World Academy of Science, Engineering and Technology
International Journal of Information and Communication Engineering
Vol:12, No:6, 2018

Quantifying Mobility of Urban Inhabitant Based on Social Media Data

Yuyun, Fritz Akhmad Nuzir, Bart Julien Dewancker

***Abstract*—**Check-in locations on social media provide information about an individual's location. The millions of units of data generated from these sites provide knowledge for human activity. In this research, we used a geolocation service and users' texts posted on Twitter social media to analyze human mobility. Our research will answer the questions; what are the movement patterns of a citizen? And, how far do people travel in the city? We explore the people trajectory of 201,118 check-ins and 22,318 users over a period of one month in Makassar city, Indonesia. To accommodate individual mobility, the authors only analyze the users with check-in activity greater than 30 times. We used sampling method with a systematic sampling approach to assign the research sample. The study found that the individual movement shows a high degree of regularity and intensity in certain places. The other finding found that the average distance an urban inhabitant can travel per day is as far as 9.6 km.

***Keywords*—**Mobility, check-in, distance, Twitter.

I. Introduction

KNOWLEDGE of human mobility patterns within cities is predominant for better urban planning. Researchers have proven that human mobility plays vital roles in planning urban infrastructure [1], urban development and human migration [2], and development of transportation facilities [3]. In previous studies, the methods to measure the mobility of citizens are usually gathered through a traditional survey or using questionnaires that attempt to capture how citizens interact with their environment [4]-[6], and the urban demographics data of where people live and work [7]. The presence of technology devices produces individual's traces and human spatial behaviors that have not been discovered before. Data on mobile phones users [8], personal digital assistant [9], and GPS devices have provided individuals' mobility information [10]. Through GPS devices, individual travel activities on the visited places can be recorded such as information of times, days, and even the types of transportation used. In addition, the smartphone can explain the human location information where the call occurred. This data becomes important due to most of the citizen has a smartphone. Thus, this device can become a sensor to explain people movement.

In recent years, the number of social networking users in the world has grown by leaps and bounds. Millions of unit's data are generated from these sites to provide knowledge for human movement. Many features are provided by social media developers to make it easier for a user in communication. Besides the status update feature, users also can attach the location embedded in the posted message. The location information shared indicates a place where a person conducts social media activity. In this study, the authors use the distribution Twitter social media data to characterize human mobility. The research will discuss the questions; what are the movement patterns of citizens? And, how far do people travel in the city?

We first identify each social media user by analyzing the people with a certain check-in (see Fig. 1). This is necessary due to the involvement of active users, making it easy for investigating human movement. Then, we calculate the distance traveled by each user (km) and identify the type of place that people visit. To recognize the name of avenue, we used the text that the user posted on Twitter as a key to determine the name of the location

Fig. 1 The check-in distribution of all locations of 546 users during a 30-day period

II. Material and Method

A. Data Collection

One significant feature of Twitter is the ability of the system to display a location map that reveals the time and place of where the status was posted. When people update their status, the system will record their geographical information by specifying the area or location in order to find their longitude and latitude coordinates at that moment. Thus, with this feature individual activity can be tracked. In this case, Makassar City,

Yuyun is with the Faculty of Environmental Engineering, the University of Kitakyushu, 808-0135, Japan and with the STMIK Handayani Makassar, Makassar, 90231, Indonesia (phone: +8190-8668-0584; e-mail: yuyunwabula@gmail.com).

Fritz Akhmad Nuzir is with the Architecture Department, Faculty of Engineering, Bandar Lampung University, Bandar Lampung 35142, Indonesia (e-mail: fritz@ubl.ac.id).

Bart Julien Dewancker is with the Department of Architecture, Faculty of Environmental Engineering, the University of Kitakyushu, 808-0135, Japan, e-mail: bart@kitakyu-u.ac.jp).

Indonesia is selected as the case study to conduct our analysis. The dataset consists of 30 days (four weeks), starting from September 1st to 30th, 2016, with a 201,118 check-ins and 22,318 users. The dataset used in this study accessed from streaming Application Program Interface (API) Twitter. It is a window application provided by Twitter for developers to access the data programmatically. The REST APIs give access to read and write Twitter data. As an example; a new tweet, author profile, follower data, time zone and location information that indicate where the tweet is posted [11].

TABLE I
DATASET DETAIL

Original dataset	**Number**
Number of Check-ins	38185
Number of users	546
Research Sample	**Number**
Number of Check-ins	2570
Number of Users	54

B. Method

To accommodate individual mobility, the authors only analyzed users with more than 30 check-ins. The next step, was the filtering process to obtain the 38,185 check-ins with 546 users (see Table I) used in this study. To determine the number of samples in the research, the author used the formula S=1/10*P, (S is sample and P is population), producing 54 users.

To spread the population evenly, the author used sampling method with a systematic sampling approach. The technique takes a sample based on alphabetical sequences of the Twitter username. For example, every user who has the first letter M will be taken twice. If the first letter of each sequence of the alphabet is processed two times, then *(26*2=54)*. This amount is equal to the number of research samples (see Table II).

III. Data Processing and Analysis

To analyze the individual movement of the study sample, we split the check-in activity into five groups. Each group contains a places activity and their mobility distance during the study period, in this case, how far they travel when they take a trip. Then in each group, we also identify the types of places visited. Fig. 2 shows the users deployment and check-ins based on groups.

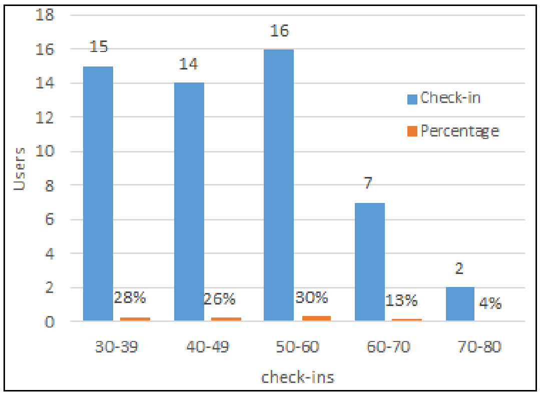


Fig. 2 Check-ins and users percentage for each group

A. Group 1

Referring to Table II of group 1, users in this group have check-in activity of around 30 to 39 times with 15 users. We analyze that in the first week, the individual's average travel reaches 48.03 km. The authors observed that their movement distance varies from 1 km to 48 km in the 1st week, 1 km to 159.7 km in the 2nd week, 1 km to 54 km in the 3rd week, 4 km to 132 km in and the 4th week. Then we analyzed the user average distance per week, which was 48.0 km in week 1, 85.0 km in week 2, 52.9 km in week 3, and 84.8 km in week 4. From the results, we conclude that the daily average mobility of people in group 1 was about 9.2 km.

In this group, we identify the type of location that people visit. Almost all of the users show check-in activity at places such as university 25%, school 19%, hotel 17%, home 13%, dormitory 9%, café 9%, and McDonalds 8% (see Fig. 4 (a)).

B. Group 2

In this group, the average check-in activity is 40 to 49 times. We first observe the individual's mobility per week. From the results of the analysis (see Table II, group 2), we see that the maximum movement of user reaches 81.2 km with the following characteristic: in week 1, the average total journey length was about 45.12 km with the shortest distance being 1.6 km and the furthest distance being 48.3 km. In week 2, the total average journey distance was 68.25 km with the shortest and the longest journey length being 9.7 km and 77 km, respectively.

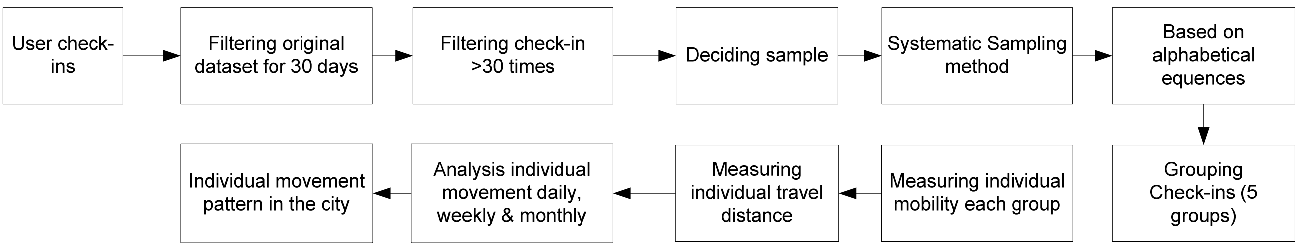
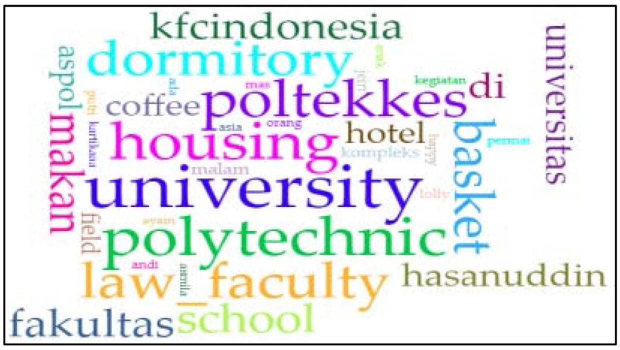


Fig. 3 Data flow diagram of method used

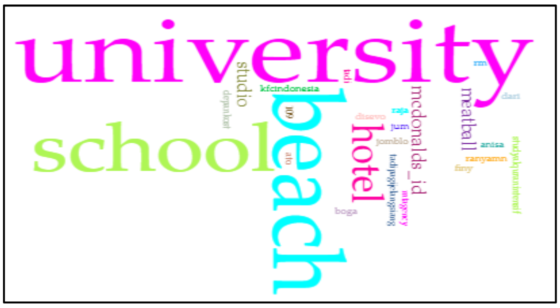
(a) group 1

(b) group 2

(c) group 3



(d) group 4



(e) group 5

Fig. 4 Analysis of user texts posted by different groups

Meanwhile for week 3, the average total movement reached 78.97 km with the shortest distance being 2.5 km and the longest 81.2 km. While in week 4, the total average length of journeys for group 2 reached 82.03 km, with the shortest and longest distance being 21.3 km and 64.4 km, respectively. It was concluded that the average length of individual trips per day for four weeks in group 2 is approximately 9.15 km.

At the same time, the authors identify the kinds of places that people visited. Fig. 4 shows the words frequency percentage of check-in venues. Analysis of the results shows that, generally user check-ins at places such as tour and travels shop 20%, cinema 12%, restaurant 13%, coffee shop 11%, dormitory 11%, dinner 9%, mall 9%, stadium 8%, and gym 7% (see Fig. 4 (b)).

C. Group 3

As shown in Fig. 4 (c), we analyze the places of group 3. The majority of tweet activity in this group is covered by location, for instance, the mall 22%, office 20%, faculty 11%, hotel 8%, nursing college 8%, KFC 7%, university 7%, hajj dormitory 6%, and cinema 8%. We observe that this group is dominated by places activity e.g., mall, office, and university faculty.

In this group, we analyze 40-49 times check-in by 16 users. Referring to Table II of group 3, we found that the highest mobility was 150 km with the following comparison: The average mobility of subjects was 84.34 km, with the shortest and longest distance being 3.3 km to 84.4 km, respectively. Then in the second week, mobility increased with the shortest and longest distance being 13 km to 94.3 km. While in week 3 and week 4, the shortest and longest distance reached 6.7 km to 150 km and 9.2 km to 94.9 km. respectively. We concluded that the average travel length per day during the four week period was 14 km.

D. Group 4

In general, the user activity in this group was closely related to the individual's activity within the university (see Fig. 4 (d)). Due to their activities around the university, it was concluded that their status was as a student. The results showed that about 53% of user check-ins were sourced from the polytechnic school 14%, home 13%, dormitory 12%, law faculty 12%, basketball court 10%, high school 10%, KFC 9%, and coffee shop 7%. The inclusion of places such as the basketball court, coffee shop and dormitory are the kinds of activity conducted around the college.

This group displays the spatial distribution of users of between 60-69 check-in instances by seven users. Table II of group 4 shows that their highest spatial movement was 123.2 km with the following characteristics: in the first week, the lowest distance of individual journey was 23.7 km and the highest was 84 km, and for the second week, the minimum trip distance was 15.8 km and the maximum was 78.9 km. Meanwhile, the shortest and longest distances for week 3 and week 4 were 19 km to 123 km and 7.9 km to 103.7 km, respectively. From the results, it was concluded that the average daily travel distance was 6.56 km.

E. Group 5

As shows in Fig. 4 (e), the majority of tweet activity was conducted at university and beach. The places percentage was dominated by activities at university 30%, beach 22%, high school 18%, hotel 10%, McDonald 7%, culinary shop (meatball) 7%, and photo studio 7%.

The check-in activity for this group was between 70-79 check-ins with two users. From Table II of group 5, it was observed that the individual with the highest mobility distance was 166 km. For week 1 and week 2, the average user journey length was 30.59 km and 78.7 km, respectively. While for week 3 and week 4, the average user journey length was 14.66 km and 13.05 km, respectively. We concluded that the total of the individual average distance traveled daily of group 5 was 9.63 km.

IV. Conclusion

The mobility dataset used in this study was collected through the Twitter Streaming Application Program Interface (API). The study focused on data that showed the check-in (specific location), time stamp, and user's status text or post activities. From this, the study measured the displacement distance of each user (daily and weekly periods) from one point to another point based on the check-in parameters. In this analysis, we used a systematic sampling approach to decide the number of research sample from Twitter user population. This paper presents a method for analyzing human mobility in Makassar city.

Analysis of the results determines that individual movement shows a great level of regularity and intensity in a specific location and at a certain time. Individuals tend to check-in at locations where their daily activities take place. For example, almost every day, the participants of this study use social media at university, which can be seen in the user text activities posted in each group (see Figs. 4 (a) and (c)-(e)). Secondly, the tendencies of the subjects were almost the same; aside from university, the next most visited destination was a shopping mall. It is worth noting that for this activity, the authors cannot be sure if the purpose of the visit was to shop or engaged in another activity (e.g., meeting friends at a coffee shop). In general, the movement pattern of the subjects in the study is: university – mall – home, university – dormitory, office – mall – home, and office – home and other. An interesting finding was that average daily mobility was 9.6 km. Thus, the results of this analysis can provide additional data for city planners to address, in particular, problems related to public transportation and traffic congestion.

Appendix

TABLE II
Distribution of Individual Weekly Travel Distances

	Trips per week (km)				
User Code	**Week 1**	**Week 2**	**Week 3**	**Week 4**	**Check-ins**
Group 1					
A54	48.1	n/a	n/a	76.4	31
A39	26.8	159.7	n/a	46.9	31
A25	12.9	43.6	22.7	18.2	31
A21	22.0	18.5	30.5	39.7	31
A3	1.30	9.3	15.8	7.30	31
A44	4.4.0	69.7	37.7	38.0	32
A34	15.0	25.3	23.3	24.7	32
A52	48.0	37.2	54.7	21.3	34
A42	42.4	n/a	n/a	132.6	34
A2	9.60	51.5	49.0	19.8	34
A26	29.2	66.2	30.4	4.40	35
A33	14.7	37.6	30.9	27.6	36
A7	46.6	8.40	12.7	64.6	36
A53	6.90	44.0	37.8	2.40	37
A20	8.20	24.1	24.7	69.4	38
Group 2					
A18	1.60	38.1	59.0	39.3	41
A16	33.3	10.2	35.1	59.1	41
A45	18.4	36.5	50.9	21.8	42
A31	25.4	9.7	26.2	25.9	42
A22	25.9	20	43.9	47.3	42
A19	36.9	54.2	13.9	56.7	42
A50	48.3	17.2	50.4	21.3	43
A30	6.00	50.7	41.4	48.2	43
A11	22.5	33.0	81.2	32.5	43
A24	36.5	39.7	28.8	23.2	44
A1	0.7	n/a	2.5	42	44
A43	25.7	45.4	53.1	47.5	45
A41	33.7	45.5	13	42.9	45
A35	1	77.6	53.4	66.4	46
Group 3					
A28	48.1	41.7	55.6	37.3	50
A23	21.2	89.8	54.5	37.8	50
A48	14.3	13.0	27.2	94.9	51
A38	24.1	48.8	33.7	87.3	51
A46	38.5	43.6	34.5	18.5	52
A37	40.5	43.4	89.4	83.4	52
A29	3.30	70.3	15.9	77.0	53
A12	41.0	46.2	74.7	53.1	53
A27	9.10	51.0	6.70	n/a	54
A17	96.0	23.2	31.2	62.8	54
A32	84.4	15.0	22.7	34.4	56
A8	18.1	33.3	79.1	29.7	56
A49	70.1	52.1	18.2	38.7	57
A4	21.1	34.5	150.2	9.20	57
A47	22.1	61.5	71.2	49.6	59
A10	38.4	94.3	91.1	17.1	59
Group 4					
A13	23.7	47.3	52.8	103.7	64
A13	n/a	n/a	17.1	84.8	64
A14	43.2	78.9	22.4	47.5	65
A36	54.4	42.4	27.5	7.90	66
A51	82.2	15.8	123.2	28.3	67
A15	84.0	76.8	19.6	78.8	69
A9	50.0	50.7	30.6	82.4	69
Group 5					
A5	n/a	166	70.6	38.3	73
A6	63.8	48.2	32.1	53	74
	54.81	78.7	72.49	80.79	
Individual daily mobility: 9.56 km					

REFERENCES

[1] M. C. Gonzalez, C. A. Hidalgo, A. L. Barabasi, “Understanding individual human mobility patterns Nature”, *Nature,* vol. 453, pp 779-782. June 2008.

[2] A. Noulas, S. Salvatore, R. Lambiotte, M. Pontil,C. Mascolo, “A tale of many cities: universal patterns in human urban mobility,” *PloS one*, vol. 7: e37027. May 2012.

[3] C. Song, T. Koren, P. Wang, A. L. Barabasi, “Modelling the scaling properties of human mobility,” *Nature Physics*, vol. 6(10), pp. 818-823. September 2010.

[4] R. Ewing, R. Cervero, “Travel and the built environment: A synthesis”, *Transportation research record*, No. 01-3515, pp.87–113.

[5] B. Vilhelmson, “Daily mobility and the use of time for different activities”, *Geo-Journal*, vol. 48, no. 3, pp. 177–185. 1999.

[6] S. Hanson, “Perspectives on the geographic stability and mobility of people in cities”, *Proceedings of National Academy of Science of the US*A, vol. 102, no. 43, pp. 15301-15306. October 2005.

[7] E. Thomas, I. Serwicka, P. Swinney, “Urban demographics where people live and work”, *DAC beachroft*. November 2015.

[8] C. Boldrini, A. Passarella, “HCMM: Modelling spatial and temporal properties of human mobility driven by users’ social relationships”, *Computer Communications*, vol. 33, no.9, pp. 1056–1074. June 2010.

[9] C. Song, Z. Qu, N. Blumm, A. L Barabasi, “Limits of predictability in human Mobility”, *Science* 327: 1018–1021. 2010.

[10] X. Cao, G. Cong, C.S. Jensen, “Mining Significant Semantic Locations from GPS Data,” *Proceedings of the VLDB Endowment*, vol. 3, no. 1, pp. 1009-1020, Sept. 2010.

[11] Twitter. Open twitter streaming api. (2015, August 26). *https://dev.twitter.com/docs/streaming-api.*