Data driven methods for effective micromobility parking

Ricardo Sandoval[1], Caleb Van Geffen[1], Michael Wilbur[1,4], Brandon Hall[3], Abhishek Dubey[1,4], William Barbour (corresponding author)[2,4], and Daniel B. Work[2,4]

[1]Department of Electrical Engineering and Computer Science, Vanderbilt University
[2]Department of Civil and Environmental Engineering, Vanderbilt University
[3]Department of Electrical and Computer Engineering, Miami University
[4]Institute for Software Integrated Systems, Vanderbilt University; 1025 16th Avenue South, Suite 102; Nashville, TN 37212

Abstract

In this work, we propose a data-driven method to use proven clustering algorithms for establishing shared electric scooter (SES) parking locations and assessing their anticipated utilization. We first address the problem of finding locations for a given number of parking facilities, based purely on demand, that maximize the number of trips that would likely be parked at these facilities. We then formulate an enhanced version of the SES parking facility problem in which exogenous environmental factors are considered, such as sidewalk width. Parking SESs on narrow sidewalks raises accessibility concerns for other users of this infrastructure and capturing these trips in dedicated parking facilities is a valid priority to trade off with pure demand maximization. These methods are demonstrated in two case studies, which use a large SES dataset from Nashville, Tennessee, USA. We provide empirical results on how many facilities are needed to serve demand of SESs and necessary capacity allocation of the facilities. When the methodology considers sidewalk width in facility placement, the refined parking locations can address 300% more problematic trips parked along narrow sidewalks, with only a nominal sacrifice, around 13%, in the overall number of trips served.

Keywords

1 Introduction

1.1 Motivation

The launch of shared electric scooters (SESs) within cities has added a new dimension to the transportation sector, amassing 38.5 million trips in 2018 [1]. The rapid deployment and user adoption of the dockless devices has quickly created management challenges for cities, especially related to infrastructure. Since the dockless devices can in principle be parked anywhere, users must be convinced to sacrifice some convenience for orderliness. Policy approaches such as geofencing, parking fines, parking guidelines, and signage are all common elements of the strategy [2]. A study during the San Francisco scooter pilot program found that encouraging riders to park in designated areas to decreased scooter parking citations by 88.9% and gained higher civilian approval of micromobility as a whole [3]. The determination of the best locations for these parking facilities is vital, especially when the use of parking facilities is voluntary.

Poorly parked scooters are not only a source of potential injuries, but could also promote inequity. Ill-parked scooters can become obstructions that can prevent people with disabilities from freely moving around on the sidewalks [4]. This, especially, becomes a concern when scooters are parked on sidewalks and cause noncompliance with the Americans with Disabilities Act [5] standards for sidewalk clearance. For example, Figure 1 shows the approximate width of sidewalks in Nashville, TN. Sidewalks colored orange and red (i.e., less than 7 feet wide) simply do not have adequate width to accommodate even a carefully-parked scooter, while still maintaining adequate width for accessibility.

The competition for scarce sidewalk space is demonstrating a need for dedicated parking infrastructure in key locations to reduce these concerns. Determining the locations for dedicated parking is based on trip demand and, potentially, the state of existing infrastructure such as sidewalk width. The high volume of scooter trips combined with the complex nature of the location problem make determining where to place these facilities a challenging process.
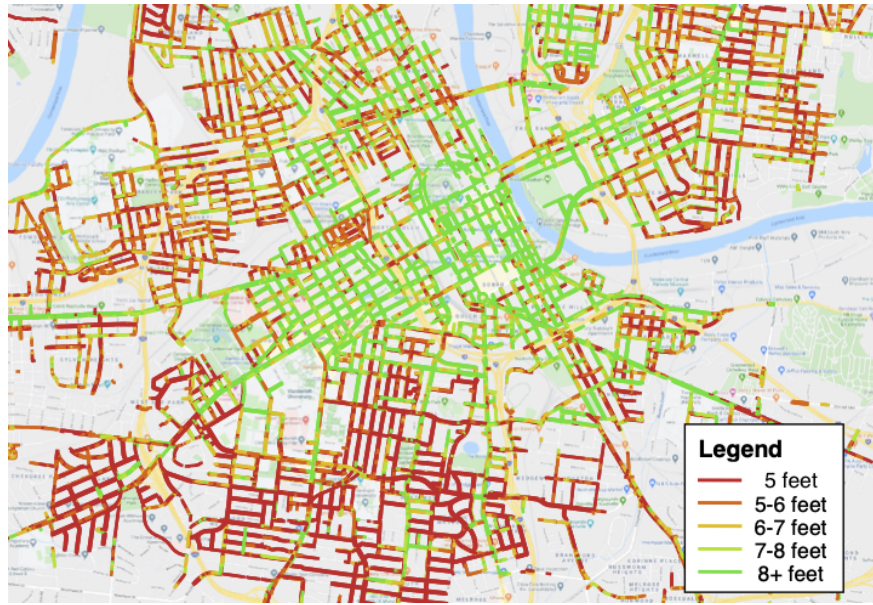
Figure 1 : Map of sidewalk widths in Nashville, TN. The color represents the width of sidewalks, where present. If scooters are not carefully parked on narrower sidewalks, they can be an obstruction to ADA compliance.

1.2   Problem Statement

The first concern of this work is identifying parking infrastructure locations for dockless e-scooters. In this context, parking areas would be marked on small areas of roads or sidewalks for e-scooters and may be instituted as encouraged practice or as mandatory. The infrastructure is desired to be effective in improving user parking behavior as broadly as possible, so therefore should be placed in the most effectual locations.

A secondary, but still vital concern for SES infrastructure is the issue of poorly parked scooters becoming obstructions along sidewalks. Scooters parked along narrow sidewalks have a larger possibility of obstructing the pathway and creating issues with ADA compliance. Therefore this work provides methods to prioritize infrastructure in areas that can mitigate this problem, while maintaining allocation of facilities to high-demand areas.

1.3 Contributions

To our knowledge, this is the first study in the United States to propose a data-driven solution for locating and prioritizing SES parking facilities in continuous space based on multiple factors, while also evaluating the efficiency and anticipated utilization of parking facilities. It also provides empirical results on how many facilities are needed to serve SES demand in a city setting and the capacity allocation of such parking facilities. In addition to demand, our study is unique in considering additional factors for facility prioritization, such as influencing characteristics of the built environment. This very well could assist cities in using their own micromobility data to place SES parking facilities in the most effective areas.

A summary of the contributions of this work are as follows:

1. We propose a data-driven solution for locating and prioritizing SES parking facilities in continuous space using straightforward clustering methods.

2. We extend our proposed parking methodology to incorporate external infrastructure for the parking placement problem through weighted clustering. As an example, we focus on sidewalk width for ADA compliance and find that a 300% increase in capture of problematic narrow-sidewalk parking can be gained with only 13% trade-off in overall trip capture.

3. We provide empirical results on how many facilities are needed to serve SES demand and the capacity allocation of these facilities. Empirical results are presented from two case studies, one focused on the Vanderbilt University area and one on Nashville, TN.

The remainder of this article is organized as follows. In Section 2, we summarize recent literature related to our research. Section 3 explains in detail the data-driven methodology for facility placement and evaluation. In Section 4, we discuss the case study dataset from Nashville, TN. Section 5 shows results from the first case study on determining efficient demand-based parking locations on Vanderbilt's campus. Section 6 contains the second case study that also considers factors related to existing infrastructure in the placement of parking facilities for the Nashville

4

metro area. In Section 7, we discuss practical application of our findings. In Section 8, the article is concluded along with an overview of possibilities for future areas of research.

2   Literature review

In this section, we present a review of relevant literature pertaining to facility location planning using optimization and data-driven methods such as clustering.

Optimization approaches are quite common in literature when determining optimal facility placements. These methods often follow the approach: maximize or minimize an objective function quantifying the desirability of a facility placement, given a set of constraints. The objectives and constraints for facility location problems vary across the available literature and include: maximizing coverage of the network [6], minimizing burden on users to find a parking facility [7], and maximizing user satisfaction [8], to name a few. Constraints may take into account a limited budget [8], availability of devices in the network [6], or the capacity of each parking location [9], for example. Researchers have also explored the reallocation of dock capacity [10] and the rebalancing of bicycles in bike sharing facilities [11] from an optimization standpoint.

A study on dockless bikes in Shanghai, China presents an optimization framework that is used for the planning of parking infrastructure utilizing the clustering algorithm DBSCAN; parking locations were geofenced parking zones with no physical boundaries [12]. Another study in Ningbo City, China, used k-means to analyze the spatiotemporal behavior with the goal of grouping stations according to similar characteristics [13].

Other studies have used clustering to determine resource allocation in bikeshare systems, grouping stations on similarities of usage patterns [8]. A micromobility study performed in Madrid, Spain used a GIS approach for docked bike share station placement [7] which minimized cost between demand points and maximized station coverage. They also evaluated practicality of the locations in regard to popular destinations.

3 Methodology

In this section we describe our approach to placement and sizing of parking facilities for shared urban mobility devices (SESs). First, we conceptually describe our methodology and present a flow chart as a visual aid. We then provide a mathematical representation of our methods. The clustering algorithms and process by which their parameters are tuned is then presented. Next, we discuss the placement of the actual facility within a cluster and how it is assessed. Finally, we explain how to incorporate external features by utilizing weighted clustering.

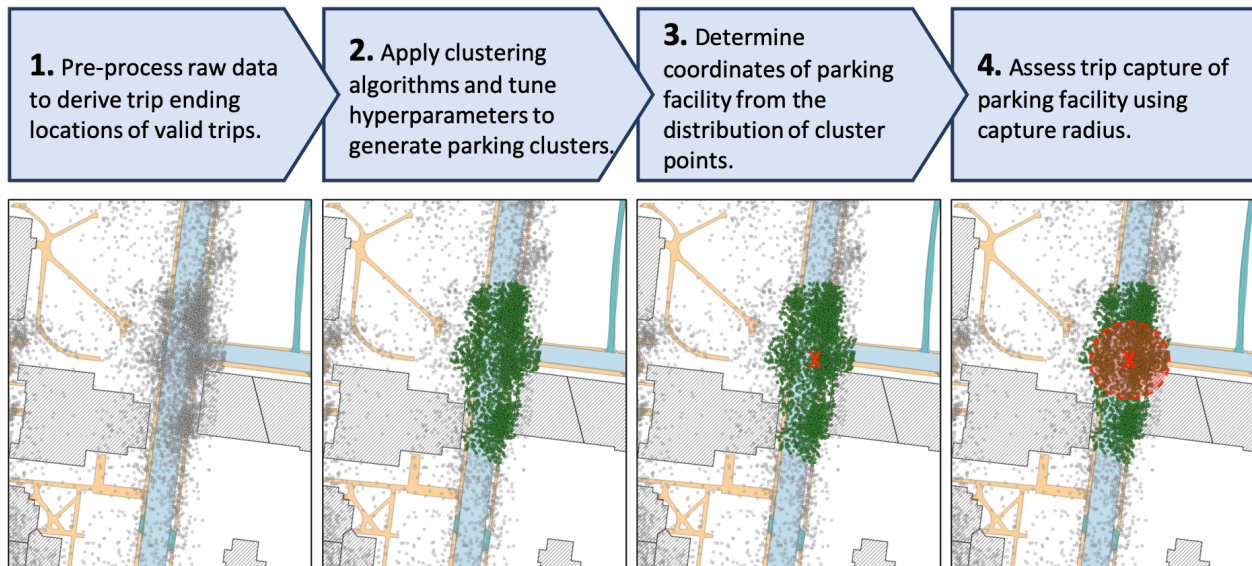## 3.1 Conceptual overview of methodology

The purpose of this analysis methodology is to find a set of point locations for micromobility parking facilities that maximize the total number of unique trips that could use them. The input is a list of trip endpoints from historical data – where each trip was parked. Clustering algorithms are used to determine high-density clusters of points where parking facilities should be located. The methodology can also take as optional input built environment factors that correspond to each trip endpoint. The output is a list of parking facilities, prioritized according to the number of trips that will potentially use the facility.

Figure 2 shows the steps of the methodology on a small area of real data in which a parking location is placed. In step 1, we collect raw scooter trip data, consisting of each trip's end locations. We preprocess data, apply temporal and geographical filters, and separate data into training and testing sets, which is discussed in Section 4. Only the training dataset is shown. In step 2, we apply clustering algorithms and perform hyperparameter tuning, explained in Section 3.3, on training data to generate clusters of trip ending points. Additional factors for facility placement may be considered within step 2 by modifying the weight of each point during clustering, which is discussed in Section 3.5. A parking facility is placed within each defined cluster of points, and is assumed to be located at a single point with a circular area of influence centered at this point. In step 3, we determine exact parking facility locations based on distributions of trip ending

6

points *within clusters*, which is discussed in Section 3.4. In step 4, we assess parking locations by counting the number of trip endpoints falling within a *capture radius*. We assume that users ending their trips within this distance from the facility will park at the facility instead of the surrounding area. A capture radius of 100 feet is used for analysis and represents a reasonable distance for a micromobility user to walk from a parking facility to their actual destination [14].

The performance of a given output – a set of facility locations – is determined by the total number of trips that fall within the capture radius from any of the parking locations. This number of trips is referred to as *trip capture*. The objective is to place a desired number of parking facilities such as to maximize trip capture. However, given that this approach does not inherently produce a global optimum, a large number of possible realizations with varying performance are possible for a desired number of clusters. This fact requires a process of hyperparameter tuning for clustering algorithms to find the highest-performing output, which we discuss in detail in Section 3.3.

We use a training dataset of trip endpoints to establish parking locations with the clustering algorithms. The performance of the parking locations is calculated and reported using a testing dataset, which does not intersect with the training dataset in order to eliminate potential bias.

1. Pre-process raw data to derive trip ending locations of valid trips.

2. Apply clustering algorithms and tune hyperparameters to generate parking clusters.

3. Determine coordinates of parking facility from the distribution of cluster points.

4. Assess trip capture of parking facility using capture radius.

Figure 2 : Overview of steps in analysis pipeline used in our methodology.

The relationship between the number of parking facilities and total trip capture works both

7

ways. A desired level of trip capture (e.g., 25% of trips) can also be used to estimate the number of parking facilities required. The number of trips captured per facility, which we refer to as *capture efficiency*, is also an informative metric for decision making. It can help quantify utility or cost-benefit of the infrastructure investment. The following summarize the performance metrics for parking locations:

- *number of trips captured*: the number of trip endpoints located a distance less than or equal to a given capture radius from the parking location coordinates.

- *percentage of trips captured*: the proportion or percentage of trip endpoints captured; applicable for a single location or totalled across a set of locations.

- *capture efficiency*: the number of trips (or mean number of trips) captured per parking location, given a set of multiple locations.

- *capacity*: the size of a parking facility in terms of the number of scooters it can accommodate.

3.2 Mathematical description of methodology

We now describe in mathematical notation the methodology of this work. This notation maps onto the conceptual overview and will be referenced later in further descriptions of the methods.

Let E be the input set of n trip endpoints, each defined by longitude (x) and latitude (y) coordinates: $E = \{(x_1, y_1), (x_2, y_2), \ldots, (x_n, y_n)\}$. The endpoints, E, may be divided into mutually exclusive training data and testing data, denoted E_{tr} and E_{ts}, respectively. Let L be a set of m parking facility locations (longitude, latitude), such that $L = \{(x_1, y_1), (x_2, y_2), \ldots, (x_m, y_m)\}$. A clustering algorithm combined with a facility placement procedure, represented loosely by the function f, produces high-demand clusters of trip endpoints and places parking facilities within each cluster. It uses training data to generate facility placement so that the testing dataset may be used to assess the performance of the facilities. The clustering algorithm is parameterized according to relevant hyperparameters p. The function f can therefore be written $f(E_{tr}, p) = L$.

Finding a set parking facilities, L, that maximizes trip capture depends on finding an ideal hyperparameter set, p, because most clustering algorithms can generate a range of possible outputs for a desired number of clusters (in this case, parking facilities). Let the trip capture for a set of parking locations, L, be based on the testing dataset of trip endpoints and be denoted by the function $c(E_{ts}, L)$. The maximum trip capture for a given number of facilities, m, can therefore be written:

$$\max_p [c(E_{ts}, f(E_{tr}, p))] \; \forall \; \{p : |f(e, p)| = m\}. \tag{1}$$

The value of p that maximizes this equation is denoted p^*. Note that in this work, p^* is determined by sampling the space of p and may not necessarily constitute the true global optimum.

3.3 Clustering algorithms

In this work, we assess the performance of three different clustering algorithms – k-means, DBSCAN, and HDBSCAN – against our objective of finding the highest performing parking facility locations. The intention of this comparison is not an exhaustive survey of possible algorithms, but rather to identify a method that has been used for these types of problems previously and select one for continued use in the case studies of this article. A factor in choosing these algorithms was their straightforward understanding and implementation.

k-means is a classical clustering algorithm [15] that has been used successfully in facility placement in other domains [16, 17, 18]. It is initialized with a specific number of clusters into which all of the training data is partitioned. DBSCAN is a density-based clustering algorithm that can discover clusters of arbitrary shape conforming to two hyperparameters and will exclude points in the dataset that are not part of a defining cluster [19]. HDBSCAN was introduced [20] as an extension [21] to the existing DBSCAN algorithm and can find clusters of varying densities using two hyperparameters [22].

The process of hyperparameter tuning in this work is an essential step to determine p^*, the hyperparameter set for an algorithm that maximizes trip capture for a given number of parking

facilities, m. Note that p^* is selected by the sampling process described here, and may not be the true global optimum.

k-means does not require hyperparameter tuning, but due to its random initialization of centroids, requires numerous runs to determine its best performance. We performed 10 runs of k-means for each value of m and select the result maximizing trip capture.

A two-stage grid search was used for DBSCAN and HDBSCAN, both of which have two parameters that can be changed. The first stage searches a wide range of values and the second searches more narrowly in an area of interest identified during stage one. From the explored hyperparameter combinations for each algorithm, a best combination, p^*, was selected for each value of m (number of clusters). The full grid space for DBSCAN hyperparameters was $\varepsilon \in [2, 150]$ feet and MinPts $\in [50, 700]$, for a total of 6,757 combinations. For HDBSCAN, the grid space was MinClusterSize $\in [50, 700]$ and MinSamples $\in [50, 1400]$, for a total of 2,145 combinations.

3.4 Parking facility placement and assessment

We now discuss how to take a cluster of points, representing a high-demand parking area, and choose a single point for the parking facility. Clusters can have non-uniform shape and density distribution, so choosing a single point within the cluster can influence trip capture. We desire a low-complexity method of determining facility placement within clusters, because in practice the placement will be dictated by factors on the ground (e.g., on street parking, sidewalk area, etc.).

In case study 1 at Vanderbilt University, we place each parking location at the latitudinal and longitudinal median coordinates of each cluster. This choice was made based on the observed shape of clusters in this case study, which was predominantly circular or ovular. In case study 2 for the broader city of Nashville, more elongated and non-convex clusters emerge, which require a different method. We evaluate the trip capture of at least 100 candidate locations by sampling a grid space within the bounds of each cluster and take the location maximizing trip capture.

More complex methods for this level of facility placement exist but, again, are perhaps secondary to the built environment considerations at this scale. It must be noted that the placement

by these methods is performed without consideration of physical constraints; for example, a parking location may occur within a building or in the middle of a street.

3.5 Weighted clustering

In this section we show how to incorporate features external to a micromobility trip dataset by performing weighted clustering. In the second case study of this work, we demonstrate this process by incorporating the sidewalk width at each location where a scooter trip was parked. This factor is of particular interest in cities where high demand is being experienced in areas where sidewalks are narrow and scooters present a potential obstruction. It may be strategic to reallocate some parking facilities from higher demand areas in order to address sidewalk constraints. Increased weighting of narrow-sidewalk trips allows us to evaluate this tradeoff, which we explore in Section 6.

By assigning each training data point in E_{tr} a weight, ω_i, we can emphasize the importance of some trip endpoints over others with respect to identified parking clusters. Consider a point j that has double the weight of another point k, such that $\omega_j = 2 * \omega_k$. In effect, this amounts to two points at the location of j relative to one point at the location of k. Weighted clustering is intuitively supported in DBSCAN via the *MinPts* hyperparameter: the weight of points in a cluster must sum to greater than or equal to *MinPts*. The default weight of each point (the "unweighted clustering" case) is, therefore, equal to 1.

For weighting trips according to the width of the sidewalk along which they are parked, we utilize a step function, given in equation (2), that assigns higher weights to trips ending along sidewalks with width less than an threshold, λ_{thresh}. The increased weight is denoted ω_{ADA} and corresponds to a threshold where ADA requirements [5] could potentially be infringed by a parked scooter: $\lambda_{\text{thresh}} = 5$ feet.

$$\omega_i = \begin{cases} 1, & \text{if } \lambda_i \geq \lambda_{\text{thresh}} \\ \omega_{\text{ADA}}, & \text{if } \lambda_i < \lambda_{\text{thresh}} \end{cases} \tag{2}$$

Table 1 : Summary statistics for scooter trips in dataset.

	Mean	Median	Standard deviation
Distance (VU)	0.97 miles	0.69 miles	0.92 miles
Duration (VU)	9.71 minutes	6.05 minutes	9.05 minutes
Speed (VU)	6.28 mph	6.31 mph	3.17 mph
Distance (Nashville)	1.15 miles	0.79 miles	1.15 miles
Duration (Nashville)	15.34 minutes	9.75 minutes	15.97 minutes
Speed (Nashville)	5.23 mph	4.98 mph	2.95 mph

In an implementation of weighted clustering where weights are treated in terms of absolute magnitude, as discussed above for DBSCAN, the weights can be normalized to the number of points in the dataset. This allows a baseline case, where all points carry a weight of one, to be more intuitively compared to a weighted case without changing hyperparameters. The normalized weight for a point i, denoted μ_i, is therefore:

$$\mu_i = \omega_i * \frac{n}{\sum_{j=1}^{n} \omega_j} \tag{3}$$

4   Micromobility dataset

The dataset of scooter trips was acquired from the Metropolitan Government of Nashville and Davidson County, Tennessee, and covers two of the scooter companies operating in Nashville as of June 2019. The historical data includes approximately 950,000 trips for these two operators made between August 30, 2018, and June 21, 2019. In order to remove potentially-erroneous data, trips are filtered based on distance (trips between 0.1 and 5 miles are kept), time (2 to 60 minutes), speed (between 2 and 15 miles per hour), and geography (start and end within Nashville limits).

The subset of data used in case study 1 includes only trips that begin or end within Vanderbilt University's campus boundary, while case study 2 uses the full Nashville dataset. Summary statistics for these two datasets can be found in Table 1.

5   Case study 1: Determination of efficient parking locations

In this section, we demonstrate the results of applying the parking facility placement methodology on the Vanderbilt subset of scooter trips. We also demonstrate facility capacity allocation and results. The Vanderbilt data subset is used in case study 1 because the distribution of demand on the campus is very localized around buildings and this provides a straightforward interpretation of location and ranking results.

5.1   Facility placement and capture potential

We explore three clustering algorithms for the placement of SES parking facilities with the primacy goal of evaluating representative candidates that are straightforward to implement and interpret.

We sort the Vanderbilt subset of the scooter trip dataset chronologically and take the first 75% of trips as training data and the latter 25% as testing data. The facility locations generated by each algorithm on the training data are evaluated with respect to the trip capture using the testing dataset and a capture radius of 100 feet. Trip endpoints that are located within the capture radius of two parking facilities are counted only once towards the total trips captured by all facilities.

The trip capture results for *k*-means, DBSCAN, and HDBSCAN algorithms are shown in Figure 3 in terms of the total number of trip endpoints that were captured out of the testing dataset. DBSCAN performed better than HDBSCAN and *k*-means across the range of cluster counts. Some noise is seen in the performance of all three algorithms such that, at times, increasing the number of clusters decreased the percentage of trips captured. This is due to the variability of the non-exhaustive hyperparameter tuning process. The performance gap between algorithms stayed relatively consistent across cluster counts.

At low numbers of clusters, *k*-means finds large, sparsely distributed clusters and facility placement by the median-based approach is not well-suited to these malformed clusters. For this reason we find density based methods such as DBSCAN and HDBSCAN provide more consistent results regardless of the number of clusters.
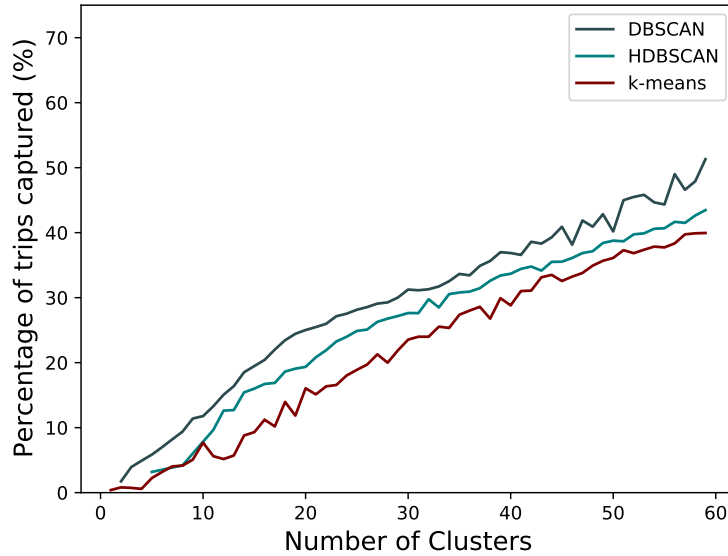
Figure 3 : Comparison of clustering algorithms for the placement of parking facilities, according to number of trips captured within 100 feet.

A real-world decision on the number of parking facilities to implement could consider both the total proportion of trips captured across all facilities (percentage of trips captured) and the number of trips captured per parking facility (capture efficiency). The overall trend in the incremental addition of a facility is that it increases overall trip capture, but at a decreasing marginal benefit. The thirtieth facility to be added captures fewer trips than the tenth, for example. Therefore, capture efficiency is highest with small numbers of facilities and decreases thereafter. This knowledge is useful in determining the number of facilities to be installed that would potentially maximize the return on investment. In a dockless micromobility system, it is not possible to capture all trips in a dedicated parking facility, so a cutoff point must be decided.

This efficiency tradeoff for DBSCAN is explored in Figure 4. We show the total percentage of trips captured (left y-axis) as a series of points that are colored according to their marginal per-facility percentage of trips captured (colorbar). The black line is the same per-facility percentage of trips captured metric, extracted for clarity (right y-axis). A hinge point is apparent in the percentage of trips captured curve; around 20 trip clusters, the per-facility capture percentage begins to decrease sharply. This indicates decreasing efficiency of the parking facility placement.

The specific locations of parking facilities generated under the highest-efficiency facility

14
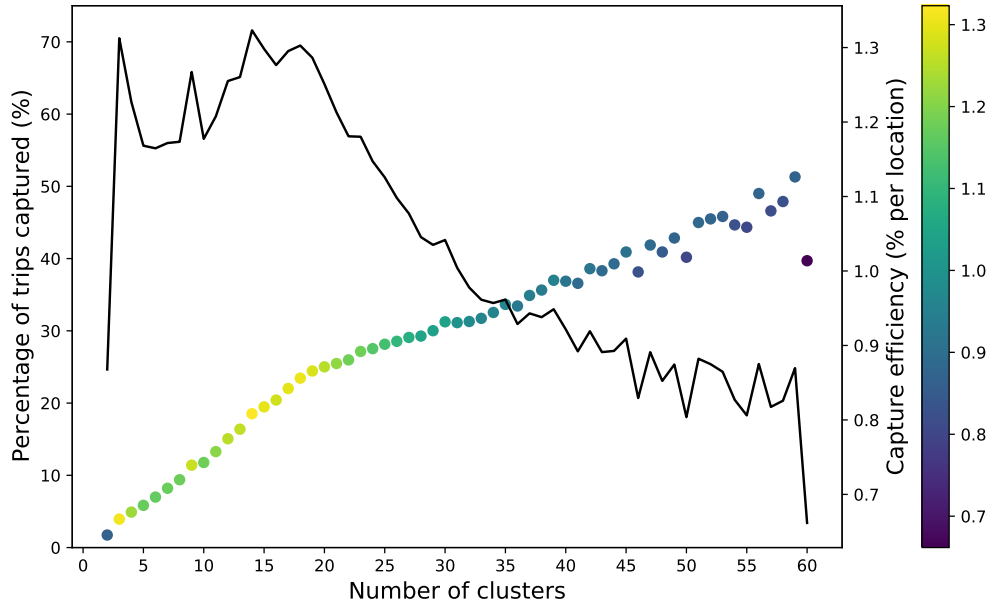
Figure 4 : Relationship between total trip capture (colored points) and per-facility percentage of trips captured (i.e., capture efficiency) (black line) for DBSCAN facility placement. *Note: capture efficiency is given on the secondary y-axis.*

placement realization, for 19 clusters, are shown in the map in Figure 5. The campus extents of Vanderbilt University are shown in orange and the campus buildings are hatched black and white. The cluster assignment of trip endpoints is shown by the color of points (unassigned points are in grey). The selected facility location for each cluster is denoted by a red X and a 100-foot capture radius is overlaid on each. A subset of parking locations, representing unique areas of interest on campus described in Table 2, is selected and marked with letters A-E on the map in Figure 5. While there appear to be areas of concentrated but un-clustered points, these areas fail to meet the mathematical cluster requirements for this run of DBSCAN. Indeed, it is difficult to discern visually from a map of tens of thousands of points where the most significant concentrations lie. This is an advantage of our methodology when using DBSCAN: clear delineation between clusters meeting criteria and all other points.
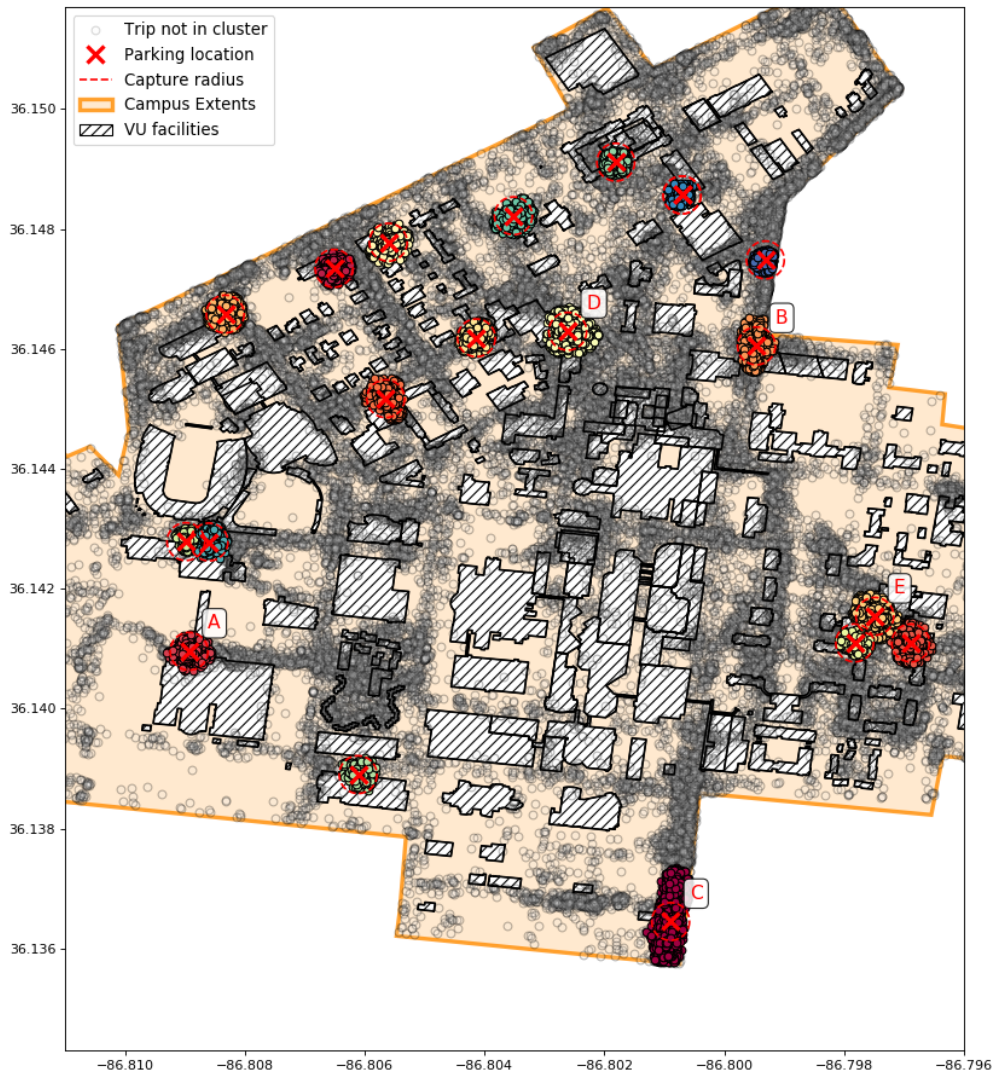
Figure 5 : Map of 19 parking locations (red X's) and capture radii of 100 feet (red dashed lines) generated by DBSCAN. The colored points at each parking location are the trip endpoints belonging to each cluster. Trip endpoints not belonging to a cluster are shown as light grey points. Vanderbilt University campus facilities and campus extents are also shown.

Table 2 : Parking location names and descriptions

Location	Area name	Description
A	Rec. Center	Campus recreation and wellness center
B	21st Ave. (North Side)	Commercial lunch time dining area
C	Hillsboro Village	Commercial area for students to eat and shop
D	Rand Hall	Main campus dining hall
E	Commons Center	Dining hall in the center of first-year student housing

5.2   Parking facility capacity

In order to evaluate the expected capacity requirements for parking locations, a derivative dataset is constructed that consists of periods of time and locations where devices are idle. By matching trips made on the same device, we determine the dwell period and location of devices between trips.

We again examine the five parking facilities marked A-E in the 19-cluster efficient realization. For each location, we compute the number of devices dwelling there during every minute covered by the training dataset. Only for the capacity evaluation, we use a capture radius of 300 feet, as opposed to the 100-foot capture radius for parking location determination. We are intentionally making an higher estimation of trip capture in order to air on the side of placing facilities that will be able to fulfill a larger demand and accommodate possible changing user behavior in the future. Times at which no scooters are present at the location are removed and the remaining minutes are arranged in descending order by number of dwelling devices. This results in a distribution of the number of scooters present at a given location at any time. The required capacity for each facility is calculated at specific percentile values for the number of parked scooters: 75, 80, 85, 90, 95, and 98. For example, the 98th percentile capacity value refers to the number of parked scooters that is greater than 98% of all other non-zero counts of parked scooters at a particular location.

The distributions of number of parked scooters for locations A-E are shown in Figure 6, and percentile values are shown by colored horizontal lines. The number of scooters in the parking location is shown on the y-axis and the number of minutes during which this capacity was present is given by the x-axis and clipped at 12,000 minutes. The total area under each distribution is indicative of the amount of scooters parked over time. Locations C and D have the largest area, indicating that they would more often contain multiple scooters. The percentile lines are given for integer values of number of scooters, so they overlap with each other in some cases. At the 98th percentile capacity, location D would require the largest size, capable of accommodating 8 scooters. Only 2% of its occupied time – approximately 50 total hours over the course of 6 months covered by the training data – would the facility be over-capacity. Interestingly, while the 98th percentile capacity is highest for location D, we see more frequently for location C periods of time during

17

which 10-14 scooters are present. Given the relatively small size required for a scooter parking facility, building to a 98th percentile capacity or higher would not be unreasonable.

These distributions of the number of parked scooters do not take into account the actual time of day corresponding to the demand. Looking at the usage of the selected parking locations by the hour of the day and by the day of the week, as shown in Figure 7, we can see varied trends of peak behavior. Location D (Rand Hall, main campus dining) has high utilization during the middle of the day, but strictly across weekdays. Location C (Hillsboro Village, commercial area) exhibits the same mid-day peak, but is heavily weighted towards weekend activity. Location A (Rec. Center) has a unique evening peak demand, stable throughout the week.

6 Case study 2: Parking considering sidewalk infrastructure

In this section, we present a case study on applying factors secondary to user demand for parking facility placement. The secondary factor considered is sidewalk width nearest the parked scooter, which is important in evaluating whether the sidewalk is at risk of being blocked by parked scooters. Even if a sidewalk meets width design standards, it can still be blocked by a device that was not parked carefully. We begin by describing how sidewalk width is calculated for each parked trip and incorporated into the clustering algorithm. We then analyze the effects different weights have on the formation of clusters and capture efficiency of the systems.

6.1 Sidewalk width and data preprocessing

The width of the sidewalk associated with each trip is calculated specifically for its parked location, because sidewalk width can vary significantly even along the length of a city block. We begin with the dataset of scooter trip endpoints, and spatial layers of sidewalk geometry.

For each trip endpoint, we determine the nearest point on a sidewalk segment. At this exact point we calculate the perpendicular width of the sidewalk geometry. From these calculated sidewalk width values, we remove trips where a sidewalk was not near a trip endpoint or points

18
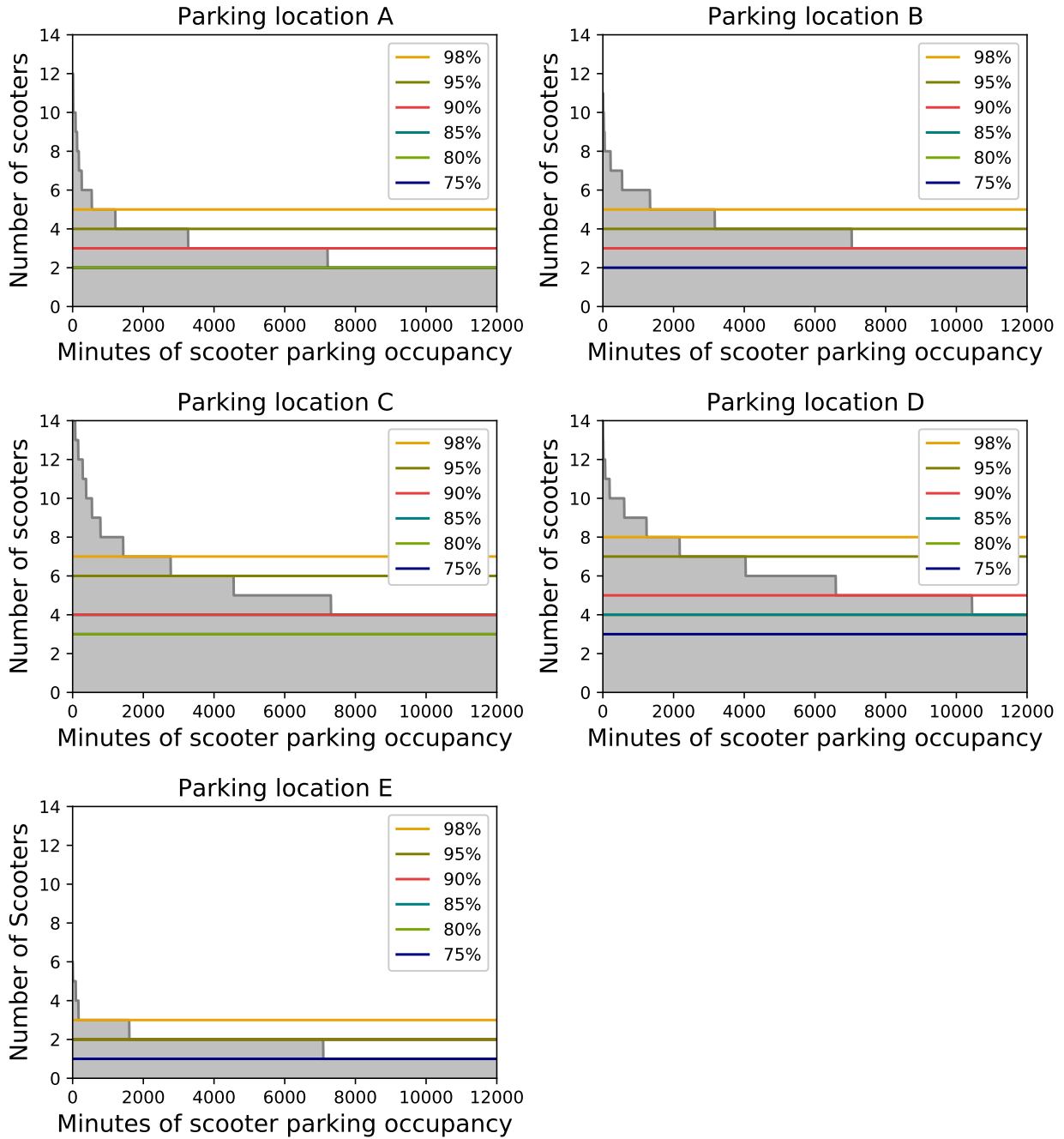
Figure 6 : Distribution of the number of parked scooters based on duration of time that occupancy level is present. This is calculated only for times when one or more devices are present. The 75th, 80th, 85th, 90th, 95th, and 98th percentile capacity values are shown for each. *Note that some percentile lines are overlaid on top of each other due to discrete capacity values. Also note that the x and y axis have been standardized across the subplots to 14 scooters and 12,000 minutes.*

19
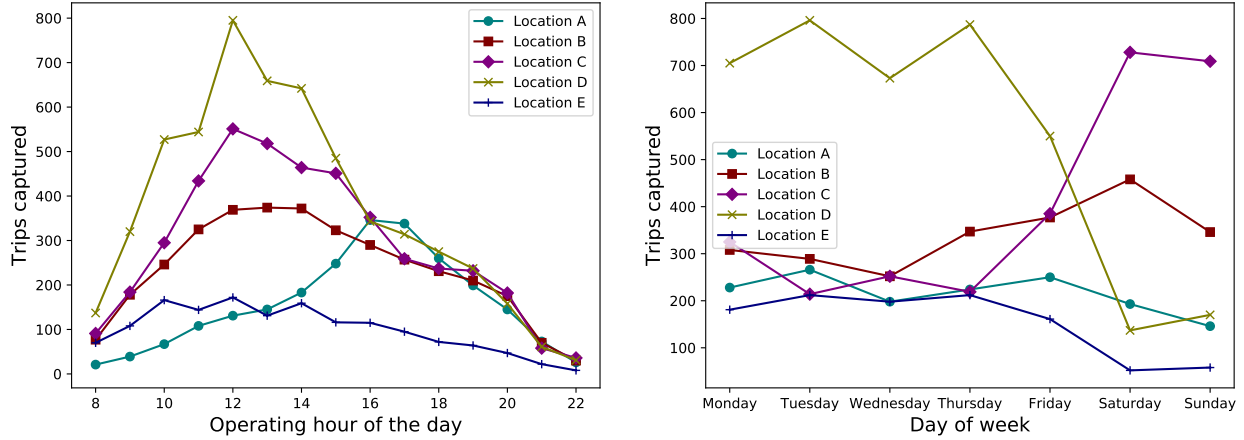
Figure 7 : Distribution of scooter trips captured for parking locations A-E across hours of the operating day (left) and days of the week (right).

with associated sidewalk width less than 3 feet. Sidewalks narrower than 3 feet are not compliant with the 2010 ADA Standards for Accessible Design [5], so we assume that this was an anomaly in sidewalk width processing since no sidewalk should be built at this width.

We separate the training and testing datasets for case study 2 by stratified random sampling (as opposed to a temporal split in case study 1). Stratification is based on sidewalk width for trips, in order to ensure that the more rare narrow-sidewalk trips are proportionally represented in training and testing data. We take 10% of trips as training data (approximately 95,000 trips) and the remaining 90% as testing data. The comparatively low volume of training data is due to computational constraints imposed by running clustering on such a large number of points.

6.2 Assessment of weighted clustering

In order to evaluate the effects of increasing the weight of narrow-sidewalk trips, we select three representative hyperparameter combinations, listed in Table 3 that generate varying number of parking locations in the baseline case. These three hyperparameter combinations are then used to train DBSCAN on the various degrees of sidewalk width weighting.

Recall from Section 3.5 that ω_i represents the raw weight of trip i and that the weights are normalized across all trips to sum to n, the total number of trips. The normalized weight for trip i

**Table 3** : DBSCAN hyperparameter combinations and their generated number of clusters.

Combination label	ε	$MinPts$	Number of clusters, $\omega = 1$
Combination 1	0.00011	80	59
Combination 2	0.000123	120	44
Combination 3	0.00012	145	24

is μ_i. For non-weighted clustering, which we refer to as the *baseline* case, all ω_i values are equal: $\omega_i = 1$. Sidewalk width for each trip i is denoted λ_i and we increase the weight of narrow-sidewalk trips relative to wider sidewalk trips.

We use these representative hyperparameter combinations to explore the effects of assigning a higher value of ω (pre-normalized) to narrow-sidewalk trips. A step function is applied to trip sidewalk widths, as described in Section 3.5: trips with respective sidewalk width less than $\lambda_{\text{thresh}} = 5$ feet are *problematic trips* and given a larger ω value, which we denote ω_{ADA}; trips above the threshold are given $\omega = 1$.

Figure 8 shows how percentage of trips captured and capture efficiency are related as we increase ω_{ADA}. In both Figure 8a and Figure 8b we can see how ω_{ADA} has an inverse relationship between percentage of trips captured and total capture efficiency, while it has a, general, positive relationship with both percentage of problematic scooter trips captured and capture efficiency of problematic scooter trips. This is expected, based on the increasing emphasis on problematic trips.

All of these metrics show a rapid change in performance for lower values of ω_{ADA}, but the rate at which they change diminishes as we approach higher values of ω_{ADA}. An increase in capture of problematic trips by over 300% relative to the baseline case is achieved with only a 13% decrease in total trip capture. This result was averaged across three hyperparameter combinations and it should be noted that the number of parking facilities increased by an average of 3 (from 42 to 45) in the comparison. While the percentage of problematic scooter trips captured increases up to $\omega_{\text{ADA}} = 40$ or higher, the capture efficiency of problematic scooter trips reaches maximum efficiency at lower values, around $\omega_{\text{ADA}} = 12$. This is because, generally, the number of parking locations increases as ω_{ADA} increases. Even though additional problematic scooter trips are captured, marginal capture diminishes and causes efficiency to decrease.
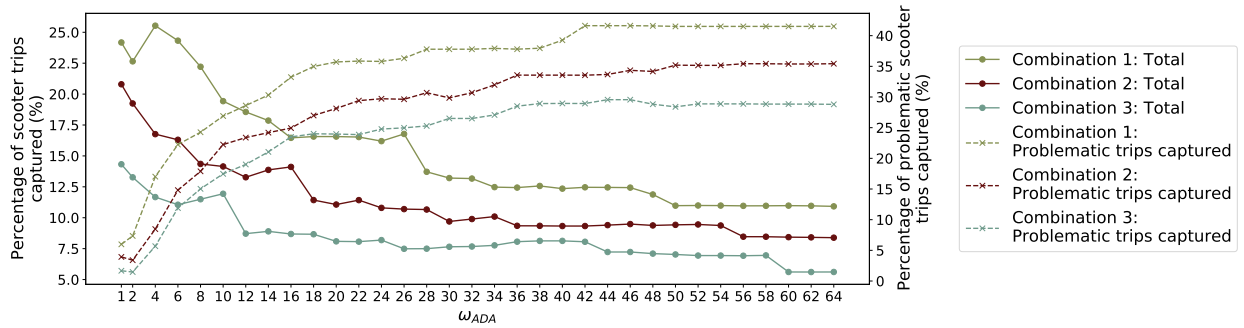
(a) Relationship between total percentage of trips captured (solid lines) and percentage of problematic scooters captured (dashed lines). *Note: percentage of problematic scooters captured is given on the secondary y-axis.*
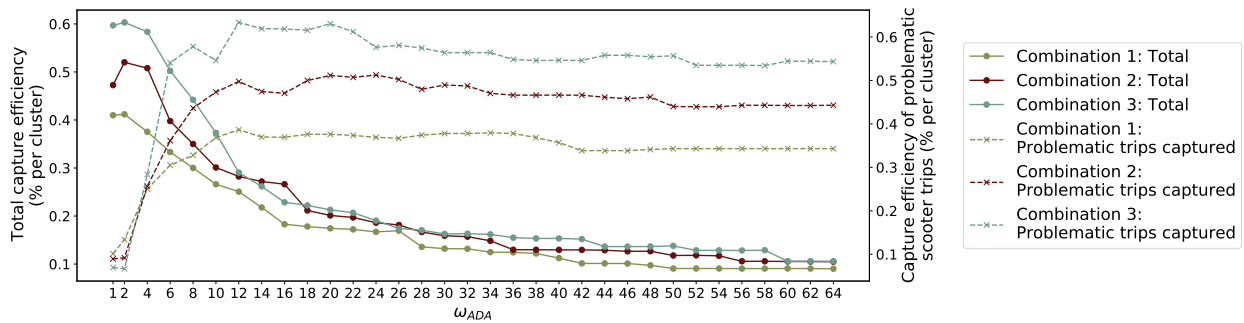

(b) Relationship between total scooter trips capture efficiency (solid lines) and capture efficiency of problematic scooter trips (dashed lines). *Note: capture efficiency of problematic scooter trips is given on the secondary y-axis.*

Figure 8 : Performance of each hyperparameter and ω_{ADA} combination. Subfigure (a) shows the effect that increasing ω_{ADA} has on percentage of trips captured, while (b) shows what happens to capture efficiency as ω_{ADA} increases.

Total capture efficiency decreases almost immediately as the value of ω_{ADA} increases because this shifts the emphasis from areas with sheer volume of trips towards those with narrow-sidewalk trips, and starts to form new clusters on these areas. This effect of shifting parking location emphasis can be seen in Figures 9a and 9b, which compares locations for the baseline case with the case where $\omega_{ADA} = 36$. With a higher value of ω_{ADA}, however, clusters make a significant shift from Downtown Nashville to Midtown Nashville, where trip demand is lower but sidewalks are far narrower. Generally, more clusters start to form as we increase the value of ω_{ADA} — 59 clusters for the baseline case and 100 clusters when $\omega_{ADA} = 36$, in this example. Figure 9c is a heatmap of all scooter trips in the city, which are heavily concentrated in Downtown Nashville. Figure 9d is a heatmap of only the problematic trips, which are less localized and more frequent in areas outside of Downtown Nashville.
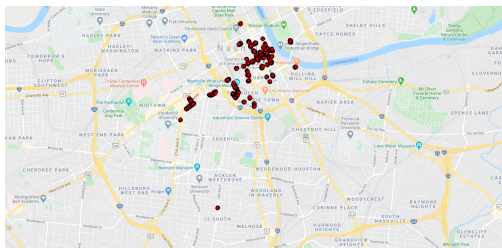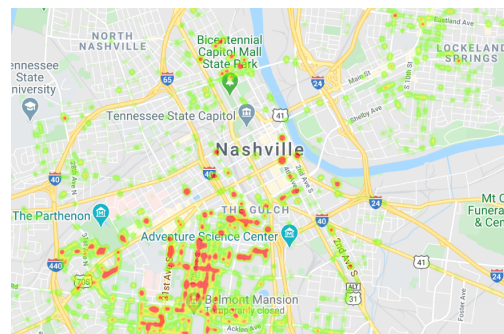

(a) Parking locations for baseline case ($\omega_{ADA} = 1$).
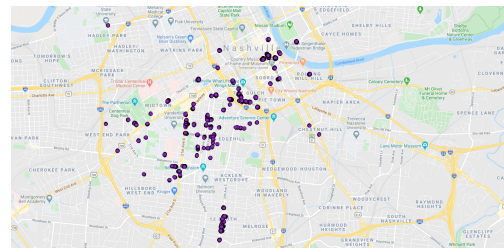

(b) Parking locations for $\omega_{ADA} = 36$.
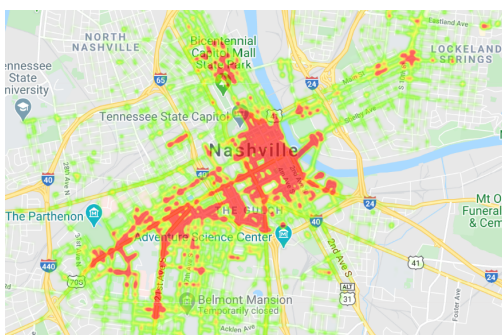
(c) Frequency of all trips end locations

(d) Frequency of ADA non-compliant trips end locations

Figure 9: Distribution of clusters and trips around Nashville based on whether trips are problematic due to sidewalk width. Subfigures (a) and (b) show the affect of applying a large ω_{ADA} value, which weights problematic trips far higher than non-problematic trips during cluster formation. Parking locations for the baseline case are concentrated in Downtown Nashville, but move further out into Midtown Nashville as problematic trips are emphasized. The spatial distribution of all trips (c) and distribution of problematic trips (d) validates this effect.
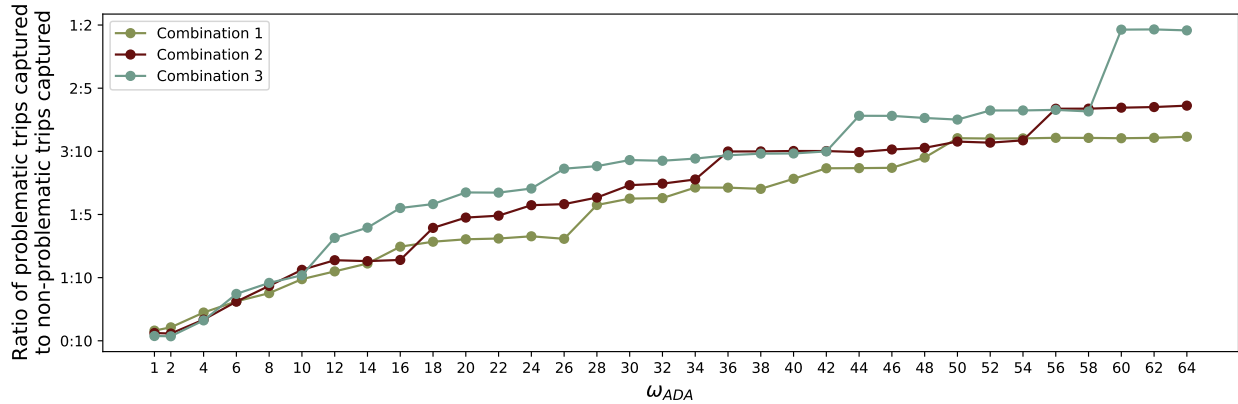
Figure 10 : Change from the baseline of the ratio between percentage of problematic scooter trips captured and percentage of trips captured.

Cities indicate by their legislation that the need for parking facilities is as much about corralling the trips as it is alleviating the pressure on the sidewalks from problematic trips. Cities that want to achieve a balance that fulfill the objectives of corralling high number of trips while also alleviating the pressure of problematic scooters incur on sidewalks should choose ω_{ADA} value that balances both objectives. Figure 10 shows the ratio between problematic scooter trips captured and non-problematic trips captured. As is expected, larger values of ω_{ADA} shift this ratio towards problematic trips. All three hyperparameter combinations begin in the baseline case by capturing only one problematic trip for approximately every 60 non-problematic trips (1:60). Some variation is seen between the hyperparameter combinations at large values of ω_{ADA}, but around $\omega_{\text{ADA}} = 30$ we see all three results capturing problematic trips at a ratio of approximately 1:4.

7   Discussion

Many practical considerations arise with respect to the real-world placement of parking facilities within cities. We consider one such influencing factor, sidewalk width or ADA compliance, in this work and how taking it into account will ultimately influence the number or percentage of trips in the area that are captured within parking facilities. The degree to which a city is willing to sacrifice trip capture for other factors will vary and could be evaluated on a case-by-case basis.

Other practical considerations for facility placement could include distance between facilities, prohibited parking areas, and density of parking facilities within an area. There are valid reasons why a city may elect to establish a dense set of parking facilities in one area, as opposed to a wider range of facilities that may capture more trips. Reaching a critical density of parking facilities may condition users to expect their presence and use them regularly. Maintaining a smaller distance between facilities will make it less likely that a user ignores parking facilities because they are too far from his or her destination. Additionally, placing parking locations close to prohibited parking areas may also increase compliance and decrease frustration with "geofencing". We have observed, qualitatively, that the actual allocation of parking facilities in Nashville has followed a broader approach. Parking areas have been allocated in a variety of neighborhoods rather than focusing strictly on downtown, which has by far the highest demand on a neighborhood level. These considerations may be made to fit into the framework of weighted clustering under a similar mechanism that we used for sidewalk width.

8 Conclusions

This article proposes the data-driven placement of dedicated SES parking facilities, so as to maximize their impact in terms of potential capture of trips. These parking facilities are sorely needed in highly congested areas that are prone to clutter of SESs and other devices on sidewalks. Data-driven placement increases the likelihood that, under an optional usage model, they are effective at convincing users to park devices because of proximity to their destinations. Parking facility locations are determined through the application of clustering algorithms. Results from a case study focused on Vanderbilt University demonstrate a point of diminished efficiency for facilities in excess of approximately 20. At this point, approximately 1 in 4 trips can reasonably be captured by a parking facility. The capacity of such facilities would need to be approximately 8 scooters, in order to serve the 98th percentile highest demand period.

The proposed methods can be customized using weighted clustering to encourage particular

25

characteristics of facility placement. We explore the emphasis of scooter trips ending near narrow sidewalks as a second case study, which is relevant for reducing conflicts between parked scooters and pedestrians or ADA sidewalk requirements. The case study results show that 300% more problematic narrow-sidewalk trips could potentially be captured with only a 13% reduction in overall trip capture.

Future work includes accounting for irregularly-shaped (non-circular) clusters of trip endpoints that should probably be served by multiple parking facilities when capture radius does not cover the entire area of dense points sufficiently. Location determination should consider truly feasible facility locations that are free from obstructions such as buildings and roadways. Capture zones that are non-circular in shape due to occlusion by buildings could be also assessed. We could also compare the clustering methods employed in this work to methods of continuous density approximation, such as kernel density estimation or Gaussian mixture models. Furthermore, a model could be created for the likelihood of SES users to park in optional dedicated facilities, parameterized by the distance from the facility to their destinations, the area in which the facility is located, and other factors.

9 Acknowledgements

References

[1] Shared Micromobility in the U.S.: 2018. Technical report, National Association of City Transportation Officials.

[2] Caroline Janssen, William Barbour, Erin Hafkenschiel, Mark Abkowitz, Craig Philip, and Daniel B Work. City-to-city and temporal assessment of peer city scooter policy. *Transportation Research Record*, 2674(7):219–232, 2020.

[3] Powered Scooter Share Mid-Pilot Evaluation. Technical report, San Francisco Municipal Transportation Agency, 2019.

[4] Anne Brown, Nicholas J Klein, Calvin Thigpen, and Nicholas Williams. Impeding access: The frequency and characteristics of improper scooter, bike, and car parking. *Transportation research interdisciplinary perspectives*, page 100099, 2020.

[5] Department of Justice. ADA standards for accessible design, September 2010.

[6] Ines Frade and Anabela Ribeiro. Bike-sharing stations: A maximal covering location approach. *Transportation Research Part A: Policy and Practice*, 82:216–227, 2015.

[7] Juan Carlos García-Palomares, Javier Gutiérrez, and Marta Latorre. Optimizing the location of stations in bike-sharing programs: A GIS approach. *Applied Geography*, 35(1-2):235–246, 2012.

[8] Leonardo Caggiani, Rosalia Camporeale, Mario Marinelli, and Michele Ottomanelli. User satisfaction based model for resource allocation in bike-sharing systems. *Transport Policy*, 80:117–126, 2019.

[9] Nanjing Jian, Daniel Freund, Holly M Wiberg, and Shane G Henderson. Simulation optimization for a large-scale bike-sharing system. In *2016 Winter Simulation Conference (WSC)*, pages 602–613. IEEE, 2016.

[10] Daniel Freund, Shane G Henderson, and David B Shmoys. Minimizing multimodular functions and allocating capacity in bike-sharing systems. In *International Conference on Integer Programming and Combinatorial Optimization*, pages 186–198. Springer, 2017.

[11] Eoin O'Mahony and David B Shmoys. Data analysis and optimization for (citi) bike sharing. In *Twenty-ninth AAAI Conference on Artificial Intelligence*, 2015.

[12] Yongping Zhang, Diao Lin, and Zhifu Mi. Electric Fence Planning for Dockless Bike-Sharing Services. *Journal of Cleaner Production*, 206:383–393, 2019.

[13] Xinwei Ma, Ruiming Cao, and Yuchuan Jin. Spatiotemporal clustering analysis of bicycle sharing system with data mining approach. *Information*, 10(5):163, 2019.

[14] The US Green Building Council (USGBC). Bicycle facilities.

[15] David Arthur and Sergei Vassilvitskii. k-means++: The Advantages of Careful Seeding. In *Proceedings of the Eighteenth ASnnual ACM-SIAM Symposium on Discrete Algorithms*, page 1027, 2006.

[16] Ke Liao and Diansheng Guo. A clustering-based approach to the capacitated facility location problem 1. *Transactions in GIS*, 12(3):323–339, 2008.

[17] Pranavi Singanamala, Dharma Reddy, and P Venkataramaiah. Solution to a multi depot vehicle routing problem using k-means algorithm, clarke and wright algorithm and ant colony optimization. *International Journal of Applied Engineering Research*, 13(21):15236–15246, 2018.

[18] Akanksha Shukla, Kusum Verma, and Rajesh Kumar. Consumer perspective based placement of electric vehicle charging stations by clustering techniques. In *2016 National Power Systems Conference (NPSC)*, pages 1–6. IEEE, 2016.

[19] Martin Ester, Hans-Peter Kriegel, Jörg Sander, Xiaowei Xu, et al. A density-based algorithm

for discovering clusters in large spatial databases with noise. In *KDD*, volume 96, pages 226–231, 1996.

[20] Ricardo Campello, Davoud Moulavi, and Joerg Sander. Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining*, volume 17, pages 160–172, 2013.

[21] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. Dbscan revisited, revisited: why and how you should (still) use dbscan. *ACM Transactions on Database Systems (TODS)*, 42(3):1–21, 2017.

[22] Leland McInnes, John Healy, and Steve Astels. HDBSCAN: Hierarchical density based clustering. *The Journal of Open Source Software*, 2(11):205, 2017.