International Journal of
Geo-Information

MDPI

Article

A Novel Dynamic Dispatching Method for Bicycle-Sharing System

Dianhui Mao, Zhihao Hao *, Yalei Wang and Shuting Fu

Beijing Key Laboratory of Big Data Technology for Food Safety, School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, China; maodh@th.btbu.edu.cn (D.M); yaleiwang15159@gmail.com (Y.W); wantingshu188@163.com (S.F.)
* Correspondence: hao_zhihao@126.com

Received: 6 January 2019; Accepted: 24 February 2019; Published: 28 February 2019

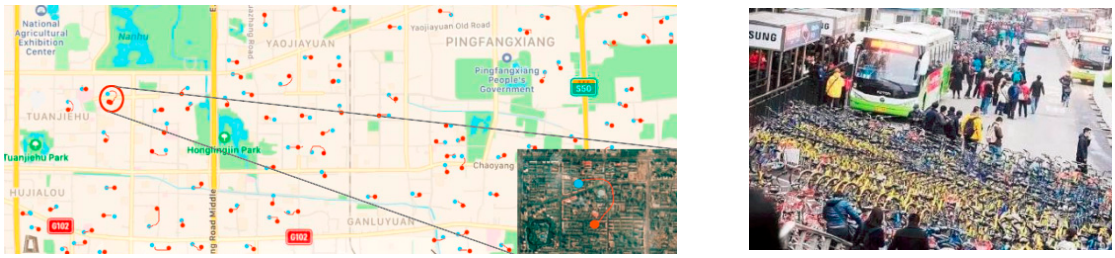
Abstract: With the rapid development of sharing bicycles, unreasonable dispatching methods are likely to cause a series of issues, such as resource waste and traffic congestion in the city. In this paper, a new dynamic scheduling method is proposed, named Tri-G, so as to solve the above problems. First of all, the whole visualization information of bike stations was built based on a Spatio-Temporal Graph (ST**G**), then Gaussian Mixture Mode (**G**MM) was used to group individual stations into clusters according to their geographical locations and transition patterns, and the Gradient Boosting Regression Tree (**G**BRT) algorithm was adopted to predict the number of bikes inflow/outflow at each station in real time. This paper used New York's bicycle commute data to build global STG visualization information to evaluate Tri-G. Finally, it is concluded that Tri-G is superior to the methods in control groups, which can be applied to various geographical scenarios. In addition, this paper also discovered some human mobility patterns as well as some rules, which are helpful for governments to improve urban planning.

Keywords: bike-sharing; Spatio-Temporal Graph; gaussian mixture mode; dispatching method

1. Introduction

Bike-sharing refers to the provision of short-term bicycle rental services in unattended urban locations. It is of flexible mobility to help reduce congestion and fuel use, which has made tremendous progress in China. To rent a bicycle in China, one simply uses a mobile app to scan the QR code on the bike, and any of the millions of bikes scattered on the sidewalks can be used by the users. China's billion-dollar bike-sharing revolution has already transformed the look and feel of cities around the country, with more than 100 million apps downloaded and billions of rides taken on many millions of bikes. Now it is going global. Some companies, such as Mobike and OFO, have expanded their business from China to the United States, from Britain to Japan, from Singapore to Bangkok. Bike-sharing has also been greatly welcomed by many metropolises, such as Beijing and New York. In Beijing, there are 700,000 shared bikes and 11 million registered users, which is nearly half the capital's population. Compared with New York, New York's Citi Bike has 10,000 bikes and 236,000 subscribers, which is the largest operation in the United States [1].

The explosive growth of users further illustrates the popularity of shared bicycles. On the other hand, it has also caused many problems. (as shown in Figure 1) For example, in China, unreasonable dispatching in the city may result in users being unable to find bicycles when they have to use them. In addition, unreasonable delivery can also lead to bicycles blocking sidewalks [2], accumulating in shopping malls, subway stations, road intersections, and even office buildings [3]. Unwanted or broken bicycles are piled up on highways, under construction, or under bridges [4].



(a) (b)

Figure 1. Delivery example: The distribution of shared bicycles in a certain area of China. (**a**) Examples of inflow/outflow used in a bike-sharing system. The red dots indicate the starting stations and the blue dots indicate the terminal stations. We have enlarged a comprehensive chart in lower right corners respectively. (**b**) The distribution of bike-sharing at a certain time in a certain place.

These problems tend to bring great difficulties to companies which provide bike-sharing services and the management of the city. Many bike-sharing companies made use of the dispatching methods to improve their operational ability and service. Meanwhile, the central government welcomed the bike-sharing as part of the "Green Urban Transport System" and urged local governments to "ensure the rational allocation of bicycles and avoid oversupply in certain areas". However, there are still some challenges:

(1) The spatial distribution of bike-sharing changes with time. In order to understand the situation of bike-sharing, Zhou et al. [5] analyzed the correlation of origin-destination (OD) flows by visualizing the large-scale movement data of the bicycle's origin-destination. Wang et al. [6] applied time series analysis to activity patterns, a hierarchical clustering algorithm using Dynamic Time Warping distances as features, and visualization on station-based data, and then employed a random forest algorithm to analyze the factors affecting bike-sharing. They use two-dimensional visualization to show the status of bike-sharing at a certain moment. However, a disadvantage of these methods is the lack of a 3D model to analyze the trend of bike-sharing with time.

(2) Currently, the dynamic scheduling method of bike-sharing is affected by many factors, especially depending on the relationship between the stations. Therefore, the division of the inflow/outflow stations is very important. For instance, Feng et al. [7] proposed a hierarchical traffic prediction model for predicted bike check-out/in number of each station cluster. Ouyang et al. [8] develop CompetitiveBike, a system to predict the particular contest among bike-sharing apps leveraging multi-source data. It utilizes Random Forest model to forecast the future competitiveness. However, the clustering methods used in these prediction models are hard clustering; that is, directly classifying a station into a cluster and ignoring the possibility that it may belong to other clusters.

In order to solve the problems mentioned above, this paper proposes a novel bike-sharing dynamic dispatching method from the perspective of strong spatial-temporal correlation of the data between the bicycle stations and the urban population movement. This method is named Tri-G. We first divided a city into many regions and clustered the regions based on the idea of dynamic programming, which used a Spatio-Temporal Graph (ST**G**) to display the whole information of the bicycle storage stations in this city. Then, the stations were clustered inside the region according to the Gaussian Mixture Model (**G**MM). Finally, the Gradient Enhanced Regression Tree (**G**BRT) algorithm was adopted to predict the real-time mobility of bicycles between stations.

Therefore, the contributions of this article are three-fold:

1. Tri-G is proposed to realize the visualization of global information to facilitate management. This method uses a three-dimensional model to show the dynamic movement of the bike-sharing, which takes the stations' transition patterns and geographical locations into consideration in an iterative approach.

2. The prediction algorithm has been improved. This article introduces a method of soft clustering (GMM), which is to analyze the possibility that a certain station belongs to any clusters. It is more regular and easier to predict than that at an individual station. This method integrates multiple similarities to be predicted based on historical periods. In addition, Tri-G has real-time performance because it runs online.
3. This method is verified through experiments with real data from New York City, so as to provide ideas for urban governance. The results indicate that Tri-G shows better performance than other methods.

2. Related Work

In recent decades, the development of information technology has made urban management more efficient. The core of a smart city is to help decision makers make intelligent and effective city-related decisions with the help of a variety of smart devices that can provide them with enough information at the right time in the right place [9]. For example, Intelligent Transportation Systems (ITS) can effectively improve the mobility and safety of traffic and reduce the impact on the environment [10]. Shared bikes have become one of the most convenient and popular tools in ITS with the growth of the sharing economy. The analysis and collection of data on shared bicycles can be used for urban planning [11,12]. For instance, if there are a large number of shared bicycles between two locations, there will be a lot of traffic pressure, and the government has to build more transportation facilities to ease the burden.

Bike-sharing has undergone four generations of development [13]. The first generation was first introduced in 1965, which provided free bicycles to borrow and return from anywhere [14]. In 1995, the second generation of bike-sharing was born in Copenhagen, with many improvements compared with the previous generation. Bicycles can be borrowed and returned from any self-service bike station throughout the city, which was also accessible using coins or smart cards. It is called City Bikes or Bycyklen [15]. In the third generation, Global Positioning System (GPS) was employed to provide a reasonable location, which was helpful to better track the bike [16]. It also incorporated a Geographic Information System (GIS) [17] that was able to analyze the distribution of bicycles and makes bicycle scheduling easier. With the advent of the Narrow Band Internet of Things (NB-IoT) [18], the fourth-generation bike-sharing is expected to greatly reduce the energy consumption of the bicycle sharing system. Moreover, analyzing the big data generated by bicycles while riding is also of instructive significance for decision makers. For example, Sun et al. [19] used crowdsourced data from bicycle sharing systems to infer the relationship between different riding goals and air pollution exposure. The results show that cyclists for the purpose of recreation and fitness are more susceptible to air pollution than those who ride bikes for commuting. This is sure to help decision makers make rules to improve cycling infrastructure and secure roads.

In order to better manage bike-sharing, visualization technology was introduced. Visualization technology can help people understand data by placing bicycle data in a visual context and can intuitively show trends of bicycle movement. Yan et al. [20] constructed a tensor based on the spatial, temporal, and user information of the bike-sharing data, and employed tensor factorization to extract latent user activity patterns. Yang et al. [21] used visualization technology to model mobility and used this new mobility modeling and prediction approach to improve the bike-sharing system operation algorithm design and pave the way for rapid deployment and adoption of bike-sharing systems. Zhang et al. [22] analyzed the characteristics of each station—the stations are modeled from the perspective of individuals and clustered by means of different models. The visualization method provides an effective technical means for prediction. However, these methods are all visualizing local areas and cannot get an overall trend.

Besides, in order to maximize the use of the bike-sharing system, the operator of the system needs to dynamically schedule bicycles, so a dynamic dispatching method is required. Many researchers have proposed lots of algorithms for dynamic scheduling of bicycle sharing systems. Huang et al. [23]

proposed an algorithm called Bimodal Gaussian Inhomogeneous Poisson (BGIP) to predict the number of bikes. It can help to optimize the repository of bikes. Olfert et al. [24] analyzed a system that cited parts of an attractive city environment, including bikes and pedestrians, in which the existing buildings are used to plan the urban area. Jiang et al. [25] proposed a method to study the effectiveness of the design method as well as the workflow of the cyclic infrastructure from the perspective of architecture and design. Harder et al. [26] proposed a combination of data sources, which included questionnaires, interviews with major participants, visual analysis of websites, counting, etc. All these works have emphasized many important factors, such as the relevance of quick connections, the aesthetic value of streetscapes, and the safety of cyclists. However, both methods ignore the contributions to the transportation network because these two methods mainly focus on the bike movement on one road. Compared with the impact of the transport network, it is unilateral only to calculate the influence in one road. Zheng et al. [27] revealed a low percentage of bicycle travel, which was caused by inadequate bicycle lanes and parking facilities. Ahillen et al. [28] combined a variety of factors and public bicycle sharing programs to obtain a city plan based on mathematical analysis. He et al. [29] made use of the simple line clustering method to analyze origin-destination data and its application for bike-sharing. In addition, Shen et al. [30] proposed the Internet of Shared Bicycles (IoSB), so as to find a feasible solution to those technical problems of the shared bicycle. However, these works are all performed off-line, and they used historical data sets to predict the trajectory, so the predicted results are not in real-time. This paper has been improved on the basis of these works. We use the visualization method to realize the real-time dynamic dispatching of bicycles based on the global information.

3. Definitions and Framework

Tri-G is a data-driven method which aims to understand the behavior patterns of bike-sharing, and it can tackle urban challenges. To better illustrate Tri-G, the definitions and framework are as follows.

3.1. The Definition of Concept

The following definitions are summarized based on [31].

Definition 1. ***(STG)*** *The Spatio-Temporal Graph (STG) G= (V, E) is a directed graph, where V and E represent the set of complete vertices and edges in G ($|E| \geq 1$), respectively. Each vertex $a \in G$; V has a geospatial position and each edge $v \in G.E$ has a spatial length. Each vertex/edge is associated with attributes varying in time, e.g., the inflow and outflow.*

Definition 2. ***(Inflow/Outflow)*** *As for each edge $v \in G.E$, the inflow of v in time interval t is $p_{in}(v,t) = \sum_{v_1 \in (G.E-v)} p(v_1, v, t)$, where $p(v_1, v, t)$ is the flow from v_1 to v in t. Likewise, the outflow of v in t is $p_{out}(v,t) = \sum_{v_1 \in (G.E-v)} p(v, v_1, t)$. Suppose S is a subgraph of an STG G, and S.E. $\subseteq$ G.E is the collection of edges in S. The inflow of S is $p_{in}(S,t) = \sum_{v \in s, E \wedge v_1 \in (G.E-S.E.)} p(v_1, v, t)$, and the outflow is $p_{out}(S,t) = \sum_{v \in s, E \wedge v_1 \in (G.E-S.E.)} p(v, v_1, t)$.*

Definition 3. ***(Actual Flow)*** *In an STG, the actual flow f_a of an edge e is e; $f_a = p_{in}(v,t) - p_{out}(v,t)$ and that of a subgraph S is $S.f_a = p_{in}(S,t) - p_{out}(S,t)$.*

Definition 4. ***(Inflow/Outflow Prediction)*** *Given a set of historical trips $H_T = \{H_{t_1}, H_{t_2}, \ldots, H_{t_T}\}$, this thesis intends to predict the inflow/outflow of each station S_j, j = 1, 2, . . . , n (cluster C_j, j = 1, 2, . . . , m) during a future period, which is set as 1 h in this work.*

3.2. Framework

Figure 2 is the framework of our model, consisting of two parts: Part A is the visualization of STG, and Part B is the prediction of the inflow/outflow of stations.

Part A expresses the spatio-temporal relationship of users' historical data and forms the STG, and then the STG index is constructed to manage data quickly and efficiently. First, the bicycle trajectory received in the most recent time interval is mapped to the road network according to the data set, and then the inflow/outflow of each road segment in the interval is calculated. Then, a city is divided into different grid cells, each of which covers several road segments, and STG indexes are built to facilitate data management. Based on the STG index, the threshold is selected as the criterion to measure the actual flow state of the bicycle in the grid, and then color the grid by analyzing the actual flow directions. Following this step, a hierarchical map structure is used to visualize the STG information.

Part B uses GMM for station clustering and the station clusters according to the geographic locations and transformation modes. Clustering stations in groups can deal with the irregular fluctuations in each station. Then, we use GBRT to predict traffic flow between stations in real time.

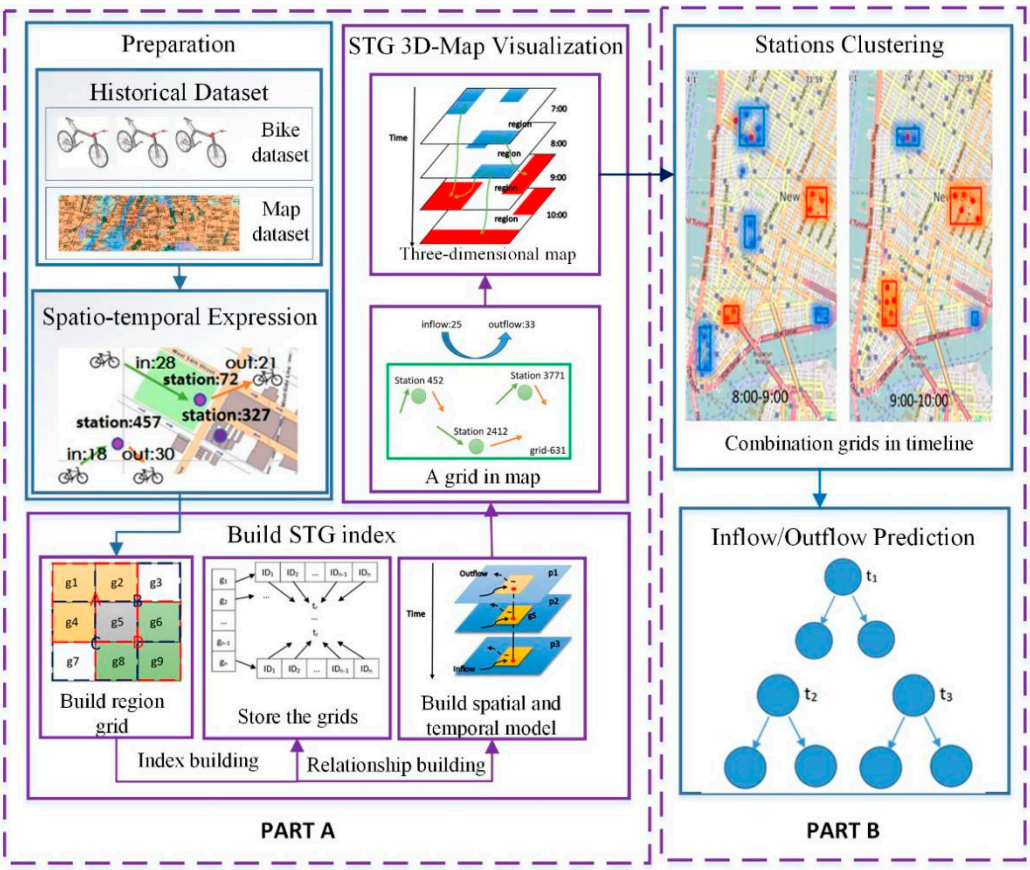


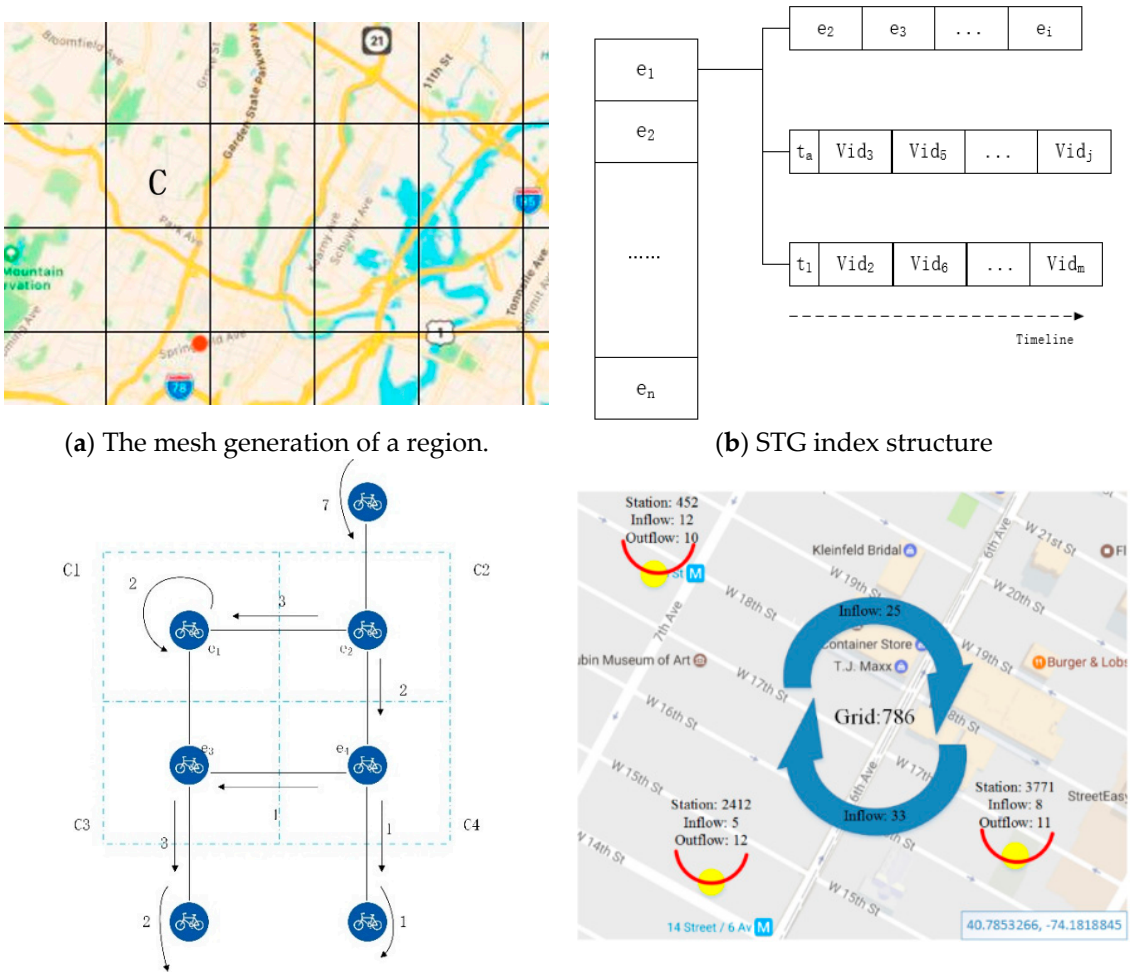
Figure 2. Framework of the paper.

4. Details Implementation

4.1. The Implementation of STG Visualization

We need to get the bike number and store the road segment information in the most recent time interval. We first build a global STG index. The index is used to quickly search information about the STG, including the dynamic flows on each edge, as well as the spatial relationships between different edges [32]. The region is divided into grids, shown in Figure 3a, and a STG index maintains the road segments, arrival time, and departure time in each grid (as shown in Figure 3b); t_l refers to the time when one bike departs from the station grid, while t_a refers to the arrival time. In the meantime,

we build an adjacency list to manage the structure and dynamic flow of an STG. For each road segment e in the adjacency list, three lists are maintained. The first is a list of road segments that are directly connected to e in the road network. The second is a list of vehicle IDs sorted by their arrival time t_a at e, and the third is a list of vehicle IDs sorted by their departure time t_l from e. As shown in Figure 3b, the list of road segments that are directly connected to e_1 is $<e_2, e_3, \ldots, e_i>$; $<Vid_3, Vid_5, \ldots, Vid_j>$ is the list of vehicle IDs at e_1 when time is t_a, and the list is $<Vid_2, Vid_6, \ldots, Vid_m>$ when time is t_l. The first list is static, while the latter two sorted lists will be updated after each round of map-matching. In the real application, we only need to store the vehicle IDs of the most recent time intervals, e.g., 1–2 h.



(**a**) The mesh generation of a region. (**b**) STG index structure

(**c**) Example of region C (**d**) Grid 786 in this region (red dot in (**a**))

Figure 3. Establishment of STG index.

To calculate the actual flow of each grid cell, the candidate cell selection algorithm is useful to check the positive actual flow of a region in a time interval, and all we need to do is to count the number of vehicles whose arrival time is within the time interval while the departure time is beyond the interval. The road segment ID is connected to its corresponding records in the adjacency list via a hash function. Since a road segment may cross two or more grid cells, the STG index can help to avoid redundant storage and index updates compared with directly storing everything in one grid cell.

The candidate cell selection algorithm to find the maximum flow are in the four adjacent grids. We define a flow upper bound $UB(g)$ of a grid cell in Equation (1).

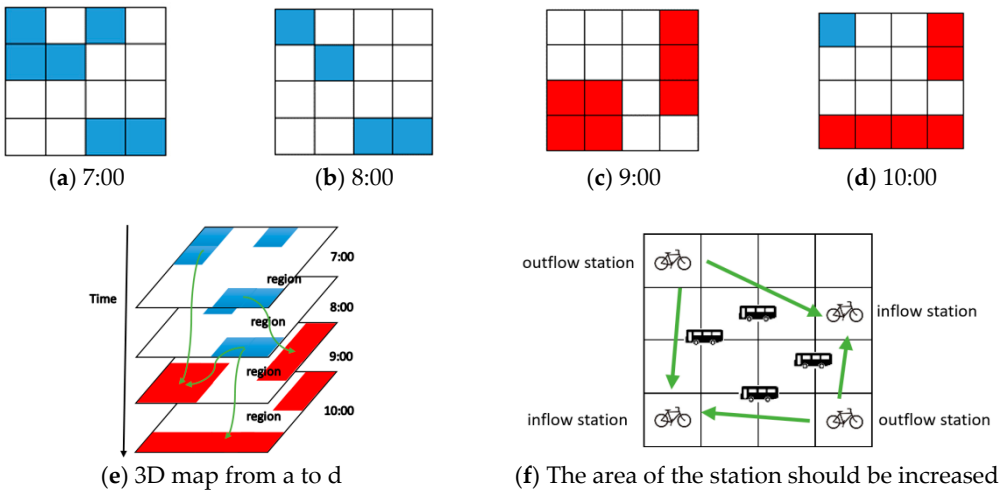
$$UB(g) = Max(A(g), B(g), C(g), D(g)), \tag{1}$$

Here A, B, C, and D, respectively, refer to the positive actual flow in four directions. If positive actual flows of the four grid cells are still less than the given flow threshold, and if UB is bigger than the threshold, g will be marked as a candidate cell. All the grids flow can be calculated in Equation (2).

$$f_a = f_{in} - f_{out}, \tag{2}$$

f_a refers to the difference in inflow and outflow. As shown in Figure 3c, the region C in Figure 3a can be divided into four sub-areas, C_1, C_2, C_3, and C_4. The inflow of C is 7 and the outflow is 3, f_a is 4. The positive result means the area has more inflow bikes, while a negative result means it has more outflow bikes. Figure 3d shows a grid flow, which is made up of three stations. Overall, there are 25 inflow-bikes and 33 outflow-bikes in this period. Thus, f_a has negative eight at that time. Meanwhile, we formulate that the threshold is ± 5 based on the initial experience, which is an outflow grid shown in Figure 3d.

A 3D map is used to display the process, as shown in Figure 4. In this paper, we calculate the flow situation according to the time line from 7:00 to 10:00. A region is divided into 4×4 grids, each of which represents a specific area. If the grid is blue, it indicates the outflow is relatively large (the net outflow > 0). If the grid is red, it indicates more inflow (the net inflow > 0). If the grid is white, it means the net inflow is 0 and the area is in equilibrium. Figure 4a–d is a comparison of all periods. Based on all this information, a three-dimensional STG is formed, as shown in Figure 4e.



(**a**) 7:00 (**b**) 8:00 (**c**) 9:00 (**d**) 10:00

(**e**) 3D map from a to d (**f**) The area of the station should be increased

Figure 4. Contrary prediction model from 7:00 to 10:00.

Furthermore, a threshold is set to estimate the movement pattern of the bikes in the grid. The grid color is marked based on real-time actual flows. For each station, the actual flow is calculated compared with the threshold if grids with the same pattern need to be merged together. Algorithm 1 shows the process of the above [33]. We name Algorithm 1 as Dpmergegrids(G).

In Figure 4e, the green arrow indicates the overall movement of the bicycles in this area at that moment. According to Figure 4a–e, the grid of the first row intersecting the first column and the grid of the fourth row intersecting the fourth column are regions with the largest outflow. The areas where the fourth row intersects the first column and the second row intersects the fourth column are regions with a large inflow. It is possible to increase the number of public transport tools (such as buses) from the area with a large outflow to the area with a large inflow, so as to improve the traffic capacity of the area (as shown in Figure 4f)

Algorithm 1 Dpmergegrids(G)

Input: The weight of edges and all nodes contained in the directed graph.
Output: dad()/*The path can be obtained according to the *dad* array. */

```
initialize: all dilg() values to ∞;/*dilg() is the superposition of grid input and output traffic with the same properties*/
Let V' be the set of vertices with indegree=0; /*Set the set V', the in-degree of points belong to V' are 0. */
for each vertex v in V' do
    dilg(v)=0;
end for
for each v ∈ (V − V') in Topological Sorting order do /* For the points other than V' in V, the longest path is sequentially obtained in the order of topological ordering and saved. */
    dilg(v) = max(u, v) ∈ E {dilg(u) + w(u, v)} /* w(u, v) represents the weight on the edge. The weight here is the amount of bicycle inflow and outflow. */
    Let (u, v) be the edge to get the maximum value;
    dad(v) = u; /* Each step records the parent node of V in the dad array. */
end for
Return the dilg(.) with maximum value.
```

4.2. Prediction Methods

In fact, it is not necessary to predict the inflow/outflow of each individual station. Knowing about the inflow/outflow of each cluster is enough for a bike's real location, because users usually use bikes at a random station closest to their origins or destinations. Besides, if there are many events that may affect the use of the bicycles, they usually tend to affect a whole area, not just a single station. We consider both the geographic location and actual flow of the station in terms of the forecasting method for two reasons:

(1) For the convenience of users, stations in a cluster should be geographically close to each other. Therefore, for users near their starting point or destination but with no bicycle available, it is acceptable to walk to another station in the same cluster to use the bicycle.
(2) Since it is necessary to predict the net flow between clusters, in order to improve its accuracy, it is hoped that stations in a cluster should have an actual flow similar to all clusters.

4.2.1. Cluster Method

There are many stations in the grid, each of which has a different real-time inflow and outflow status, because a lot of bikes reach or depart from a station and there is no activity for some bikes at the station. However, the descriptions obtained from station occupancy data cannot account for such differences. In addition, the density of the station is also an important indicator for evaluating the bicycle sharing system [34]. In order to handle the irregular fluctuation of each station, a Gaussian mixture model is introduced in this thesis to cluster the stations into groups. A Gaussian mixture model is a probabilistic model that assumes all the data points are generated from a mixture of a finite number of Gaussian distributions with unknown parameters. The model can also deal with the differences in movement patterns. Therefore, stations in a cluster should not only be geographically close to each other, but also have a transition pattern similar to all clusters.

In this method, the number of grids is provided as K Gaussian distribution. Every Gaussian distribution can influence the bike station, and the predominant factor is π_k; x refers to the number of the bike station, and θ refers to parameters of the Gauss model [35,36]

$$\theta = \{\pi_1, \ldots, \pi_k, \theta_1, \ldots, \theta_k\}, \sum_{k=1}^{K} \pi_k = 1, \pi_k \in [0, 1] \tag{3}$$

$$p(x|\theta) = \sum_{k=1}^{K} \pi_k p_k(x|\theta_k) \tag{4}$$

$$p_k(x|\theta_k) = N(x|\mu_k, \sum k) \tag{5}$$

Here, we estimate the model parameters π_k for each class of impact factors, which can be calculated in Equation (6); μ_k is calculated in Equation (7); $\sum$k refers to the covariance matrix shown in Equation (8).

$$\pi_k = \frac{\sum_{i=1}^{M} Q^i\left(z_k^{(i)}\right)}{M}. \tag{6}$$

$$\mu_k = \frac{\sum_{i=1}^{M} x^{(i)} Q^i\left(z_k^{(i)}\right)}{\sum_{i=1}^{M} Q^i\left(z_k^{(i)}\right)}. \tag{7}$$

$$\sum k = \frac{\sum_{i=1}^{M} (x^{(i)} - \mu_k)(x^{(i)} - \mu_k)^T Q^i\left(z_k^{(i)}\right)}{\sum_{i=1}^{M} Q^i\left(z_k^{(i)}\right)}. \tag{8}$$

After that, the expectation maximization [37] (EM) algorithm is used to calculate the cluster's results, which is shown in Equation (9).

$$l(\theta) = \sum_{i=1}^{M} \sum_{Z^{(i)}} Q^i\left(z^{(i)}\right) \log \frac{p\left(x^{(i)}, z^{(i)}; \theta\right)}{Q^{(i)}\left(z^{(i)}\right)} \equiv \sum_{i=1}^{M} \sum_{k=1}^{K} Q^i\left(z_k{}^{(i)}\right) \log \pi_k N(x^{(i)}|\mu_k, \sum k). \tag{9}$$

There are two steps involved when using EM algorithm in GMM, the first of which is E-step, and the second is M-step. E-step uses observed data and the existing model to predict the missing stations and prepare for the M-step [38].

$$Q^i\left(z_k^{(i)}\right) = p\left(z_k^{(i)}|x^{(i)}; \theta\right) = \frac{\pi_k N\left(x^{(i)}; \mu_k, \sum k\right)}{\sum_{k=1}^{K} \pi_k N\left(x^{(i)}; \mu_k, \sum k\right)} \tag{10}$$

Here, M-step will take the derivative of the log-likelihood to obtain estimates directly.

By means of this method, the iterations are obtained until convergence, and then the $Q^i\left(z_k^{(i)}\right)$ can be used for clustering. Later, individual stations are grouped into clusters according to their geographical locations and transition patterns.

4.2.2. GBRT

Gradient Boosting Regression Tree (GBRT) [39] is one of the most effective machine learning models for prediction, which is a non-parametric statistical learning technique for regression. It is flexible enough to fit complex nonlinear relationships. First, a sequence of simple regression trees is computed:$\{t_1(x), t_2(x), \ldots, t_r(x)\}$. The process is shown in Figure 5.

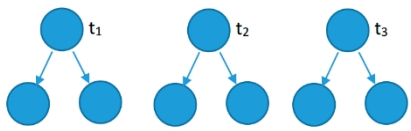


Figure 5. GBRT model.

Here, each tree is built to predict the residual of the preceding trees and is calculated in Equations (11) and (12).

$$t_i = argmin_g \sum_{t=1}^{N} L(y_t - T_{i-1}, t(x_t)). \tag{11}$$

$$T_{i-1} = \sum_{l=1}^{i-1} t_l(x). \tag{12}$$

Here, L refers to a loss function in Equation (11), and $\{x_t, y_t\}_{t=1}^{N}$ refers the training dataset.

GBRT model should accumulate all the results of the regression tree as the final one. Each tree learns the residual error from the last tree.

Predictions are made of the combining decisions of $\{t_1(x), t_2(x), \ldots, t_r(x)\}$ shown in Equation (13).

$$T(x) = t_1(x) + t_2(x) + \ldots + t_r(x). \tag{13}$$

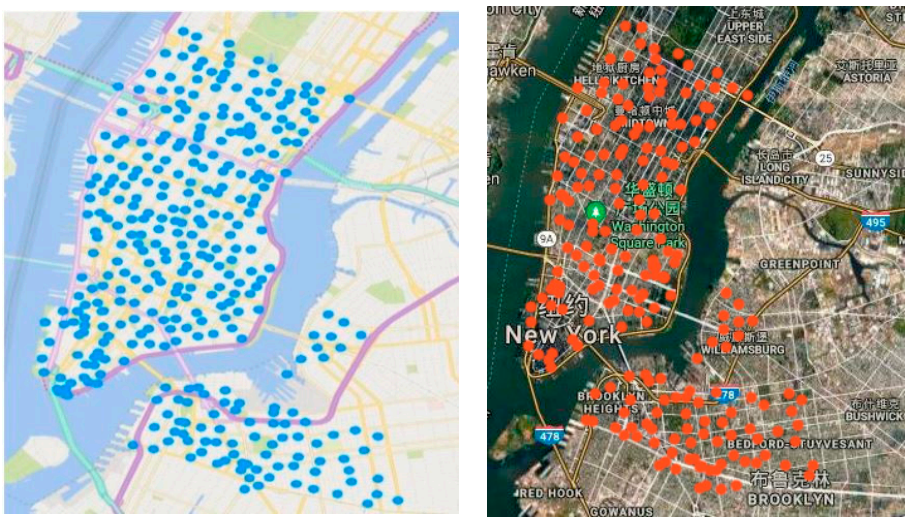
Here, x refers to the corresponding features, which has a significant influence on the entire traffic, y_t is the actual value in period t. $T(x)$ refers to the prediction results.

5. Experiments and Suggestions

5.1. Building a Spatiotemporal Map Based on the New York Dataset

Machine configuration: The operating environment for the experiment is Windows 10, equipped with dual-core CPU with 2.20 GHz dominant frequency and 8 GB memory.

Datasets: This research uses bike trip data of Citibike [40] sharing system in Manhattan, New York City. The data contain 1,037,712 trips generated by 6376 bikes and 330 stations in March 2014. Each bike trip involves start/stop time and start/end stations. Bike stations with geographical positions (i.e., vertices) and connections between them (i.e., edges) form an STG. The inflow and outflow of a bike station can be obtained by counting people returning and renting bikes at the station, respectively. Since New Yorkers make about 113,000 bike trips daily, bike trip data are reflected in part of the traffic flows in New York City. The distribution of stations is shown in Figure 6. Figure 6a shows a general distribution and Figure 6b shows the real-time distribution at a certain time.



(**a**) (**b**)

Figure 6. Station distribution: (**a**) the general distribution and (**b**) the real-time distribution.

Statistical analysis of all stations is carried out in the dataset, and some stations selected are shown in Table 1. The data was selected ranging from 8 a.m. to 9 a.m. on 12 March 2014. If the actual flow rate is positive, it means that the inflow amount exceeds the outflow amount, and if it is negative, it means that the outflow amount exceeds the inflow amount.

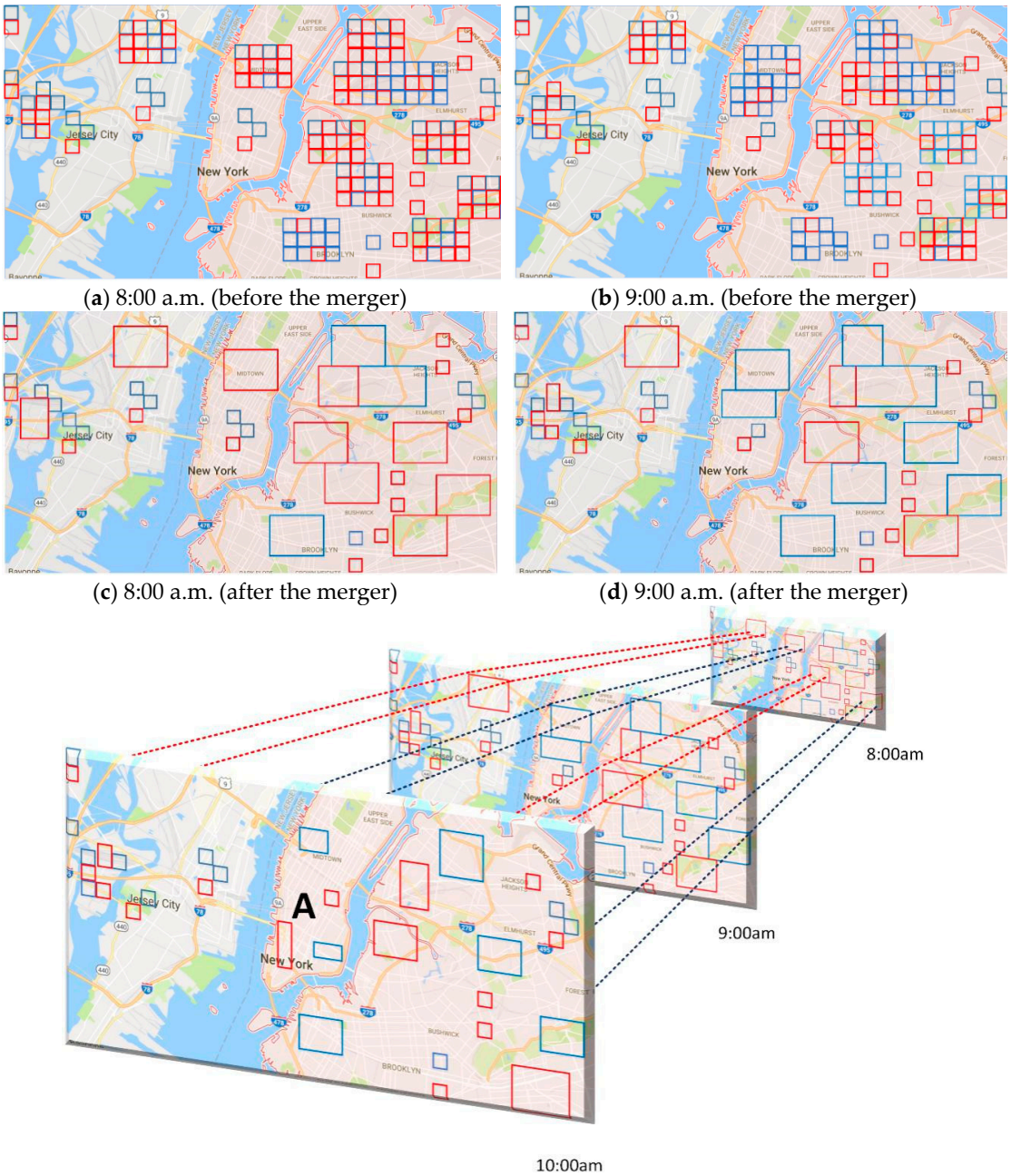
Table 2 shows some of the parameters of Tri-G. An area with a radius of $(0.5 \times \sqrt{2})$ kilometers is selected, in which Tri-G is adopted. As shown in the table, the time for building STG index is 0.27 s, the time for meshing is 4.18 s, and the time for merging grids is 0.33 s.

Table 1. Details of data.

Station	Inflow	Outflow	Actual Flow	Mark Stations
231	35	21	+14	–
343	57	73	−16	–
25	102	37	+65	blue
78	12	78	−66	red
512	33	172	−139	red

Table 2. Running Time of Tri-G.

Setting	Build STG Index	Mesh Grid	Merge Grids
$d = 0.5 \times \sqrt{2}$	0.27 s	4.18 s	0.33 s



(**a**) 8:00 a.m. (before the merger)

(**b**) 9:00 a.m. (before the merger)

(**c**) 8:00 a.m. (after the merger)

(**d**) 9:00 a.m. (after the merger)

(**e**) 3D model

Figure 7. The results of grids.

Figure 7 is obtained by analyzing real-time data about the trajectory of the bicycles. Figure 7a,b are real-time images of bicycle trajectory data at 8:00 a.m. and 9:00 a.m., respectively, in New York City. As can be seen from Figure 7a to Figure 7b, the colors of some grids vary significantly. The number of the blue grids in Figure 7b is significantly larger than that of the red grids. Then, the adjacent areas of the same color are merged. If the adjacent regions have the same color, the two grids are merged into one of the grids in Figure 7c,d. Finally, we predict the 10:00 a.m. image through Figure 7c,d, and then use a 3D map to show the trend of bicycle movement at different time intervals. The results are shown in Figure 7e.

To better analyze Figure 7e, Figure 8 shows the details of Region A in Figure 7e. As can be seen from the results, most stations are blue from 8:00 to 9:00, which means that this period is the peak of bicycle outflow. In addition, there are still two red grids during this period, and the red grid continues to expand between 9:00 and 10:00, which means more people are cycling into the region. Based on these findings, the general trend of bikes can be inferred, as shown in the third picture in Figure 8.

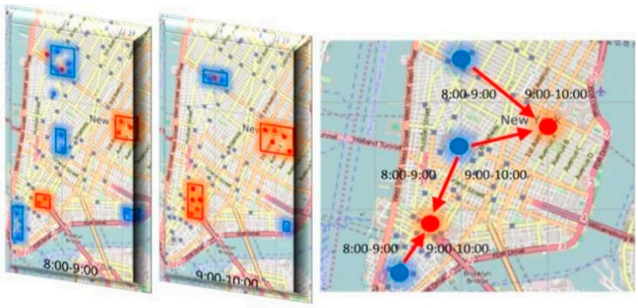


Figure 8. Details of A.

5.2. Prediction Results

5.2.1. Preparations

Tri-G is a dispatching model that combines GMM and GBRT. To measure its efficiency, the clustering method (GMM) is compared with the other two clustering methods and the prediction method (GBRT) is compared with the other two prediction methods. During the experiment, this research also cross-combined the clustering method and the prediction method.

Clustering model:

Contrast method: Grid clustering cluster algorithm (GC). It is a method which divides the city into uniform grids clustering. This research divides the city into uniform grids. The stations falling into the same grid belong to the same cluster.

K-means cluster algorithm: It is a cluster algorithm, given group classification number k ($k \leq n$). The method should cluster the data into k groups and consider the distance only. Bipartite clustering [41] (BC) method uses K-means to cluster.

Predictive model:

HA: It is a method to choose the average value of the historical results.

ARMA: It is a method to understand and predict the future value in a time series.

HP-KNN: It is a method to predict the entire traffic and allocate each cluster based on the proportion across clusters.

GBRT: It is a method to make direct predictions, which is similar to entire traffic prediction.

Metric: The values of Root Mean Logarithmic Squared Error (RMLSE) and Error Rate (ER) are adopted to measure the results, as shown in Equations (14) and (15).

$$\text{RMLSE} = \frac{1}{T}\sum_{t=1}^{T}\sqrt{\frac{1}{m}\sum_{l=1}^{m}\left(\log\left(\hat{X_{C_{l,t}}}+1\right)-\log\left(X_{C_{l,t}}+1\right)\right)^2} \tag{14}$$

$$\mathrm{ER} = \frac{1}{T}\sum_{t=1}^{T}\frac{\sum_{l=1}^{m}|\hat{X_{C_{l,t}}} - X_{C_{l,t}}|}{\sum_{l=1}^{m}X_{C_{l,t}}} \tag{15}$$

Here, $X_{C_{l,t}}$ refers to the ground truth of the inflow/outflow of cluster C_l during t, while $\hat{X_{C_{l,t}}}$ refers to the corresponding prediction value.

5.2.2. Experimental Results

It is easy to understand that the larger the cluster number is, the smaller the average scope of clusters is. Thus, for each cluster, it needs to be within a reasonable bound for practical purpose. In our research, the number of clusters is chosen by experiments, and finally we cluster all the stations into clusters. This research uses New York City data to compare the detection results of different methods. The GBRT will be compared with two common methods, HA and ARMA. The inflow prediction errors are compared by clustering, which can show the prediction accuracy of our model (as shown in Table 3).

Table 3. Inflow prediction error.

Methods \ Baseline		HA	ARMA	GBRT
RMLSE	GC	37.7%	36.3%	38.2%
	BC	36.5%	35.2%	36.5%
	GMM	36.4%	35.2%	**36.3%**
ER	GC	34.7%	34.0%	30.9%
	BC	35.2%	34.4%	30.9%
	GMM	35.1%	34.2%	**30.8%**

Based on GC and BC, with the RMLSE measurement method, it is found that the errors of prediction accuracy decreased by 1.48% on average compared with GC and decreased by 0.1% compared with BC. Using the ER method, it is found that the error in prediction accuracy reduced by an average of 0.3% compared with GC and by 0.13% compared with BC. In addition, the average error rate using GBRT reduced by 4.13% compared with HA and by 3.3% compared with ARMA. The Inflow prediction of deviation using RMLSE and ER is shown in Table 3.

Table 4 shows the deviation of the prediction accuracy of the outflow. Using the RMLSE measurement method, it is discovered that the error in prediction accuracy was on average 1.94% lower than GC and 0.26% lower than BC. Besides, the error rate using GBRT reduced by 0.23% compared with HA and by 4.13% compared with ARMA. Using the ER method, our method also reduced by about 0.14% and 0.26% compared with the GC phase, BC. Moreover, the average error rate using GBRT reduced by 4.1% compared with HA and by 3.37% compared with ARMA.

Table 4. Outflow prediction error.

Methods \ Baseline		HA	ARMA	GBRT
RMLSE	GC	38.7%	37.1%	38.6%
	BC	37.2%	35.4%	36.9%
	GMM	37.1%	35.3%	**36.8%**
ER	GC	35.3%	34.6%	31.1%
	BC	35.5%	34.6%	31.4%
	GMM	35.1%	34.4%	**31.0%**

5.3. Suggestions

According to the above experimental analysis, the map is divided into three grids, A, B, and C, which are shown in Figure 9. After analysis, it can be seen that between 8 o'clock and 9 o'clock, there are many bicycles starting from A and passing B to C. According to the survey, it is found that there are many residential areas in Grid A, several schools in Grid B, and some companies in Grid C. Therefore, it can be concluded that the flow formation may be due to the fact that many people ride bicycles to send their children to school and then go to work, consequently leading to traffic congestion around schools. Therefore, we propose increasing the number of bicycle stations in B, especially around schools in B, so as to alleviate parking pressure and solve traffic congestion.

Figure 9. Suggestions on bicycle stations. Grid A in red, grid B in blue, grid C in green.

According to Figure 8, we can figure out that more people rent a bike from the blue dots and return it at 9 o'clock in the red dots. The whole map is shown in Figure 10. During this period, most people move following the direction of the red arrow from blue dots. Based on this, some recommendations can be made for New York City Planning. It is recommended that 4 and 5 bus stops should be added around the blue point in the Manhattan and Brooklyn areas to improve transportation.

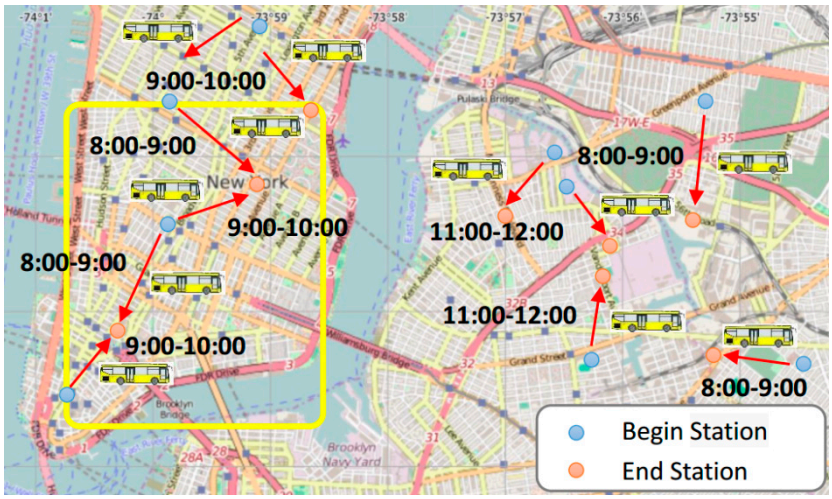


Figure 10. Suggested schematic.

6. Conclusions

This paper proposes a dynamic scheduling method of bike-sharing named Tri-G. As part of this work, New York's bike-sharing dataset was used to conduct a case study in order to test the feasibility of the system. Tri-G used the STG to construct a 3D model on bicycle flow changes, which can intuitively show the general trend. GMM and GBRT are used to make predictions and analyze these prediction results, so as to find out effective suggestions to solve problems. The experimental results provide a useful reference for New York City traffic management departments and bike-sharing operators to take appropriate measures to alleviate traffic pressure. For example, increasing the number of public transportation tools, such as buses in specific areas of Manhattan and Queens, and encouraging them to introduce more specific policies. Additionally, RMLSE and ER were used to evaluate different clustering methods and prediction methods for the purpose of analyzing the performance of Tri-G. The results show that Tri-G is of higher prediction accuracy. However, there are still some limitations. For example, the use of bike-sharing may be affected by other factors, such as the use of commute tools like private cars and weather conditions like rainy days, which have potential to cause bias in research results. In the future, we should consider more influencing factors to get more accurate experimental results.

Author Contributions: Conceptualization, D.M.; methodology, D.M.; software, Z.H.; validation, D.M., Z.H., and S.F.; formal analysis, Y.W.; investigation, Y.W.; resources, Z.H.; writing—original draft preparation, Z.H.; data curation, D.M.; visualization, Z.H.; writing—review and editing, D.M.; supervision, D.M.; project administration, D.M.; funding acquisition, D.M.

Funding: This research was funded by The National Social Science Fund (18BGL202).

Conflicts of Interest: The authors declare no conflict of interest and the founding sponsors had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. China Exports its Bike-Sharing Revolution to the US, the World. Available online: https://www.stuff.co.nz/world/asia/96478393/china-exports-its-bikesharing-revolution-to-the-us-the-world (accessed on 4 September 2017).
2. Van der Spek, S.C.; Scheltema, N. The importance of bicycle parking management. *Res. Transp. Bus. Manag.* **2015**, *15*, 39–49. [CrossRef]
3. Boldrini, C.; Incaini, R.; Bruno, R. Relocation in car sharing systems with shared stackable vehicles: Modelling challenges and outlook. In Proceedings of the 2017 IEEE 20th International Conference on Intelligent Transportation Systems (ITSC), Yokohama, Japan, 16–19 October 2017; pp. 1–8.
4. Larsen, J. Bicycle parking and locking: Ethnography of designs and practices. *Mobilities* **2017**, *12*, 53–75. [CrossRef]
5. Zhou, Z.; Meng, L.; Tang, C.; Zhao, Y.; Guo, Z.; Hu, M.; Chen, W. Visual abstraction of large scale geospatial origin-destination movement data. *IEEE Trans. Vis. Comput. Graph.* **2019**, *25*, 43–53. [CrossRef] [PubMed]
6. Wang, L.; Wu, J.; Li, W. Research on the mode and driving mechanism of public bicycles in small and medium-sized cities. *J. Earth Inf. Sci.* **2019**, *21*, 25–35.
7. Feng, S.; Chen, H.; Du, C.; Li, J.; Jing, N. A hierarchical demand prediction method with station clustering for bike sharing system. In Proceedings of the 2018 IEEE Third International Conference on Data Science in Cyberspace (DSC), Guangzhou, China, 18–21 June 2018; pp. 829–836.
8. Ouyang, Y.; Guo, B.; Lu, X.; Han, Q.; Guo, T.; Yu, Z. Competitivebike: Competitive analysis and popularity prediction of bike-sharing apps using multi-source data. *IEEE Trans. Mob. Comput.* **2018**. [CrossRef]
9. Rathore, M.M.; Ahmad, A.; Paul, A.; Rho, S. Urban planning and building smart cities based on the internet of things using big data analytics. *Comput. Netw.* **2016**, *101*, 63–80. [CrossRef]
10. Alam, M.; Rayes, A.; He, X.; Atiquzzaman, M.; Lloret, J.; Tsang, K.F. Guest editorial introduction to the special issue on dependable wireless vehicular communications for intelligent transportation systems (its). *IEEE Trans. Intell. Transp. Syst.* **2018**, *19*, 949–952. [CrossRef]

11. Faghih-Imani, A.; Hampshire, R.; Marla, L.; Eluru, N. An empirical analysis of bike sharing usage and rebalancing: Evidence from barcelona and seville. *Transp. Res. Part A Policy Pract.* **2017**, *97*, 177–191. [CrossRef]
12. Gleason, R.; Miskimins, L. *Exploring Bicycle Options for Federal Lands: Bike Sharing, Rentals and Employee Fleets*; TRB: Washington, DC, USA, 2012.
13. Shaheen, S.; Martin, E.; Cohen, A. Public Bikesharing and Modal Shift Behavior: A Comparative Study of Early Bikesharing Systems in North America. *Int. J. Transp.* **2013**, *1*, 2–18. [CrossRef]
14. Shaheen, S.A. *Public Bikesharing in North America: Early Operator and User Understanding*. Available online: https://rosap.ntl.bts.gov/view/dot/24566 (accessed on 5 January 2019).
15. Otero, I.; Nieuwenhuijsen, M.; Rojas-Rueda, D. Health impacts of bike sharing systems in Europe. *Environ. Int.* **2018**, *115*, 387–394. [CrossRef] [PubMed]
16. DeMaio, P. Bike-sharing: History, impacts, models of provision, and future. *J. Public Transp.* **2009**, *12*, 3. [CrossRef]
17. Camm, J.D.; Chorman, T.E.; Dill, F.A.; Evans, J.R.; Sweeney, D.J.; Wegryn, G.W. Blending or/ms, judgment, and GIS: Restructuring P&G's supply chain. *Interfaces* **1997**, *27*, 128–142.
18. Xu, J.; Yao, J.; Wang, L.; Ming, Z.; Wu, K.; Chen, L. Narrowband internet of things: Evolutions, technologies, and open issues. *IEEE Internet Things J.* **2018**, *5*, 1449–1462. [CrossRef]
19. Sun, Y.; Mobasheri, A. Utilizing crowdsourced data for studies of cycling and air pollution exposure: A case study using strava data. *Int. J. Environ. Res. Public Health* **2017**, *14*, 274. [CrossRef] [PubMed]
20. Yan, Y.; Tao, Y.; Xu, J.; Ren, S.; Lin, H. Visual analytics of bike-sharing data based on tensor factorization. *J. Vis.* **2018**, *21*, 495–509. [CrossRef]
21. Yang, Z.; Chen, J.; Hu, J.; Shu, Y.; Cheng, P. Mobility modeling and data-driven closed-loop prediction in bike-sharing systems. *IEEE Trans. Intell. Transp. Syst.* **2019**. [CrossRef]
22. Zhang, Y.; Wen, H.; Qiu, F.; Wang, Z.; Abbas, H. Ibike: Intelligent public bicycle services assisted by data analytics. *Future Gener. Comput. Syst.* **2019**, *95*, 187–197. [CrossRef]
23. Huang, F.; Qiao, S.; Peng, J.; Guo, B. A bimodal gaussian inhomogeneous poisson algorithm for bike number prediction in a bike-sharing system. *IEEE Trans. Intell. Transp. Syst.* **2018**. [CrossRef]
24. Olfert, C. *Urban Planning, Architecture and Bike Trails*; Sa & B Mag Sustainable Architecture & Building, Janam Publications Inc.: Toronto, ON, Canada, 2009.
25. Jiang, J.J. Bike Your City: Planning and Designing Cycling Infrastructure in the Urban Environment. Master's Thesis, Victoria University of Wellington, Wellington, New Zealand, 2012.
26. Silva, V.; Harder, H. *Urban Design Interventions towards a Bike Friendly City*; Trafikdage: Aalborg, Denmark, 2013.
27. Zheng, Y. Methodologies for cross-domain data fusion: An overview. *IEEE Trans. Big Data* **2015**, *1*, 16–34. [CrossRef]
28. Ahillen, M.; Mateo-Babiano, D.; Corcoran, J. Dynamics of bike sharing in washington, dc and brisbane, australia: Implications for policy and planning. *Int. J. Sustain. Transp.* **2016**, *10*, 441–454. [CrossRef]
29. He, B.; Zhang, Y.; Chen, Y.; Gu, Z. A simple line clustering method for spatial analysis with origin-destination data and its application to bike-sharing movement data. *ISPRS Int. J. Geo-Inf.* **2018**, *7*, 203. [CrossRef]
30. Shen, S.; Wei, Z.Q.; Sun, L.J.; Su, Y.Q.; Wang, R.C.; Jiang, H.M. The shared bicycle and its network—internet of shared bicycle (IOSB): A review and survey. *Sensors* **2018**, *18*, 2581. [CrossRef] [PubMed]
31. Hong, L.; Zheng, Y.; Yung, D.; Shang, J.; Zou, L. Detecting urban black holes based on human mobility data. In Proceedings of the 23rd SIGSPATIAL International Conference on Advances in Geographic Information Systems, Seattle, WA, USA, 3–6 November 2015; p. 35.
32. Cuadros-Vargas, A.J.; Nonato, L.G.; Tejada, E.; Ertl, T. Generating segmented tetrahedral meshes from regular volume data for simulation and visualization applications. In Proceedings of the Computat. Model. Objects Presented Images (CompIMAGE), Niagara Falls, NY, USA, 21–23 September 2018; pp. 141–146.
33. Rostami, S.; Creemers, S.; Leus, R. Precedence theorems and dynamic programming for the single-machine weighted tardiness problem. *Eur. J. Oper. Res.* **2019**, *272*, 43–49. [CrossRef]
34. Riding the Bike-Share Boom: The Top Five Components of a Successful System. Available online: https://www.itdp.org/2014/03/25/riding-the-bike-share-boom-the-top-five-components-of-a-successful-system/ (accessed on 25 March 2014).

35. Vogel, P.; Mattfeld, D.C. Strategic and operational planning of bike-sharing systems by data mining—A case study. In *International Conference on Computational Logistics*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 127–141.
36. Windmeijer, F. A finite sample correction for the variance of linear efficient two-step GMM estimators. *J. Econ.* **2005**, *126*, 25–51. [CrossRef]
37. Gebru, I.D.; Alameda-Pineda, X.; Forbes, F.; Horaud, R. Em algorithms for weighted-data clustering with application to audio-visual scene analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 2402–2415. [CrossRef] [PubMed]
38. Zhang, K.; Gonzalez, R.; Huang, B.; Ji, G. Expectation–maximization approach to fault diagnosis with missing data. *IEEE Trans. Ind. Electron.* **2015**, *62*, 1231–1240. [CrossRef]
39. Friedman, J.H. Greedy function approximation: A gradient boosting machine. *Ann. Stat.* **2001**, *29*, 1189–1232. [CrossRef]
40. Data. Available online: http://www.citibikenyc.com/system-data (accessed on 4 September 2017).
41. Cai, Q.; Liu, J. Hierarchical Clustering of Bipartite Networks Based on Multiobjective Optimization. *IEEE Trans. Netw. Sci. Eng.* **2018**. [CrossRef]