International Journal on Advanced Science Engineering Information Technology

Vol.8 (2018) No. 4
ISSN: 2088-5334

Spatiotemporal Mining of BSS Data for Characterising Seasonal Urban Mobility Dynamics

Ida Bagus Irawan Purnama

Electrical Engineering Department, Bali State Polytechnic, Badung, 80361, Indonesia
E-mail: ida.purnama@pnb.ac.id

***Abstract*— Digital traces of individual mobility can be revealed from the origin-destination sensing systems of BSS (Bicycle Sharing System). This record enables wide analysis of human mobility traits in urban area including pattern, trend, and anomalies. This study investigates and compares trip history of BSS open data from two cities, London and New York, along a year period with respect to annual weather data as explanatory factors. This aims to get insights about seasonal urban mobility dynamics both temporally and spatially. Results show that, for both cities, there are differences as well as similarities of temporal correlation level between riding behavior of BSS users and hour of the day, day of the week, season, and local weather. Practically, the most correlated factor can be further considered and used as predictive features. Meanwhile, the proposed spatial analysis shows the positive bikes imbalance occurs in the morning, mostly at inner stations because of inward flow, and vice versa. This spatial extent can be used for redistribution purpose, specifically in order to provide enough resources at the highly visited stations before peak time occurs.**

***Keywords*— data mining; knowledge discovery; open data; spatiotemporal analysis; bicycle sharing system; urban mobility dynamics.**

I. INTRODUCTION

The endorsement of healthier, cheaper, and simple mobility option in the crowded urban area leads to the enormous expansion of Bicycle Sharing System (BSS), roughly 1000 systems worldwide as of December 2016 [1]. It is also to be the desired reduction in a traffic jam and carbon pollution. In addition, this system prevents people from being troubled with private bike ownership issues such as parking, storage, routine maintenance, and theft. Most of BSS also come with mobile apps. These entire benefits place BSS are among the most prospective sustainable transportation mode [2].

Equipped with sensing systems at their pickup and return points, BSS record the pickup and return times of a one-way bike hire that reflects the individual mobility. Furthermore, BSS allows users to freely determine their trip routes as well as trip times so that they provide a fine-grained mobility trace. Bikes can be rented and returned flexibly based on users need. As a result, the flexibility of BSS offers specific efficiency for users. On the other hand, they arise from uncertainty and randomness for bike operators [3]. In this case, BSS are on-demand means that need intense monitoring to keep enough resources (bikes and docking slots) in proper locations and times [4]. Conversely, other transportation modes such as buses or trains are mass-based system with fixed schedules and routes so that dynamics resources allocation are not critical [5]. However, people tend to visit similar places in their daily activities. This makes the likely BSS use could be projected from the mobility record, which is embedded in their trip history. Accordingly, by examining some critical features such as weather, season and peak times, the mobility pattern of BSS users should be able to be understood [4].

Designed for a one-way and short trip purpose, docking stations of BSS are usually installed in homogeneously every few hundred miles. To use the system, a user may take a bike from any station then returns it to another station in the system and usually free for the first 30 minutes [6]. The problem arises if the availability of bikes and vacant docking slots cannot encounter the immediate demand. This makes users cannot use the system. They may not find bikes to hire. In another case, they may not be able to bring them back to the preferred locations [2]. This circumstance is named as imbalanced state and will degrade the service level if it happens frequently. Hence, bikes redistribution is needed and becomes an essential challenge for operators. A proper scenario based on historical users mobility pattern may help to make the redistribution task as efficiently as possible [7].

From a temporal aspect, Borgnat et al. [6] have investigated BSS data from Lyon, France, using a signal processing approach. Then, the time evolution of users' mobility dynamics is modeled. The proposed cycle station term for a temporal pattern that varies over days of the week using one-hour resolution range. They grouped this temporal pattern into weekdays showing three usage peaks, in the morning, at noon lunchtime, and late afternoon, while weekend days showing only one peak mostly in the afternoon. Meanwhile, O'Brien et al. [8] explored the station's usages footprint from 38 BSS on the globe to reveal a general view of temporal changes in bikes distribution within those stations. They also observed the dynamics of stations occupancy rates to see their fluctuation over time. Still in Europe, using Barcelona data, Froehlich et al. [9] conducted a temporal usages analysis for all stations. They identified common behaviors pattern across stations. There is a relation between shared behaviors and location and proximity (spatial), and also the hour of the day (temporal). They also exhibited that the insights of urban dynamics and aggregated users mobility can be gained using BSS data. Similar to [8], for station usages, a repeating three peaks are also found on weekdays. This shows the similar morning-evening commuting and lunchtime in Lyon and Barcelona.

From weather outlook, Corcoran et al. [10] explored the trip flows in Brisbane in correspondence with the weather. They underlined that, at the system-wide, rain level and winds speed are considerably related to the trip counts. Meanwhile, the likely weather effect investigation was also conducted by Gebhart, and Nolan [11] for trip flows in Washington DC. In this case, they used more weather features and linked them to an hourly count of uptakes and duration of renting. Their findings suggest that the impact of cold temperature and high level of rainfall as well as humidity are relatively substantial to reduce the probability of renting and the duration of use.

From a spatial extent, the buffer area around each bike station that may have a particular impact on that station was introduced by O'Brien et al. [8]. It is approximately one kilometer. This distance then can be used as initial proximity for further spatial analysis such as to observe the impact if one station is out of services. In other study, stations spatial dependencies are visually shown by Froehlich et al. [9]. They displayed that less active stations are mostly situated at the outer part of the system, while uphill stations with higher altitude incline to be vacant. Both of these spatial studies show that there is a particular spatial neighborhood correlation among stations.

This data-driven study mainly aims to investigate the temporal aspects of seasonal urban mobility dynamics within a long period of days and to find the correlation between these mobility dynamics and external factors. This study also localizes stations imbalance at peak times by spatial visualization and propose a station activities measurement for that imbalance. Ultimately, this study observes and analysis origin and destination pair between most visited stations that have common bikes exchange at peak times when usages are high.

II. Material And Method

A. Dataset

The main datasets used in this study are from one year (2017) trip history of London BSS (Santander Cycles) and New York BSS (Citibike). London data has 9.4 million trip records from 730 bike stations, whereas New York data has 16,3 million trip records from 652 bike stations. Both of these open data can be accessed from http://cycling.data.tfl.gov.uk/ for London and www.citibikenyc.com/system-data for New York. Then, weather data that are used as descriptive factors are from http://nw3weather.co.uk/ and www.wunderground.com for London and New York respectively.

B. Methodology

Temporal analysis is first conducted to reveal BSS daily usages within one-year period for both cities. This aims to investigate whether the BSS daily usages follow specific patterns, seasonal trends, and anomalies. Later, the annual weather data, which consist of daily rainfall, avg. of humidity, temperature, pressure, wind speed, and dew point, are supplied. The correlation coefficient for each feature is calculated using the Pearson test and Evans range [12] to unfold the features, which have significant impacts on BSS users. This is also intended to highlight any differences as well as similarities. Afterward, hourly usage over the course of a day is analyzed to reveal the busy hours when usages reach the peak and imbalance potentially occurs.

Using the peak hour's subset, the spatial analysis is then conducted to visualize the usages weight of bike stations during the peak hours of pickup and return. This is followed by showing the balance level to localize which stations get more bikes or have more vacant docking slots because of imbalanced pickup and return activities. This spatial insight is expected can be used for rebalancing purpose before peak times occur when demand is very high. After that, the inter-station or OD pair's analysis using circular plot is presented to observe the connection weight between stations. Knowing the common pair between stations is expected to give insights to which stations a station has intense bikes exchange. This could be useful to understand which other stations will become a new pair when the common pair is unavailable.

III. Results And Discussion

A. Daily Usages Characteristics

By aggregating individual trip on a daily basis, Fig.1 shows the patterns of daily usages that vary for weekdays (sky blue bar), weekends (red bar), and holidays (blue bar). Both cities show general trends where usages on weekdays are mostly more than on weekends and holidays showing a commuting trait. There are anomalies in London where weekends are more than weekdays which are in the second week of April and the first weekends of July. Also, in Christmas week on December are more than a normal weekday. There are a couple of days when usages are very low, below than normal average, or even zero. They are between the end of September and the first week of October for London, and in the mid of February and also March for New York. These indicate specific occasion may happen during those periods.

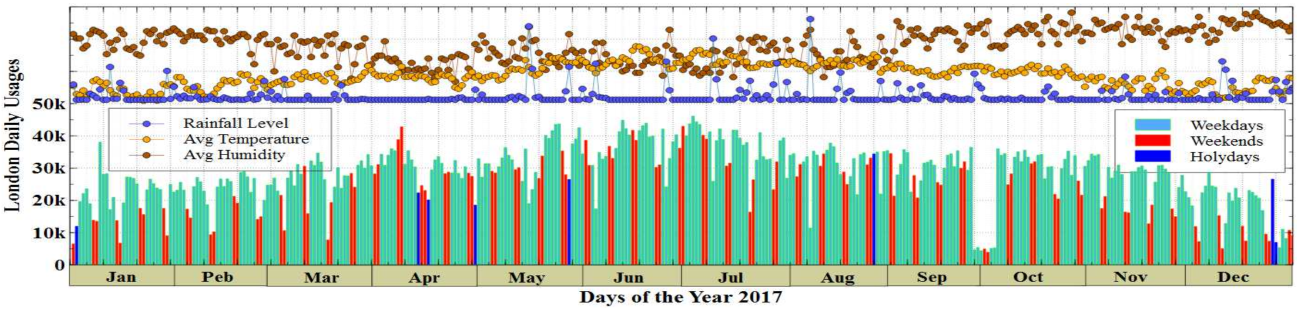
Another trend is usages increase toward summer (from January to July) and vice versa, decrease toward winter (from July to December). This could be related to the weather showing that there are specific seasonal features, which highly influence riding behavior of BSS users. Pearson's Correlation Coefficient (PCC) test is used to observe the correlation of each feature of external factors with the daily usages.

TABLE I.
WEATHER COEFFICIENTS CORRELATION BY PEARSON TEST

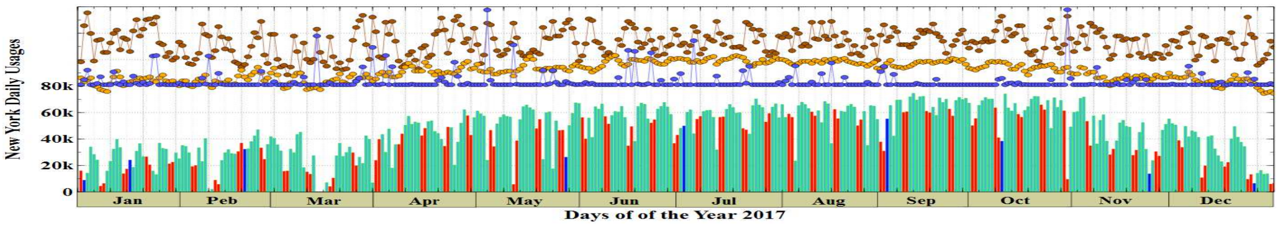
Parameters	London		New York	
	Corr (%)	Range	Corr (%)	Range
Rainfall	-33	*Weak*	-33	*Weak*
Temperature	62	*Strong*	80	*Very Strong*
Humidity	-58	*Moderate*	-2	*Very Week*
Pressure	17	*Very Week*	3	*Very Week*
Wind Speed	-20	*Week*	-56	*Moderate*
Dew Point	42	*Moderate*	67	*Strong*

Then, the Evans Range of Correlation is used for verbally classifying those levels of correlation. The result is shown in Table I where the negative value shows the relationship between the variables is negatively correlated, or if one variable decreases the paired one increases. Conversely, the positive value suggests the relationship between the variables is positively correlated, or both of them decrease or increase at the same time. Then, a class range are named for each correlation.

In London, the only feature, which has a strong correlation, is temperature followed by humidity and dew point, which have a moderate correlation. Meanwhile, in New York, the temperature has a very strong correlation followed by dew point and wind speed, which have a strong and moderate correlation respectively. Concerning previous study results in Brisbane [10] and Washington DC [11], this suggests that, in different places, BSS users have distinct riding behavior when dealing with local weather conditions.

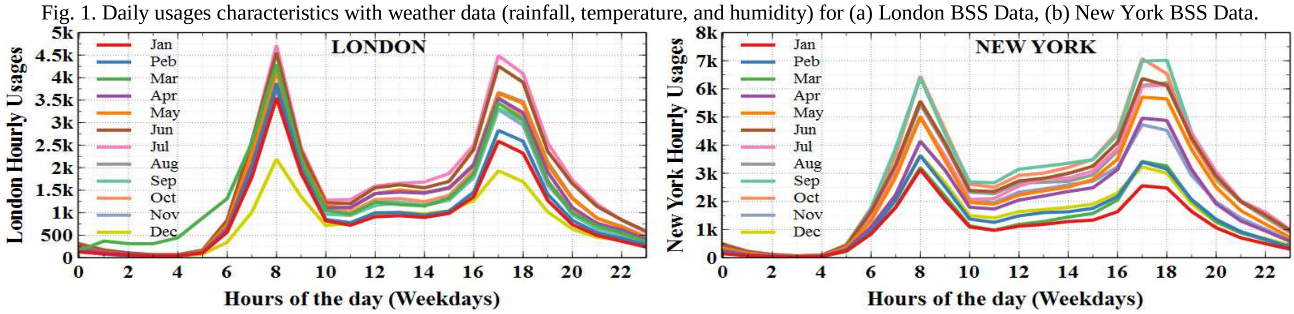


(a) London BSS Data

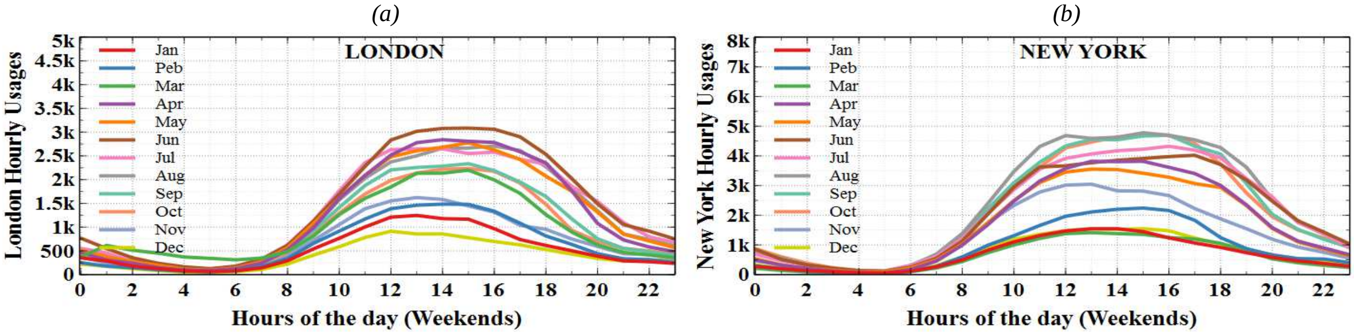


(b) New York BSS Data

Fig. 1. Daily usages characteristics with weather data (rainfall, temperature, and humidity) for (a) London BSS Data, (b) New York BSS Data.



(a) *(b)*



(c) *(d)*

Fig. 2. Hourly usages characteristics for London (a) Weekdays and (c) Weekends, for New York (b) Weekdays and (d) Weekends

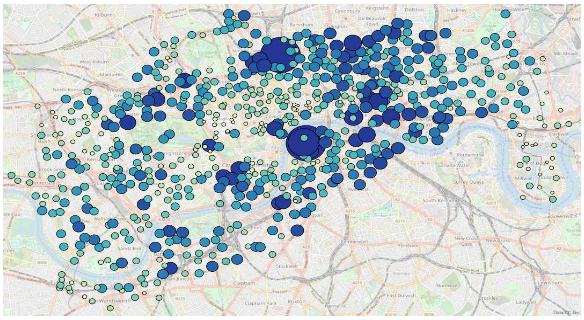
.

For instance, the temperature was found not to have an essential impact for bikers in Brisbane because this Queensland capital is located in the sub-tropical climate region. Here, the variation of temperature is relatively small [10]. However, there is a similarity between London and New York where rainfall has a similar correlation level, which is only -33%. Practically, the very strong and strong features can be used as proper predictive features for prediction. For example, to predict the next hour of pickup or return to certain stations.

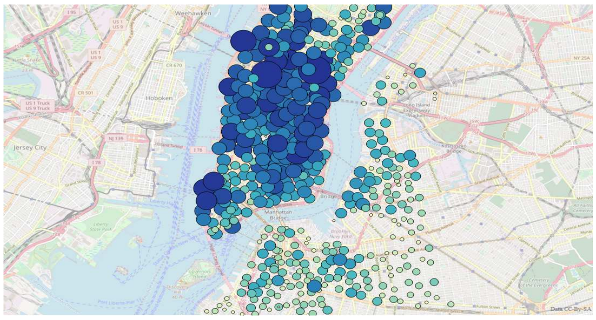
B. Hourly Usages Characteristic

Usage characteristic on hourly basis tells how it varies over the course of a day. Fig. 2 displays hourly characteristic averaged per month. It can be seen that both cities show similar patterns. Weekdays have two sharp peaks, in the morning (between 6 am to 10 am) and afternoon (between 4 pm to 8 pm) indicating busy hours. On the other hand, weekends only have one moderate peak in the middle of the day distributed between 10 am to 8 pm. Here, the peaks on weekdays correlate with times when people usually take a trip to and from their workplaces in the morning and the evening respectively. It can be stated that hourly characteristic on weekdays indicates the commuting pattern. Also, the afternoon commuting peak of all months is broader. This means that the spread of commuting hours in the afternoon is wider compared to the morning. Gaining sharp peaks in hourly usage may arise asymmetric flows in the system. This may causes imbalance in bikes availability per station if no proper rebalancing is undertaken.

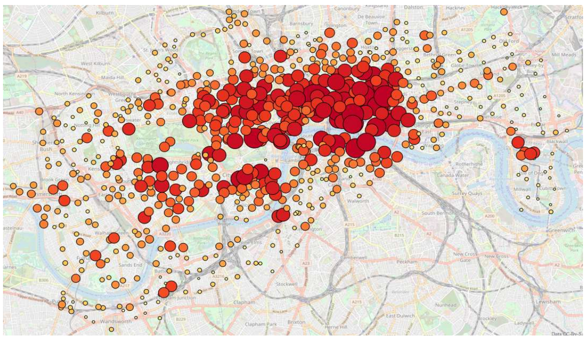
A socio-cultural aspect of a city can also be revealed from the existence of two spikes in hourly usage. In this case, not many citizens use BSS during weekday lunchtime in both London and New York. This finding is unlike a work conducted by Borgnat et al. [6] in Lyon and Froehlich et al. [9] in Barcelona. They found not only two sharp peaks during weekdays morning and afternoon but also one moderate peak at lunchtime indicating that BSS are used not only for weekdays commuting. Meanwhile, on weekends, only a moderate peak arises that concentrates in the midday showing leisure utilization. Once again, this weekend result is dissimilar from Barcelona pattern with two consecutive peaks around midday and in the afternoon [9]. Yet, it is quite similar to Lyon with one peak centralized in the afternoon.



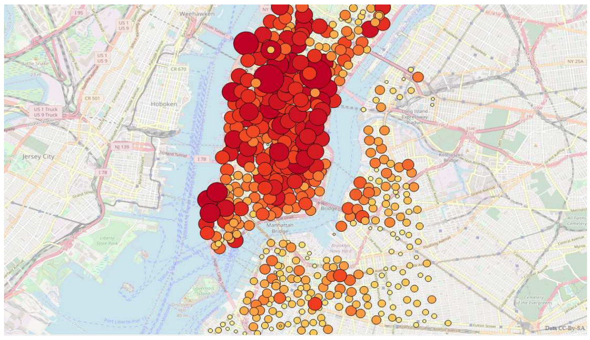
(a) Pickup distribution in the morning peak hours (London)



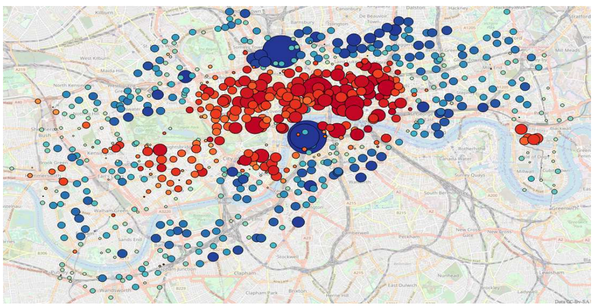
(b) Pickup distribution in the morning peak hours (New York)



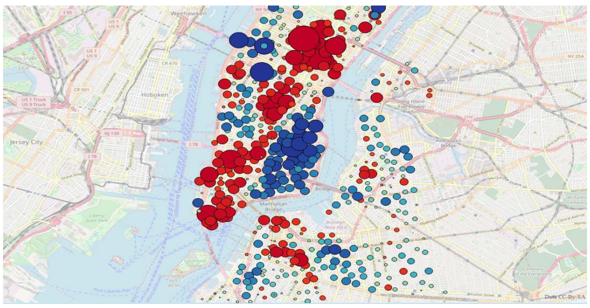
(c) Return distribution in the morning peak hours (London)



(d) Return distribution in the morning peak hours (New York)



(e) Bikes balance = Return – Pickup (London)



(f) Bikes balance = Return - Pickup (New York)

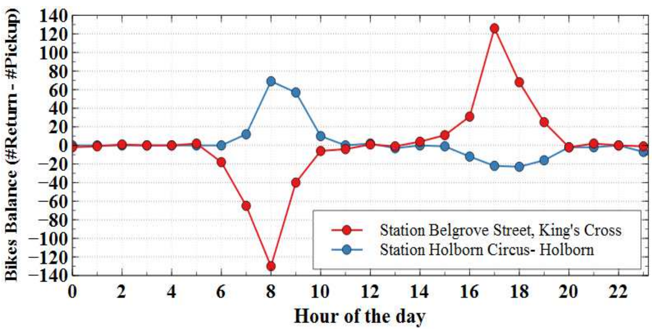
Fig. 3. Station activities, pickup, return, and bikes balance, in the morning peak hours in London (a,c,e) and New York (b,d,f)

C. Station Usages Analysis

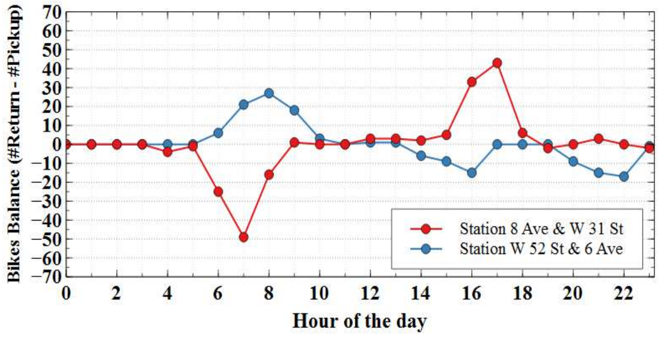
Examining BSS workflow, there are dynamic processes that determine the availability of resources (bikes and vacant docking slots) at stations. One state and three dynamics activities can categorize them. The state is available bikes as an opening balance, and activities are a pickup, return and redistribution processes that will dynamically increase or decrease the balance. For each station, its bike balance for a specified period can be calculated as opening balance + return – pickup ± redistribution. As London and New York BSS trip data only consist of pickup and return records, this section will visually investigate the weight and balance of stations solely based on these two activities.

Fig. 3 show the visualization examples of station activities during peak hour in the weekdays morning (between 6 am to 10 am) when stations have the highest usages. Circles represent the weight of stations. The big circle indicates a busy station. In London, busier stations for pickup are shown mostly in the outer areas of the city (3.a). Simultaneously, busier stations for return are coming closer to the inner areas (3.b). Therefore, this outer city to inner city bikes flow can be defined as an inward flow. As people who commute mostly travel from suburbs, this inward flow is not unexpectedly. From their home, they may use public transport to stations in the morning then use bikes for the last way of their destination (workplaces). For both cities, the outcomes of the inward flow can be seen in Fig. 3.e&f which display the number of bikes balance as a result of return minus pickup. Contrarily, such activities are supposed in the afternoon which is the opposite of the morning pattern. In this case, the flow is from inner to outer, and it can be stated as an outward flow signaling commuters leave the city.

Inward and outward flow produce imbalance. For instance, the brown circles in Fig. 3.e&f, mostly in inner stations, denotes that return is more than a pickup in those stations in the weekday morning. This indicates more bikes and lack of empty docking slots so that operator has to remove some bikes. This state is defined as a positive bikes imbalance. Simultaneously, the opposite state where the pickup is more than return appears mostly in outer stations. This is displayed by the blue circles which yields more vacant docking slots than bikes. This is called a negative bikes imbalance. Fig. 4 exhibit the example of these both bikes imbalance over an hour of the day for two stations in each city.



(a) London's stations



(b) New York's stations

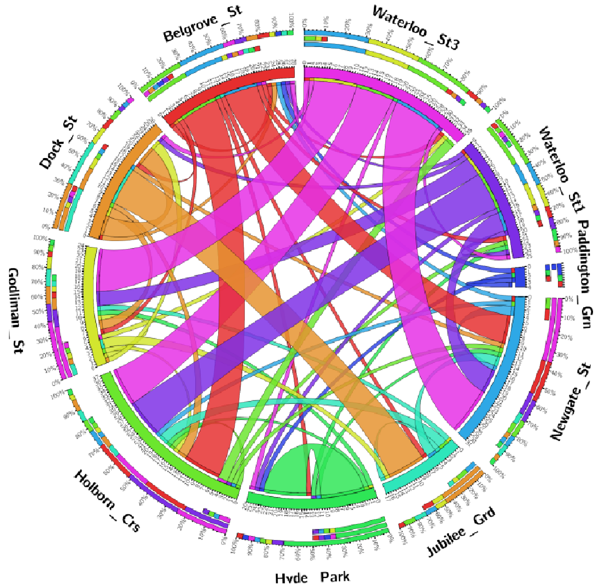
Fig. 4. Bikes balance in stations as a result of return minus pickup.

Fig. 4.a indicates Belgrove station (red circle) has more pickup than return in the morning peak times, between 5 am to 9 am. This leads to a lack of bikes in this station (negative imbalance). Conversely, there are more bikes in this station in the afternoon peak times when it has more return than pickup (positive imbalance). This pattern is similar to Station 8 Ave & W 31 St in New York, Fig. 4.b, but it is opposite to Station Holborn and Station W 52 St & 6 Ave (blue circle). This proposed method can give an overview of the level of imbalance dynamics over the course of a day for specific stations. Bikes operators are suggested to monitor such high imbalance stations frequently.

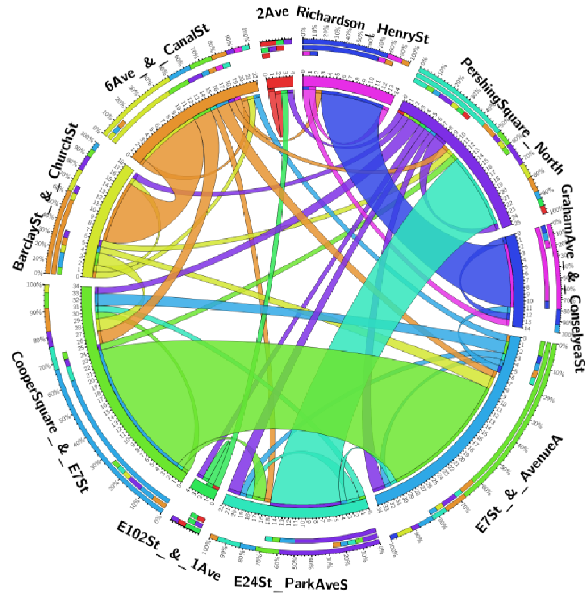
D. Origin-Destination Pairs Analysis

As a dynamics network, BSS contains OD pairs between stations that vary overtimes. One approach to analyzing the dynamicity of OD pairs is using a circular plot [13]. Here, this plot is used to display the inter-stations bounds so that their weight can be investigated. Fig. 5.a shows the morning peak time of OD pairs in London for ten stations. There are four highly visited stations as the origin (pickup) which are Dock_St (brown), Belgrove_St (red), Waterloo_St3 (magenta), and Waterloo_St1 (purple). In this case, Waterloo_St3 is the primary origin for Godliman_St, Holborn_Crs, and Newgate_St shown by Magenta fat link from Waterloo_St3 to those three destination stations. Conversely, Waterloo_St3 is the leading destination for these three stations in the afternoon shown by yellow, light green, and light blue link from these three stations coming into Waterloo_St3, Fig. 5.c.

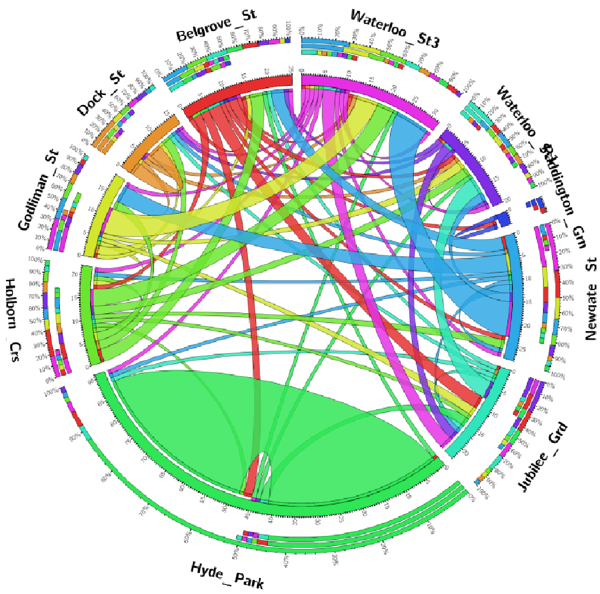
There is also a round trip anomaly where the destination station is similar to its origin station. This is is a typical case because, in BSS, the destination stations are usually different from their origin as this shared system is intended for the one-way rent mechanism. In London, this round trip occurs mostly at Hyde Park station (green), and it is much more frequent in the afternoon. This signifies a leisure use around the park, whereas, in New York, they are similar to London in which if a station becomes a primary origin in the morning, then it becomes the leading destination in the afternoon. This is shown by a link between Cooper Square _&_E7St and E7St_&_AvenueA in the afternoon (blue), Fig. 5. d, which is a reverse of their morning link (green), Fig. 5.b. This OD pairs analysis can be further used to find the simple pair among stations. If their pairs are immediately unavailable, then the effect of this change on the network can be understood. This insight may help the operator to set a backup scenario if some stations are out of services.



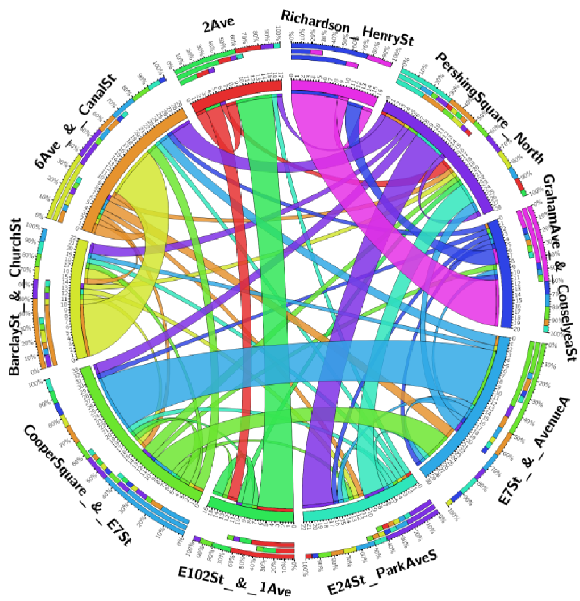
(a) Morning peak time of OD pair in London



(b) Morning peak time of OD pair in New York



(c) Afternoon peak time of OD pair in London



(d) Afternoon peak time of OD pair in New York

Fig. 5. Inter-station pair weight of the morning and afternoon peak in London (a,c) and New York (b,d) for ten stations.

IV. Conclusion

Spatiotemporal mining of BSS open data with explanatory factors has been undertaken to investigate seasonal human mobility dynamics in two big urban area, London, and New York. Some spatiotemporal knowledge of urban mobility using BSS has been revealing. It is found that daily usage is highly correlated with hours of the day, days of the week, season, and local weather. Also, hourly usages show a strong commuting trait on weekdays with two sharp peaks both in the morning and afternoon indicating busy hours. The commuting pattern is also shown by spatial analysis results in the morning peak time, where outer stations get more pickup, while inner stations get more return showing inward flow from suburbs, and vice versa. This suggests that the operator of the bike have to pay more attention before peak times occur, especially at stations which have high demand to keep enough resources. For example, from the results of bikes balance analysis, the outer stations should have more bikes in the morning and more vacant docking slots in the afternoon. Meantime, OD pairs analysis show that there are specific pairs among stations in interchanging bikes frequently. However, there is also found a roundtrip anomaly where pickup and return stations are similar.

As the trip number is related to some external factors that have been found, the future work of this paper will be continued to develop a seasonal prediction system. More importantly, how that prediction system can then help bikes operator to maintain their service level is worth to be investigated. For instance, knowing how many bikes will be rented for the next couple of hours will be practically useful.

REFERENCES

[1] D. Gutman, "Will helmet law, kill Seattle's new bike-share program?" The Seattle Times, December 19, 2016.

[2] I. B. I. Purnama, N. Bergmann, R. Jurdak, and K. Zhao, "Characterising and Predicting Urban Mobility Dynamics by Mining Bike Sharing System Data," in *the 11th IEEE UIC*, pp. 159-167, 2015.

[3] B. Chen, F. Pinelli, M. Sinn, A. Botea, and F. Calabrese, "Uncertainty in urban mobility: Predicting waiting times for shared bicycles and parking lots," in *the 16th International IEEE Conference on Intelligent Transportation Systems*, pp. 53-58, 2013.

[4] I. B. I. Purnama, "Spatiotemporal analysis of urban mobility dynamics: A case study of bicycle sharing system," Uni. of Queensland, 2018.

[5] R. Jurdak, "The impact of cost and network topology on urban mobility: A study of public bicycle usage in 2 US cities," *PloS One*, vol. 8, 2013.

[6] P. Borgnat, P. Abry, P. Flandrin, C. Robardet, J.-B. Rouquier, and E. Fleury, "Shared bicycles in a city: A signal processing and data analysis perspective," *Advances in Complex Systems*, vol. 14, pp. 415-438, 2011.

[7] P. DeMaio, "Bike Sharing: History, Impacts, Model of Provision, and future," *Journal of Public Transportation*, vol. 12, no. 4, 2009.

[8] O. O'Brien, J. Cheshire, and M. Batty, "Mining bicycle sharing data for generating insights into sustainable transport systems," *Journal of Transport Geography*, vol. 34, pp. 262-273, 2014.

[9] J. Froehlich, J. Neumann, and N. Oliver, "Sensing and predicting the pulse of the city through shared bicycling," in *International Joint Conference on Artificial Intelligence IJCAI*, pp. 1420-1426, 2009.

[10] J. Corcoran, T. Li, D. Rohde, E. Charles-Edwards, and D. Mateo-Babiano, "Spatio-temporal patterns of a Public Bicycle Sharing Program: the effect of weather and calendar events," *Journal of Transport Geography*, vol. 41, pp. 292-305, 2014.

[11] K. Gebhart and R. B. Noland, "The impact of weather conditions on bike-share trips in Washington, DC," *Transportation*, vol. 41, 2014.

[12] J. D. Evan, "Straightforward statistics for the behavioral sciences," *Pacific Grove, CA: Brooks/Cole Publishing*, 1996.

[13] N. Sander, G. J. Abel, R. Bauer, and J. Schmidt, "Visualising migration flow data with circular plots," *Vienna Institute of Demography*, 2014.