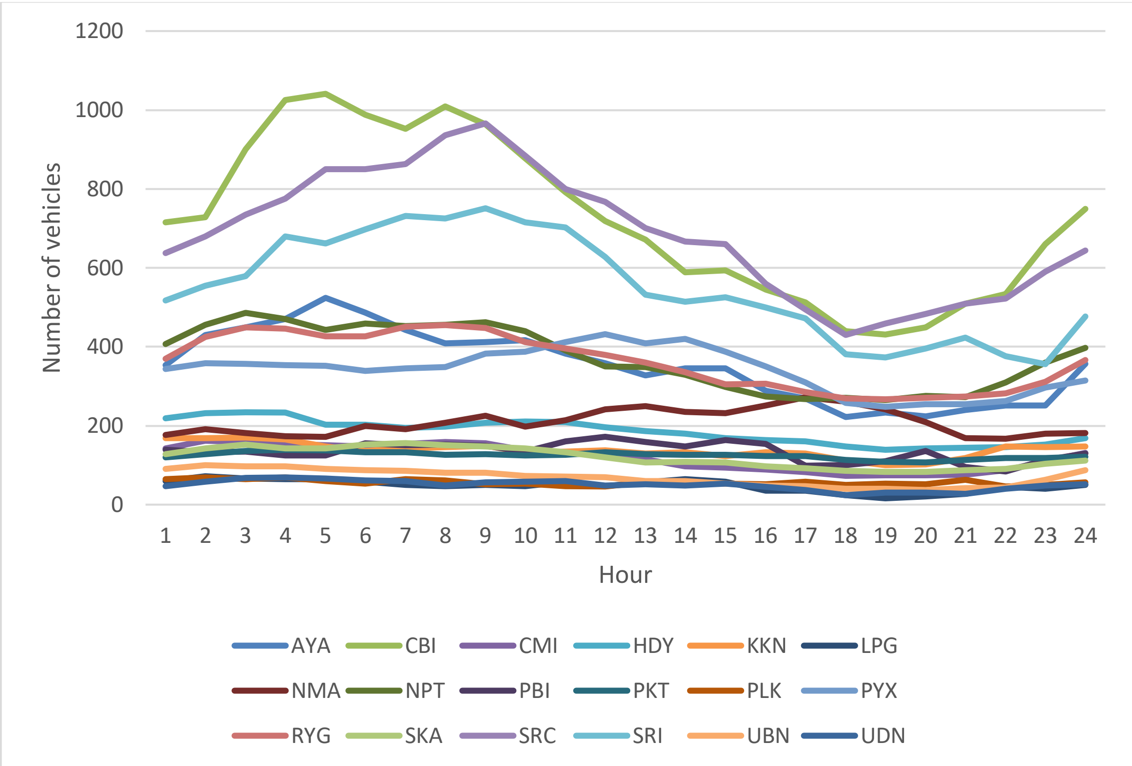
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 8. Number of vehicles per hour without Bangkok for 5-6 March 2017



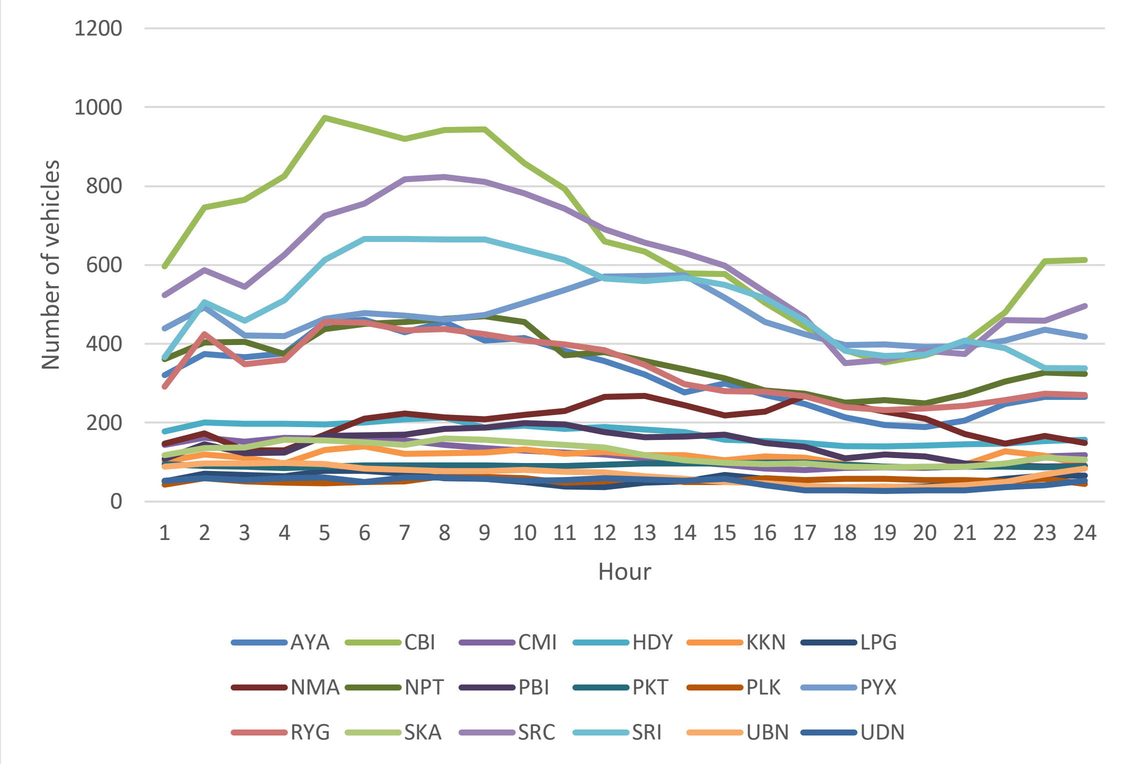
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 9. Number of vehicles per hour without Bangkok for 13-14 September 2017



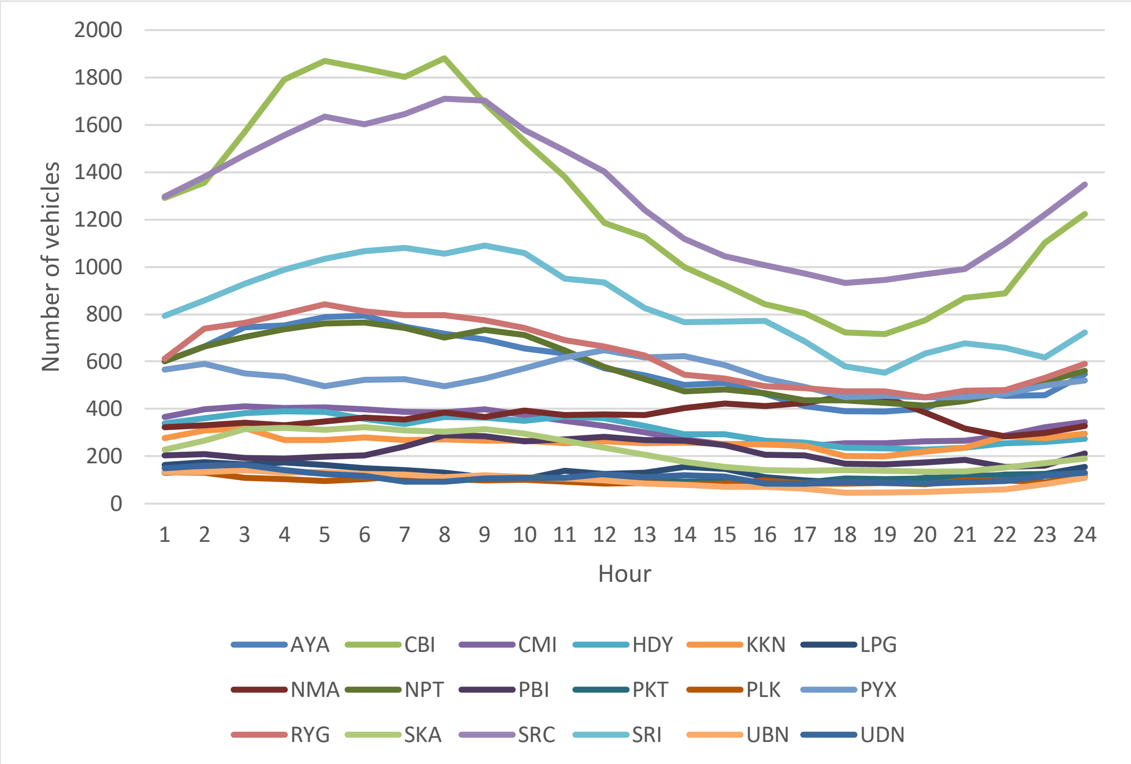
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 10. Number of vehicles per hour without Bangkok for 4-5 March 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 11. Number of vehicles per hour without Bangkok for 12-13 September 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

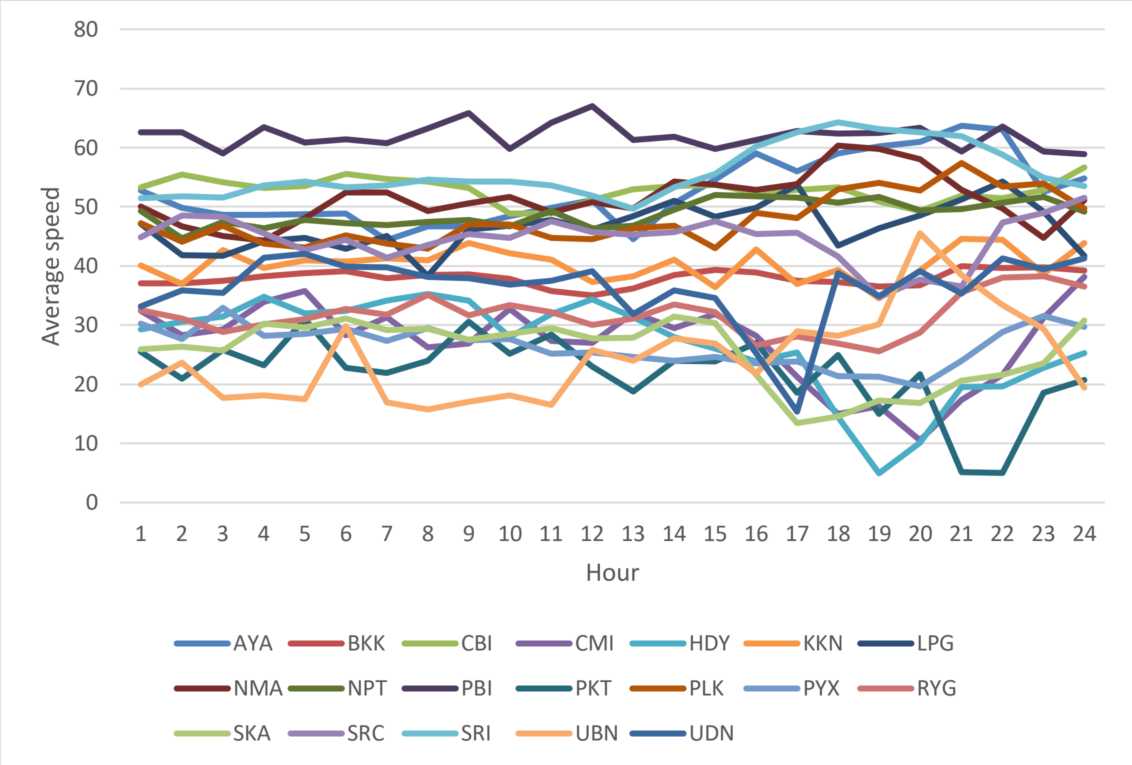
3.2.2 Average Driving Speed per Hour

The average driving speed is calculated as the mean speed of vehicles within the city boundary by the hour. Records with zero speed are removed from the mean. Therefore, the average is only for moving vehicles, and does not including parked vehicles.

In the results shown in Figure 12 to Figure 15, slower speeds were typically indicated around 16:00-19:00hrs., which is likely because of reduced speed for safety at sunset and go-home time. Some cities, such as AYA, NPT, PBI, PLK, SRI, UBN, and UDN, maintain the driving speed in the daytime, or have even faster speeds. The cities likely to benefit have highways with less risk of traffic incidents.

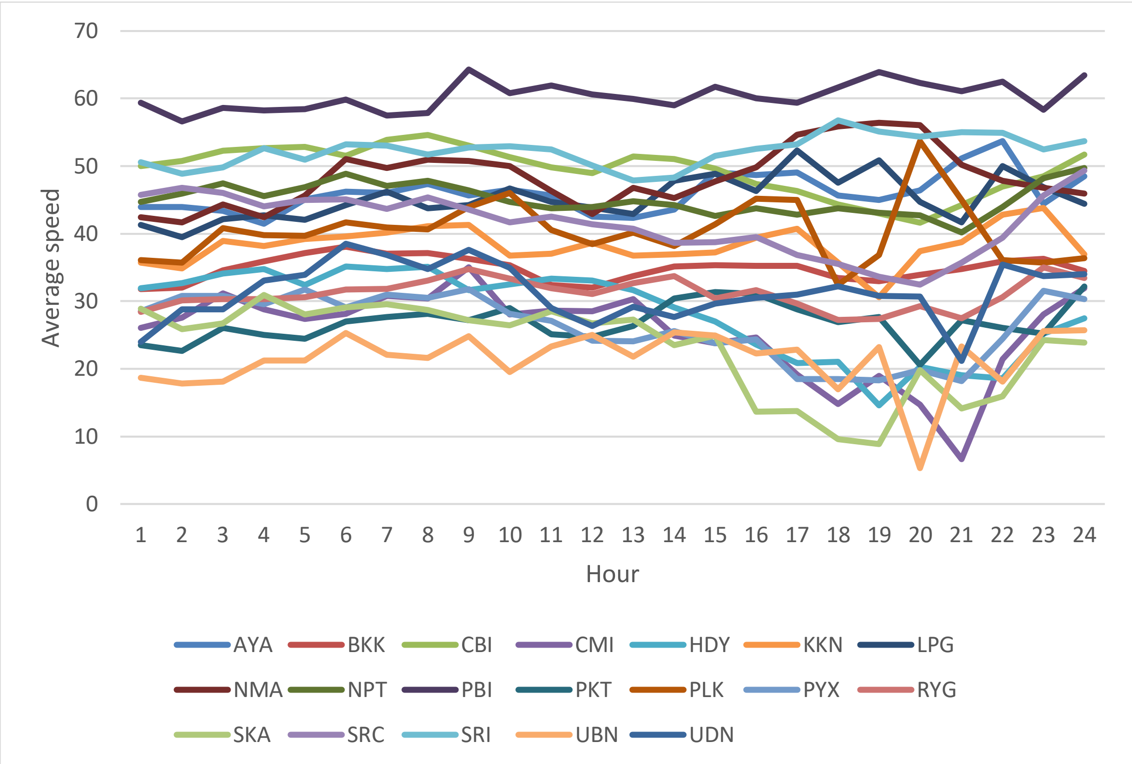
For BKK, the driving speed did not vary notably over time, because most of the records were from taxis in this crowded city.

Figure 12. Average speed per hour for 5-6 March 2017



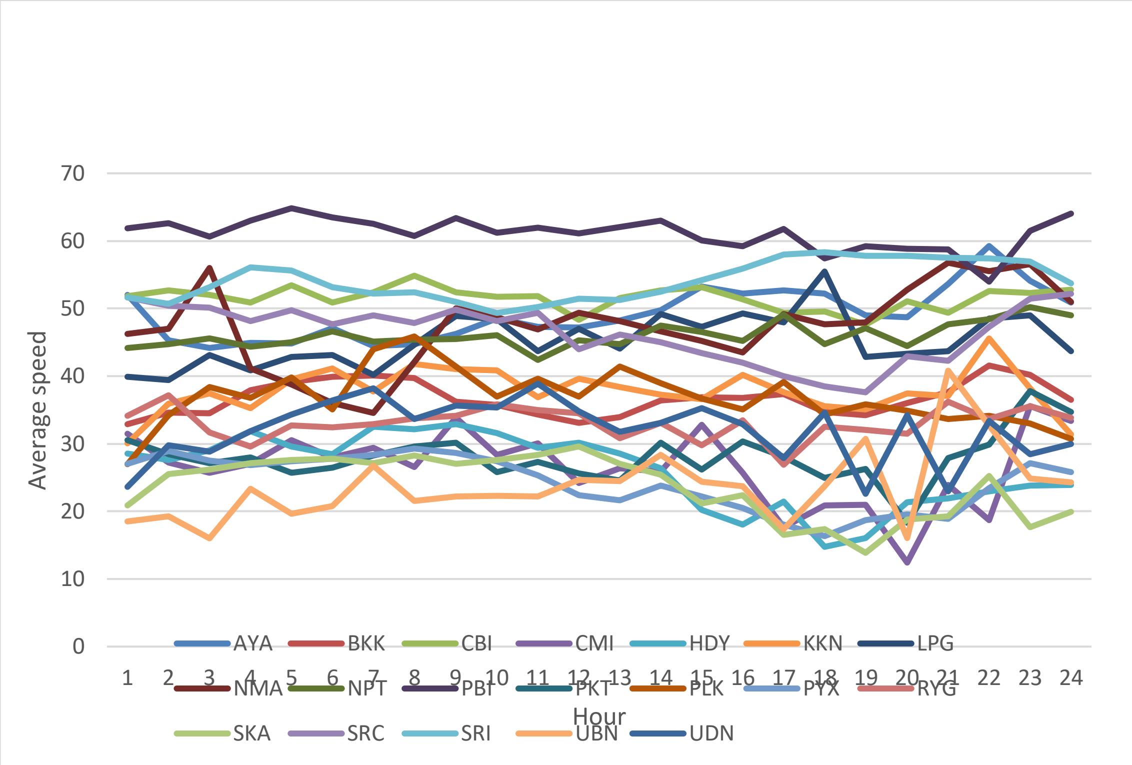
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 13. Average speed per hour for 13-14 September 2017



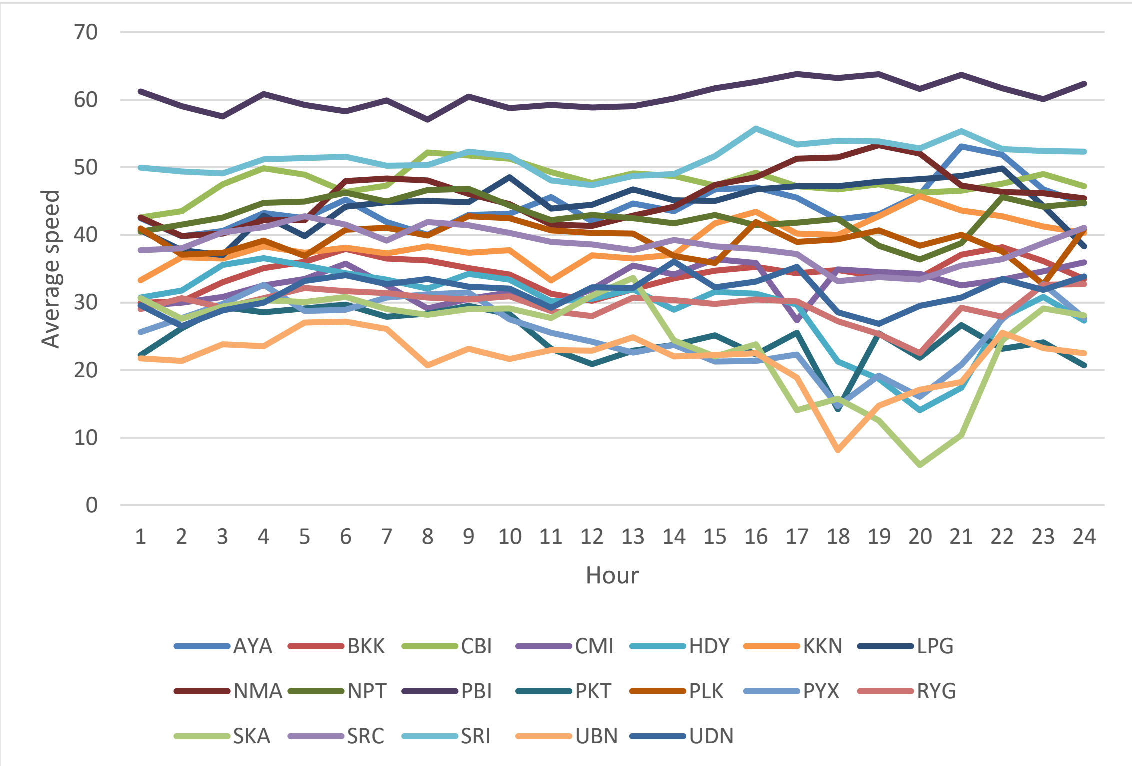
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 14. Average speed per hour for 4-5 Mach 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 15. Average speed per hour for 12-13 September 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

3.3 Inter-City Analysis

We analyzed inter-city connectivity by numbers of traveled cities within the 48 hours. In a case that a vehicle traveled a lot of cities within a 48-hour period, we regard the vehicle achieved higher connectivity. We also regard a greater number of such vehicles indicate higher connectivity among the traveled cities. In the following section, we present results of counting the vehicles per traveled cities as well as the number per city.

3.3.1 Number of Vehicles per Number of Travelled Cities

More than half of the vehicles traveled only one or two cities within the 48 hours while very few vehicles traveled more than five cities (Table 3 and Figure 16). It indicates that the half of vehicles are connecting two cities by shuttle trips.

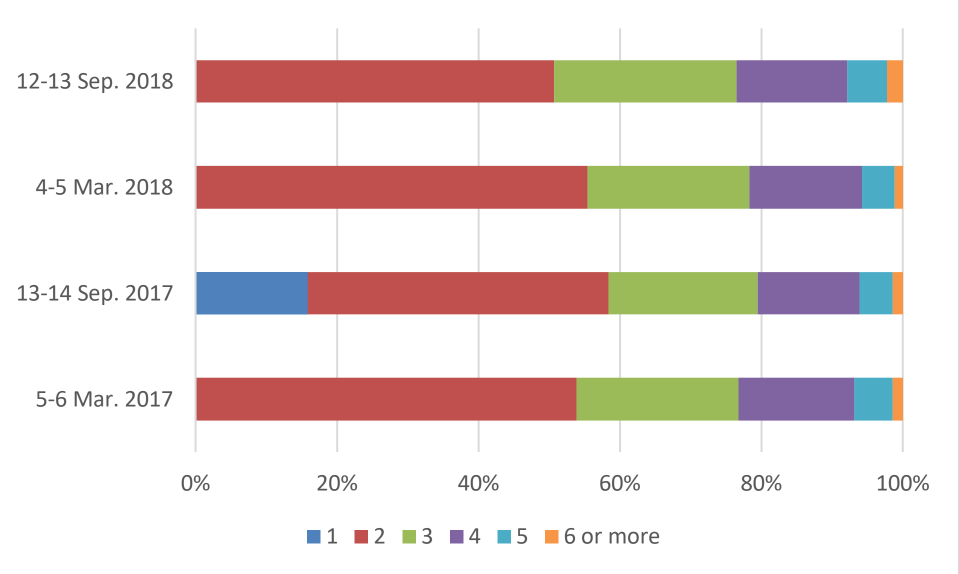
Although the other half are likely connecting the cities more than two cities, the vehicles are possibly just passing through cities between a pair of cities. Whereas further investigation on stay time in the cities needed, the indicators provide insights of potential connectivity among the cities.

Table 3. Number of vehicles per number of traveled cities

	Number of visited cities											
	1	2	3	4	5	6	7	8	9	10	11	12
5-6 Mar. 2017	-	5,424	2,315	1,645	546	80	40	14	7	2	1	1
13-14 Sep. 2017	1,906	5,108	2,537	1,727	557	112	42	9	8	3	1	-
4-5 Mar. 2018	-	6,543	2,702	1,881	540	83	32	14	7	2	-	-
12-13 Sep. 2018	-	9,681	4,916	2,982	1,068	285	102	26	11	4	1	-

Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 16. Proportion of vehicles per number of traveled cities



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

3.3.2 Number of Vehicles and Travelled Cities per City

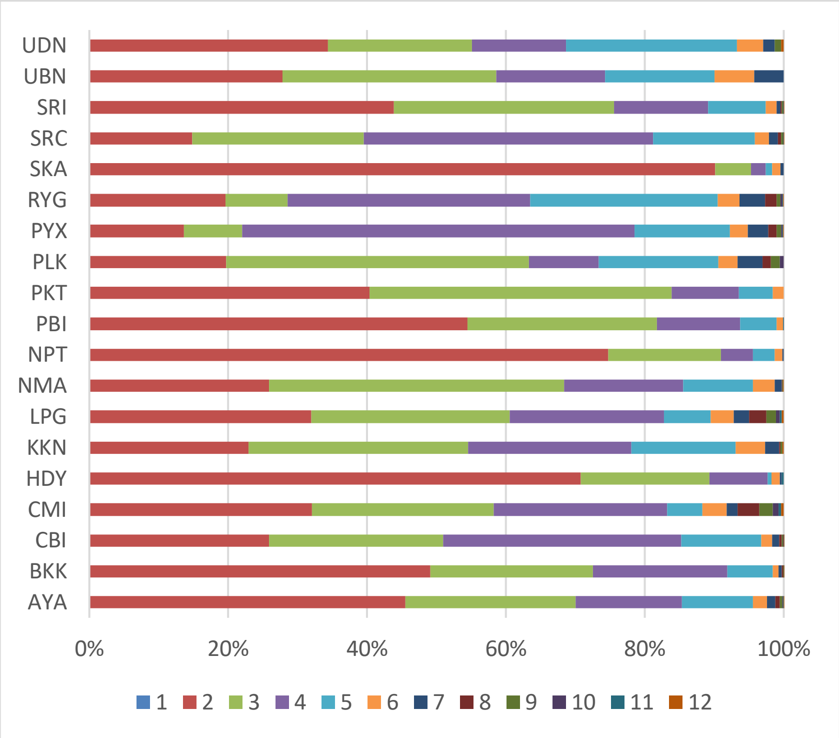
We calculated the number of vehicles with number of traveled cities per city. In a case that a city has more vehicles with more traveled cities, the city is regarded with higher connectivity. The results are shown in Figure 17 to Figure 20 and Table 4 to Table 7.

SKA and HDY have higher proportion of vehicles traveling only two cities. It is likely due to the remoteness of the cities on land. The cities might connect with the other cities rather by ocean, not so much by land.

SRC, PYX, RYG have higher proportion of vehicles with more traveled cities. The cities are on the largest industrial area; therefore, a lot of vehicles drop by the cities between contractors and retailers among the nation. As discussed in 3.3.1, stay time of the vehicles needs to be investigated for precise analysis on the connectivity.

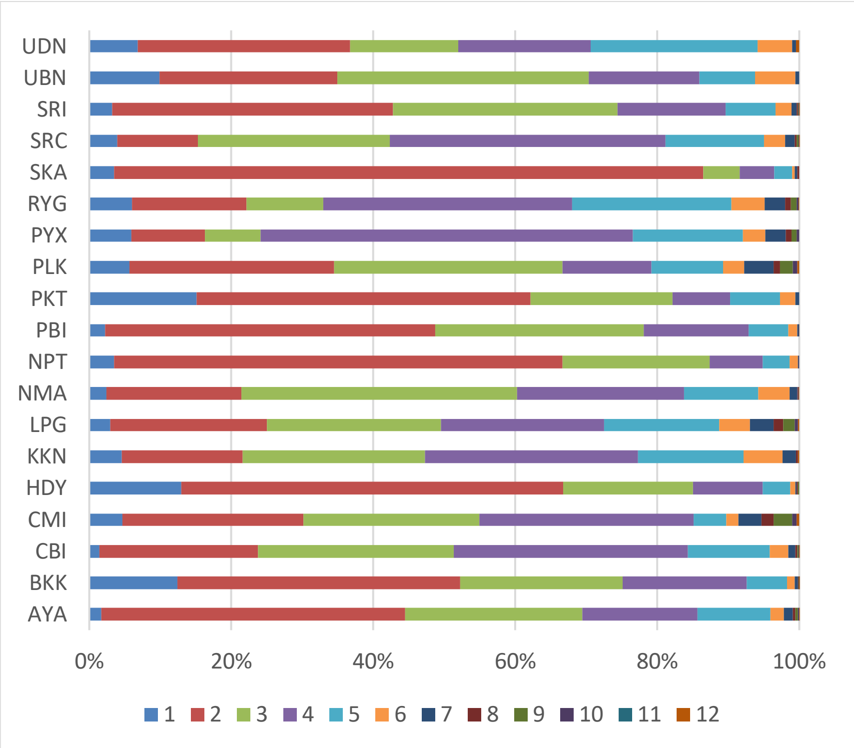
Some inland cities, such as PLK, NMA, and KKN, have notably high proportion of vehicles travelling more than two cities. The cities likely take a role of transport hub to enhance logistics network of the nation.

Figure 17. Proportion of traveled cities for 5-6 March 2017



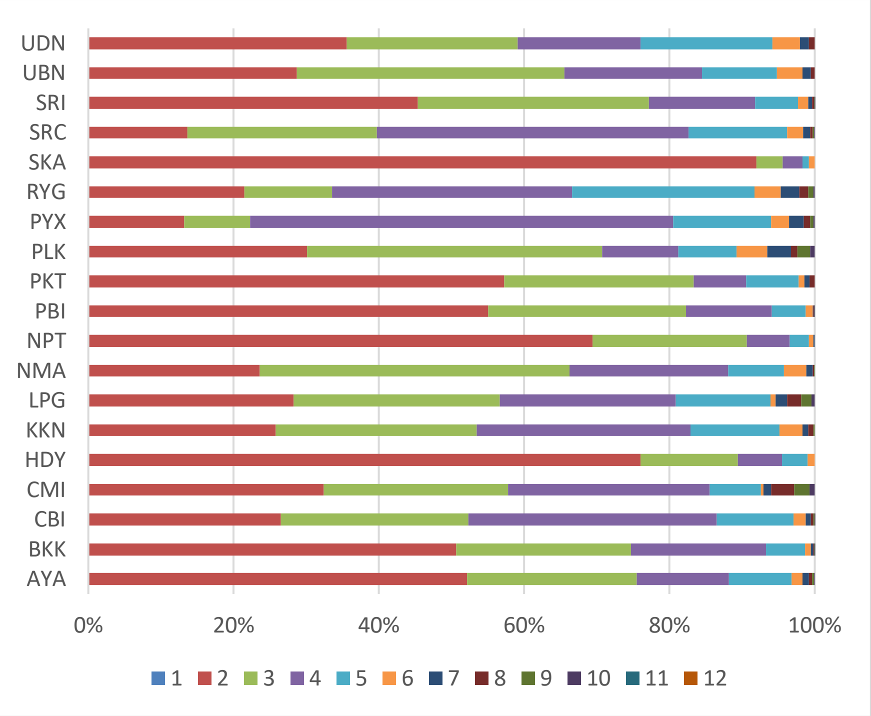
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 18. Proportion of traveled cities for 13-14 September 2017



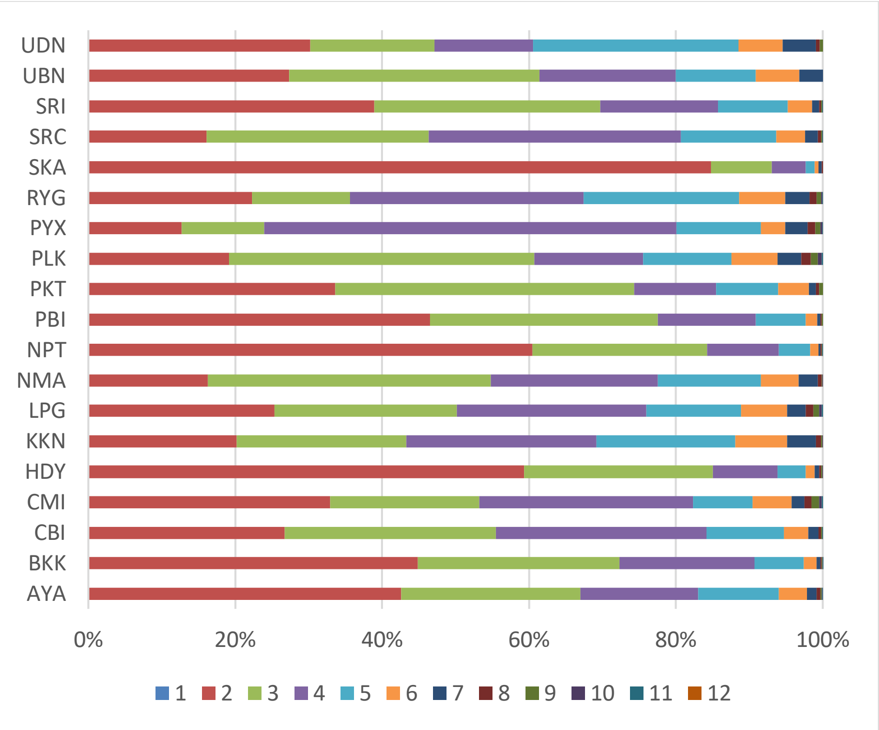
Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 19. Proportion of traveled cities for 4-5 March 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Figure 20. Proportion of traveled cities for 12-13 September 2018



Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Table 4. Number of vehicles per number of traveled cities for 5-6 March 2017

City	Number of traveled cities											
	1	2	3	4	5	6	7	8	9	10	11	12
AYA	-	904	486	303	204	40	24	13	7	2	1	1
BKK	-	3,965	1,892	1,569	523	72	34	14	7	2	1	1
CBI	-	994	965	1,316	442	62	39	14	7	2	1	1
CMI	-	82	67	64	13	9	4	8	5	2	1	1
HDY	-	244	64	29	2	4	1	-	-	-	1	-
KKN	-	136	187	139	89	25	12	1	2	-	-	1
LPG	-	115	103	80	24	12	8	9	5	2	1	1
NMA	-	328	540	217	128	39	13	1	2	-	-	1
NPT	-	930	202	58	39	13	3	-	-	-	-	-
PBI	-	732	367	161	71	12	1	-	-	-	1	-
PKT	-	25	27	6	3	1	-	-	-	-	-	-
PLK	-	65	144	33	57	9	12	4	4	2	-	-
PYX	-	147	90	609	148	28	32	13	7	2	1	1
RYG	-	170	77	303	234	27	32	14	5	2	1	1
SKA	-	211	12	5	2	3	1	-	-	-	-	-
SRC	-	444	740	1,251	439	61	39	14	7	2	1	1
SRI	-	1,233	889	383	231	45	15	7	3	2	1	1
UBN	-	39	43	22	22	8	6	-	-	-	-	-
UDN	-	81	49	32	58	9	4	-	2	-	-	1

Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Table 5. Number of vehicles per number of traveled cities for 13-14 September 2017

	Number of traveled cities											
	1	2	3	4	5	6	7	8	9	10	11	12
AYA	38	949	552	361	226	42	29	8	8	3	-	1
BKK	1,155	3,716	2,129	1,628	533	104	37	7	8	3	-	1
CBI	53	855	1,049	1,259	440	99	40	8	8	3	-	1
CMI	14	77	75	91	14	5	10	5	8	2	-	1
HDY	71	294	100	54	21	4	1	1	1	-	-	-
KKN	29	108	163	190	94	35	12	2	-	-	-	1
LPG	13	98	109	102	72	19	15	6	7	2	-	1
NMA	31	241	494	299	133	56	13	2	-	1	-	1
NPT	46	840	275	100	50	16	1	-	-	1	-	-
PBI	33	675	428	214	81	19	2	-	1	1	-	-
PKT	28	87	37	15	13	4	1	-	-	-	-	-
PLK	19	97	108	42	34	10	14	3	6	2	-	1
PYX	52	91	69	462	136	28	25	7	7	3	-	-
RYG	61	163	109	354	226	48	29	8	8	3	-	1
SKA	11	258	16	15	8	1	1	1	-	-	-	-
SRC	120	349	831	1,190	426	93	40	8	8	3	-	1
SRI	99	1,207	968	464	215	68	22	6	2	3	-	1
UBN	19	48	68	30	15	11	1	-	-	-	-	-
UDN	14	61	31	38	48	10	1	-	-	-	-	1

Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Table 6. Number of vehicles per number of traveled cities for 4-5 March 2018

	Number of traveled cities											
	1	2	3	4	5	6	7	8	9	10	11	12
AYA	-	1,189	532	291	197	34	20	9	7	2	-	-
BKK	-	4,825	2,294	1,773	514	78	28	12	7	2	-	-
CBI	-	1,129	1,095	1,455	450	73	30	13	7	2	-	-
CMI	-	92	72	79	20	1	3	9	6	2	-	-
HDY	-	375	66	30	17	5	-	-	-	-	-	-
KKN	-	149	160	170	71	18	5	4	1	-	-	-
LPG	-	120	120	103	55	3	7	8	6	2	-	-
NMA	-	317	575	294	103	42	11	3	1	-	-	-
NPT	-	996	305	84	39	8	3	-	-	-	-	-
PBI	-	941	465	202	80	16	2	1	1	1	-	-
PKT	-	79	36	10	10	1	1	1	-	-	-	-
PLK	-	101	136	35	27	14	11	3	6	2	-	-
PYX	-	161	110	709	164	30	25	10	6	2	-	-
RYG	-	221	124	340	258	37	26	13	7	2	-	-
SKA	-	332	13	10	3	3	-	-	-	-	-	-
SRC	-	441	843	1,385	438	72	31	13	6	2	-	-
SRI	-	1,481	1,039	479	191	48	16	9	2	1	-	-
UBN	-	50	64	33	18	6	2	1	-	-	-	-
UDN	-	86	57	41	44	9	3	2	-	-	-	-

Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

Table 7. Number of vehicles per number of traveled cities for 12-13 September 2018

	Number of traveled cities											
	1	2	3	4	5	6	7	8	9	10	11	12
AYA	-	1,693	968	636	436	155	49	22	8	4	1	-
BKK	-	6,857	4,210	2,814	1,024	268	83	26	11	4	1	-
CBI	-	1,907	2,057	2,045	756	237	98	26	10	4	1	-
CMI	-	255	157	225	63	41	14	7	8	3	1	-
HDY	-	630	273	94	40	13	7	3	1	1	-	-
KKN	-	257	295	329	241	90	50	9	2	1	-	-
LPG	-	250	246	255	127	62	25	10	8	4	1	-
NMA	-	367	868	514	316	116	59	12	2	1	-	-
NPT	-	1,401	552	226	99	26	7	4	1	1	1	-
PBI	-	1,278	853	365	188	42	12	6	3	-	-	-
PKT	-	72	87	24	18	9	2	1	1	-	-	-
PLK	-	119	258	92	75	39	20	8	6	3	1	-
PYX	-	180	159	794	163	47	43	15	9	4	1	-
RYG	-	414	248	590	393	117	62	18	10	4	1	-
SKA	-	552	54	30	8	3	3	1	-	-	-	-
SRC	-	915	1,716	1,954	734	226	97	24	10	4	1	-
SRI	-	1,971	1,558	814	479	170	50	13	7	2	1	-
UBN	-	78	97	53	31	17	9	-	-	-	-	-
UDN	-	160	90	71	148	32	24	3	2	-	-	-

Source: Author based on Probe Data and Shapefile prepared in Chapter 1.

4 Conclusions

In this chapter, we demonstrated analysis on intra- and inter-city connectivity using the vehicle probe data for selected 48-hour slots in March and September in 2017 and 2018. We conducted the analysis by aggregating the probe data of commercial vehicles with overlay to geographical extents of the major cities identified by night-time light satellite image data. Key findings are as below.

1) **Intra-city analysis using number of vehicles per hour –** the cities were classified into more vehicles in the daytime or night time. The types were likely associated with drivers’ preference on traffic conditions by the time.
2) **Intra-city analysis using average speed per hour –** Some cities indicated notable changes of driving speeds by the time. This is likely owing to traffic condition with people’s commuting as well as transport infrastructure, such as highways.
3) **Inter-city analysis by counting vehicles per traveled cities –** More than half of the vehicles were traveling only two cities within the 48-hour periods, which were possibly shuttle trips between two cities. Some cities in the large industrial areas and inland cities indicated high proportion of vehicles were travelling more than two cities, indicating contribution to connectivity among the city.

We did not consider logistics networks between cities and stay time in the city; therefore, we did not capture the details of travels, such as travel distances and activities in the cities. For the future works, we will investigate the remained issues by applying transportation analysis methods, such as origin-destination matrixes.

1.5 SQL Code for Calculating the Numbers

The numbers are calculated from the probe data by use of SQLite database management system with spatial data extension, called Spatialite, which is freely available opensource software. The codes are shown as below. “:TABLE” shall be replaced with the name of table with the probe data.

```
-- For calculating the numbers for Figure 4 to Figure 11
.output city_hour_n_vehicle_:TABLE.txt
SELECT city, t_hour, sum(1) AS n_vehicle
FROM (
   SELECT imei,strftime("%H",clocktime) AS t_hour,city
   FROM :TABLE
   LEFT JOIN cities ON ST_Intersects(:TABLE.geom, cities.geometry)
   WHERE city is NOT NULL
   GROUP BY imei,strftime("%H",clocktime),city
) AS t1
GROUP BY city, t_hour
ORDER BY city, t_hour
;
-- For calculating the average speed for Figure 12 to Figure 15
.output city_hour_speed_:TABLE.txt
SELECT city, t_hour, AVG(avg_speed)
FROM (
   SELECT imei,strftime("%H",clocktime) AS t_hour,city,AVG(speed) AS avg_speed
   FROM :TABLE
   LEFT JOIN cities ON ST_Intersects(:TABLE.geom, cities.geometry)
   WHERE city is NOT NULL AND speed > 0.0
   GROUP BY imei,strftime("%H",clocktime),city
) AS t2
GROUP BY city, t_hour
ORDER BY city, t_hour
;


-- For calculating the number of travelled cities

.output city_n_city_n_vehicle:TABLE.txt
```

```
CREATE TABLE imei_n_city_:TABLE AS
SELECT imei, sum(1) AS n_city
FROM (

SELECT DISTINCT imei,city
FROM :TABLE
LEFT JOIN cities ON ST_Intersects(:TABLE.geom, cities.geometry)
WHERE city is NOT NULL

) AS t1
GROUP BY imei
;

.output city_n_vehicle_by_n_city_:TABLE.txt
SELECT n_city, sum(1) AS n_vehicle
FROM imei_n_city_:TABLE
GROUP BY n_city
;

SELECT city, n_city, sum(1) AS n_vehicle
FROM (

SELECT DISTINCT :TABLE.imei, city, n_city
FROM :TABLE, imei_n_city_:TABLE, cities
WHERE city is NOT NULL
AND imei_n_city_:TABLE.imei = :TABLE.imei
AND ST_Intersects(:TABLE.geom, cities.geometry)

) AS t2
GROUP BY city, n_city
;
```

References

Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Tomasini, M. (2018). Human Mobility: Models and Applications.

Chen, Y., Wang, B., Liu, X., & Li, X. (2018). Mapping the spatial disparities in urban health care services using taxi trajectories data. Transactions in GIS, 22(2), 602–615. http://doi.org/10.1111/tgis.12333

Liu, X., Gong, L., Gong, Y., & Liu, Y. (2015). Revealing travel patterns and city structure with taxi trip data. Journal of Transport Geography, 43, 78–90. http://doi.org/10.1016/j.jtrangeo.2015.01.016

Liu, Z., Ma, T., Du, Y., Pei, T., Yi, J., & Peng, H. (2018). Mapping hourly dynamics of urban population using trajectories reconstructed from mobile phone records. Transactions in GIS, 22(2), 494–513. http://doi.org/10.1111/tgis.12323

Malleson, N., Vanky, A., Hashemian, B., Santi, P., Verma, S. K., Courtney, T. K., & Ratti, C. (2018). The characteristics of asymmetric pedestrian behavior: A preliminary study using passive smartphone location data. Transactions in GIS, 22(2), 616–634. http://doi.org/10.1111/tgis.12336

Qian, X., & Ukkusuri, S. V. (2015). Spatial variation of the urban taxi ridership using GPS data. Applied Geography, 59, 31–42. http://doi.org/10.1016/j.apgeog.2015.02.011

Tang, J., Liu, F., Wang, Y., & Wang, H. (2015). Uncovering urban human mobility from large scale taxi GPS data. Physica A: Statistical Mechanics and Its Applications, 438, 140–153. http://doi.org/10.1016/j.physa.2015.06.032

Yang, Z., Franz, M. L., Zhu, S., Mahmoudi, J., Nasri, A., & Zhang, L. (2018). Analysis of Washington, DC taxi demand using GPS and land-use data. Journal of Transport Geography, 66, 35–44. http://doi.org/10.1016/j.jtrangeo.2017.10.021

Zong, F., Wu, T., & Jia, H. (2018). Taxi Drivers' Cruising Patterns-Insights from Taxi GPS Traces. IEEE Transactions on Intelligent Transportation Systems, 1–12. http://doi.org/10.1109/TITS.2018.2816938