B&B: Planning Bus Routes with Sharing-bikes in the City

Yue Yu, **Tong Xia** and **Yong Li***

Abstract

Recently, the emergence of sharing-bikes exerts a significant impact on citizens' daily travel. However, when planning bus routes, the current system still does not take them into consideration. In this paper, we propose B&B, a data-driven system to plan bus routes with the consideration of sharing-bikes. Aiming to maximize the travel flow coverage, we design a heuristic approach to extend bus routes. Meanwhile, we propose a function to determine whether a road segment should be designed as a bus route or it is more suitable for passengers to ride sharing-bikes in that area. Moreover, to ensure the efficiency of bus routes, a constraint is set up on the directness of bus routes. Extensive evaluation of two large-scale datasets in New York City demonstrates that our system achieves the best coverage on both flows and areas, which outperform by 7.05% and 15.55% over the baselines without consider sharing-bikes. Our system lays a solid foundation for planners to plan the city transportation systems overall, especially concerning the design of the bus routes and sharing-bike lanes.

1 Introduction

With the development of the city and the advocacy of green-life, sharing-bikes are available almost everywhere in many cities and countries, especially in China, which aim to solve the '*last mile*' problem [Chu *et al.*, 2019], as citizens could ride a sharing-bike to take the public transportations including metro and buses. Since it is difficult to change operation routes of metros due to the high cost of construction, the bus route planning system with the consideration of sharing-bikes to provide more efficient service is needed. Previous work has studied bus route planning [Chuah *et al.*, 2016; Liu *et al.*, 2015; Chen *et al.*, 2014] and bike lane design [Bao *et al.*, 2017] by data-driven methods. However, they are investigated separately, and no work planned bus routes with the sharing-bikes. In this paper, we intend to explore the bus route planning problem with sharing-bikes to meet people's requests of travel.

Despite of its great significance, it is non-trivial to plan bus route with the consideration of sharing-bikes due to several challenges: *1*) How to design bus routes to meet the travel demand in the city while facilitating bike-ridings to bus stations. If we design the bus routes in a small area with dense traffic, it is more convenient for passengers but the coverage of overall bus service would be small and the travel efficiency would decrease. *2*) How to consider the design of bus routes and bike services together. To reduce the travel time and serve more traffic flow, it is reasonable to design more bus routes but the efficiency of the bus system would be discounted, as riding a bike is more preferred in short trips. *3*) For each bus route, how to make the trade-off between meeting more travel demand and providing satisfactory service is also challenging.

To overcome these challenges, we propose a data-driven approach, named B&B, to plan **B**us routes with sharing-**B**ikes. For the first challenge, we detect some intensive pick-up and drop-off locations over the city as bus station candidates and extend the bus routes among them to serve the majority of travel demand. While for travel between near places, it is more convenient for passengers to ride sharing-bikes. By doing so, requests for both long and short distance trips can be satisfied at the same time. For the second challenge, we define a decision function to determine when to plan a road segment as the bus route based on the patterns of sharing-bike trips, which is utilized in the dynamic expansion process as short-distance while small-flow lanes would be filtered for passengers to ride bikes, while lanes with longer distance or larger flow would be selected as bus routes. To solve the third challenge, we define a directness bound for bus route to maximize the flow covered by the route network with limited number and length of bus routes as well as restricted directness. Our contributions can be summarized as follows:

- We investigate the problem of planning bus routes with sharing-bikes. To the best of our knowledge, this is the first study that attempts to design the bus route with the consideration of sharing-bikes by a data-driven method.
- We formulate the bus route planning problem as an optimization problem with three practical constraints, which maximize the coverage of passengers' travel demand all over the city with the sharing-bikes. Moreover, we propose a three-stage heuristic approach to solve the formulated problem.
- We evaluate our system using two large-scale datasets of New York City. Compared with the state-of-the-art baselines, our method achieves the best performance across different parameters. Specifically, it outperforms the benchmark by 7.05% and 15.55% in terms of traffic flow and area coverage without considering sharing-bikes.

2 Related Work

Data-Driven Urban Planning: With the availability of large-scale mobility data, data-driven urban planning become increasingly popular [Zheng *et al.*, 2014]. [Yuan *et al.*, 2012; Xia *et al.*, 2019; Xu *et al.*, 2018] demonstrated the existence of different functions regions in a city through GPS trajectories. [Yao *et al.*, 2018] built a network representation of

*Corresponding author. E-mail: liyong07@tsinghua.edu.cn.

human movement based on vehicle tracks. With the popularity of sharing-bikes, researchers investigated the problem of improving urban transportation systems with sharing-bikes. For example, [Nair *et al.*, 2013] analyzed bicycle data from Paris and uncovered the relationship between bicycle usage and multimodal trips, and [Liu *et al.*, 2016; Li *et al.*, 2018b] studied the dynamic reposition of sharing-bike systems. In this paper, we provide a data-driven approach to design the bus route of a city using OD data.

Spatio-temporal Trajectory Mining: A significant amount of literature aim to mine the spatio-temporal trajectories since the mobility data has recently become widely available. Related works about origin-destination travel time estimation [Li *et al.*, 2018a], spatio-temporal representation learning [Yao *et al.*, 2018; Yu *et al.*, 2020] and traffic flow prediction [Shi *et al.*, 2020; Feng *et al.*, 2020] have been published. Additionally, [Liu *et al.*, 2011] utilized real taxi trajectories in Beijing to detect outliers in spatio-temporal traffic, and [Zheng *et al.*, 2011] used the GPS trajectories of taxis in urban areas to evaluate the effectiveness of the planning. In this work, since taxi trajectories reflect citizens' travel demand, we can utilize the origin and destination of each taxi trajectory to design bus routes.

Bus Route Designing: Urban transportation network design problem includes two parts, one is the road network design, another is the public transit network design [Farahani *et al.*, 2013]. In this paper, we mainly focus on the latter problem. Traditional methods optimized an objective function with several constraints and heuristic methods [Chen *et al.*, 2014], genetic algorithms [Chakroborty, 2003] and other mathematical approaches [Brata *et al.*, 2015] have been used to solve the optimization problem. Different from those works, we utilize heuristic while dynamic method to design bus route, and most importantly, we take the sharing-bike into consideration during the route expansion process.

3 Problem Formulation

In this section, we attempt to model and define the problem of planning bus routes with the consideration of sharing-bikes. To effectively plan the bus routes, we assume that bus stations are selected from those hot-spots, i.e. locations with highly intensive ODs. Given the identified hot-spots, we build a weighted graph $\boldsymbol{G} = (\boldsymbol{V}, \boldsymbol{E})$ to simulate travel demand, where the vertex set $\boldsymbol{V}$ denotes the station candidates derived from hot-spots and the distance matrix $\boldsymbol{D}$ denotes the distance among all candidate stations. For a pair of nodes $v_i, v_j \in \boldsymbol{V}$, we say there exists an edge between v_i and v_j in $\boldsymbol{E}$, if there is trajectory records between v_i and v_j, and the flow matrix $\boldsymbol{F}$ represents the intensity of such traffic flow. For readability, we summarize the major notations used throughout the paper in Table 1.

Based on the flow network, our system aims to optimize the total flow served by both buses and sharing-bikes under several practical constraints. Here, we define the set of bus routes as $\boldsymbol{L} = \{L_1, L_2, ..., L_i, ...\}$, where the total number of the bus routes is denoted as N_L. The i-th route is denoted as $L_i = (v_1^i, v_2^i, ..., v_s^i)$, where s is the number of stops in L_i, and $v_j^i \in \boldsymbol{V}(j = 1, 2, ..., m)$ stands for the j-th station in route i. Therefore, the total length of L_i is

$$l_i = \sum_{j=1}^{s-1} \boldsymbol{D}[v_j^i, v_{j+1}^i] \tag{1}$$

Notation	Description
$\boldsymbol{V}$	The set of all candidate bus stations.
$\boldsymbol{F}$	The flow matrix for all candidate stations.
$\boldsymbol{D}$	The distance matrix for all candidate stations.
$\boldsymbol{L}$	The set of all the designed bus routes.
$\boldsymbol{B}$	The set of all trips fulfilled via sharing-bikes.
$r(v_m, v_n)$	The directness between station m and n.
B_m	The length budget for every bus routes.
l_m	The maximum length for sharing-bike trips.
N	The maximum number of bus routes.

Table 1: List of commonly used notations.

As for sharing-bikes, the bike services are denoted as $\boldsymbol{B} = \{B_1, B_2, ..., B_i, ...\} = \{(v_1^1, v_2^1), (v_1^2, v_2^2), ..., (v_i^1, v_i^2), ...\}$, which means sharing-bikes would be utilized for travelling between v_i^1 and v_i^2. $(i = 1, 2, ..., n, ...)$. Below are some general criteria for planning routes:

Criterion 1: The operation as well as the management cost for the bus system will increase rapidly with the growing number of routes. Thus, there should be an upper bound N for the number of bus routes.

Criterion 2: There should be an upper bound on the length of each route since the volume of fuel in each bus is limited. Moreover, if the route is too long, then the operation order of the bus lines are susceptible to traffic jam and traffic accidents. The maximum length of each route is set to be B_m.

Criterion 3: To make the transit operation of buses more efficiently, there should be restrictions on bus routes to prevent them from taking zigzag routes. We use the ratio $r(v_m^i, v_n^i)$ to measure the directness between station m and n in route i, which can be calculated as follows,

$$r(v_m^i, v_n^i) = \frac{\sum_{j=m}^{n-1} \boldsymbol{D}[v_j^i, v_{j+1}^i]}{\boldsymbol{D}[v_m^i, v_n^i]} \quad (m < n). \tag{2}$$

The directness ratio should not be too large as the deviation of the route will make it lose passengers due to the caused relatively long commute time.

Optimization Goal. The ultimate goal is to design optimal bus routes to maximize the overall coverage of the traffic flow with the sharing-bikes. Based on the criteria above, we define the total flow f as the following: We assume that the passengers are willing to use the service of bus and sharing-bikes when their travel origin v_o and destination v_d are connected by buses and sharing-bikes and the route are direct enough (i.e. $r(v_o, v_d) \le \alpha$). With the definition, the optimization problem can be expressed as below:

$$\begin{aligned} \max \quad & \sum_{i=1}^{N_L} \sum_{v_o^i, v_d^i \in \boldsymbol{L}||\boldsymbol{B}, o<d} (\boldsymbol{F}[v_o^i, v_d^i] + \boldsymbol{F}[v_d^i, v_o^i]), \\ s.t. \quad & N_L \le N, \\ & l_i \le B_m, \quad \forall i \\ & r(v_o^i, v_d^i) \le \alpha, \quad \forall i, o, d(o < d). \end{aligned} \tag{3}$$

Our goal is to find a set of bus routes $\boldsymbol{L}$ and a set of sharing-bike lanes B to maximize the total flow in Equation 3 under the constraints 1-3. The first two criteria are aiming to con-

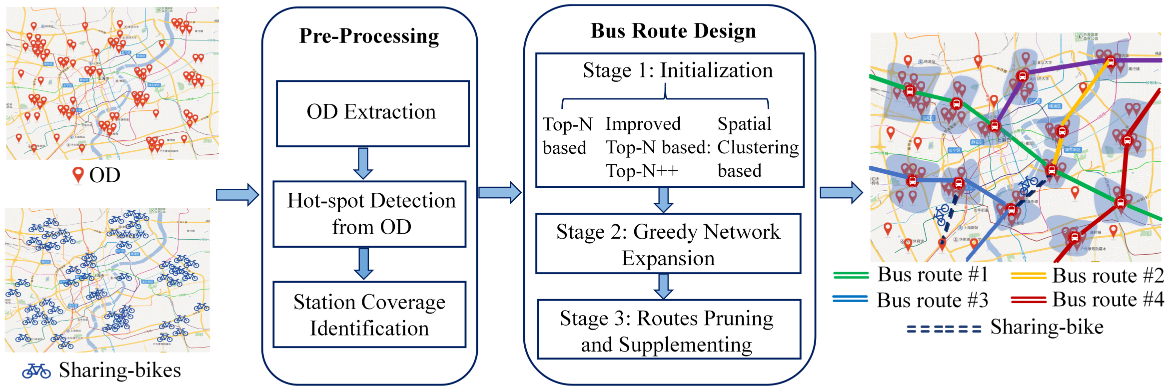


Figure 1: System overview

struct an economic bus system, while the last criterion is proposed to make the bus system rationalized and humanized.

4 System

In order to take the long and short distance of travel demand into consideration altogether and differentiate between buses and sharing-bikes, we propose our B&B system and give an overview of it in Figure 1.

4.1 Pre-processing.

In the pre-processing, We conduct a two-step clustering to formulate the flow network from OD data. The two steps are designed as follows:

Step 1: Hot-spot Detection. There are two small steps in detection. Firstly, the OD points are clustered by DBSCAN method [Ester *et al.*, 1996] to detect the density centers, which are taken as hot-spots and bus station candidates. Then, clusters are exaggerated by aggregating nearby points to identify the coverage of each station candidates. The number of the bus station is usually 200-300, and the distance between two bus station is usually 3km-5km in a modern city*.

Step 2: Bus Station Identification: After deriving several hot-spots, we expand the coverage of each cluster by aggregating OD to their nearest cluster centers. With the identified clusters, we establish the flow network based on hot-spots. To be specific, for the distance matrix $\boldsymbol{D}$, the distance between the center of clusters represents for the distance between those hot-spots. For the passenger flow matrix $\boldsymbol{F}$, we match our extracted OD-pairs with stations to determine the passenger flow between hot-spots.

4.2 Bus Route Design.

The bus route design is known to be a complex, nonlinear, non-convex, multi-objective NP-hard problem [Liu *et al.*, 2001]. To address this problem efficiently, We propose a three-stage heuristic approach. The overall algorithm consists of 3 main components: Firstly, we initialize the bus routes as a balance between flow and station coverage. Then, we propose a greedy network expansion algorithm considering sharing-bikes. Finally, we prune and supplement the existing bus routes further improve the coverage of the flow as well as the hot-spots.

*http://kns.cnki.net/kns/detail/detail.aspx?FileName=SCSDGB50220-1995&DbName=SCSD

Route Initialization. In our problem, we first need to find N road segments with each one stands for a route as a start. Selecting N routes with maximum passenger flow per unit distance is a natural way for initialization, but those road segments are often concentrated on particular areas, which may miss some important areas and reduce the coverage of hot-spots. To solve this problem, the hierarchical clustering method [Wardjr, 1963] is applied to split the hot-spots into N tiny parts. Then we find the road segment with the maximum passenger flow per unit distance in each smaller part as initialization. It has the advantage that the road segments belong to different part of the city, thus they are not connected and avert the deficiencies mentioned before.

Greedy Route Expansion. Based on the route initialization, we propose a heuristic algorithm to add road segments with higher flow gain on passenger flow per unit distance to the route network consecutively. Since we aim to consider the riding of sharing-bikes during bus route design process, the critical point is to determine when to select a road segment as a bus route and when passengers would ride bikes.

We first define some crucial notations in expansion: c is the added road segment, l_c is the length of c, i_c is the ID of the route in which the road segment is added to, $v_o^{i_c}, v_d^{i_c}$ stand for the origin and destination for c respectively, $p(l)$ represents for the relationship between the distance and the probability of riding sharing-bikes, E_c stands for the expected increase on number of people taking bus after adding c, Δf_c stands for the flow gain of adding c and $\Delta g_c, \Delta g'_c$ represents the gain of overall flow and bus passengers after adding c respectively.

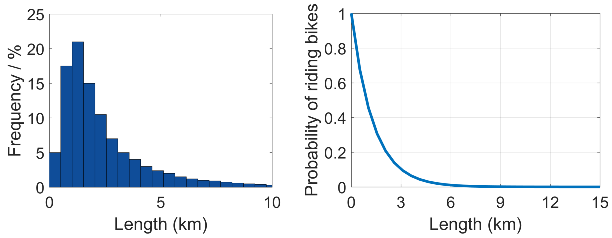
First we need to study the travel patterns of sharing-bike users. Figure 2(a) shows the travel distance distribution $f(l)$ of sharing-bikes. Based on the statistical result of sharing-bike trips, we define the relationship between the travel distance l and the probability p of riding a bike as follows,

$$p(l) = Pr(l' \geq l) = 1 - \int_0^l f(l')\, dl'. \tag{4}$$

The diagram of function $p(l)$ is shown in Figure 2(b). As there is an inverse relationship between the people taking buses and riding bikes, we use the probability function $p(d)$ to measure E_c as Equation 5.

$$E_c = \Delta f_c \cdot (1 - p(l_c)). \tag{5}$$

Considering the mobility pattern of the sharing-bike users, there should be a maximum distance l_m for the trip of sharing-bikes. Also, when picking the optimal road seg-



(a) Bike Trip Length Distribution of New-York

(b) The Relation between the Prob. of Riding Bikes and Trip Length

Figure 2: Mobility Pattern about Sharing-bike Users.

ments, the segment's length and its relationship with the current route should not be overlooked when seeking for maximum gain. As for directness factor, the directness of the route L_{i_c} in which the road segment is added is calculated as $r(v_1^{i_c}, v_d^{i_c})$. Then, we can derive the expression for Δg_c as Equation 6,

$$\Delta g_c = \frac{\Delta f_c}{l_c \cdot r(v_1^{i_c}, v_d^{i_c})}. \tag{6}$$

When considering sharing-bikes, the score of gain w.r.t. bus passengers $\Delta g_c'$ is shown as Equation 7,

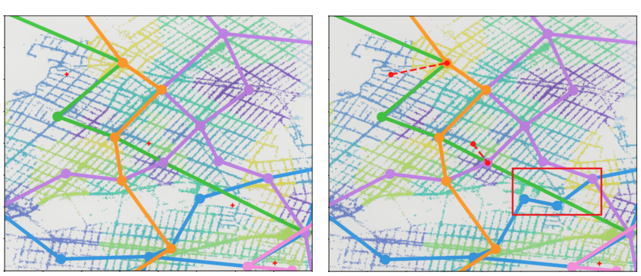
$$\Delta g_c' = \Delta g_c \cdot (1 - p(l_c)) = \frac{E_c}{l_c \cdot r(v_1^{i_c}, v_d^{i_c})}. \tag{7}$$

The detail for the algorithm is shown in Algorithm 1. In our algorithm, we first find all the road segments that serve as the extension of the current routes and satisfy the criteria in Equation 3. Next, we divide the road segments into two categories $\boldsymbol{l}, \boldsymbol{s}$ as the road segments in $\boldsymbol{s}$ are shorter than l_m and the segments in $\boldsymbol{l}$ are longer than l_m. For each category, we calculate the optimal road segments based on Equation 6 as c_l, c_s, the maximum score for flow gain Δg_c as $\Delta g_{max,s}, \Delta g_{max,l}$, the length of optimal road segments as l_{c_l}, l_{c_s} respectively.

When comparing the optimal result in two categories, if the Δg_c for c_s is greater than that of c_l, but $\Delta g_c'$ for c_s is smaller, it implies that there is sufficient demand to construct the route, but few people are willing to take the bus. Under this circumstance, bike rides are utilized to provide transportation services. Otherwise, bus routes are planned regularly, as we compare the value of $\Delta g_{max,s}, \Delta g_{max,l}$, and choose the corresponding segment with higher score.

Routes Pruning and Supplementing.

To make further improvement in the coverage of hot-spots and passenger flows, we could add some stations which are



(a) After expansion

(b) After pruning and supplementing

Figure 3: An example of bus route design process.

Algorithm 1 Bus Route Expansion with Sharing-bikes

Input: Initialized bus routes $\boldsymbol{L} = \{L_1, L_2, ..., L_N\}$, Length Budget B_m, Directness Parameter α, Maximum distance l_m.
Output: Result bus routes $\boldsymbol{L}$, Sharing-bike rides $\boldsymbol{B}$.
begin
  Initialization: Sharing-bike services $\boldsymbol{B} \leftarrow \emptyset$;
  The budget of each bus route $B_i \leftarrow B_m - \boldsymbol{D}[v_1^i, v_2^i]$.
  // Routes Expansion
  do
    Calculate passenger flow f based on Equation 3.
    Determine candidate set for road segments C_r.
    Reset $c_l, c_s \leftarrow \emptyset$, $\Delta g_{max,s}, \Delta g_{max,l} \leftarrow 0$, $l_{c_s}, l_{c_l} \leftarrow 0$.
    for $c \in C_r$ **do**
      Retrieve overall routes $\boldsymbol{R}$ based on $\boldsymbol{L} \cup \boldsymbol{B} \cup \{c\}$.
      Calculate the score Δg_c based on Equation 6.
      if $l_c < l_m$ & $\Delta g_c > \Delta g_{max,s}$ **then**
        $\Delta g_{max,s} \leftarrow \Delta g_c, l_{c_s} \leftarrow l_c, c_s \leftarrow c$.
      else if $\Delta g_c > \Delta g_{max,l}$ **then**
        $\Delta g_{max,l} \leftarrow \Delta g_c, l_{c_l} \leftarrow l_c, c_l \leftarrow c$.
    if $\Delta g_{max,l} > \Delta g_{max,s}$ **then**
      $L_{c_i} \leftarrow L_{c_i} \cup \{c_l\}, B'_{c_i} \leftarrow B'_{c_i} - l_{c_l}$
    else
      Calculate the $\Delta {g'}_c$ **for both** c_s **and** c_l**.**
      if $\Delta g_{c_s} > \Delta {g'}_{c_l}$ **then**
        $L_{c_i} \leftarrow L_{c_i} \cup \{c_s\}, B'_{c_i} \leftarrow B'_{c_i} - l_{c_s}$.
      else
        $\boldsymbol{B} \leftarrow \boldsymbol{B} \cup \{c_s\}$ *// Flow will be covered by bikes.*
  While $C_r \neq \emptyset$
  // Routes Pruning and Supplementing
  for $v \in \boldsymbol{V}$ **do**
    if $v \notin \boldsymbol{L}$ & $v \notin \boldsymbol{B}$ **then**
      for $L_i \in \boldsymbol{L}$ **do**
        Finding the nearest station to v **in** L_i: $j \leftarrow \text{argmin}_j\, [d(v_j^i, v) + d(v_{j+1}^i, v)]$;
        Calculate the length increase: $\Delta l \leftarrow \boldsymbol{D}[v_i^j, v] + \boldsymbol{D}[v_i^{j+1}, v] - \boldsymbol{D}[v_i^j, v_i^{j+1}]$;
        if $\Delta l > (\alpha - 1)\boldsymbol{D}[v_i^j, v_i^{j+1}]$ & $\Delta l < B_i$ **then**
          $L_i \leftarrow L_i \cup \{v\}$; *//Adding* v *to the bus route*
          $B_i \leftarrow B_i - \Delta l$;
          Break
  return $\boldsymbol{L}, \boldsymbol{B}$

not added to the bus routes currently, but the constraints on both budget and directness would not be violated after adding them to the whole system. The rule of pruning and supplementing is shown in Algorithm 1.

To better exemplify the role of pruning and supplementing, we choose several bus routes shown in Figure 3. In Figure 3(a), the route is relatively straight but it fails to include passed stations. In the red rectangle labeled area of Figure 3(b), passed stations are supplemented to satisfy more traffic flow, but the directness of the whole route slightly increases, which is the trade-off between meeting travel demand and travel efficiency.

As for sharing bikes, when the distance between a particular station and its nearest bus station is within l_m, then the corresponding travel demand between it and stations in routes can be satisfied via riding sharing-bikes to nearest bus stations and taking buses then. In Figure 3(b), sharing-bike travels are shown by red dashed lines, which can further elevate the coverage of flows and stations of our system.

5 Evaluation

5.1 Experimental Settings

Datasets. We utilize two kinds of dataset: *Green Taxi Trip Data*† and *Citi Bike Trip Data* [Li *et al.*, 2015] from New York City, one of the largest city in the world, for evaluation. The key characteristics of these datasets are summarized in Table 2. We filter the OD pairs with the travel time less than 1 minutes or the travel length less than 500m, which are noise records caused by driver's mis-operation. After that, more than 1 million taxi OD records are obtained. Such two large-scale datasets in New York City guarantee the accuracy of our evaluation. To speed up the pre-processing steps, we spatially divide them into $10m \times 10m$ grids and only the geographical center of the grid is reserved. Then, we detect the hot-spots from grid OD and identify the bus station candidates via our pre-processing method. Finally, we build the flow network for bus route expansion. We try different parameters for clustering to detect hot-spots as bus station candidates. The optimum neighbor radius ϵ = 100m and the minimum number of points in neighborhood is 110. As a result, there are 246 clusters in total.

Service	Time	# of Records
Taxi Trip Data	Aug. 1st - 31st, 2014	2,639,500
Bike Trip Data	Aug. 1st - 16th, 2014	473,621

Table 2: Summary of basic statistics of two representative datasets in New York City.

Baselines and Metrics. We compare our B&B method with several baselines.For the state-of-the-art solution [Chuah *et al.*, 2016], it proposed a system to build a loop line among ODs as follows.

L&NB: Planning bus routes by loop line method without considering the sharing-bike. The loop lines are also common when designing bus routes.

L&B: Planning bus routes by loop line method with the sharing-bikes.

Besides, we also take some variants of our system into consideration overall:

B&NB: Planning bus routes by our system without the sharing-bikes.

B&B+EP: Planning bus routes by our system with the sharing-bikes only in expansion stage.

B&B+SP: Planning bus routes by our system with the sharing-bikes only in supplementing stage.

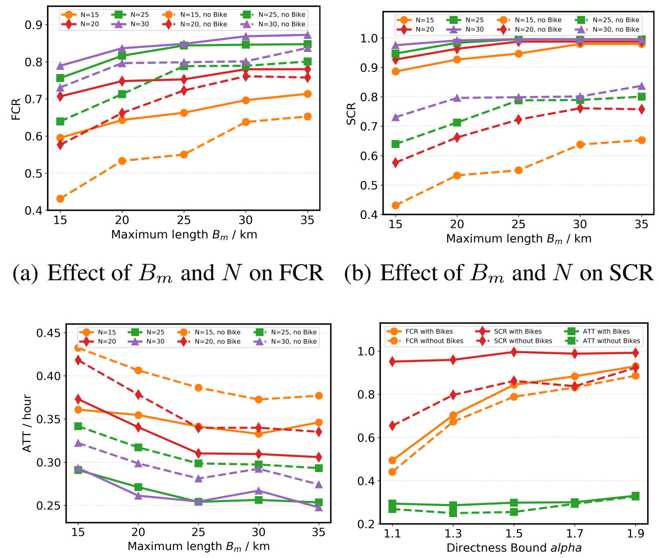
When studying the effect of different factors on our planning algorithm, the default parameters are set as Table 3. Besides, the default way of initialization is agglomerative clustering algorithm. To evaluate how our system satisfy the goal we aim to achieve, we use the following three metrics:

The Flow Coverage Rate (FCR): The ratio of the flow covered by our system to the total flow extracted from OD.

The Station Coverage Rate (SCR): The ratio of the number of bus station included by our routes to the number of all bus station candidates.

The Average Travel Time (ATT): The average travel time from all origin to destination flow satisfied by our bus routes.

†https://data.cityofnewyork.us/Transportation/2014-Green-Taxi-Trip-Data/2np7-5jsg



(a) Effect of B_m and N on FCR (b) Effect of B_m and N on SCR

(c) Effect of B_m and N on ATT (d) Effect of α

Figure 4: The effect of parameters in our algorithm.

5.2 The Optimal Solution for Planning Bus Routes

For a large city like New York, to plan the bus routes efficiently and economically, we mainly tune two parameters: the maximum length B_m for routes and the number of routes N by grid search to find the optimal solution. We tune B_m from 15km to 35km, N from 15 to 30 respectively. In this case, directness bound α is set to 1.5.

Figure 4 gives the result of the comparison with different budgets. We have the following conclusions:

1) With the increase of the B_m and N, FCR and SCR both rise and ATT declines. However, the changes of the them slow down subsequently as B_m increases, which indicates that a relatively small budget would be enough to cover the traffic demand. The reason behind it is that the road segments with the highest flows are chosen first in our method. What's more, most of the stations will be included into the route when B_m are large, thus making the travel efficient enough and making it unnecessary to add new stations.

2) The performance of our methods with sharing-bikes is better than those without them, but the gap between the result has narrowed when B_m becomes larger. It is because that when the budget becomes larger, more bus stations will be covered. Therefore, the excessive travel demand completed

Parameter	Value
Maximum trip length l_m of Sharing-bikes	$2km$
The average speed of buses v_b	40km/h
The average speed of Sharing-bikes v_{sb}	15km/h
The average speed of walking v_w	5km/h

Table 3: The Default Parameters in our experiments.

by sharing-bike rides will be less in both the expansion stage and the supplementing stage.

From the discussion above, we find that B_m=25km and N=25 would be an appropriate setting for the planning of the bus routes in New York City. When N is fixed and B_m increases 40%(from 25km to 35km), the gain on FCR is merely 0.4%, on ATT is merely 0.2% and SCR just not change. Similarly, when B_m is fixed and N increases 20%(from 25 to 30), the gain on FCR and ATT is merely 0.5% and 0.2% and SCR still not change. Those are enough to demonstrate that this setting can achieve a trade-off between providing efficient service for planners while being econimic in real planning. Moreover, compared with the method without sharing-bikes, the result under this setting shows that FCR has been promoted by 7.05%, SCR has been improved by 15.55% and ATT has been reduced by 17.32%, which justifies the efficacy of the implementation of sharing-bikes.

5.3 The Benefits of Considering Sharing-bikes

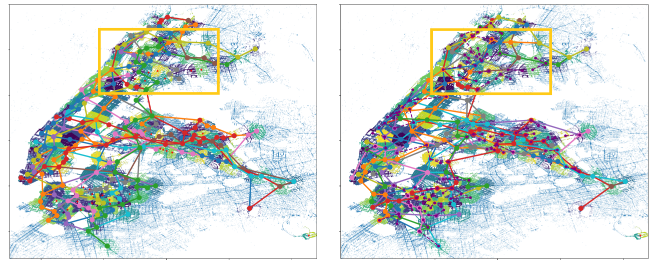
We compare our system B&B with B&NB, B&B+EP and B&B+SP to further evaluate the benefits of considering sharing-bikes in different stages of our system. Results under different groups of parameters are shown Table 6 with $\alpha = 1.5$, as we observe that our B&B method always outperforms B&NB, B&B+EP, and B&B+SP. Specifically, FCR has been improved by about 2.20% (on average, the same below) and SCR has been improved by 1.20% about when sharing-bikes are considered in routes expansion, which means our decision function plays an important role in trading-off between bus and bike. Moreover, FCR has been improved by about 7.54% and SCR has been improved by 21.77% about when sharing-bikes are considered in routes supplementing, which means the designed bus routes can meet more traffic demand when considering sharing-bikes. By combining these two aspects together, our B&B method improves FCR by 9.74% and SCR by 22.97%, which can serve more traffic flow under the same conditions.

To visualize the effect of sharing-bikes, we show the route planning results for New York City with the derived optimal settings in Figure 5. Although the designed bus routes would cover most areas in the city to meet passengers' travel demand, it is worth noting that the result is more reasonable when taking sharing-bikes into consideration. Take the yellow rectangle labeled area as an example. Bus stations are unable to cover all stations in Figure 5(a). Actually, these short-distance routes should be improved, because sharing-bikes can better meet such travel demand as Figure 5(b) shown. In this case, both traffic flow coverage and station coverage are significantly improved by consideering sharing-bikes.

5.4 Comparison of Different Expansion Methods

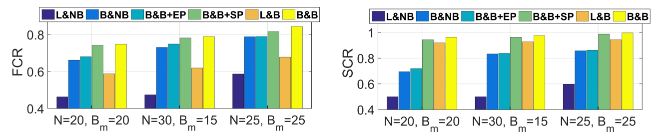
We also compare the results with and without considering sharing-bikes. The length limit of loop lines is relatively higher. In this case, we set the limit to $B_{m,loop} = 50km$.

The results are also shown in Figure 6, which verifies our algorithm. Obviously, our B&B method outperform others on both FCR and SCR with the different set of budget constraints. In addition, the increase of FCR exceeds 20% in almost all the cases regardless of sharing-bikes, which demonstrates that our expansion method can meet more traffic flow.



(a) Without sharing-bikes (b) With sharing-bikes

Figure 5: A case result for route planning for New York City, where dots represent bus stations, lines different colors represent different bus routes and the short purple lines represent that passengers would ride sharing-bikes to their nearest bus station.



(a) Comparison of FCR with different expansion methods (b) Comparison of SCR with different expansion methods

Figure 6: Comparison of different expansion methods.

5.5 The Influence of Directness Parameter

The parameter α plays a crucial part in the design of the bus route. When α increases, the restriction on directness is relaxed and the number of candidate road segments in expansion increases, resulting in covering more stations. However, when α is higher, passengers will take a longer time from the source to destination. From the figure 4(d), we observe that our FCR and SCR are always higher than that without bikes, while our ATT is always lower, which demonstrates that our system can increase the satisfaction for traffic flow and decrease the travel time at the same time. It is also worth noting that when α increases, the travel time first decreases than increases, which means that we can achieve a trade-off between providing efficient and satisfactory service for a passenger.

6 Conclusion

In this paper, we investigated the problem of bus routes planning with the sharing-bikes. Our goal is to maximize the flow coverage under three practical constraints: 1) the number of routes, 2) the length of each route, and 3) the directness of each road. Based on the flow extracted from large-scale OD data, we propose a heuristic approach to expand bus routes. Experiments on two large-scale datasets show the gains of our system and demonstrate that our system enables the city planners to design better public transportation systems.

Although we do not take current bus routes into consideration, it is still reasonable to put our design into real practice. To be specific, we can compare our design with the current public transportation system: if the bus route is contained into the system, we suggest that it should be reserved. Meanwhile, if the designed bus route is not included in the current system, we recommend that this route be replenished to current routes. In this way, our proposed B&B can be utilized to enhance the travel convenience in metropolis, which also sheds light on intelligent urban planning with the consideration of multiple forms of transportation.

References

[Bao *et al.*, 2017] Jie Bao, Tianfu He, Sijie Ruan, Yanhua Li, and Yu Zheng. Planning bike lanes based on sharing-bikes' trajectories. In *Proc. SIGKDD*, pages 1377–1386, 2017.

[Brata *et al.*, 2015] Komang C. Brata, Deron Liang, and Sholeh H. Pramono. Location-based augmented reality information for bus route planning system. *International Journal of Electrical and Computer Engineering*, 5(1):142–149, 02 2015.

[Chakroborty, 2003] Partha Chakroborty. Genetic algorithms for optimal urban transit network design. *Computer-Aided Civil and Infrastructure Engineering*, 18(3):184–200, 2003.

[Chen *et al.*, 2014] Chao Chen, Daqing Zhang, Nan Li, and Zhi Hua Zhou. B-planner: Planning bidirectional night bus routes using large-scale taxi gps traces. *IEEE Transactions on Intelligent Transportation Systems*, 15(4), 2014.

[Chu *et al.*, 2019] Junhong Chu, Yige Duan, Xianling Yang, and Li Wang. The last mile matters: Impact of dockless bike sharing on subway housing price premium. *Available at SSRN 3195004*, 2019.

[Chuah *et al.*, 2016] Seong Ping Chuah, Huayu Wu, Yu Lu, Liang Yu, and Stephane Bressan. Bus routes design and optimization via taxi data analytics. In *Proc. CIKM*, pages 2417–2420, 2016.

[Ester *et al.*, 1996] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, number 34, pages 226–231, 1996.

[Farahani *et al.*, 2013] Reza Zanjirani Farahani, Elnaz Miandoabchi, W.Y. Szeto, and Hannaneh Rashidi. A review of urban transportation network design problems. *European Journal of Operational Research*, 229(2):281 – 302, 2013.

[Feng *et al.*, 2020] Jie Feng, Ziqian Lin, Tong Xia, Funing Sun, Diansheng Guo, and Yong Li. A sequential convolution network for population flow prediction with explicitly correlation modelling. In *Proc. IJCAI*, 2020.

[Li *et al.*, 2015] Yexin Li, Yu Zheng, Huichu Zhang, and Lei Chen. Traffic prediction in a bike-sharing system. In *Proc. SIGSPATIAL*, 2015.

[Li *et al.*, 2018a] Yaguang Li, Kun Fu, Zheng Wang, Cyrus Shahabi, Jieping Ye, and Yan Liu. Multi-task representation learning for travel time estimation. In *Proc. SIGKDD*, pages 1695–1704. ACM, 2018.

[Li *et al.*, 2018b] Yexin Li, Yu Zheng, and Qiang Yang. Dynamic bike reposition: A spatio-temporal reinforcement learning approach. In *Proc. SIGKDD*, pages 1724–1733. ACM, 2018.

[Liu *et al.*, 2001] Chao-Lin Liu, Tun-Wen Pai, Chun-Tien Chang, and Chang-Ming Hsieh. Path-planning algorithms for public transportation systems. In *Proc. ITSC*, pages 1061–1066. IEEE, 2001.

[Liu *et al.*, 2011] Wei Liu, Yu Zheng, Sanjay Chawla, Jing Yuan, and Xing Xie. Discovering spatio-temporal causal interactions in traffic data streams. In *Proc. SIGKDD*, 2011.

[Liu *et al.*, 2015] Yanchi Liu, Chuanren Liu, Nicholas Jing Yuan, Lian Duan, Yanjie Fu, Hui Xiong, Songhua Xu, and Junjie Wu. Exploiting heterogeneous human mobility patterns for intelligent bus routing. In *Proc. ICDM*, 2015.

[Liu *et al.*, 2016] Junming Liu, Leilei Sun, Weiwei Chen, and Hui Xiong. Rebalancing bike sharing systems: A multi-source data smart optimization. In *Proc. SIGKDD*, pages 1005–1014. ACM, 2016.

[Nair *et al.*, 2013] Rahul Nair, Elise Miller-Hooks, Robert C. Hampshire, and Ana Bušić. Large-scale vehicle sharing systems: Analysis of vélib'. *International Journal of Sustainable Transportation*, 7(1), 2013.

[Shi *et al.*, 2020] Hongzhi Shi, Quanming Yao, Qi Guo, Yaguang Li, Lingyu Zhang, Jieping Ye, Yong Li, and Yan Liu. Predicting origin-destination flow via multi-perspective graph convolutional network. In *Proc. ICDE*, 2020.

[Wardjr, 1963] Joeh. Wardjr. Hierarchical grouping to optimize an objective function. *Publications of the American Statistical Association*, 58(301):236–244, 1963.

[Xia *et al.*, 2019] Tong Xia, Yue Yu, Fengli Xu, Funing Sun, Diansheng Guo, Depeng Jin, and Yong Li. Understanding urban dynamics via state-sharing hidden markov model. In *The World Wide Web Conference*, page 3363–3369, 2019.

[Xu *et al.*, 2018] Fengli Xu, Tong Xia, Hancheng Cao, Yong Li, Funing Sun, and Fanchao Meng. Detecting popular temporal modes in population-scale unlabelled trajectory data. *Proc. IMWUT*, 2(1):1–25, 2018.

[Yao *et al.*, 2018] Zijun Yao, Yanjie Fu, Bin Liu, Wangsu Hu, and Hui Xiong. Representing urban functions through zone embedding with human mobility patterns. In *Proc. IJCAI*, pages 3919–3925, 2018.

[Yu *et al.*, 2020] Yue Yu, Tong Xia, Huandong Wang, Jie Feng, and Yong Li. Semantic-aware spatio-temporal app usage representation via graph convolutional network. *Proc. IMWUT*, 4(3), September 2020.

[Yuan *et al.*, 2012] Jing Yuan, Yu Zheng, and Xing Xie. Discovering regions of different functions in a city using human mobility and pois. *Proc. ACM SIGKDD*, 2012.

[Zheng *et al.*, 2011] Yu Zheng, Yanchi Liu, Jing Yuan, and Xing Xie. Urban computing with taxicabs. In *Proc. Ubicomp*. ACM, 2011.

[Zheng *et al.*, 2014] Yu Zheng, Licia Capra, Ouri Wolfson, and Hai Yang. Urban computing: Concepts, methodologies, and applications. *ACM Trans. Intell. Syst. Technol.*, 5(3), September 2014.