Trip Distribution of Urban Worker at Sudirman Station

View the article online for updates and enhancements.

Trip Distribution of Urban Worker at Sudirman Station

F Yunita[1], H Setiadi[1] and A Gamal[2]

[1]Department Geography, Faculty Mathematics and Natural Sciences, Universitas Indonesia
[2]Department of Architecture, Faculty of Engineering, Universitas Indonesia
Jl. Margonda Raya, Kelurahan : Pondok Cina, Kecamatan : Beji, Kota: Depok, Provinsi : Jawa Barat, Kode Pos : 16424, Indonesia.

Abstract. Most researchers using questionnaire survey in determining the trip distribution pattern in general especially to find out information on origin and destination. In Indonesia, a transportation survey is conducted by the Ministry of Transportation every five years, which mean the data will go outdated not long after it is published. There is a new way to get updated data by utilizing new data sources that are obtained through communication and information technologies such as geolocation feature on Twitter. This study tried to infer the trip distribution of urban worker at Sudirman Station using survey and geolocation feature on Twitter. The results of this study indicated that the comparison of trip distribution which was obtained through the survey showed more varied results than new data sources such as social media Twitter. The variation of the trip distribution was assessed by the differences of precise locational information which was obtained by both approaches. Both approaches could be used to describe the trip distribution is based on the structure of the city. This study also contributed to explore the usage of geolocation feature for estimating the trip distribution of urban worker at urban transit area by using additional phase during the generating twitter data process.
Keywords: *trip distribution; urban; geolocation*

1. Background

Nowadays, the development of residential, offices and business district in urban areas is oriented towards transport nodes. Several cities in Europe, the United States, Australia, and Japan have implemented Transit Oriented Development (TOD), a regional

development concept that encourages residential, retail, and office development around transport nodes [1]. The Transit Oriented Area seeks to maximize the usage of land utility around the transport node. The development of transit-oriented areas in Indonesia is directed based on rail transport network, but currently, it's still stuck in the planning process [2]. Transport node that gives great effect to the development of the city is train stations and airports [3]. The development of transit-oriented areas also has constraints related to estimate the trip generation, traffic conditions, and availability of parking lots [2].

In Jakarta Metropolitan Area, most people treat the train station as a place of interchange of mode because they need to continue their commute by other means of transportation in order to reach their destination. After the emergences of the online motorbike taxi, there is an increase in the usage of 11% train stations as a place of interchange of mode has increased 11% [4]. This finding indicates that the transportation node, especially train station, has dynamic movement. According to Commuterline Jabodetabek Agency, there are ten train stations that have a large number of users per day in 2016 [5]. Sudirman Station is one of them and it is also currently being developed as a transit-oriented area. Therefore, in supporting the development of transit-oriented areas, it is deemed necessary to have an updated description of the trip distribution of urban worker.

Urban workers are chosen as the representations of the trip distribution at transport nodes because they dominate the urban mobility in general. This statement is in accordance with the study in Hong Kong Metropolitan Area that reveals that the urban movement in Asian cities is ubiquitously shaped by travel patterns and the selection of modes for working purposes [6]. The determination of trip distribution predominantly uses a survey questionnaire with a relatively large sample size to find out the origin and destination locational information [7]. Modelling the trip distribution also uses a survey approach to household interviews for specifying the origin and destination of movement [7]. According to quoted information from the official portal of the Ministry of Transportation Republic of Indonesia, the survey of origin and destination for transportation that is done every five years [8]. This fact indicates the possibility of data was relatively outdated. Both findings indicate a challenge in obtaining origin and destination information based on the survey. The challenge of obtaining information of origin and destination through surveys is currently tackled by the increasing use of transportation-related information new data sources which is obtained through communication and information technologies such as big data [9].

Big data can be generated by social media, smartphones, Satellite Navigation (SatNav) as well as other technologies [9]. Several studies related to urban mobility have utilized geolocation information for academic research. The urban mobility of residents in US cities [10] , while a similar approach to predicting domestic commuters' patterns in the UK [11](Mc. Neil et al, 2017) also conducted by using geolocation information for predicting urban mobility. Twitter is one of several social media platforms which provides geolocated information. It also allows people to download the tweets from the Internet. The credibility of Twitter data in mobility studies has been validated in the study

of Lenormand in 2014, which compares data from Twitter, mobile phones, and official data (census), which they concluded that the three sources offer comparable results [12].

Indonesia is the fifth largest country as a Twitter user and Jakarta is the city with the largest user contributors [13]. Therefore, the use of Twitter data is considered very potential to infer the locational origin and destination of urban workers in the Jakarta Metropolitan Area. Based on a review of the challenges in estimating trip distribution at transit-oriented development areas as well as emerging methods such as survey approaches and the availability of new data sources such as Twitter, this study attempts to describe urban worker distributions at transport nodes based on survey approaches and social media (Twitter) which is integrated through visualization of map.

2. Materials and Methods

Data collection and processing in this study were divided into two stages. The first stage was the collection and processing of primary data which was obtained through the survey. The second stage was the collection and processing of Twitter data. Data collection through survey was conducted by distributing survey questionnaires to 68 respondents in Sudirman Station which were selected randomly with the consideration of the average number of Sudirman Station users per day. The size of the sample survey was generated based on PT. KAI Commuterline Jabodetabek data that revealed the number of Sudirman Station users was 35.493 per day in 2016 (Kumparan,2017). Then the survey was conducted using a confidence level of 90% with a 10% confidence interval. It was chosen by referring to the general transit user survey, i.e. significant intervals of 95% or 90% and at ± 5-10% [14]. The results showed the address information of the origin and destination of the urban workers which was obtained through the survey are then digitized. Digitization of locational origin and destination was done using Google Earth software.

Data collection via Twitter was done by utilizing the Twitter Public API which was accessed via Python 2.6. The parameters that were used to identify Sudirman Station users such as coordinates with a radius of 1km and the keyword 'stasiun sudirman'. Furthermore, the data were classified based on (1) the type of tweet obtained and (2) the type of tweet that has the geolocation or keywords information were considered valid. In determining the origin and destination of twitter users based on tweets that contain geolocated information on the last 1000 tweets. Tweets that have the largest number of geolocation similarities are assumed to be where they come from, whereas tweets which had more than five similar geolocation tag were considered as destinations which are assumed to be routinely visited as the destination of movement [11]. The number of tweets used for comparing the trip distribution, equal to the number of respondents being asked in the direct survey.

Comparison of trip distribution based on survey and Twitter used quota sampling based on a number of crawling results obtained through twitter. In this study the survey method using nonproportional quota sampling, which is to adjust the minimum number of results which was obtained through Twitter during April 2018. The nonproportional quota

sampling is performed on the percentage of quotas not matching the proportion of sub-groups present in the minimum population [15].
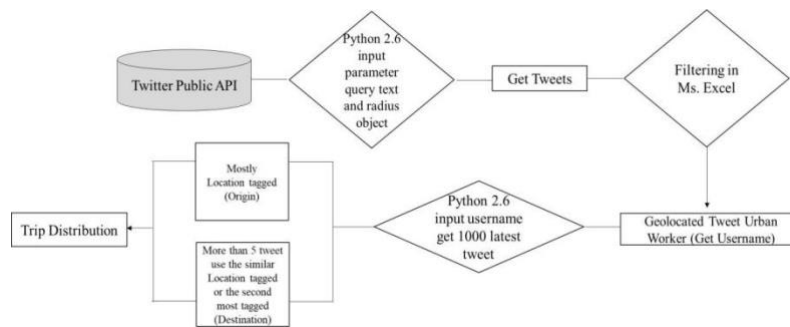

Figure 1. Twitter Data Extraction Flow

3. Results

The trip distribution results which were obtained through survey consist of origin and destination addresses of the urban worker. The information contains detailed information so that it could be classified by the administrative city. Figure 2 shows the trip distribution of urban workers in Sudirman Station was dominated by travel from the suburbs of Jakarta Metropolitan City such as Bekasi City, Bogor District, South Tangerang City, and Depok City to Central Jakarta and South Jakarta Administration City. The suburbs become the dominant origin of the trip to the city center because there were many residential areas.
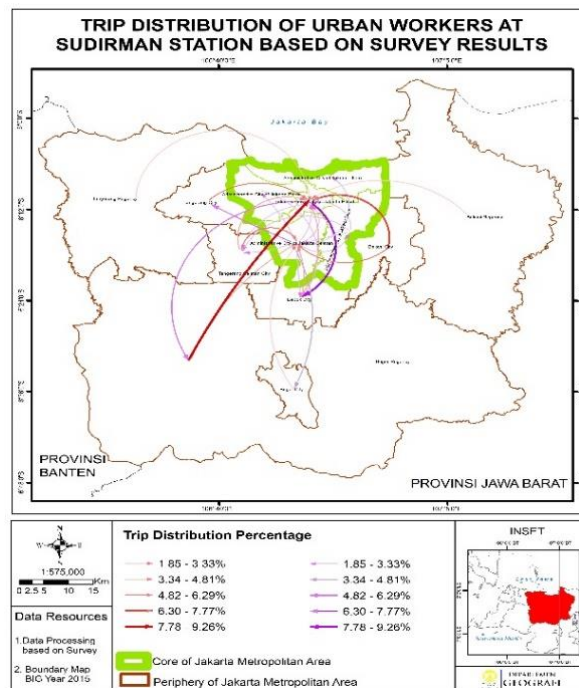
Figure 2. Urban Worker Transport Distribution based on Survey results

Based on Twitter data which was obtained in four weeks in April 2018, there were 223 tweets from 223 users containing the text 'stasiun sudirman'. Based on the total tweet, there was only 4% of the total tweets that included geolocation information with the Sudirman Station keyword, while 96% make a tweet only with the keywords. Tweets which contain the keywords were assumed that the Twitter user is also a user of Sudirman Station. The results of 223 tweets were then filtered and traced back by identifying the last 1000 tweets of each user. The acquisition of 1000 tweets per user is then derived from the user's origin and destination information by assuming similarities geolocation. The result was only 24% of the total users matched the criteria. In Figure 3 shows the comparison of invalid tweets and valid tweets that can be used to determine the distribution of the trip. The results show that only 14% of tweets per week could be identified as the origin and destination information. The results show that there were only 54 identifiable users that can be assumed as origin and destination.
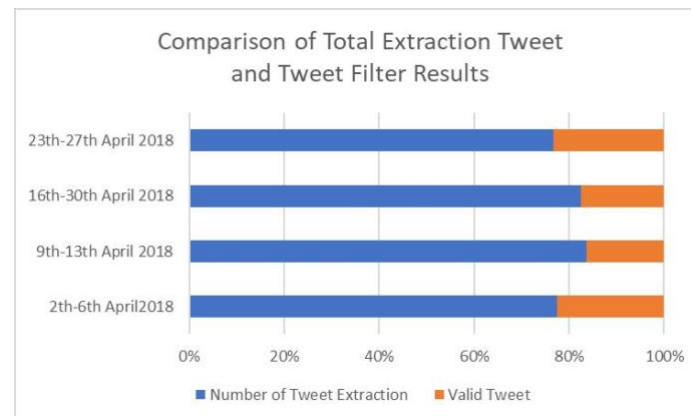
Figure 3. Total Compression Data Extraction and Filtering Results Tweet
Source: Processing Results (2018)

The geolocated tweet has minimal granularity of place types to return and must be one of neighbourhood, city, admin or country [16]. Regarding to the data processing, the geolocated tweet that is possible to be extracted is district locational information. The exact coordinates may appear accurately if the users activate the data location sharing feature. Based on the result, it shows that the only 20% tweets contained the geolocation and the keywords, and 80% tweets only contained the keywords. Tracing urban mobility based on twitter can also consider using several keywords which were related to activity on a particular location or places [17]. In this study the main keyword used for determining the locational origin is sleep, but then the extraction results show other keywords that are also relevant and the represent the origin location. There were several keywords such as '*tidur*', '*berangkat*', '*bangun*', '*sleep*', '*ronda*', '*nyoblos*', '*alarm*'. Meanwhile the main keywords which were used for determining the destination location is work, but then the results of extraction indicated other keywords that are also relevant and represent the location of the destination. There were also several keywords such as, office, meeting, work, and job.

Table 1. Comparison of Percentage of Origin and Destinations Estimation		
Origins		
Survey	**Twitter**	
Address of travel origin (sub-district)	The most frequent geolocated tweet	The most frequent geolocated tweet and keywords
100%	55.56%	44.44%
Destinations		
Survey	Twitter	

Address travel destination (sub-district)	The second most frequent geolocated tweet	The second most frequent geolocated tweet and keyword
100%	72.22%	27. 78%

Table 1 shows the comparison in determining the origin and destination based on surveys and twitter. Through the survey, we obtained information that could represent district location which were reclassified based on their address. Meanwhile, the Twitter results were identified based on geolocated tweet and keywords. Based on Table 1, it can be considered that the response rate survey was higher than Twitter. In order to determine the origin and destination, geolocated indicator were predominantly obtained rather than geolocated and keywords indicator. The indicator of the most frequent geolocated tweet was relatively higher for identifying the destination. This indicated that Twitter users in Jakarta Metropolitan Area frequently shared locational information on their destination area.
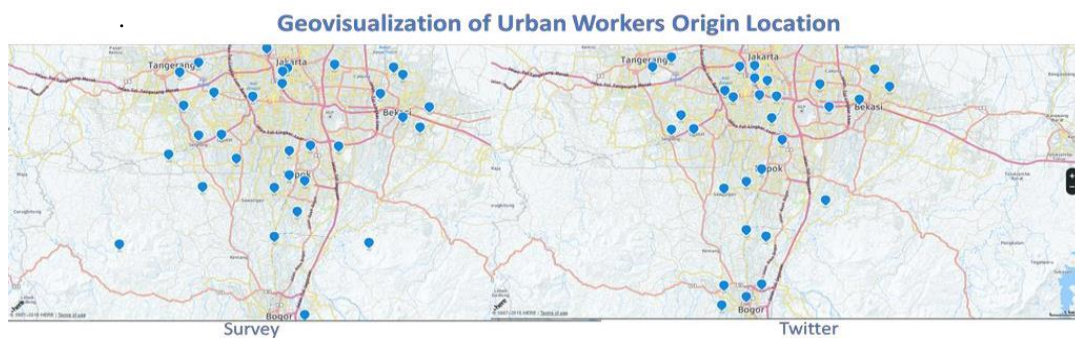

Figure 4. The Comparison of Urban Worker Origin Location

This map shows the area of Jakarta Metropolitan Area. The blue dot represents the origin district of the urban worker. This figure represent that the origin location of urban workers based on surveys and twitter shows the same characteristics which districts dominated by residential areas in suburb of Jakarta Metropolitan Area.
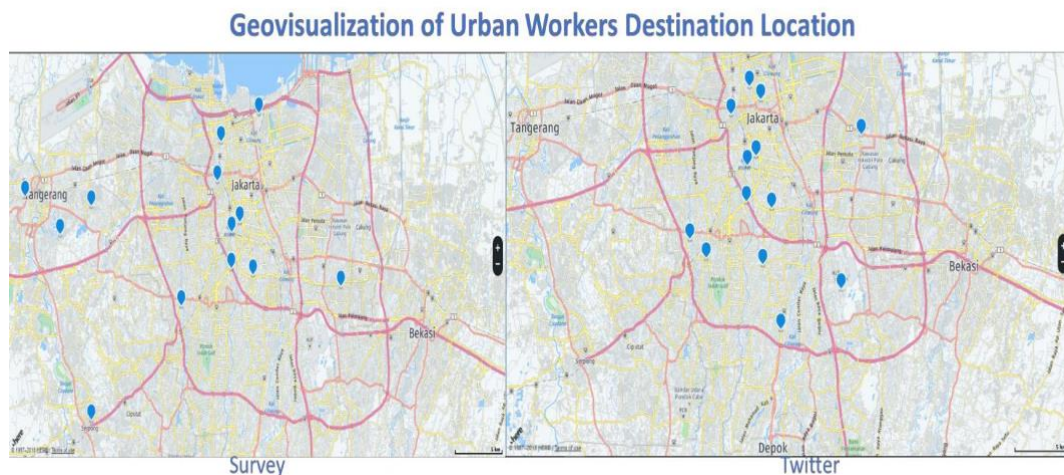
Figure 5. The Comparison of Urban Worker Destination Location

Figure 5 shows that the destination of urban worker which emerge on Twitter was dominated from district in urban area. Meanwhile based on survey the destination district of the urban worker came from regency in suburb of Jakarta Metropolitan Area. The frequency of information sharing via Twitter related to work objectives is dominated by regions has proximately to the city center. This is probably caused by two things. The fact that more office workers are in the city center. Then, the social prestige causes that sharing information on destination locations in city center is more often done by urban workers, compared to urban workers with destinations in the suburb.

The results of the trip distribution based on Twitter results were reclassified from the district locational level into the administrative city level. Figure 5 and 6 shows the comparison of trip distribution which were obtained via survey and Twitter. The trip distribution that dominated Sudirman Station users come from the suburbs, which were Bogor Regency, Bogor City, Bekasi City, Tangerang City, and South Tangerang City for doing activity in the city center such as the Central Jakarta Administration City, South Jakarta Municipality, and City East Jakarta Administration.

Based on figure 6 shows that the comparison of trip distribution percentage based on survey and Twitter shows only the same value of trip distribution from Bogor Regency, Bogor City, and Bekasi City to South Jakarta Administration City and from South Tangerang City to Central Jakarta Administration City. There were also trip distribution that were not covered by the survey but were represented by twitter and vice versa.
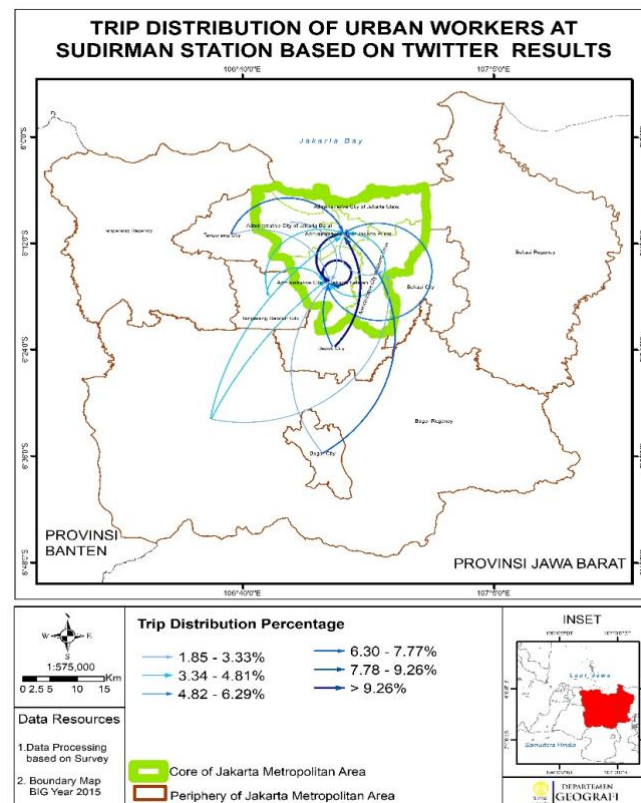
Figure 6. The Trip Distribution of Urban Workers at Sudirman Station based on Twitter Results

Source: Data Processing Results 2018

Based on structure of the city, 64.81% of trip distribution was obtained via Twitter dominated from the suburbs to the metropolitan city center, while in the city center there was 35.17% of trip distribution. The surveys shows more detailed variations, 33.33% of trip distribution from downtown to suburban areas, while 53.70% of trip distribution from the outskirt city to the city center, and the 12.96% of trip distribution in the city center.

The trip distribution results which was obtained through surveys and twitter indicated that the two data sources have quite different information characteristics. The locational origin and destination information which were obtained through the survey contains more details and then were classified based on the district information. In contrast, the locational origin and destination information which were obtained through Twitter only contains the location information of the district, thus it does not represent the distribution of the location of origin and destination in detail. This indicates that the use of survey

data sources can be used for more detailed description needs, while the information obtained through twitter can be used to describe for general purpose.

4. Conclusions

The comparison of trip distribution which was obtained through surveys and Twitter showed the different distributions of emerging travel. The survey results showed three types of trip distribution, i.e. (1) from the metropolitan suburbs to the metropolitan city center, (2) from the center and to the metropolitan city center, and (3) from the center to the metropolitan suburbs. The trip distribution which was obtained through Twitter shows only two types of trip distribution, i.e. (1) trip distribution from the metropolitan suburbs to the city center and (2) from the metropolitan city center to the metropolitan city center. The variation of the distribution was assessed by the differences of precise locational information which was obtained by both approaches, meanwhile both approaches could be used to describe the trip distribution in general. The use of surveys could be better to represent travel information from the outskirts to the city center and vice versa. Based on these findings, Twitter can represent the distribution of workers' trips that are dominated by activities in the city center.

5. References

[1] Mungkasa (2012) Housing Development in Implementation of 'Compact City' Model in DKI Jakarta. Spatial and Land Bulletin. Jakarta: National Planning and Development Agency.

[2] Syabir, Ibnu (2013). Policy and Strategy Development Connectivity Jakarta Metropolitan Area area with Transit Oriented Development Approach Region. Final Report of the National Priorities Research Master Plan for the Acceleration and Expansion of Indonesian Economic Development 2011-2025. SAPPK ITB.

[3] Wang, Shirui (2015). Spatiotemporal Visual Analysis of Traffic Flow Patterns Related to Transport Hubs from Floating Car Data. Master Thesis. Department of Civil, Geo and Environmental Engineering Technische Universität München

[4] ITDP Indonesia (2017). 8 Guidance Points and Recommendations for Sustainable Transport Policy In DKI Jakarta. Jakarta: ITDP Indonesia. Executive Summary

[5] Kumparan. (2018). *Infografis: Berapa Jumlah Penumpang Commuter Line?*. [online] Available at: https://kumparan.com/@kumparannews/infografis-volume-penumpang-commuter-line [Accessed 6 Feb. 2018].

[6] Lau, J. and Chiu, C. (2004). Accessibility of workers in a compact city :. Habitat International, 28 (1), pp.89-102.

[7] Banos and Thévenin (2011). Geographical Information for Urban Transport System. London: Wiley

[8] <asep.muhamad[at]torche.co.id>, A. (2018). KEMENHUB LAKUKAN SURVEI ATTN 2011. [online] Dephub.go.id. Available athttp://dephub.go.id/berita/baca/kemenhub-lakukan-survei-attn-2011-7404/?cat=QmVyaXRhfHNlY3Rpb24tNg [Accessed 7 Mar. 2018].

[9] van Wee B (2016) Accessible accessibility research challenges. Journal of Transport Geography 51: 9-16

[10] Beiró MG, Panisson A, Tizzoni M, Cattuto C (2016) Predicting human mobility through the assimilation of social media traces into mobility models. EPJ Data Sci 5 (1): 17

[11] Mc. Neil, G., Bright, J., Hale, SA (2017) Estimating local commuting patterns from geolocated Twitter data. EPJ Data Science Open Springer Journal

[12] Moya-Gómez, B., Salas-Olmedo, M., García-Palomares, J. and Gutiérrez, J. (2017). Dynamic Accessibility using Big Data: The Role of the Changing Conditions of Network Congestion and Destination Attractiveness. Networks and Spatial Economics

[13] Mononimbar and Herman. (2017). Indonesia Fifth-Largest Country in Terms of Use Users accessed through http://www.jakartaglobe.beritasatu.com/news/indonesia-fifth-largest-country-in-terms-of-twitter-users/ January2th 2018 12:43

[14] Transportation Services University Winconsins (2012). 2012 On-Board Campus Bus Survey. Madison: University Winconsins. Final Report

[15] Alvi, Mohsin (2016). A Manual for Selecting Sampling Techniques in Research. University of Karachi, Iqra University

[16] Developer.twitter.com. (2018). GET geo/search. [online] Available at: https://developer.twitter.com/en/docs/geo/places-near-location/api-reference/get-geo-search [Accessed 14 Dec. 2018].

[17] Yuyun, Akhmad Nuzir, F. and Julien Dewancker, B. (2017). Dynamic Land-Use Map Based on Twitter Data. Sustainability, 9 (12), p.2158.