Analysis of Public Transportation Patterns in a Densely Populated City with Station-based Shared Bikes

Di Wang
Joint NTU-UBC Research Centre of Excellence in Active Living for the Elderly (LILY Research Centre)
Nanyang Technological University
Singapore
wangdi@ntu.edu.sg

Evan Wu
School of Computer Science and Engineering (SCSE)
Nanyang Technological University
Singapore
wuevan1993@gmail.com

Ah-Hwee Tan
School of Computer Science and Engineering (SCSE) &
LILY Research Centre
Nanyang Technological University
Singapore
asahtan@ntu.edu.sg

ABSTRACT

Densely populated cities face great challenges of high transportation demand and limited physical space. Thus, in these cities, the public transportation system is heavily relied on. Conventional public transportation modes such as bus, taxi and subway have been globally deployed over the past century. In the last decade, a new type of public transportation mode, shared bike, emerged in many cities. These shared bikes are deployed by either government-regulated or profit-driven companies and are either station-based or station-less. Nonetheless, all of them are designed to better solve the last-mile problem in densely populated cities as complements to the conventional public transportation system. In this paper, we analyse the public transportation patterns in a densely populated city, Chicago, USA, using comprehensive datasets covering the transportation records on shared bikes, buses, taxis and subways collected over one year's time. Specifically, we apply self-regulated clustering methods to reveal both the majority transportation patterns and the irregular ones. Other than reporting the autonomously discovered transportation patterns, we also show that our method achieves better clustering performance than the benchmarking methods.

CCS CONCEPTS

• **Computing methodologies** → **Cluster analysis**; *Neural networks*;
• **Applied computing** → **Transportation**;

KEYWORDS

Self-regulated clustering, shared bike, densely populated city, public transportation pattern

ACM Reference Format:
Di Wang, Evan Wu, and Ah-Hwee Tan. 2018. Analysis of Public Transportation Patterns in a Densely Populated City with Station-based Shared Bikes. In *The 3rd International Conference on Crowd Science and Engineering (ICCSE'18), July 28–31, 2018, Singapore, Singapore.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3265689.3265697

ICCSE'18, July 28–31, 2018, Singapore, Singapore

ACM ISBN 978-1-4503-6587-1/18/07…$15.00
https://doi.org/10.1145/3265689.3265697

1 INTRODUCTION

Cities exist to provide access to people, goods, services and information that the better and more efficient this access is, the greater the social and economic benefits urban living brings. Due to various reasons, a successful city always attracts more and more residents and thus becomes more and more densely populated. To achieve better sustainability, a densely populated city must employ effective management and design strategies to minimize overcrowding. As such, public transportation systems play the most crucial role in the provision of spatial accessibility, which provide the residents an affordable and efficient way of transportation. Over the past century, conventional public transportation modes such as buses, taxis and subways have been globally deployed. In the last decade, a new type of public transportation mode, shared bikes, emerged in many cities along with the emerging trend of green transport.

Bike-sharing system is a type of service wherein bicycles are available for shared usage among various individuals on a short-term rental basis. Although some shared bikes are occasionally rented for sightseeing or other entertaining purposes, for most of the time, the shared bikes are rented by those people who need to travel a considerably long distance, which may be too time-consuming and tedious by walking. Therefore, bike-sharing problems are often considered as the last-kilometer transportation problems.

The shared bikes are deployed by either government-regulated or profit-driven companies and are either station-based or station-less. Regardless of the regulating company of the shared bikes, the rental cost is made affordable to attract more users. The station-based shared bikes are relatively better managed as rented bikes have to be returned to vacant slots. However, the fixed bike stations may cause inconvenience due to supply-demand imbalance in terms of location difference and time variance. On the other hand, station-less shared bikes are much more flexible in terms of rental and return as they do not specify the fixed locations before and after the usage. Thus, station-less bikes provide much easier access to the users. However, due to the loose management, more and more bike abuse cases were observed in cities with station-less bikes deployed.

In this paper, we analyse the public transportation patterns in a densely populated city, Chicago, USA. The reason of choosing Chicago as the city of interest is because its public transportation datasets are comprehensively available. Specifically, usage records (in varying level of details) of the station-based shared bikes, buses, taxis and subways in Chicago are all public available. We include all the aforementioned public transportation datasets over year 2016

in our study. For the knowledge discovery tool, we select a self-regulated clustering method named Interest-Focused Clustering based on Adaptive Resonance Theory (IFC-ART) [24] to reveal both the majority transportation patterns and the irregular ones. It is encouraging to find that IFC-ART is able to discover certain patterns associated to, but not embedded in, the underlying datasets, e.g., weather conditions across different days. Other than reporting the autonomously discovered transportation patterns, we also show that our method achieves better clustering performance than the benchmarking methods in most cases.

The rest of the paper is organized as follows. In Section 2, we review the related literature. In Section 3, we present the clustering method used for knowledge discovery. In Section 4, we introduce the design of the case studies and discuss the findings. Finally, in Section 5, we conclude this paper and propose future work.

2 RELATED WORK

In this section, we first review the recent studies conducted on the conventional public transportation systems. We then review the more recent studies on the emerging transportation mode: shared bikes. We finally review the studies that consider both the shared bikes and the conventional public transportation modes.

2.1 Conventional Public Transportation Systems

In this big data era, urban bus companies have collected a tremendous amount of travel data from their passengers over the past years. Thus, these data enable a series of studies regarding public bus scheduling and optimization. For example, Wang et al. [26] proposed a bus scheduling model considering the time-dependent traffic and demand. The experimental results show that their proposed model can significantly reduce the bus waiting time, thus enables the provision of similar averaged waiting time by despatching lesser number of buses. Moreover, Gokasar and Cetinel [11] found that bus stop deployment characteristics affect bus dwelling time. Hence, dwell patterns of a new bus stop can be estimated by comparing dwell pattern of a similar bus stop. Therefore, planners can avoid long dwelling time of a new or old bus stop via reallocation.

As a convenient mode of transportation, taxis have been widely used in densely populated cities. Different from bus and subway, a taxi only carries a limited number of passengers, but it provides a faster and more direct means of transportation with relatively higher expenses. Please note that private taxis such as Uber are not considered in this paper and many existing studies mainly due to the availability of their data. Rao et al. [15] found that the pattern of empty taxis searching for new passengers is highly similar to the search pattern of animal's food-hunting in the wild. Yazici et al. [28] investigated the influence of weather on the demand of taxis. They found that the impact of weather is not consistent among different time periods during the day. More interestingly, inclement weather increases travel time, but effect due to weather conditions decreases as the transportation network gets congested.

Due to their rapid speed and large capacity to carry many passengers at once, subway systems play a crucial role in densely populated cities. Xiong et al. [27] found that the conventional operational strategies may not dynamically meet the ever increasing complexity of the current subway systems. Moreover, they proposed a new management and control system for subways to provide monitoring, forecasting, warning, incident management and real-time scheduling functionalities. As such, the proposed system may improve the reliability, efficiency, safety and service level of complex subway systems. On the other hand, Deng and Xu [6] investigated the characteristics of the ridership in different subway stations and found that the ridership is highly proportional to the surrounding land use.

The afore-reviewed literature focus on the investigations on the individual usage of the conventional public transportation systems. In this subsection, we do not plan to review a combination of them, but will do so after we review the shared bike studies in Section 2.2.

2.2 Recent Studies on Shared Bikes

Most existing academic studies on bike-sharing were mainly focused on the forecast of bike usages and better dynamic balance between supply and demand. Wang [25] studied the CitiBike[1] dataset and found that in 2015 there was a dramatic increase in bike-sharing users that required the expansion of bikes and their docking stations. Thus, the effectiveness of bike rebalancing strategies became much more crucial. Moreover, Wang [25] found that long-term subscription users were less affected by time, weather and temperature difference but short-term users tend to be highly affected by those factors.

Biehl et al. [1] investigated the Divvy dataset (the same dataset used in our study, see Section 4) from June 2015 to May 2016 and found that bike-sharing functions well in communities with well-integrated transportation systems. Moreover, they found stronger elasticity effects at the community level for three key socio-spatial variables, namely the percentage of Black and Hispanic residents and the duration of the bike-sharing station's presence. As such, they suggested that the bike-sharing regulators should pay close attention to the diffusion of bike-sharing across different areas of the city. Hyland et al. [13] found that the shared bike usage in Chicago increases with the increase of bike-sharing stations within the 1-5 km radius for members and 2-8 km for non-members. Moreover, they found that the usage decreases if the bike-sharing stations are too close in proximity, i.e., within 0.8 km for members and 1.6 km for non-member. Sun et al. [17] found that in Chicago neither traffic accidents nor traffic congestion influences the shared bike usage. However, bus accessibility is positively associated with the shared bike's usage. Furthermore, both on-street and off-street violent crimes tend to decrease the usage. In this paper, we also thoroughly investigate the Divvy dataset, but focus more on the relationship of the shared bike usage with other public transportation modes.

2.3 Studies on Both Shared Bikes and Others

Brand et al. [2] stated that transport planners and services operators often fail to include the entire trip of individual commuters into consideration. Therefore, they investigated both the bus and shared bike usage data and found that bus systems with higher frequencies and speeds can attract twice the amount of cyclists on the access (entry) and egress (exit) sides. At the same time, passengers accept longer access and egress distances with higher speed and higher frequency bus services. Fishman et al. [9] found two key barriers preventing high usage of bike-sharing, namely car convenience and

[1] URL: https://www.citibikenyc.com/system-data

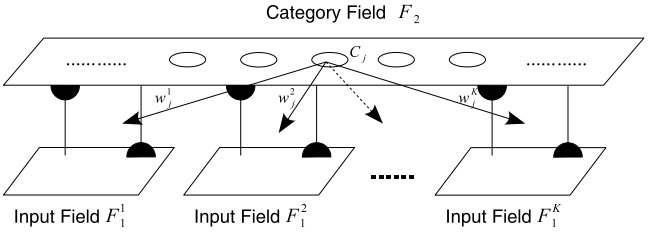


Figure 1: Network structure of IFC-ART.

docking station inconvenience (for station-based shared bikes). The first barrier may be highly correlated to the residential density in the area and the road congestion conditions. The second barrier mainly relates to the distance from the docking station to the passenger's targeted destination. Shaheen et al. [16] conducted a survey and found that bike-sharing reduces the number of bus rides in several cities. Moreover, a large decrease in car usage was observed in all cities. Campbell and Brakewood [3] found that in New York city, USA, approximately 50-70% of bus trips have been substituted by shared bikes. Furthermore, the denser the bike-sharing network is, more passengers will choose cycling over taking other public transits. The afore-reviewed literature studied the impact of the emerging bike-sharing on the conventional public transportation modes. In this paper, we focus on the usage of the shared bikes and their relationship with other public transportation modes in Chicago.

3 EMPLOYED CLUSTERING METHOD

In terms of the travel pattern analysis model, we select the Interest-Focused Clustering based on Adaptive Resonance Theory (IFC-ART) [24] clustering technique. IFC-ART extends the original Fusion-ART model [18] by allowing users to identify the interesting features based on the preliminary understanding of the underlying dataset.

As depicted in Figure 1, IFC-ART consists of a high-level category field and K number of low-level feature channels (or input fields), where K varies according to different clustering tasks and different datasets. Each committed code in the category field represents certain learned association among the input features in its weight vectors. The operations of IFC-ART follow the dynamics of Fusion-ART [18], which are introduced in the following subsection.

3.1 Fusion-Adaptive Resonance Theory

Before we present the dynamics of Fusion-ART [18], we introduce all terms involved in its operations as follows.

Input vectors: Let $\mathbf{I}^k = (I_1^k, I_2^k, \ldots, I_L^k)$ denote the input vector, where I_l^k denotes input l to channel k, for $l = 1, 2, \ldots, L$ and $k = 1, 2, \ldots, K$, where L denotes the length of $\mathbf{I}^k$ and K denotes the total number of input fields.

Input fields: Let F_1^k denote an input field that receives $\mathbf{I}^k$ and let $\mathbf{x}^k = (x_1^k, x_2^k, \ldots, x_L^k)$ denote the activation vector of F_1^k receiving $\mathbf{I}^k$. Please note that normalization is performed on $\mathbf{I}^k$ to obtain $\mathbf{x}^k$, such that $\mathbf{x}_l^k \in [0, 1]$. If fuzzy ART operations (see (1) and (3)) [4] are used, $\mathbf{x}^k$ is further augmented with a complement vector $\overline{\mathbf{x}}^k$, where $\overline{x}_l^k = 1 - x_l^k$. This augmentation is named complement coding, which is applied to prevent the code proliferation problem [23].

Category field: Let F_2 denote the category field and let $\mathbf{y} = (y_1, y_2, \ldots, y_J)$ denote the activation vector of F_2, where J denotes the number of codes in F_2.

Weight vectors: Let $\mathbf{w}_j^k$ denote the weight vector of the jth code C_j in F_2 for learning the input patterns in F_1^k, where $j = 1, 2, \ldots, J$.

Parameters: The dynamics of Fusion-ART are regulated by the parameters associated with each input field, namely choice parameters $\alpha^k > 0$, learning rate parameters $\beta^k \in [0, 1]$, contribution parameters $\gamma^k \in [0, 1]$, where $\sum \gamma^k = 1$, and vigilance parameters $\rho_j^k \in [0, 1]$.

ART involves a bottom-up processing of the external information and a top-down modulation of the internal knowledge. Specifically, the bottom-up processing consists of the *code activation* and *code competition* processes and the top-down modulation consists of the *template matching*, *template learning*, and *knowledge readout* processes. All these five processes are introduced as follows.

Code activation: Given $\{\mathbf{x}^k|_{k=1}^K\}$, for each F_2 code j, the corresponding activation T_j is computed as follows:

$$T_j = \sum_k \gamma^k \frac{|\mathbf{x}^k \wedge \mathbf{w}_j^k|}{\alpha^k + |\mathbf{w}_j^k|}, \tag{1}$$

where the fuzzy AND operation $\wedge$ is defined by $p_i \wedge q_i \equiv min(p_i, q_i)$ and the norm $|.|$ is defined by $|\mathbf{p}| \equiv \sum_i p_i$.

Code competition: Given $\{T_j|_{j=1}^J\}$, the F_2 code with the highest activation value is named the winner, which is indexed at j^*, where $j^* = \arg\max_j T_j$.

Template matching: Given the winner code C_{j^*}, the match between input pattern and weight vector of C_{j^*} is computed as follows:

$$M_{j^*}^k = \frac{|\mathbf{x}^k \wedge \mathbf{w}_{j^*}^k|}{|\mathbf{x}^k|}. \tag{2}$$

If C_{j^*} satisfies the vigilance criteria such that $\forall M_{j^*}^k \geq \rho_{j^*}^k$, a resonance occurs in which leads to the subsequent learning or readout process. Otherwise, a mismatch reset occurs in which $T_{j^*} \leftarrow 0$ until a resonance occurs at another F_2 code.

Template learning: If learning is required, once found C_{j^*} that satisfies the vigilance criteria, its corresponding weight vectors are updated by the following learning rule:

$$\mathbf{w}_{j^*}^{k(\text{new})} = (1 - \beta^k)\mathbf{w}_{j^*}^{k(\text{old})} + \beta^k(\mathbf{x}^k \wedge \mathbf{w}_{j^*}^{k(\text{old})}). \tag{3}$$

Knowledge readout: If readout is required, C_{j^*} presents its weight vectors to the input fields, such that $\mathbf{x}^{k(\text{new})} = \mathbf{w}_{j^*}^k$.

In terms of clustering, the dynamics of Fusion-ART can be summarized as follows. Based on the similarity measures (see (1)), a winner cluster can be identified. If the input pattern satisfies the vigilance criteria of the winner cluster (see (2)), it will be added into the identified cluster (see (3)). Otherwise, Fusion-ART will select another winner until the vigilance criteria are satisfied and learn accordingly. At the end of the autonomus clustering process, each committed code in the category field represents one formed cluster.

3.2 ART Clustering with Focused Preferences

To adaptively regularize the vigilance parameters while maintaining the focus on the interesting or important features, IFC-ART self-regulates its vigilance parameters associated with each feature and

each cluster under a generic framework. In addition to the standard set of parameters used in Fusion-ART, IFC-ART employs two additional parameters, namely the magnifying parameter ϕ, which is used to define the difference between the normal and interesting features, and the regularization parameter δ, which is used to adaptively tune the vigilance parameters during the clustering process.

The algorithm of IFC-ART is summarized as follows. Based on the basic understanding of the underlying dataset and the clustering task, first of all, among the K number of input fields of IFC-ART, we can identify the set of interesting features IF. Then the remaining ones form the set of normal features NF, such that $|IF| + |NF| = K$. Given the initial vigilance parameter value $\rho_0 \in [0, 1]$, the vigilance of an uncommitted code is computed as follows:

$$\rho_J^k = \begin{cases} \min\{(1+\phi)\rho_0, 1\}, & \text{if } k \in IF, \\ \rho_0, & \text{otherwise,} \end{cases} \tag{4}$$

where ϕ refers to the magnifying parameter and $\phi \in (0, 1)$. As such, the difference between the interesting or preferred features and the normal ones is preserved during the initial formation of all clusters.

Furthermore, during the clustering process, the vigilance parameters are also self-regulated according to the learned weight vectors and the input patterns. Specifically, when a mismatch reset occurs during template matching, other than setting the activation value T_{j^*} to 0 during the presence of the current input pattern, for every input field that violates the vigilance criterion, the corresponding vigilance parameter is adjusted as follows:

$$\rho_{j^*}^{k(\text{new})} = M_{j^*}^k + \delta, \tag{5}$$

where δ is a significantly small number and $\delta > 0$. This particular regularization of the vigilance parameters is named match tracking. Its rational is to minimize the conflict between the clusters to a minimum degree defined by δ.

The overall algorithm of IFC-ART is summarized in Algorithm 1. For more details, please refer to [24].

Algorithm 1 Interest-Focused Clustering based on ART

Require: $\rho_0 \in [0, 1]$, $\phi \in [0, 1]$, $\delta > 0$, $\alpha^k > 0$, $\beta^k \in [0, 1]$, $\gamma^k \in [0, 1]$, and $\sum \gamma^k = 1$, where $k = 1, 2, \ldots, K$ and K denotes the number of input fields

Compute ρ_1^k for the uncommitted code in F_2 {see (4)}
Initialize IFC-ART with K, α^k, β^k, γ^k, and ρ_1^k
for all $\{\mathbf{I}^k|_{k=1}^K\}$ in the given dataset **do**
  obtain $\{\mathbf{x}^k, \overline{\mathbf{x}}^k|_{k=1}^K\}$, such that $\mathbf{x}_l^k \in [0, 1]$, where $l = 1, 2, \ldots, |\mathbf{I}^k|$, and present the pattern to F_1^k
  for all C_j in F_2, where $j = 1, 2, \ldots, J$ and J denotes the number of clusters in F_2 **do**
    compute T_j {see (1)}
  end for
  loop
    identify j^*, such that $j^* = \arg\max_j T_j$
    compute $M_{j^*}^k$ {see (2)}
    if $\forall M_{j^*}^k \geq \rho_{j^*}^k$ **then**
      exit loop
    else
      $T_j \leftarrow 0$
      for all $M_{j^*}^k < \rho_{j^*}^k$ **do**
        $\rho_{j^*}^{k(\text{new})} \leftarrow M_{j^*}^k + \delta$ {see (5)}
      end for
    end if
  end loop
  if $j^* = J$ {winner is an uncommitted cluster} **then**
    $\mathbf{w}_J^k = \{\mathbf{x}^k, \overline{\mathbf{x}}^k\}$
    $J \leftarrow J + 1$ {create a new uncommitted cluster}
    $\mathbf{w}_J^k \leftarrow (1, 1, \ldots, 1)$ and $\overline{\mathbf{w}}_J^k \leftarrow (1, 1, \ldots, 1)$
    assign ρ_J^k {see (4)}
  else
    update $\mathbf{w}_{j^*}^k$ {see (3)}
  end if
end for

4 ANALYSIS OF SHARED BIKE USAGE

Clustering methods have been widely applied to discover the intrinsic patterns embedded in various type of data, such as data analytic in general [14, 20], financial predictions [19, 22], lifestyle differentiation [24], disease diagnosis [21], etc. In this paper, we investigate the public transportation data collected in Chicago, USA, across the whole year of 2016. Specifically, we obtain the shared bike data from the Divvy dataset[2] and the bus, taxi and subway data from Chicago Transit Authority data portal[3]. In the following subsections, we first introduce the general setups and then present the various case studies by investigating the aforementioned dataset(s).

4.1 Case Study Setups

In all the aforementioned datasets, the location is always represented in terms of latitude and longitude. Therefore, we adopt the haversine formula to compute the distance between two locations:

$$d_h = 2r \arcsin\left(\sqrt{\sin^2\left(\frac{\tau_2 - \tau_1}{2}\right) + \cos(\tau_1)\cos(\tau_2)\sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right), \tag{6}$$

where r denotes the radius of the earth, τ_i denotes the latitude of location i, and λ_i denotes the longitude of location i. Furthermore, the bearing between the two locations is computed as follows:

$$\theta = \arctan(\sin(\Delta\lambda)\cos(\tau_1)\cos(\tau_2)\sin(\tau_2) - \sin(\tau_1)\cos(\tau_2)\cos(\Delta\lambda)), \tag{7}$$

where $\Delta\lambda$ and $\Delta\tau$ represent the difference between the two locations in terms of latitude and longitude, respectively.

Because there are no ground truth in travel patterns, the clustering results cannot be evaluated using accuracy and entropy types of measures. For the internal evaluation of clustering results, we select the Davies–Bouldin Index (DBI) [5] as the measuring metric:

[2] URL: https://www.divvybikes.com/system-data

[3] Bus: https://data.cityofchicago.org/Transportation/CTA-Ridership-Bus-Routes-Daily-Totals-by-Route/jyb9-n7fm; Taxi: https://data.cityofchicago.org/Transportation/Taxi-Trips/wrvz-psew; Subway: https://data.cityofchicago.org/Transportation/CTA-Ridership-L-Station-Entries-Daily-Totals/5neh-572f

$$DBI = \frac{1}{J}\sum_{i=1}^{J}\max_{i\neq j}\left(\frac{\sigma_i + \sigma_j}{d(m_{C_i}, m_{C_j})}\right), \quad (8)$$

where σ_i denotes the average distance of all elements in the ith cluster to its centroid m_{C_i} and $d()$ computes the Euclidean distance between two vectors. DBI combines the measure of both intra-cluster similarity (numerator of the max term in (8)) and inter-cluster similarity (denominator). A smaller DBI value suggests better performance.

In each case study, four clustering methods of the ART family are applied. With regard to Algorithm 1, the overall algorithm implements IFC-ART. If Step 1 is bypassed, the rest procedures implement Fusion-ART-WMT (with match tracking). If Step 16 is bypassed, the rest procedures implement Fusion-ART-WIF (with interesting features). If both steps are bypassed, the rest procedures implement Fusion-ART. Other than the four ART-type clustering methods, K-means [12], Spectral Clustering [7], BIRCH [29], DBSCAN [8] and Affinity Propagation [10] are also used for benchmarking purpose. However, only K-means and the ART-type clustering methods are applied to every case study. The other clustering methods are not able to handle datasets comprise more than tens of thousands samples.

For all the ART-type clustering methods, the following common parameter values are used invariantly: $\alpha^k = 0.01$, which is mainly used to avoid NaN in (1), and $\gamma^k = 1/K$, which equally assigns the contribution of each feature. For IFC-ART, we simply assign ϕ in (4) to γ^k and assign δ in (5) to 0.001. Furthermore, we apply a grid search method to obtain the best clustering results. Specifically, we vary the value of ρ_0 (see (4)) from 0.1 to 0.7 (0.8 and 0.9 normally require significantly longer run time with inferior results, see [24]) with an increment of 0.1 and $\beta^k \in \{0.001, 0.005, 0.01, 0.05, 0.1, 0.5\}$ (0.001 and 0.005 are only used for datasets with 500,000+ samples).

Commuters' travelling patterns highly depend on the day type, i.e., weekday, weekend and holiday. Therefore, for all the case studies (see the following subsections), we separate the workdays and non-workdays and perform respective investigations.

4.2 Case Study 1: Daily Shared Bike Usage

First of all, to get a feel of the usage of shared bikes on a daily basis, we extract two features, namely the total number of bike rental transactions and the averaged trip duration within the day. Moreover, the number of transactions is selected as the interesting feature because it reflects the utilization of the shared bikes. The best clustering of all benchmarking models are reported in Table 1.

Although it is clearly shown in Table 1 that DBSCAN obtains the best DBI scores, it discards a huge amount of data samples, i.e., 55% and 72% of the data samples are treated as outliers, respectively, and are removed from the final clustering results. Therefore, in this case study, Fusion-ART-WIF performs the best.

The clustering results obtained by Fusion-ART-WIF are reported in Table 2. Although we only use the Divvy dataset in this case study for pattern discoveries, we link the obtained clustering results to the daily weather conditions[4]. First of all, we found through correlation analysis that among all the weather conditions, such as humidity, wind, visibility, rain, fog, thunderstorm, hail, etc., only temperature and snow are highly correlated to the daily shared bike usage. It is encouraging to learn that our clustering algorithm may identify partial weather conditions merely based on the daily shared bike usage that all the snowing days are grouped within one cluster only.

Furthermore, the ability of our proposed clustering method to autonomously identify the outlier(s) is well illustrated in the last row of Table 2. It is shown that one particular day, 16th July 2016, which has similar trip duration (the weather conditions as well) with Cluster #2 but much higher number of shared bike rental transactions, is considered as a cluster itself. Upon closer inspection, quite a number of events[5] were organized on that Saturday, such as "2016 Blackhawks Convention", "Summer on Southport Art Festival", "Celebrate Clark Street Festival", and many more. These popular summer events contributed significantly more number of shared bike rental transactions.

Table 1: Performance Comparison on Daily Shared Bike Usage

Model	#cluster	DBI	#cluster	DBI
	Workdays		**Non-Workdays**	
K-means	2	0.45	2	0.40
Spectral Clustering	5	0.76	5	0.95
BIRCH	3	0.52	3	0.51
DBSCAN	3	**0.24**	2	**0.08**
Affinity Propagation	10	4.89	6	0.88
Fusion-ART	2	0.45	2	0.38
Fusion-ART-WMT	2	0.45	2	0.38
Fusion-ART-WIF	2	**0.44**	3	**0.33**
IFC-ART	2	0.45	3	**0.33**

Table 2: Clustering Results in Case Study 1

ID	#	trip count	duration	temperature	snow%	depth
Workdays (255 samples)						
1	141	5961.29 (2625.90)	685.30 (82.74)	3.66 (8.41)	24%	1.40 (3.54)
2	114	15014.87 (2007.49)	950.36 (82.43)	21.50 (4.85)	0%	0 (0)
Non-Workdays (111 samples)						
1	57	3509.27 (2485.37)	842.60 (183.05)	2.46 (7.39)	25%	1.00 (2.93)
2	53	15531.56 (2928.56)	1343.93 (107.69)	20.21 (4.65)	0%	0 (0)
3	1	24989 (-)	1406.70 (-)	19 (-)	0%	0 (-)

# denotes the number of data samples in the cluster; the number within () denotes standard deviation; snow% denotes the percentage of the days that snowed; and depth denotes the averaged depth of the snow in cm.

4.3 Case Study 2: Rental and Return Behaviours

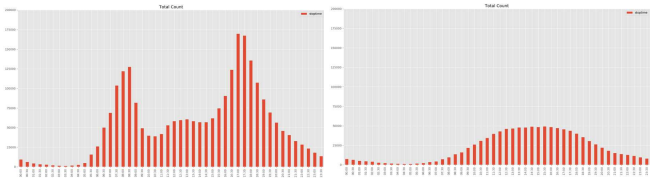
Normally, densely populated cities have morning and afternoon peak hours, which are highly correlated to the office hours during workdays. In this subsection, we investigate whether such peak hours also exist in shared bike rentals. Specifically, we extract three features from the Divvy datasets, namely the transaction time (for both rentals and returns) and the location (both latitude and longitude

[4]URL: https://www.wunderground.com/history/daily/KORD

[5]URL: http://www.chicagonow.com/show-me-chicago/2016/03/chicago-summer-festivals-2016-calendar-of-summer-festivals-and-events/

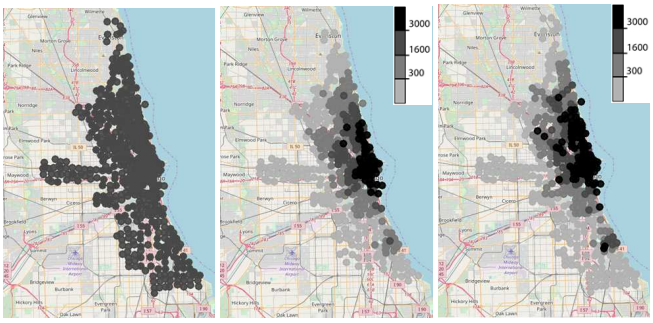
Table 3: Performance Comparison on Rentals and Returns

Model	#cluster	DBI	#cluster	DBI
	WDays–Rental		**WDays–Return**	
K-means	2	**0.43**	2	**0.43**
Fusion-ART	3	1.99	4	2.81
Fusion-ART-WMT	4	1.86	4	2.36
Fusion-ART-WIF	3	1.50	4	2.15
IFC-ART	5	1.75	5	1.94
	Non-WDays–Rental		**Non-WDays–Return**	
K-means	2	**0.62**	2	**0.62**
Fusion-ART	6	2.31	3	1.85
Fusion-ART-WMT	6	2.75	4	2.45
Fusion-ART-WIF	6	1.78	3	2.08
IFC-ART	4	2.76	4	2.48



(a) Workdays **(b) Non-Workdays**

Figure 2: Histogram for shared bike returns.



(a) Layout of shared bike stations in Chicago **(b) Heat-map in the morning (Cluster #2, Workdays–Rental in Table 4)** **(c) Heat-map in the evening (Cluster #1, Workdays–Rental in Table 4)**

Figure 3: Layout of shared bike stations and heat-maps.

for rentals and returns). Moreover, the transaction time is selected as the interesting feature. The best clustering results are reported in Table 3. Please note that except K-means and the ART-type methods, the other clustering methods are not able to handle this huge amount of data samples (3,595,383 in total).

Although K-means and Fusion-ART-WIF obtain the best overall DBI scores, they do not produce consistent number of clusters between rental and return (these behaviours should be coherent). In view of this, IFC-ART produces more meaningful clusters (see Table 4). Please note that two features, namely distance between the rental and return stations and the rental duration , which are not used in the clustering process, are also included for discussions.

Table 4: Clustering Results in Case Study 2

ID	#transactions	time (hh:mm)	distance (km)	duration (s)
	Workdays–Rental (2,546,877 samples)			
1	1,224,103	17:16 (2:51)	1.98 (1.51)	926.06 (1930.74)
2	813,347	09:19 (3:04)	1.94 (1.49)	857.65 (1513.09)
3	508,182	14:28 (4:31)	1.87 (1.43)	780.27 (1435.24)
4	1,144	07:19 (1:53)	1.27 (1.05)	655.45 (1613.40)
5	101	03:29 (2:21)	1.69 (1.20)	2598.97 (9369.72)
	Workdays–Return (2,546,877 samples)			
1	1,241,427	17:33 (2:57)	1.99 (1.52)	925.07 (1701.20)
2	1,074,467	10:32 (3:51)	1.90 (1.43)	847.31 (1758.32)
3	230,800	14:52 (4:32)	1.93 (1.51)	734.31 (1554.36)
4	93	01:13 (1:15)	2.05 (3.10)	1059.99 (1711.61)
5	90	02:51 (2:09)	2.04 (2.06)	3328.56 (10878.01)
	Non-Workdays–Rental (1,048,506 samples)			
1	596,001	16:14 (3:43)	2.06 (1.60)	1267.35 (2347.46)
2	450,416	12:15 (4:02)	2.08 (1.62)	1297.67 (2069.88)
3	2,036	14:45 (3:41)	1.18 (1.02)	1326.05 (3948.85)
4	53	02:02 (1:32)	2.35 (3.55)	4324.74 (14843.32)
	Non-Workdays–Return (1,048,506 samples)			
1	766,911	16:11 (3:27)	2.08 (1.63)	1284.20 (2140.64)
2	280,373	10:54 (4:23)	2.06 (1.56)	1270.00 (2475.90)
3	1,107	17:27 (2:44)	1.30 (1.37)	1451.10 (4077.47)
4	115	02:32 (1:56)	2.21 (2.55)	1871.45 (7564.65)

As shown in Table 4, the clustering results are quite meaningful and coherent, e.g., for the largest clusters in workdays (rental and return, respectively), they are close in size (1,224,103 vs. 1,241,427) and the time of rental plus the duration equal to the time of return (17:16+0:15=17:33). Moreover, we can observe that although the distance between rental and return stations does not significantly vary with the day type, i.e., workday or non-workday, the duration of the rental varies significantly. This finding suggest that the shared bikes are more often rented for leisure usages on non-workdays rather than commuting purposes on workdays. The histogram on shared bike returns (see Figure 2) also supports the same finding and shows the peak returns in workdays are observed around 8:00–8:30 in the morning and 17:00–17:30 in the afternoon, as observed in other public transportation modes as well. Furthermore, the shared bike station deployments in Chicago, the major morning and afternoon

transaction patterns in workdays are visualized in Figure 3. The center areas of Chicago are well highlighted in Figures 3(b) and 3(c) (along the coast). Moreover, the shared bike rental patterns in the morning and evening peak hours are quite coherent.

4.4 Case Study 3: Travel Pattern Analysis on All Public Transportation Modes

In this subsection, we report the travel pattern analysis on the combination of all public transportation methods[6], namely shared bike, bus, taxi and subway. To investigate the impact of shared bikes on all the other public transportation modes, we first find their common routes. Specifically, for each pair of the shared bike stations, between which a rental transaction has been recorded, we find whether the same route exists for taxi, bus and subway trips. As long as both the starting location (i.e., taxi pick-up location, boarding bus stop, and boarding subway station) and ending location are within 300m radius of the shared bike stations, respectively, we consider it as the same route. The distance of 300m is intuitively selected because it can be approximately converted to five minutes' walking, which is tolerable and comfortable for most commuters. Furthermore, we remove the shared bike trips of less than 1km. By doing this, we try to remove those rental usage for leisure purposes, e.g., in certain cases, the distance between the rental and return stations is or near to 0km. Furthermore, a relatively longer distance is required in the investigation of the complementary or substitution effects of shared bikes on the conventional public transportation methods, e.g., for a distance of less than 1km, few people may choose to take a taxi.

Because the travel behaviours are different between peak and non-peak hours, we divide the shared bike trips according to their rental hours. Specifically, we apply the threshold of 70,000 transactions to determine the peak hours. As such, the workdays-AM-peak and workdays-PM-peak trips took place during 7:00–9:00am and 15:30–19:30pm, respectively (see Figure 2(a)). The remaining ones on workdays are grouped together as the workdays-non-peak trips. Furthermore, the non-workdays trips are not divided as there is no significant surge in shared bike usage (see Figure 2(b)).

After obtaining the grouped trips, we first try to investigate the complementary impact of the shared bikes on conventional public transportation modes. The investigation results show that the shared bikes indeed function as a novel convenient mode of public transportations that provide affordable-and-easy trips that are not often found before. For example, the comparison between shared bike and taxi trips on the same routes (see Table 5) shows that for a majority number (over 87%) of trips, people may choose to rent a bicycle regardless of the time period.

To further investigate the substitution impact of shared bikes on all the conventional public transportation modes, we only keep the trips sharing the same routes among shared bikes, buses, taxis and subways. Due to the limitation of the available datasets, for bus and subway rides, we only know the daily bus ridership (without the knowledge of boarding and alighting stations) and daily subway station entry counts (without the knowledge of exit information), respectively. Therefore, in this case study, we alternatively choose

[6]Due to the page limitation, in this paper, we do not include the travel pattern analysis of shared bikes with one or more other public transportation method(s). However, interested readers may refer to Evan's Final Year Project report, URL: https://repository.ntu.edu.sg/handle/10356/74040.

Table 5: Trips on the Same Routes for both Bikes and Taxis

Time period	#Bike trips	#Taxi trips	Relative difference
Workdays-AM-Peak	278,557	28,751	89.68%
Workdays-PM-Peak	555,489	62,957	88.67%
Workdays-Non-Peak	612,634	69,166	88.71%
Non-Workdays	425,523	54,878	87.10%

Table 6: Performance Comparison on All Modes

Model	#cluster	DBI	#cluster	DBI
	Workdays-AM-Peak (3,200 samples)		**Workdays–PM-Peak** (5,971 samples)	
K-means	4	3.20	3	**1.81**
Spectral Clustering	5	13.99	5	2.62
BIRCH	3	13.83	3	35.44
DBSCAN	42	6.44	55	9.62
Fusion-ART	7	2.10	14	3.18
Fusion-ART-WMT	7	**1.71**	12	3.04
Fusion-ART-WIF	7	2.10	6	2.37
IFC-ART	7	**1.71**	10	2.66
	Workdays-Non-Peak (6,070 samples)		**Non-Workdays** (2,118 samples)	
K-means	5	2.35	3	4.43
Spectral Clustering	5	2.43	5	4.23
BIRCH	3	8.70	3	4.55
DBSCAN	59	13.03	50	11.19
Fusion-ART	7	**2.32**	7	4.14
Fusion-ART-WMT	6	2.37	7	4.14
Fusion-ART-WIF	7	2.44	7	**3.60**
IFC-ART	7	2.39	7	3.63

the most relevant features that we can extract besides the route information, namely the number of bus services sharing the same route and whether a subway transit (between different subway lines) is required for the same route. Specifically, the following four features on all the same routes for various public transportation modes are used in this case study: (i) the number of shared bike trips recorded, (ii) the number of taxi trips recorded, (iii) the number of different bus services available, and (iv) whether a commuter needs to transit a different subway line. Moreover, features (ii) and (iii) are identified as interesting features, as they produce more meaningful clusters. The best clustering results are reported in Table 6 (Affinity Propagation cannot well handle these datasets that it always generates more than 500 clusters and produces overly large DBI scores).

Due to the page limitation, we choose not to show the detailed clustering results in this paper. However, we find the statistical analysis on the transportation behavioural difference based on whether a trip requires a subway transit interesting to report. Specifically, we present the difference in averaged shared bike usage, taxi trips, and the available bus services per same route that either requires or does not require a subway transit in Table 7. Moreover, all the entries are statistically significantly different between zero subway transit and one or more transits, by applying ANOVA.

Different from Table 5, Table 7 represents those well established routes that have more available bus services, much more taxi trips, and relatively much lesser shared bike usage. It is interesting to learn

Table 7: Pattern Analysis on Different Transportation Modes

Transport	transit×	transit√	transit×	transit√
	Workdays-AM-Peak		**Workdays-PM-Peak**	
Bike trips	1.20	1.35	1.28	1.45
Taxi trips	38.04	22.47	59.19	42.95
Bus services	7.42	2.63	5.76	2.54
	Workdays-Non-Peak		**Non-Workdays**	
Bike trips	1.34	1.26	1.65	1.42
Taxi trips	106.34	74.53	45.60	52.80
Bus services	6.65	2.77	4.83	2.67

that for peak hours (upper portion of Table 7), there are always more numbers of shared bike trips if at least one subway transit is required. On the contrary, for non-peak hours (lower portion of Table 7), the corresponding numbers are lesser. This finding suggests that during peak hours, commuters may be more willingly to use shared bikes if they would spend more time on the transition of subway lines. Therefore, the regulators of the shared bikes should consider to deploy more bike stations or increase the amount of available bikes (either statically or dynamically) across those areas, in which the commuters cannot be served by one single subway line. As such, they may generate more revenues by increasing the number of rental transactions and at the same time, may better alleviate the burden of other public transportation modes.

5 CONCLUSION

In this paper, we briefly introduce the emerging mode of public transportation in densely populated cities, i.e., shared bikes. Furthermore, we investigate the public transportation patterns with the focuses on shared bike usage in Chicago, USA, using all the public transportation data collected in 2016. Specifically, we apply ART-type clustering methods to reveal both the majority transportation patterns and the irregular ones. Our clustering methods are shown to produce more meaningful clustering results. Moreover, the reported results may be especially useful to the regulators of shared bikes.

Going forward, we aim to investigate the much more dynamic behaviours of another type of shared bikes: station-less ones. Moreover, we also plan to search for comprehensive datasets that comprise the detailed bus and subway boarding and alighting information.

ACKNOWLEDGMENT

This research is supported by the National Research Foundation, Prime Minister's Office, Singapore under its IDM Futures Funding Initiative.

REFERENCES

[1] A. Biehl, A. Ermagun, and A. Stathopoulos. 2018. Community mobility MAUPing: A socio-spatial investigation of bikeshare demand in Chicago. *Journal of Transport Geography* 66 (2018), 80–90.

[2] J. Brand, S. Hoogendoorn, N. van Oort, and B. Schalkwijk. 2017. Modelling multimodal transit networks integration of bus networks with walking and cycling. In *Proceedings of International Conference on Models and Technologies for Intelligent Transportation Systems*. IEEE, 750–755.

[3] K. B. Campbell and C. Brakewood. 2017. Sharing riders: How bikesharing impacts bus ridership in New York city. *Transportation Research Part A: Policy and Practice* 100 (2017), 264–282.

[4] G. A. Carpenter, S. Grossberg, and D. B. Rosen. 1991. Fuzzy ART: Fast Stable Learning and Categorization of Analog Patterns by an Adaptive Resonance System. *Neural Networks* 4, 6 (1991), 759–771.

[5] D. L. Davies and D. W. Bouldin. 1979. A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1, 2 (1979), 224–227.

[6] J. Deng and M. Xu. 2015. Characteristics of subway station ridership with surrounding land use: A case study in Beijing. In *Proceedings of International Conference on Transportation Information and Safety*. IEEE, 330–336.

[7] W. E. Donath and A. J. Hoffman. 1973. Lower bounds for the partitioning of graphs. *IBM Journal of Research and Development* 17, 5 (1973), 420–425.

[8] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *Proceedings of International Conference on Knowledge Discovery and Data Mining*. 226–231.

[9] E. Fishman, S. Washington, N. Haworth, and A. Mazzei. 2014. Barriers to bikesharing: An analysis from Melbourne and Brisbane. *Journal of Transport Geography* 41 (2014), 325–337.

[10] B. J. Frey and D. Dueck. 2007. Clustering by passing messages between data points. *Science* 315, 5814 (2007), 972–976.

[11] I. Gokasar and Y. Cetinel. 2017. Evaluation of bus dwelling patterns using bus GPS data. In *Proceedings of International Conference on Models and Technologies for Intelligent Transportation Systems*. IEEE, 867–871.

[12] J. A. Hartigan and M. A. Wong. 1979. Algorithm AS 136: A K-means clustering algorithm. *Journal of the Royal Statistical Society, Series C* 28, 1 (1979), 100–108.

[13] M. Hyland, Z. Hong, H. K. R. de Farias Pinto, and Y. Chen. 2018. Hybrid cluster-regression approach to model bikeshare station usage. *Transportation Research Part A: Policy and Practice* (2018). In press.

[14] M. Parmar, D. Wang, X. Zhang, A.-H. Tan, C. Miao, J. Jiang, and Y. Zhou. 2018. REDPC: A residual error-based density peak clustering algorithm. *Neurocomputing* (2018). To appear.

[15] F. Rao, X. Zhang, Y. Dong, C. Chen, and H. Dong. 2014. Understanding the mobility pattern of passenger-searching taxis. In *Proceedings of International Conference on Intelligent Transportation Systems*. IEEE, 290–295.

[16] S. Shaheen, A. Cohen, and E. Martin. 2014. Public Bbikesharing in North America: Early operator understanding and emerging trends. *Transportation Research Record: Journal of the Transportation Research Board* 2387 (2014), 82–92.

[17] Y. Sun, A. Mobasheri, X. Hu, and W. Wang. 2017. Investigating impacts of environmental factors on the cycling behavior of bicycle-sharing users. *Sustainability* 9, 6 (2017). Article No. 1060.

[18] A.-H. Tan, G. A. Carpenter, and S. Grossberg. 2007. Intelligence through interaction: Towards a unified theory for learning. In *Proceedings of International Symposium on Neural Networks*. 1094–1103.

[19] D. Wang, X. Qian, C. Quek, A.-H. Tan, C. Miao, X. Zhang, G. S. Ng, and Y. Zhou. 2018. An interpretable neural fuzzy inference system for predictions of underpricing in initial public offerings. *Neurocomputing* (2018). To appear.

[20] D Wang, C. Quek, and G. S. Ng. 2004. Novel self-organizing Takagi Sugeno Kang fuzzy neural networks based on ART-like clustering. *Neural Processing Letters* 20, 1 (2004), 39–51.

[21] D. Wang, C. Quek, and G. S. Ng. 2014. Ovarian cancer diagnosis using a hybrid intelligent system with simple yet convincing rules. *Applied Soft Computing* 20 (2014), 25–39.

[22] D. Wang, C. Quek, and G. S. Ng. 2016. Bank failure prediction using an accurate and interpretable neural fuzzy inference system. *AI Communications* 29, 4 (2016), 477–495.

[23] D. Wang and A.-H. Tan. 2015. Creating autonomous adaptive agent in a real-time first-person shooter computer game. *IEEE Transactions on Computational Intelligence and AI in Games* 7, 2 (2015), 123–138.

[24] D. Wang and A.-H. Tan. 2016. Self-regulated incremental clustering with focused preferences. In *Proceedings of International Joint Conference on Neural Networks*. IEEE, 1297–1304.

[25] W. Wang. 2016. *Forecasting Bike Rental Demand Using New York Citi Bike Data*. Ph.D. Dissertation. School of Computing, Dublin Institute of Technology.

[26] Y. Wang, D. Zhang, L. Hu, Y. Yang, and L. H. Lee. 2017. A data-driven and optimal bus scheduling model with time-dependent traffic and demand. *IEEE Transactions on Intelligent Transportation Systems* 18, 9 (2017), 2443–2452.

[27] G. Xiong, D. Shen, X. Dong, B. Hu, D. Fan, and F. Zhu. 2017. Parallel transportation management and control system for subways. *IEEE Transactions on Intelligent Transportation Systems* 18, 7 (2017), 1974–1979.

[28] M. A. Yazici, C. Kamga, and A. Singhal. 2013. Weather's impact on travel time and travel time variability in New York city. In *Proceedings of Annual Meeting on Transportation Research Board*. 1–18.

[29] T. Zhang, R. Ramakrishnan, and M. Livny. 1996. BIRCH: An efficient data clustering method for very large databases. In *Proceedings of ACM SIGMOD International Conference on Management of Data*. 103–114.