Trip Route Planning for Bicycle-Sharing Systems

Jiawei Zhang
University of Illinois at Chicago
Chicago, IL, USA
jzhan9@uic.edu

Philip S. Yu
University of Illinois at Chicago
Chicago, IL, USA
Institute for Data Science
Tsinghua University, China
psyu@cs.uic.edu

***Abstract*—Bicycle-sharing systems (BSSs) offering shared bike usages to the public are becoming more and more popular nowadays. In bicycle-sharing systems, people can borrow and return bikes at any bike stations in the service region, but the free-ride time is usually limited. Therefore, for long-distance bike trips, individuals need to pre-schedule the bike trip route in advance and change the bike within the free-ride time so no over-time fees will be charged. In this paper, we will study the trip route planning problem for individuals when using the bicycle-sharing systems. Given the trip origin and destination, we aim at identifying the optimal trip route from the origin to the destination through the bike stations. To address the problem, we conduct a thorough analysis about an existing BSS, Divvy, launched in Chicago. Based on the analysis results, a novel bike route planning framework "BSSs based Trip rOute Planning" (STOP) is proposed in this paper, which can identify the optimal trip route by mapping the problem into a minimum-cost network flow problem. Extensive experiments conducted on real-world bicycle-sharing system datasets demonstrate the effectiveness of STOP.**

***Index Terms*—Route Planning, Bicycle-Sharing Systems, Geographic Information Systems, Vehicle Networks and Applications**

I. INTRODUCTION

Bicycle-sharing systems (BSSs) [26], are the public transportation service systems launched in the urban areas, in which the bikes are available for shared use to the public. The bikes together with the stations as well as bike trip routes among the stations can form a kind of intelligent transportation network. Generally, people can borrow bikes from the nearby stations and return the bikes to any stations in the city. Without concerns about the parking issues, BSSs have become an important short-distance travel option for both local residents and tourists. Due to the needs of green and low-carbon public transport vehicles nowadays, BSSs have been launched in many big cities (e.g., Chicago[1], New York[2], San Francisco[3]), and achieved a remarkable success. A detailed analysis about the Chicago Divvy BSS is available in [26].

In many of the BSSs, the bicycles are equipped with sophisticated electronic sensors and real-time GPS tracking devices, which can monitor traffic conditions and average vehicles speed. Via the information exchange, people can know the general trip time and design their traveling route

[1]https://www.divvybikes.com
[2]https://www.citibikenyc.com
[3]http://www.bayareabikeshare.com

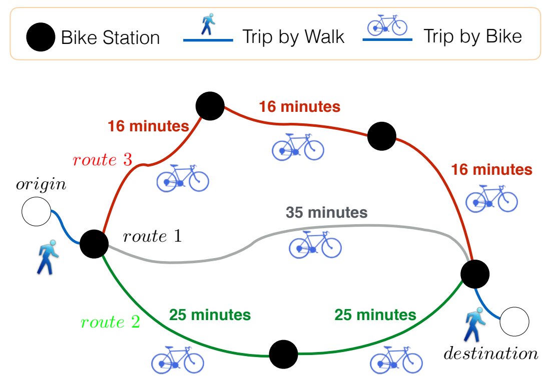


Fig. 1. An example of trip routes in BSSs.

wisely. In addition, to attract usages from the public, BSSs usually provide the services either for free or with very low service charges. Depending on the specific travel needs, customers can purchase either *annual membership* or *one-day pass* tickets, both of which will allow unlimited 30-minute trips, and trips longer than 30 minutes will lead to over-time fees. To avoid unreasonable long-time occupation of the bikes by individuals, constraining the bike riding time is a common practice used by the majority of the existing BSSs in service. Therefore, for long-distance trips taking more than 30 minutes, e.g., a full-day bike trip for sightseeing, customers need to pre-plan the trip route in advance and change their bikes at stations for every 30 minutes regularly, if they don't want to pay the unnecessary over-time fees.

Problem Studied: In this paper, we will study the trip route planning problem for BSSs to help the customers determine the optimal bike route (i.e., a sequence of stations for borrowing/exchanging/returning the bikes). Formally, the problem is defined as the ROUTEPLANNING problem. Given the trip origin and destination, ROUTEPLANNING aims at identifying the optimal bike route with the minimum costs between the origin and destination without over-time fees.

To illustrate the problem more clearly, we also provide an example in Figure 1, where the trip origin and destination together with 5 bike stations are shown. Between the origin and destination, 3 different bike routes sharing common start/end bike stations are given, which take different amount of time. The free-ride time of the bikes in the system is 30 minutes. Via route 1, customers ride directly to the end station and it takes 35 minutes, which exceeds the time limit already,

TABLE I
PROPERTIES OF THE DIVVY DATASET

datasets	trip	station
2013 Q3-Q4	759,788	300
2014 Q1-Q2	905,699	300
2014 Q3-Q4	1,548,935	300
2015 Q1-Q2	1,096,239	474

and can introduce over-time fees. Both route 2 and route 3 have no overtime fees, but customers need to change the bikes during the trip. Route 2 one bike change and takes 50 minutes, while route 3 requires two bike changes at two intermediate stations and takes 48 minutes. Among these three bike routes, the optimal one without over-time charges is to be identified in ROUTEPLANNING.

The ROUTEPLANNING problem is a new problem and has never been studied for BSSs before. The ROUTEPLANNING problem is different from traditional route direction works [21], [25], [19], [4], [23], [11], as these existing works mainly focus on discovering the route paths based on the traffic road networks for general vehicles without time limit concerns. Instead of the whole traffic road network, the input of ROUTEPLANNING is merely a set of *isolated bike stations*, and the objective is to identify the optimal *bike route* with the *free-ride time constraint*.

The ROUTEPLANNING problem is a difficult problem consisting of many open challenges:

- *Trip Route Cost*: Different trip routes can lead to certain commute costs, like the total travel distance, the number of bike changes, and the time cost, etc. A proper measure of the trip cost is needed.
- *Free-ride Time Constraint*: Due to the free-ride time constraints, the trip route segments between bike stations cannot exceed 30 minutes. Pre-pruning of the "illegal" route segment is a necessary step.
- *Optimal Route Detection*: Given the trip origin and destination, extraction of the optimal trip route with the minimum costs between them consisting of the feasible route segments only is a challenging problem.

To address these above challenges, a new bike route extraction method "BSSs based Trip rOute Planning" (STOP) is introduced in this paper. In STOP, the costs introduced by the bike route are measured with the *time*, *number of bike changes* and *trip distances* simultaneously. To maintain the time constraints, we will introduce the *bike route network* concept and propose to prune "illegal" trip route segments from the network. Since these "illegal" trip routes actually violate the time constraint, they will not involved in any feasible solutions at all. Therefore, the pruning step will not removing any feasible solutions but can also shrink the search space effectively. Based on the pruned network, STOP applies a minimum-cost network flow based model to extract the optimal bike trips.

The remaining part of this paper is organized as follows. At the very beginning, we will first introduce and analyze

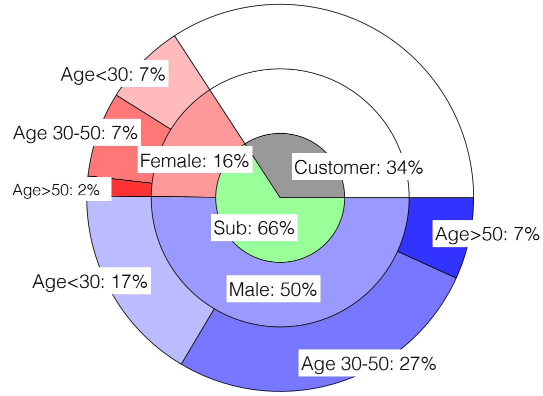


Fig. 2. Percentages of trip belong to different categories of users (Red: Female Subscribers; Blue: Male Subscribers; Green: Subscribers; Gray: Customers).

the Divvy bicycle-sharing system launched in Chicago in Section II. After that, in Section III, we will introduce the concept definitions and formulation of the ROUTEPLANNING problem. The method is described in Section IV, which is evaluated in Section V. Finally, Section VI talks about the related works and Section VII concludes the paper.

II. BICYCLE-SHARING SYSTEM ANALYSIS

To gain a more comprehensive knowledge about the BSSs, in this section, we will analyze a real-world BSS launched in Chicago, i.e., the Divvy Bike. In the following parts, we will first introduce the datasets about Divvy, and then show some brief analysis results about the user composition, bike usage temporal patterns, as well as the station distribution and station usage spatial patterns.

A. Divvy Dataset Description

Before analyzing the individuals' travel behaviors, we will introduce the dataset about a real-world bicycle-sharing system first in this section. The dataset used in this paper is about the Divvy bicycle-sharing system initially launched in the Chicago city on June 28, 2013. At the very beginning, Divvy had about 750 bikes at 75 stations (operating in an area spanning from the Loop north to Berwyn Avenue, south to $59th$ Street, west to Kedzie Avenue, and east to the Lake Michigan coast). A quick expansion has been made at early 2015, and Divvy now operates $4,760$ bicycles at 474 stations (in an area bounded by $75th$ Street on the south, Touhy Avenue on the north, Lake Michigan on the east, and Pulaski Road on the west).

The Divvy bicycle-sharing system datasets are public and new datasets are released every two quarters, which can be downloaded at its official website[4]. We downloaded the Divvy bicycle-sharing system data on November 2, 2015, which contains 4 separate datasets time ranging from the middle of 2013 to the middle of 2015. The downloaded datasets include the complete historical trip records as well as the station information, whose statistical information and detailed descriptions are available in Table I and as follows.

[4]https://www.divvybikes.com/data

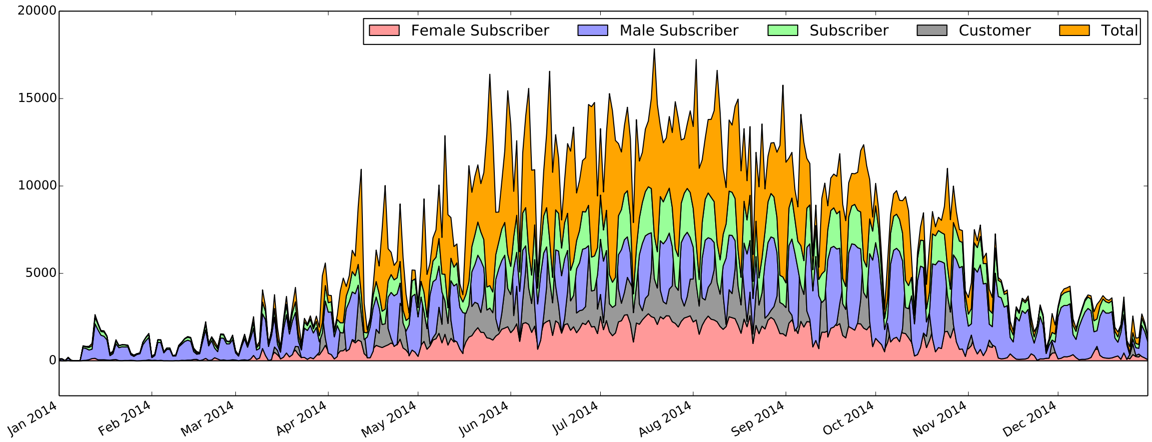


Fig. 3. Statistics of trips on each day of the 2014 year (X axis: each day of 2014; Y axis: number of trips in one day).

- *Trip*: Each trip record in the datasets has a unique ID. From the trip record data, we can know the trip start and end time as well as the corresponding origin and destination bike stations. The trip record also indicates whether the user is an annual membership holder or just an one-day pass holder, who are called the "subscriber" and "customer" respectively. For the annual membership subscribers, the trip record data also includes their gender and birth year information, which is helpful for categorizing the users (into male vs female, as well as youth vs senior) and allows us to study the bike-usage behaviors of different categories of people.
- *Station*: For each station, we can know its ID, name as well as its specific location, which is represented as a (latitude, longitude) coordinate pair in the dataset. At stations, bikes are locked at the docks and the numbers of docks available at the stations are called the station capacities, which are also available in the datasets.

As shown in Table I, generally, the Chicago people like to use the Divvy bike a lot and, on average, $179,610$ trips were taken in each month during the past two years. Meanwhile, the number of stations doesn't change in the first 3 datasets (which are all 300), and increases to 474 in the last dataset because of the scheduled expansions during the spring and early summer in 2015.

In the following subsections, we will study the datasets in great detail to analyze the user composition, individuals' temporal travel patterns, and spatial travel patterns respectively. Based on the analysis results, we will introduce the ROUTEPLANNING problem based on these stations and the STOP model to address the problem.

B. Divvy User Composition

The Divvy Bike can attract the usage from a very diverse group of people in Chicago. From Figure 2, we observe that the majority of trips are actually taken by the registered "subscribers" (i.e., the green area marked with "Sub"), which account for about 66% in the total trips, while those finished by the "customers" with one-day pass (i.e., the gray area) account for 34% in all.

No extra information is available for the "customers", as they just buy one-day pass and no personal information is recorded. Meanwhile, for the "subscribers" with formal membership registrations, we can know more (e.g., gender and age) about them and can further study their compositions.

As shown in Figure 2, the "subscribers" area is further divided into the "male" and "female" subscribers. Among all these Divvy bike trips, "male subscribers" (i.e., the blue area) finish about 50% of them, and "female subscribers" (i.e., the red area) have taken 16% of the trips.

In addition, we also count the trips finished by people belonging to 3 different age groups, which include young people: age$<$30; mid-aged people: 30$\leq$age$<$50; and senior people: age$\geq$50, which are denoted by the red/blue color of different saturations in Figure 2. From the result, we observe that among the 50% bike trips finished by the "male subscribers", the ratio of trips taken by the young, mid-aged and senior people account for 17%, 27% and 7% respectively. Meanwhile, the trip finished by the female subscribers belonging to these 3 groups are 7%, 7%, and 2% respectively. Therefore, the Divvy bike is preferred and frequently used by the young and mid-aged people, who together finish about 58% of the total trip.

In summary, based on the analysis results, we can partition the users into several categories (e.g., "customers" vs "subscribers", "male" vs "female", young vs mid-aged vs senior). In the following sections, we will study the temporal and spatial travel patterns of different categories of users in detail.

C. Temporal Travel Patterns

1) Bike Usage in One Year: The Divvy bicycle-sharing system provides bike sharing services throughout the whole year. To have a look at the bike usage within a year in the dataset, we count up the trip records on each day in 2014 taken by "customers", "male subscribers" and "female subscribers" respectively, whose results are available in Figure 3.

From Figure 3, we observe that people use the Divvy bike everyday, but the majority of the trips concentrate within the months ranging from April to November, and the number of trips taken during the winter seasons is quite limited. Such a phenomenon can be correlated to the weather in Chicago and, to support such a statement, we also check

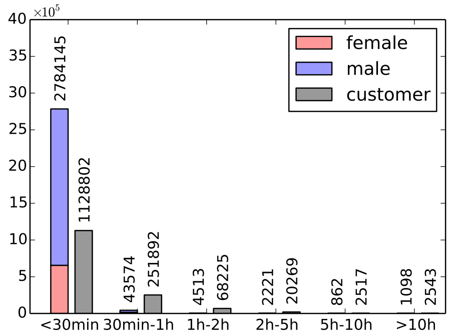


Fig. 4. Trip Duration (X axis: trip time duration, Y axis: # trips).

the historical weather within the Chicago area in 2014 from Weather Underground[5]. According to the historical weather data in Chicago, the average temperatures during January and February of 2014 were below $20°F$, and the average temperature in November and December of 2014 were below $40°F$ respectively. Meanwhile, over 20 days snowed in January 2014, and the numbers of snowing days during the February, November and December were all larger than 10. In this kind extreme weather conditions, traveling by bike is almost infeasible. Meanwhile, as the weather gets better, Divvy bike usage increases steadily.

Besides the weather reasons, some other factors can also influence the Divvy bike usages, like various events celebrated in Chicago. For instance, from Figure 3, we observe that people's bike riding activities reach the peak on July 19-20, 2014 (Saturday and Sunday) in Chicago. According to Chicago event schedule[6], at the same time, various events were taken place at Chicago including the "Pitchfork Music Festival" (tens of thousands of music fans are involved and gathered together), "Taste of River North", "Chicago Craft Beer festival", etc. To attend these celebration festivals, Divvy bikes with no worries about the parking issues are the ideal travel options for people. Viewed in this perspective, the Divvy bike riding activities are also correlated with the offline events.

2) Bike Trip Time Length Distribution: In addition, to study the bike usages in the real-world, we calculate the average trip time duration in the whole dataset to be 17.76 minutes. To study the detailed distributions of trip time length, in Figure 4, we partition the trip length into 6 bins: $\{<30$ minutes, 30 minute-1 hour, 1 hour-2 hours, 2 hours-5 hours, 5 hours-10 hours, >10 hours$\}$ and count the number of trips belonging each time bin. From Figure 4, we can observe that the number of trips which are shorter than 30 minutes ridden by "subscribers" and "customers" are $2,784,145$ and $1,128,802$, which together accounts for 90.77% of the total bike trips. In other words, the majority of users will return the bike within the free-ride time (i.e., 30 minutes) and don't want to pay the over-time charges. However, about $397,714$ bike trips are still longer than 30 minutes, the majority of which are taken by "customers". The number of over-time trips ridden

[5]http://www.wunderground.com

[6]http://www.choosechicago.com/articles/view/CHICAGO-EVENTS-FESTIVALS-2014-CALENDAR-HIGHLIGHTS/1243/

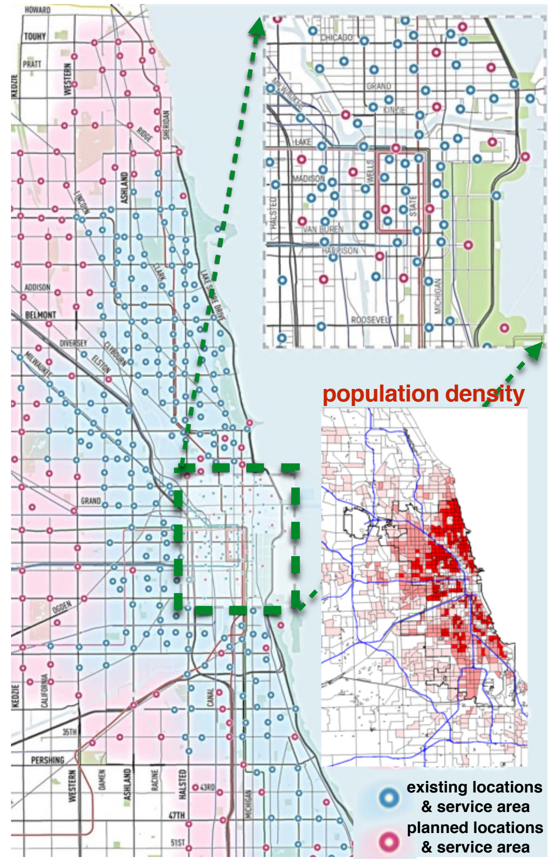


Fig. 5. Divvy station distribution and population distribution at Chicago [1].

by "customers" is $345,446$, and accounts for 86.85% of the total over-time trips.

D. Spatial Travel Patterns

1) Spatial Distribution of Bike Stations: In Figure 5, we show the distribution of the Divvy stations at the Chicago city, where the blue area and blue dots are the existing service region and the existing Divvy station locations. Due to the vast travel demands from the public, Divvy is expanding its service region to broader areas by adding new stations to both new and existing service regions, i.e., the red area and the red dots. By comparing the number of stations in the existing and planned service regions (i.e., the blue and red areas), we observe the station distribution is denser in the blue region, which also corresponds to the densely populated area at the Chicago city. The most prosperous area at Chicago should be the Loop area, which is also the region that the Divvy bike was initially launched at. To have a clear view about the stations available at that area, in Figure 5, we also zoom in the area (i.e., marked in the green dashed square), from which we can observe divvy stations within the Loop region is extremely dense and many new stations are to be added.

2) Bike Trip Geo-Distance Distribution: For bike trips taken by different categories of people among the stations, their trip distance can vary a lot. In Figure 6, we show the distributions about the distance (in kilometers (KM)) of trips taken by "customers", "male subscribers" and "female subscribers" respectively. Here, *Manhattan Distance* [24] is used as the distance measure, as the roads in Chicago are

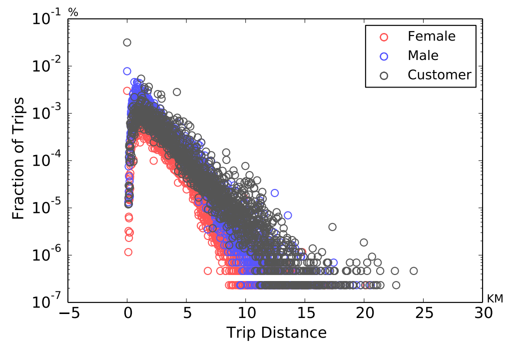


Fig. 6. Trip Geo-Distance (X axis: geo-distance (unit: KM), Y axis: fraction of trips).

very similar to those in Manhattan. We observe that the distribution curves don't follow the power law distribution [5] exactly, where the majority of trips are within distance about $0.5 - 5$KM, while those shorter than 0.5KM or longer than 5KM only account for a very small proportion. What's more, based on the historical trip data, we calculate the average distance of trips ridden by "customers" and "subscribers" (both male and female subscribed users) to be 2.12KM and 1.91KM respectively. In other words, trips finished by the customers are slightly longer.

3) Frequently Travelled Station Pairs: Generally, different Divvy bike trips are for different purposes, and the purpose can be captured more clearly by considering the origin and destination stations at the same time. For example, if the bike trip departs from residential region and the destination is a campus, the rider is likely to be a student and uses Divvy bike to commute from home to schools; while if the trip origin and destination stations are both attraction sites, then the rider mainly uses the Divvy bike for sightseeing. Motivated by this, we show the top 5 frequently traveled station pairs of "male subscribers", "female subscribers" and "customers" in Figure 7, where the origin and destination stations are listed and marked on the map.

From Figure 7, we observe that the top ranked Divvy station pair for "male subscribers" is "Station 283 $\rightarrow$ Station 174", where station 283 is at the Chicago loop region (i.e., the Chicago city center area full of office buildings) and station 174 is just next to the "Ogilvie Transportation Center". Therefore, the divvy trip for male users from station 283 and station 174 can be for catching up transportation vehicles from their workplaces.

Meanwhile, the top ranked station pair for "female subscribers" is "Station 284 $\rightarrow$ Station 255", where station 284 is next to "The Art Institute of Chicago" and station 255 is next to various spots, e.g., "The Field Museum", "Chicago Shedd Aquarium" and "Chicago Adler Planetarium". In addition, between station 284 and station 255, there exist an exercise trail for jogging and bike-riding along the Lake Michigan coast, and Chicago people like to go there for relax a lot. Therefore, the divvy trip for female users from station 284 and station 255 can be for either museum visiting or personal exercises.

For the "customers", we observe that the top 5 Divvy trips

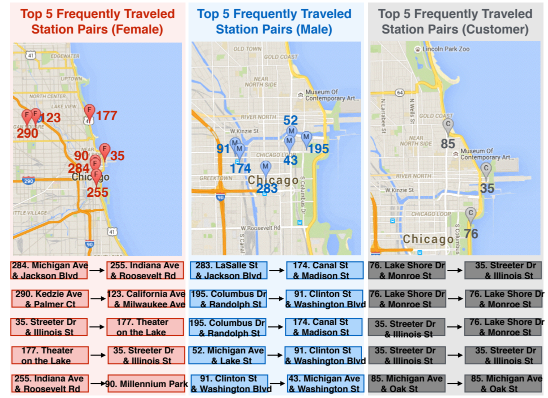


Fig. 7. Top 5 most frequently traveled stations.

for them are actually among only 3 stations, which are "Station 35" (a station next to Chicago Navy Pier), "Station 76" (a station next to Millennium Park) and "Station 85" (a station next to the Oak Street Beach). Actually, these 3 stations are all close to attraction spots and are very popular for tourists. In the trip station pair list, we notice that customers will depart and arrive at the same Divvy stations (e.g., Station 76 $\rightarrow$ Station 76, and Station 35 $\rightarrow$ Station 35), which means they borrow the bike from a station to wander around the nearby places and return the bike back to the same station. However, such observations (i.e., borrowing and returning bikes at common stations) are not common for the subscribed users.

III. Problem Formulation

Based on the previous introduction about the Divvy BSS and the analysis results, in this section, we will introduce several important terminologies used in this paper, and give the formal definition of the RoutePlanning problem.

A. Terminology Definition

In BSSs, bike stations are distributed in the service region, and we can represent the stations as set $\mathcal{S} = \{s_1, s_2, \cdots, s_n\}$, where n is the station number. The problem studied in this paper is the trip route planning based on the BSSs (i.e., stations in $\mathcal{S}$), and trip from the origin and destination can be represented with the *BSSs based trip* concept as follows:

Definition 1 (BSSs based Trip): Given the trip origin o and destination d, the whole trip route from o to d based on the BSSs actually consists of three sub-trips: (1) trip from the origin o to the start station $s_i \in \mathcal{S}$ by walk, (2) the bike trip from the start station s_i to the end station $s_j \in \mathcal{S}$ via bike, and (3) the trip from end station s_j to the destination d by walk.

Depending on the distance between the trip origin o and destination d, the bike sub-trip in the above definition can involve either one single or several trip segments by bike.

Definition 2 (Bike Sub-Trip): A bike trip t can be formally represented as a sequence of consequential bike trip segments $t = s_i \rightarrow s_k \cdots \rightarrow s_j$, where s_i and s_j are the start and end bike stations respectively, while other stations involved in t are the intermediate bike stations.

To simplify the representations, we will misuse the set theory notations that $s_k \in t$ and $s_i \rightarrow s_k \in t$ to denote station

s_k and segment $s_i \rightarrow s_k$ is one part t in this paper. In addition, via the bike sub-trip t, the whole *travel trip* from origin o to destination d can be represented as sequence $o \rightarrow t \rightarrow d$, where $o \rightarrow t = o \rightarrow s_i$ and $t \rightarrow d = s_j \rightarrow d$.

B. *Problem Formulation*

Based on the definitions introduced above, we can formally define the ROUTEPLANNING problem as follows:

Problem Definitions (ROUTEPLANNING): Given the trip origin o and destination d on the map, the ROUTEPLANNING problem aims at identifying the optimal trip route from o to d based on the BSSs, i.e., bikes at stations $\mathcal{S}$. Let $r : o \rightarrow t \rightarrow d$ be a potential trip route, which connects o and d via the bike trip t. The costs introduced by r can be quantified as function cost(r) (whose concrete representation is available in the following section). The optimal trip route to be identified in ROUTEPLANNING can be represented as

$$r^* = \arg\min_{r \in \mathcal{R}} \text{cost}(r)$$
$$s.t. \quad \text{time-constraint}(s_i \rightarrow s_j), \forall s_i \rightarrow s_j \in t,$$

where $\mathcal{R}$ represents the set of all potential trip routes from o to d. Time constraints on trip segments $s_i \rightarrow s_j \in t$ are added to ensure their time costs are within the reasonable range. For simplicity reasons, we will not distinguish different group of users when addressing the ROUTEPLANNING problem in this paper. In the following section, we will introduce the framework STOP to address the ROUTEPLANNING problem.

IV. PROPOSED METHODS

In this section, we will talk about the framework STOP in detail. The bike trip cost measure will be introduced in Section IV-A, and after that in Section IV-B, we will introduce the *bike route network* concept and propose to prune "illegal" trip route segments from the network to maintain the time constraints. Finally, in Section IV-C, we will talk about the optimal trip route extraction from the pruned route graph that introduces the minimum costs.

A. *Trip Route Cost*

The trip route cost can be measured based on various types of information about the route, including the distance, time and number of bike changes in the trip.

Time Cost of Trips

Let $r : o \rightarrow t \rightarrow d$ be a trip route from the origin to the destination on the map, where $t = s_i \rightarrow s_k \rightarrow \cdots \rightarrow s_j$ denote the bike sub-trip. The time cost of the trip route r can be divided into three parts: (1) time taken to go to start station s_i from o, (2) time cost of bike sub-trip from s_i to s_j, and (3) time cost of heading to destination d from s_j , which can be represented as

$$\text{cost}_t(r) = \text{time}(o \rightarrow t) + \text{time}(t) + \text{time}(t \rightarrow d).$$

Meanwhile, considering that the trip route r consists of the sub-trips both by walk and by bike, in the real scenarios, it takes different amount of efforts for people to walk and

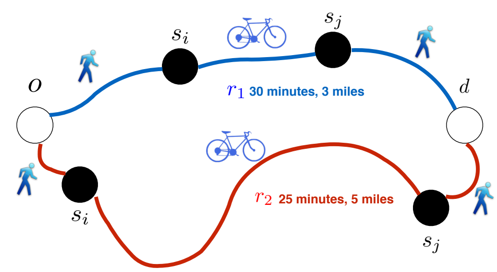


Fig. 8. An example about trip route time and distances.

ride bikes for the same time length. As pointed out in [6], the energy needed for people to walk for 10 minutes would be much less than that needed for them to ride bikes for 10 minutes. Therefore, the time costs introduced by different sub-trips should be assigned with different weights:

$$\text{cost}_t(r) = \alpha \cdot \text{time}(o \rightarrow t) + \text{time}(t) + \alpha \cdot \text{time}(t \rightarrow d),$$

where parameter $\alpha \in [0, 1]$ represents the weight of the time cost of sub-trips by walk.

Distance Cost of Trips

Besides the time consumption, the whole trip distance is another important factor that should be considered in define the trip route cost function. Trip route distance has some correlations with the time cost, but they are totally different measures. As shown in Figure 8, for two trip routes r_1 and r_2 connecting the trip origin o and destination d, the distance of routes r_1 is much shorter than r_2 but it actually takes longer time. Meanwhile, as proposed in [6], it consumes less energy for people to ride bikes for 10 miles than by walk at a regular speed. Therefore, we propose to define the weighted distance based cost function by weighting the bike sub-trip distance with parameter $\alpha \in [0, 1]$:

$$\text{cost}_d(r) = \text{dist.}(o \rightarrow t) + \alpha \cdot \text{dist.}(t) + \text{dist.}(t \rightarrow d),$$

Bike Change Cost of Trips

In addition to the time and distance based cost measures, the number of bike changes is another unique factor in measuring the quality of BBSs-based trip routes. Routes which are short in the geographical distance and time but need the customers to change bikes too often may not be good options for the customers. Viewed in this perspective, we propose to define the trip route cost based on the number of needed bike changes in route r to be:

$$\text{cost}_c(r) = \text{length}(t) - 1 = |t| - 1.$$

Joint Commute Trip Cost Function

By combining the above cost measures together, we can define the joint cost function for trip route r as follows

$$\text{cost}(r) = \beta \cdot \text{cost}_t(r) + \theta \cdot \text{cost}_d(r) + (1 - \beta - \theta) \cdot \text{cost}_c(r),$$

where the linear combination parameters β and θ denote the weights of the time and distance costs respectively.

B. Bike Route Network and Pre-Pruning

Generally, in the real-world, customers can ride the bikes between any pairs of stations in the system service region. However, some of the trip route segments can be either "illegal" (violating the time constraint) or not regular routes (e.g., through dangerous blocks) that few people will take them actually. In this paper, we will introduce the concept of *bike route network*, and propose to pre-prune the network by removing the infeasible routes.

Bike Route Network

Based on the bike stations $\mathcal{S}$, all the potential pairwise direct bike trip routes among the stations can be represented as set $\mathcal{L} = \mathcal{S} \times \mathcal{S} - \{(s_i, s_i)\}_{s_i \in \mathcal{S}}$, where the trips starting and ending at the same station are not counted as they contribute nothing for the travel from origin o to destination d. Here, the concept of *direct bike trip routes* denotes the unit of bike sub-trips, and all the bike sub-trips will consist of the direct bike trip routes in $\mathcal{L}$. The bike stations $\mathcal{S}$ together with the direct trip routes $\mathcal{L}$ will form the following bike route network.

Definition 4 (Bike Route Network): The *bike route network* consisting of stations $\mathcal{S}$ can be represented as $G = (\mathcal{S}, \mathcal{L})$, where $\mathcal{L} = \mathcal{S} \times \mathcal{S} - \{(s_i, s_i)\}_{s_i \in \mathcal{S}}$.

According to the definition, the *bike route network* is actually a clique of size $|\mathcal{S}|$. For each direct bike route in graph G, we can represent its time cost and geographical distance as $\text{time}(t)$ and $\text{dist.}(t)$ respectively, which can be obtained with both Google APIs[7] and the historical bike trip record data.

Pruned Bike Route Graph

However, according to the definition of the ROUTEPLANNING problem, there exists a time constraint for each trip route segment in the bike sub-trip. Many of the *direct bike trip routes* in $\mathcal{L}$ violating the constraint and will not be selected in the final route r. The bike trip time cost can be estimated with either the historical bike trips between the origin and destination stations, or the real-time GPS tracking system equipped with the bicycles together with the current traffic conditions. In this paper, we make use of the historical bike trip time length information to estimate the potential time cost of each trip origin and destination pairs. By analyzing the customers' historical bike riding records, their trip length are usually within a reasonable range, which can be represented as $[\mu_t - \sigma_t, \mu_t + \sigma_t]$, here μ_t and σ_t denote the average trip time length and the standard deviation of the trip length distribution. According to the bike trip datasets (to be introduced in Section V), we can obtain the mean trip duration as well as the standard deviation to be $\mu_t = 761.56$ and $\sigma_t = 429.73$. The *direct bike trip routes* in $\mathcal{L}$ that are not within the range will hardly appear in customers' trip routes actually.

In this paper, we propose to prune the direct bike trip routes that are either greater than 30 minutes or not within the normal trip length range from the route network, which can not only maintain the time constraints but also greatly shrink the search space of the potential bike routes. Formally, we can represent the obtained the *pruned route graph* $\overline{G} = (\mathcal{S}, \overline{\mathcal{L}})$, where $\overline{\mathcal{L}} = \{t | t \in \mathcal{L} \wedge \mu_t - \sigma_t \leq \text{time}(t) \leq \min\{\mu_t + \sigma_t, 60 \times 30\}\}$. All the *direct bike route* available in $\overline{G}$ are the "legal" trip route segments.

[7]https://developers.google.com/maps/

C. Framework STOP

Given the trip origin and destination o and d, the optimal bike trip t connecting o and d will consist of stations and direct commute routes from the pruned commute graph $\overline{G}$ only. In this paper, we will address the problem with the minimum cost network flow model.

1) Network Flow Graph Construction: Given the pruned bike route graph $\overline{G} = (\mathcal{S}, \overline{\mathcal{L}})$, trip origin o and trip destination d, we propose to construct the network flow graph node set $\mathcal{V} = \mathcal{S} \cup \{o, d\}$. What's more, besides the existing bike trip routes connecting bike stations in $\overline{G}$, we further introduce two groups of links connecting the trip origin o and stations in $\mathcal{S}$ and stations in $\mathcal{S}$ to the destination node d, as well as a back propagation link (d, o). All these links can be used to construct the network flow graph link set $\mathcal{E} = \overline{\mathcal{L}} \cup (\{o\} \times \mathcal{S}) \cup (\mathcal{S} \times \{d\}) \cup \{(d, o)\}$. In addition, each link $l \in \mathcal{E}$ is associated with a specific weight $\text{cost}(l) = \beta \cdot \text{cost}_t(l) + \theta \cdot \text{cost}_d(l) + (1 - \beta - \theta) \cdot \text{cost}_c(l)$. As introduced in Section IV-A, depending on the link type, the concrete representation of $\text{cost}(l)$ will be

$$\begin{cases} \beta \cdot (\alpha \cdot \text{time}(l)) + \theta \cdot \text{dist.}(l) + 0, \text{ if } l \in \mathcal{E} \setminus (\overline{\mathcal{L}} \cup \{d, o\}); \\ \beta \cdot \text{time}(l) + \theta \cdot (\alpha \cdot \text{dist.}(l)) + (1 - \beta - \theta) \cdot 1, \text{ if } l \in \overline{\mathcal{L}}; \\ 0, \text{ otherwise.} \end{cases}$$

Based on the above remarks, we can define the network flow graph as follows:

Definition 5 (Network Flow Graph): Based on the pruned bike route graph $\overline{G} = (\mathcal{S}, \overline{\mathcal{L}})$, trip origin o and trip destination d, the network flow graph can be represented as $H = (\mathcal{V}, \mathcal{E}, \text{cost})$, where $\text{cost}: \mathcal{E} \to \mathbb{R}$ maps the links in $\mathcal{E}$ to the corresponding trip cost.

The optimal trip route can be effectively extracted from H with the minimum cost network flow model.

2) Optimal Commute Trip Route Extraction: In the network flow model, each link in the constructed network flow graph is associated with certain amount of network flow, which is constrained by specific bounds.

Bound Constraint of Network Flow In the network flow model, the amount of network flow associated with links in $\mathcal{E}$ can be denoted as $\{x_{u,v}\}_{(u,v) \in \mathcal{E}}$. Meanwhile, the amount of network flow of each edge needs to meet certain flow constraint, which can be generally represented as

$$\underline{f}_{u,v} \leq x_{u,v} \leq \overline{f}_{u,v}, \forall (u, v) \in \mathcal{E},$$

where $\underline{f}_{u,v}$ and $\overline{f}_{u,v}$ represent the upper bound and lower bound of edge (u, v) respectively. For different types of edges, we will introduce the specific constraints for them as follows.

More specifically, to extract the optimal bike trip route based on the network flow model, we only allow the network flow of amount 1 to propagate within the *network flow graph*. To achieve such a objective, we set both the upper and lower bounds of the back-propagation edge (d, o) to be exactly 1, i.e.,

$$x_{d,o} = 1.$$

Meanwhile, for the other links in the network flow graph, the network flow of at most amount 1 can propagate across each link. We propose to represent the lower and upper bounds of these links to be 0 and 1 respectively and allow $x_{u,v}$ to take integer values only:

$$x_{u,v} \in \{0, 1\}, \forall (u, v) \in \mathcal{E} \setminus \{(d, o)\},$$

where $x_{u,v} = 1$ denotes that the link is selected as one segment of the optimal trip route.

Mass Balance Constraint

Besides the flow constraint of each link, in the network flow model, for each node (e.g., u) in the graph, the amount of network flow going into u should be equal to that going out from u. In other words, the network flow through the network should meet the *mass balance constraint*:

$$\sum_{v \in \mathcal{V}, (v,u) \in \mathcal{E}} x_{v,u} = \sum_{w \in \mathcal{V}, (u,w) \in \mathcal{E}} x_{u,w}.$$

Minimum Cost Network Flow

For all potential trip routes from origin o to destination d, we aim at discovering the optimal one that introduces the minimum cost. The cost introduced by all the potential links in the network flow graph H can be represented as

$$\text{cost}(H) = \sum_{(u,v) \in \mathcal{E}} x_{u,v} \cdot \text{cost}(u, v),$$

where only the costs of the selected links will be counted.

Furthermore, the optimal trip route that leads to the minimum trip cost can be obtained by addressing the following optimization objective function

$$\begin{aligned} & \min \sum_{(u,v) \in \mathcal{E}} x_{u,v} \cdot \text{cost}(u, v) \\ s.t. \quad & \sum_{v \in \mathcal{V}, (v,u) \in \mathcal{E}} x_{v,u} = \sum_{w \in \mathcal{V}, (u,w) \in \mathcal{E}} x_{u,w}, \\ & x_{d,o} = 1, \\ & x_{u,v} \in \{0, 1\}, \forall (u, v) \in \mathcal{E}. \end{aligned}$$

The objective equation is actually an integer programming problem, which can be resolved with some open source toolkits, e.g., PuLP[8] and SciPy[9]. We will not talk about how to solve the problem here due to the limited space. The links corresponding to the variables which are assigned with value 1 (except the back propagation link (d, o)) in the results will be selected to form the optimal trip route result.

[8]https://pypi.python.org/pypi/PuLP

[9]http://docs.scipy.org

V. EXPERIMENTS

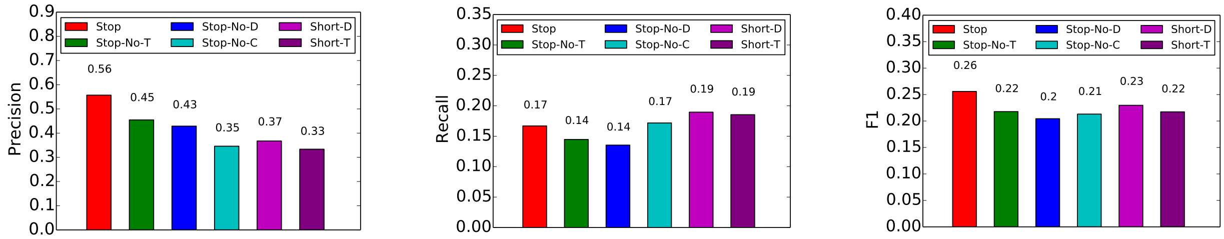
To test the effectiveness of the proposed bike trip route planning framework STOP, in this paper, we conduct extensive experiments on real-world bicycle-sharing system dataset (i.e., the Divvy BSS dataset introduced in Section II). In this section, we will first describe the experiment settings in detail. After that, we will provide the experiment results with detailed explanations and give some case studies.

A. Experiment Settings

1) Experiment Setups: From the historical bike trip records, we extract a set of consequential trip segments taken by the same people, which are used as the trip instances for evaluation. We can represent the trip instance set as $\mathcal{T}$, where $t \in \mathcal{T}$ denotes a specific trip route consisting of several consequential trip segments. The geographical coordinates of the start and end stations of each trip route can be used as the input of framework STOP. By addressing the objective equation in the network flow model, we can obtain the set of trip segments selected from the pruned bike route network, which will be outputted as the optimal trip route of the input coordinate pairs. In the experiments, the parameter α is set as 0.8.

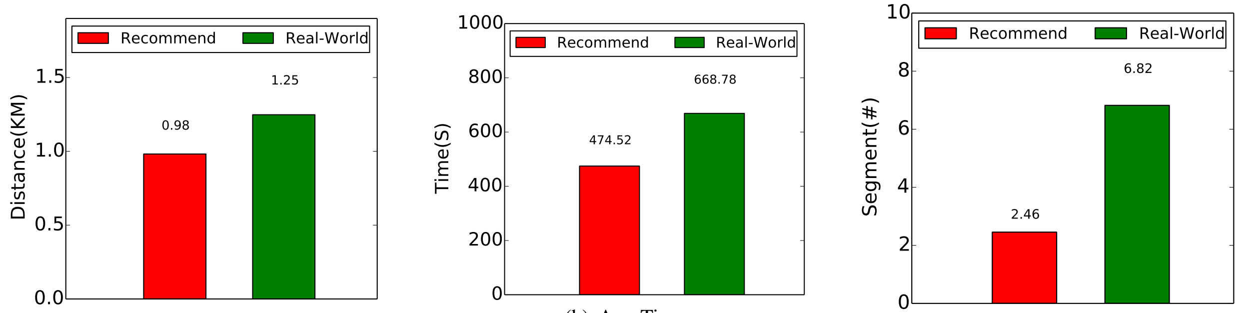
2) Comparison Methods: Different from traditional trip route planning methods, the bicycle-sharing system introduces much more constraints about the potential routes to be identified, e.g., the stations are fixed, the free-ride time limits. To demonstrate the advantages of STOP, we compare STOP with many baseline methods, and the comparison methods used in the experiments include:

- STOP: The framework STOP is the bike trip route planning method introduced in this paper. STOP defines the trip cost based on *time*, *distance* and *bike changes* information (where weights $\beta = \theta = \frac{1}{3}$), and extract the optimal trip route with the network flow model.
- STOP-NO-T, STOP-NO-D, STOP-NO-C: To demonstrate the motivation of involving the *time*, *distance* and *bike changes* in the trip cost definition, several variants of STOP are introduced in the experiments, which include (1) STOP-NO-T which neglects the time information in cost definition (i.e., parameters $\beta = 0$, $\theta = 0.5$), (2) STOP-NO-D which doesn't consider the distance information (i.e., parameters $\beta = 0.5$, $\theta = 0$), and (3) STOP-NO-C which fails to consider the bike changes (i.e., parameters $\beta = 0.5$, $\theta = 0.5$).
- SHORT-D: In addition, the trip route direction problem can also be transformed into the traditional shortest path problem. We also apply the traditional Dijkstra's algorithm to find the shortest paths from the route network [19] where the edge weights are defined as the geographical distance.
- SHORT-T: Similar to method SHORT-D, we also address ROUTEPLANNING problem by finding the shortest path from the route network where the edge weights are the time cost [19].



(a) Avg Precision (b) Avg Recall (c) Avg F1

Fig. 9. Bike Trip Route Planning Result of Comparison Methods.



(a) Avg Distance (b) Avg Time (c) Avg Bike Change #

Fig. 10. Comparison of Bike Trip Route Planned by STOP with Real-world Trip Route without Planning.

3) Evaluation Metrics: For each trip origin and destination pairs, we compare the planned trip route segments connecting the origin and destination of the comparison methods with the ground truth segments and calculate its Precision, Recall and F1 scores of the results. The average Precision, Recall and F1 scores of all the trip origin and destination pairs obtained by the different comparison methods are used as the evaluation metrics.

In addition, compared with the ground truth, trip segments outputted by the methods can be of different length (i.e., number of segments) and different quality (i.e., time and distance). In this paper, we also evaluate the performance of STOP by calculating the average trip route distance, average trip time and average number of segments involved in the planned trip routes, which is also compared with real-world average distance, time and segment number as well.

B. Experiment Result

The experiment results are available in Figures 9-10.

In the plots of Figure 9, we show the average Precision, Recall and F1 scores achieved by different comparison methods. From the results, we can observe STOP utilizing all the information can outperform other baseline methods with significant advantages when the evaluation metrics are average Precision, and average F1. For instance, the average Precision achieved by STOP is 0.56, which is 24% higher than the average Precision obtained by STOP-NO-T, 30% higher than that achieved by STOP-NO-D, 60% higher than that obtained by STOP-NO-C, 51% higher than that achieved by SHORT-D, and 69% higher than that of SHORT-T. For the average Recall measure, method STOP also has comparable performance to SHORT-D and SHORT-T, which can obtained the average Recall scores of 0.17, 0.19 and 0.19 respectively.

In addition, by comparing the trip routes planned by STOP with the ground truth trip route segments, the planned route segments are much better in trip distance, trip time, and number of bike changes. According to the results given in Figure 10, the average Distance of the trip routes planned by STOP is 0.98 KM(kilometer), which is 21.6% shorter than the average distance of the trip routes without planning; the average Time cost of the planned trip route takes about 474.5 seconds, which account for only about 70% of the trip route time cost without planning; the average Segment number is only 2.46, which is only 36% of the trip route segment without planning.

In sum, for the planned trip routes by STOP, a large number of the segments are already taken by the customers in the real-world. Meanwhile, from the whole trips perspective, the planned trips are generally much better (in both bike changes, distances, and time costs) than the real-world trip routes without pre-planning.

C. Case Study

In addition, we also do a case study to illustrate the advantages of framework STOP, and the result is shown in Figure 11. Given the trip origin at the "South Loop" area in Chicago and the destination at the "North Avenue Beach", we show the planned route by STOP and the real trip route without pre-planning. From the plot, we can observe that the planned route is shorter than the one without planning. In addition, the planned trip route requires two bike changes during the trip, while the other one changes the bike 4 times in all.

VI. RELATED WORKS

Traffic route planning and direction is a traditional research problem, which has been studied for many years, and dozens of research works have been published on this topic [12], [9], [18], [14], [23], [4], [13], [3], [11]. To generate mission-adaptable routes in an accurate and efficient manner, Szczerba et al. [21] propose a novel route planning approach computed

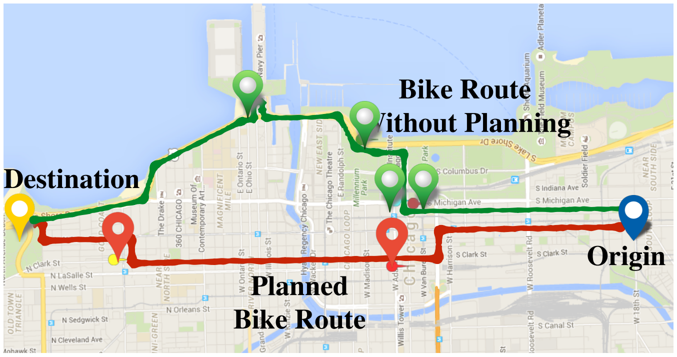


Fig. 11. A case study about the planned bike route vs bike trip route without planning.

in real-time, which can take into account various mission constraints, like minimum route leg length, etc. Simmons et al. observe that driving is a routine activity where people drive to the same destinations using the same routes. Based on such an intuition, Simmons et al. [20] propose to learn individuals' driving patterns from the historical records and try to predict the drive route and destination in advance.

Meanwhile, most popular route planning systems (Windows Live Local, Google Maps, etc.) generate driving directions using a static library of roads and road attributes. They ignore the preferences of the drivers they serve. To overcome such a serious shortcoming, Letchner et al. [15] presents a set of methods for including driver preferences and time-variant traffic condition estimates in route planning. Time complexity issue is a very serious problem for most route planning algorithms. To speed up the route planning algorithm, a fast route planning framework is introduced by Sanders et al. in [19]. More information about existing engineering trip route planning algorithms is available in [7], [2].

Bicycle-sharing has received increasing attention in recent years. A detailed analysis about the Chicago Divvy BSS is available in [26]. DeMaio gives a complete introduction about the history, impacts, models of provision, and future of bicycle-sharing systems in [8]. Midgley provides a complete overview work about the bicycle-sharing schemes, management, policies, and challenges as well as opportunities in [17]. A large number of other review and case-study works on bicycle-sharing systems have appeared so far [16], [27], [22], [10], which study the bicycle-sharing systems from different aspects and directions.

VII. Conclusion

In this paper, we have studied the RoutePlanning problem for BSSs with various types of constraints, and a new route planning framework Stop is introduced in this paper. Stop maps the RoutePlanning problem into a minimum cost network flow problem, where the costs of route segments are defined based on their corresponding time, distance and bike change information. We have conducted experiments on a real-world BBSs dataset, Divvy, and the experiment results have demonstrated the outstanding performance of Stop compared with other route planning methods.

VIII. Acknowledgement

This work is supported in part by NSF through grants IIS-1526499 and NVIDIA Corporation with the donation of the Titan X GPU used for this research.

References

[1] Divvy: Chicago's newest transit system. http://divvybikes.tumblr.com/post/95362930905/expansion-map-2015. [Online; accessed 24-November-2015].
[2] H. Bast, D. Delling, A. Goldberg, M. Müller-Hannemann, T. Pajor, P. Sanders, D. Wagner, and R. Werneck. Route planning in transportation networks. 2014.
[3] M. Baum, J. Dibbelt, A. Gemsa, D. Wagner, and T. Zündorf. Shortest feasible paths with charging stops for battery electric vehicles. In *SIGSPATIAL*, 2015.
[4] M. Baum, J. Dibbelt, T. Pajor, and D. Wagner. Energy-optimal routes for electric vehicles. In *SIGSPATIAL*, 2013.
[5] A. Clauset, C. Shalizi, and M. Newman. Power-law distributions in empirical data. *SIAM Review*, 2009.
[6] M. Cohen and M. Heberger. Driving vs. walking: Cows, climate change, and choice. *Pacific Institute*, 2008.
[7] D. Delling, P. Sanders, D. Schultes, and D. Wagner. Algorithmics of large and complex networks. chapter Engineering Route Planning Algorithms. 2009.
[8] P. DeMaio. Bike-sharing: History, impacts, models of provision, and future. *Journal of Public Transportation*, 2009.
[9] J. Eisner, S. Funke, and S. Storandt. Optimal route planning for electric vehicles in large networks. In *AAAI*, 2011.
[10] C. Etienne and O. Latifa. Model-based count series clustering for bike sharing system usage mining: A case study with the vÉlib system of paris. *ACM TIST*, 2014.
[11] S. Funke and S. Storandt. Personalized route planning in road networks. In *SIGSPATIAL*, 2015.
[12] J. Gilmore and A. Semeco. Terrain navigation through knowledge-based route planning. In *IJCAI*, 1985.
[13] M. Goodrich and P. Pszona. Two-phase bicriterion search for finding fast and efficient electric vehicle routes. In *SIGSPATIAL*, 2014.
[14] K. Haigh, J. Shewchuk, and M. Veloso. Route Planning and Learning from Execution. In *Working Notes of the AAAI Fall Symposium*, 1994.
[15] J. Letchner, J. Krumm, and E. Horvitz. Trip router with individualized preferences (trip): Incorporating personalization into route planning. In *IAAI*, 2006.
[16] P. Midgley. The role of smart bike-sharing systems in urban mobility. *Land Transport Authority JOURNEY*, 2009.
[17] P. Midgley. Bicycle-sharing schemes: Enhancing sustainable mobility in urban areas. *United Nations Department of Economic and Social Affairs*, 2011.
[18] E. Nikolova and D. Karger. Route planning under uncertainty: The canadian traveller problem. In *AAAI*, 2008.
[19] P. Sanders and D. Schultes. Engineering fast route planning algorithms. In *WEA*, 2007.
[20] R. Simmons, B. Browning, Y. Zhang, and V. Sadekar. Learning to predict driver route and destination intent. In *ITSC*, 2006.
[21] R. Szczerba, P. Galkowski, I. Glicktein, and N. Ternullo. Robust algorithm for real-time route planning. *IEEE Transactions on Aerospace and Electronic Systems*, 2000.
[22] P. Vogel and D. Mattfeld. Strategic and operational planning of bike-sharing systems by data mining: A case study. *Computational Logistics*, 2015.
[23] D. Wilkie, C. Baykal, and M. Lin. Participatory route planning. In *SIGSPATIAL*, 2014.
[24] D. Wilson and T. Martinez. Improved heterogeneous distance functions. *Journal of Artificial Intelligence Research*, 1997.
[25] S. Winter and A. Grubacher. Modeling costs of turns in route planning. *GeoInformatica*, 2002.
[26] J. Zhang, X. Pan, M. Li, and P. Yu. Bicycle-sharing system analysis and trip prediction. In *MDM*, 2016.
[27] L. Zhang, J. Zhang, Z. Duan, and D. Bryde. Sustainable bike-sharing systems: Characteristics and commonalities across cases in urban china. *Journal of Cleaner Production*, 2015.