THREE ESSAYS ON SHARED MICROMOBILITY

A Dissertation
Submitted to the Graduate Faculty
of the
North Dakota State University
of Agriculture and Applied Science

By

Ali Rahim-Taleqani

In Partial Fulfillment of the Requirements
for the Degree of
DOCTOR OF PHILOSOPHY

Major Department:
Transportation, Logistics, and Finance

March 2020

Fargo, North Dakota

# NORTH DAKOTA STATE UNIVERSITY

Graduate School

**Title**

THREE ESSAYS ON SHARED MICROMOBILITY

**By**

Ali Rahim-Taleqani

The supervisory committee certifies that this dissertation complies with North Dakota State University's regulations and meets the accepted standards for the degree of

DOCTOR OF PHILOSOPHY

SUPERVISORY COMMITTEE:

Dr. Jill Hough
Chair

Dr. Raj Bridgelall

Dr. Bruce Maylath

Dr. Kendall Nygard

Dr. Joe Szmerekovsky

Approved:

22 July 2020
Date

Dr. Tim O. Peterson
Department Chair

# ABSTRACT

Shared micromobility defines as the shared use of light and low-speed vehicles such as bike and scooter in which users have short-term access on an as-needed basis. As shared micromobility, as one of the most viable and sustainable modes of transportation, has emerged in the U.S. over the last decade., understanding different aspects of these modes of transportation help decision-makers and stakeholders to have better insights into the problems related to these transportation options.

Designing efficient and effective shared micromobility programs improves overall system performance, enhances accessibility, and is essential to increase ridership and benefit commuters. This dissertation aims to address three vital aspects of emerging shared micromobility transportation options with three essays that each contribute to the practice and literature of sustainable transportation.

Chapter one of this dissertation investigates public opinion towards dockless bikes sharing using a mix of statistical and natural language processing methods. This study finds the underlying topics and the corresponding polarity in public discussion by analyzing tweets to give better insight into the emerging phenomenon across the U.S.

Chapter two of this dissertation proposes a new framework for the micromobility network to improve accessibility and reduce operator costs. The framework focuses on highly centralized clubs (known as k-club) as virtual docking hubs. The study suggests an integer programming model and a heuristic approach as well as a cost-benefit analysis of the proposed model.

Chapter three of this dissertation address the risk perception of bicycle and scooter riders' risky behaviors. This study investigates twenty dangerous maneuvers and their corresponding frequency and severity from U.S. resident's perspective. The resultant risk matrix and regression model provides a clear picture of the public risk perception associated with these two micromobility options. Overall, the research outcomes will provide decision-makers and stakeholders with scientific information, practical implications, and necessary tools that will enable them to offer better and sustainable micromobility services to their residents.

# ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my academic advisor, Dr. Jill Hough, for her support, encouragement, and guidance throughout my Ph.D. journey at NDSU. I am also grateful to many other academic members and advisors. I would like to convey my great appreciation to my committee members, Dr. Kendall Nygard, Dr. Raj Bridgelal, Dr. Joseph Szmerekovsky, and Dr. Bruce Maylath. They have served as both professional and personal advisors to me. Their guidance has made my graduate studies both productive and enjoyable.

This research would not have been possible without some other key contributors to my research. My deepest acknowledgments also go to Dr. Chrysafis Vogiatzis, who never stopped challenging me. He is a true mentor and has helped me develop research skills that are necessary to complete this level of research. I am also grateful to Dr. Eunsu Lee for his support and for placing his trust and confidence in my abilities early in my Ph.D. It is a great honor working with both of you.

Last but not least, I want to express my humble gratitude to my parents, Afzal Adeli-Gilani and Samad Rahim-Taleqani, my sisters, Sima, Sepideh, and Samira. Most significantly, I would like to acknowledge the love of my life, Ameneh Forouzandeh-Shahraki, and my daughter, Diana, for their never-ending understanding and love. This journey would be impossible without their unconditional support, compassion, and love.

# DEDICATION

To my Mom and Dad,

Couldn't have done it without your genetic material.

To my wife, Ameneh, and my lovely daughter Diana,

Without whom this journey would have been completed a year earlier.

To caffeine and YouTube

My companions through many a long night of research and coding.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# 1. INTRODUCTION

## 1.1. Background

Shared micromobility is the category of transportation modes that encompasses shared-use fleets of bikes, e-bikes, e-scooters and other small vehicles. The vehicles might be fully or partially human-powered. The primary objective of this mode of transportation is to address the first-mile/last-mile problem mile and shrink transit deserts in cities. In other words, shared micromobility may increase the accessibility to other modes of transportation, including mass transit services. Mobike, a China-based dockless bike sharing system, nearly doubled accessibility to jobs, education, and health care by filling the transit gaps in Beijing [1]. Micromobility not only improves connection to mass transit, but also could be an alternative to personal car trip as more than 50% of car trips made in the United States are less than five miles in length[2]. The business model, either docked or dockless format, has gained tremendous attention in the last decade. But questions concerning the ultimate size, scope of the system, technology acceptance, enforcement strategies, and associated risks have also emerged.

## 1.2. Research Objectives

The emergence of shared micromobility in the United States requires that public and private stakeholders must understand and evaluate the aspects of this mode of transportation. Shared micromobility enhances public transportation visibility, reduces the use and dependence on fossil fuel, and promotes healthy lifestyles. Hence, it is essential for private and public stakeholders to have a better understanding of system operations, communities' feedback, and issues related to these systems. Therefore, it is worthwhile for researchers and practitioners to recognize the underlying factors associated with adoption, change, and abandonment of such modes of transportation. A sustainable transportation mode that ensures low-cost, low-risk and high social benefits requires a proactive planning and decision-making process. Thus, in three essays, we examined the bike and scooter sharing program and addressed some of the common problems with these shared-use mobility systems. As these systems evolve, we tried to address the following questions:

1. What do U.S. residents think about new modes of transportation, given the dominant car culture? What is the best method to learn about people's opinions/views?

2. How could we modify the transportation network to improve people's accessibility to these systems, increase ridership, and reduce operators' costs? What is the alternative model to address the old(docked-based bike and scooter sharing system) and the new and emerging one (dockless system)?

3. What are the risky behaviors associated with shared micromobility? Given the lack of comprehensive enforcement, what are peoples' perceptions of risky riding behaviors? How could we quantify these risks?

### 1.3. Research Methodologies and Contributions

The following three studies will provide the original contributions to address these questions and understand some aspects of shared micromobility in the United States.

Chapter 2 addresses the first set of questions, and the final product was published in Transportation Research Record with the title "PUBLIC OPINION ON DOCKLESS BIKE SHARING: A MACHINE LEARNING APPROACH." Structured interviews or surveys can be affected by response bias, including the demeanor of the surveyor, the phrasing of questions, the way an experiment is conducted, and the desires of participants to act as a good experimental subject and to provide socially desirable responses. Hence, opinion mining on social media data could be a complement to traditional methods, as users voluntarily give their opinions on the subjects of interest when compared to surveys conducted where no interview is involved. People use Twitter as a platform to express their emotions and the problems they face in using any mode of transportation. We used natural language processing methods to determine the polarity of sentiment and topics in their discussion.

**Impact**: The findings help operators, city authorities, and public stakeholders to refine insights into public needs, preferences, experiences, and opinions on their services. The findings also suggest that opinion mining on social media data could be a complement to traditional methods, as users voluntarily give their feedback on the subjects of interest when compared with surveys conducted where no interview is involved. Additionally, the user profile might influence the perception of others towards the user's tweet. Unlike traditional surveys, in which other participants are anonymous, online platforms, to some extent, affect people's opinions.

Chapter 3 describes a novel model to tackle the problems described in the second sets of questions. The final product was published in the journal of Advanced Transportation with the title "MAXIMUM CLOSENESS CENTRALITY K-CLUBS: A STUDY OF DOCKLESS BIKE SHARING." Shared-use mobility options have undergone a series of developments. The most recent is referred to as the dockless or free-floating system. Recently, it has become essential to accommodate the free-floating system. This change comes with an increase in the coordination cost, as bicycles/scooters are no longer checked in and out from bike sharing stations that are fully equipped to handle the volume of requests. Instead, they can be checked in and out from virtually anywhere. Hence, we proposed a new framework for combining traditional bike stations with locations that can serve as free-floating bike sharing stations. The framework is based on identifying highly centralized-clubs (i.e., connected subgraphs of restricted diameter). The restricted diameter reduces coordination costs as dockless vehicles can only be found in specific locations. Also, we use closeness centrality as this metric allows for quick access to dockless bike sharing while, at the same time, optimizing the reach of service to bikers/customers.

**Impact**: First, we used the k-club concept, combined with closeness centrality, to identify candidate locations that could be geo-fenced. We also allow for weight at each node of the network and gravity model. This modification enhances the speed of the k-club formation through the heuristic algorithm devised. Second, we analyzed the exact optimization model and the heuristic devised and compared them in computational time and solution obtained. In each k-club obtained for varying values of k riders (commuters) can reach to any other neighboring sites within a fixed distance (controlled by k), implying that the virtual locations provide better accessibility to demand points. Last, we presented potential strategies for operators to further manage the inventory by applying incentives and making bicycle collection and rebalancing more cost-effective.

Chapter 4 explains the risky behaviors and the associated underlying factors associated. The final product is under review by the Journal of Safety Research with the title “RISK PERCEPTION OF BICYCLE/SCOOTER RIDERS RISKY BEHAVIORS.” The shared mobility industry shows no sign of stopping. However, there are public debates on the road behavior of users. Some display reckless behaviors like fast cornering, ignoring red lights, taking shortcuts at zebra crossings, speeding, and risky maneuvering. These scofflaw rides are not only irresponsible and a danger to other road users, but they also damage bikes/scooters, causing increased maintenance costs, service interruptions, and customer dissatisfaction for bike/scooter share programs. We conducted two separate surveys through the Qualtrics® platform. Participants were asked to rate the severity and frequency of 20 risky behaviors of riders on five Likert scales. The risk matrix is built based on the magnitude and frequency of each risk, and ordered logistic regression is applied to identify significant factors. Regression analysis revealed that age and income are significant factors shared between both survey groups. Level of education and living in urban areas are two statistically significant factors explaining the different risky behaviors with bicycles or scooters. In general, the survey results show that participants perceive that there is a low risk associated with reckless behaviors. It may imply that they are exposed to fewer incidents, or the media exaggerates the news about the incidents. Further research on other aspects of risk, such as network geometry and safety education, would help better understand the underlying factors

**Impact**: We extended the literature by developing a risk-matrix associated with bike/scooter riding behaviors. The findings offer several insights for practitioners to develop new enforcement policies and safety education programs to enhance scooter/bicycle sharing programs and provide a safe environment for all road users.

# 2. PUBLIC OPINION ON DOCKLESS BIKE SHARING: A MACHINE LEARNING APPROACH

## 2.1. Abstract

Dockless bike sharing is an emerging paradigm. Like many other technologies, it brings advantages and disadvantages to communities. Further investigation into public opinion will shed light on the impact of this technology on communities and provide input to city authorities for transportation planning. Transportation planning processes can be enhanced by engaging the community through social media technologies. Social media like Twitter, Facebook, and other microblogging media have been used for planning, but have not been extensively evaluated for that purpose. This study examined approximately $32,000$ posts on Twitter to assess public opinion on dockless bike sharing systems. Using a mix of text mining and statistical techniques, we examined relevant posts to determine the sentiment polarity of tweets, the underlying topics in the tweets, and the extent of engagement and impact on the decision-making process. Results given by two different sentiment algorithms show that there is more positive than negative polarity across the algorithms. Also, the findings show that the underlying topics in tweets include electric scooters, private e-hailing companies, and blockage of sidewalks, among others. The results indicate that the dockless shared micromobility models are potentially useful in generating participation, but faced substantial technical, analytical, and communication barriers to influencing decision-making [1].

## 2.2. Introduction

Bike sharing is a sustainable mode of transportation in many cities around the world. It brings advantages to existing transportation networks, including increased personal mobility, reduced traffic congestion, decreased fossil fuel use and dependence, increased public transit visibility, health benefits, and increased public environmental awareness [3]. The rapid growth of bike sharing in the United States continued in 2017, showing 25% more ridership than in 2016. Two primary factors contributed to this growth:

1. increasing ridership in existing systems; and

[1] The material in this chapter was co-authored by Ali Rahim Taleqani, Jill Hough, and Kendall Nygard. Ali Rahim Taleqani had primary responsibility for data collection, analysis, and interpretation of results. Ali Rahim Taleqani also drafted and revised all versions of this chapter. This chapter appears in Transportation Research Record (Rahim Taleqani, Ali, Jill Hough, and Kendall E. Nygard, 2019)

2. the launch of several new and different bike share systems.

Some of these systems have a new bike sharing customer interface, commonly known as dockless bike sharing. By the end of 2017, after several unpermitted and pilot programs, there were 25 cities and suburbs hosting dockless programs operated by five companies: Jump, Lime, Mobike, Ofo, and Spin, as well as some smaller companies, for example, Pace, Donkey Republic, VBike, LennyBike, and Riide. By the end of 2017, dockless systems accounted for almost 44% of all bike sharing bikes in the United States. However, only an estimated 4% of trips were attributed to dockless bike systems [4].

Introduced in China, dockless bike sharing is a system in which a commuter unlocks a bike with a cell phone, rides it, parks it, and relocks it. The intent of the system is to alleviate a prominent problem with station-based bike sharing programs, called the rebalancing problem. The practice offers enhanced convenience, mobility, and reduces the negative externalities (such as reduced space available for street parking) when compared with the traditional bike sharing model. However, various problems with dockless bike sharing systems have been reported by news agencies [5, 6]. The major problem is that piles of bikes from bike sharing companies are increasingly becoming a public nuisance. Parking or leaving bikes in improper locations is aesthetically objectionable and, in some cases, clogs sidewalks, rendering them unwalkable. In the absence of regulations, this trend may result in blocked right of ways and increases in trash, vandalism, and many other negative consequences [7]. This study aims to evaluate the consistency of the problems reported on by news corporations and expressed in public opinion.

Researchers, practitioners, and public and private stakeholders have often conducted surveys to capture public opinions. With the advent of the Internet and the emergence of social media, systems such as Twitter have been used to harvest and extract subjective information from text available online. With the brisk growth of social networks, increasing numbers of people actively engage in microblogs, Facebook, Instagram, Twitter, and other outlets. These outlets are a communication tool for people to share their emotions, opinions, news, experiences, desires, and expectations. This information can have a significant impact on others in decision-making. Hence, the analysis of such information becomes an integral part of any public or private enterprise [8]. The vast amount of data available in text form is relatively cumbersome to analyze and eventually requires statistical techniques to retrieve a general summary of opinions/views. Examining such content provides valuable information and insight about a specific topic. There are a plethora of studies on text analysis [9]. Sentiment analysis, as a subarea of text analytics, is a computational study that includes detection, extraction, and classification of opinions, sentiments, emotions, and attitudes concerning different topics expressed in textual inputs [10]. Another useful method to analyze a large volume of unlabeled text is topic modeling, a statistical method for discovering the key and pervading themes in a vast and unstructured collection of documents on social media networks [11].

Microblogging is an online broadcast medium typically smaller than a traditional blog. Twitter, as a common form of microblogging, launched on July 13, 2006. Each "tweet" or "short blog" is limited to 280 characters, equivalent to the size of a newspaper headline. Twitter could also be a good source for collecting public opinion because of the heterogeneity of users. Twitter users are from different social backgrounds, ranging from ordinary people to professionals, and include organization representatives, celebrities, and politicians. Thus, the tweets collected are the words of users with varying interests, which makes twitter feeds a valuable online source of opinion. This study describes a system aimed at discovering public opinions about dockless bike sharing programs as well as identifying underlying topics within the tweets. Tweets available in the English language are used to determine public opinions and concerns about dockless bike sharing. This study is likely the first to examine microblogging in regard to dockless bike sharing. The primary objective of this study is to assess public opinion toward dockless systems with the aid of text mining. This study uses text-mining techniques in conjunction with statistical methods to:

- examine public sentiments toward dockless bike sharing for better city planning; and
- identify potential topics in public opinion about dockless bike sharing.

This study sheds light on the current literature to address these two issues by studying public opinion with regard to dockless bike sharing systems using Twitter. The findings will help city officials understand the public perception toward an emerging, and sometimes nuisance, phenomenon for better planning and development and as a model to address other challenges.

The paper is organized as follows: First, related works are reviewed. The research methodology and conceptual framework as well as data collection and statistical methods used in the analysis are then discussed. The Results and Findings section presents the results of public sentiments about dockless systems. This section also includes discussion about the clustering results of public opinion on dockless bike sharing. Then the topic modeling and the findings are described. The final section concludes the paper with discussions and recommendations for future research. Limitations of this study are also discussed.

## 2.3. Literature Review

Effective public engagement in transportation planning leverages the sense of community by bringing citizens with common goals together for better decision-making. Evans-Cowley and Griffin described various involvement techniques such as public meetings, surveys, Facebook, and Twitter. Each of these techniques has different levels of citizen involvement and effectiveness. They concluded that although public meetings have a higher opportunity cost - a typical meeting takes an hour of time plus travel - they often result in low levels of citizen involvement. On the other hand, Twitter offers the lowest opportunity cost - it takes only a few seconds - and results in medium levels of citizen involvement [12]. Online social networking platforms provide a type of tool that supports interaction between groups of people sharing common interests. For example, this study uses online social networks to discover public opinions toward dockless bike sharing. One prominent feature of social networking tools is microblogging. A microblog is a posting of a short message that can include phrases, comments, images, audios, videos, and URLs. It allows a user to share information with their social networks. According to a Transit Cooperative Research Program (TCRP) report, transport agencies use social media plat-forms for one or more of the following reasons:

1. timely updates;
2. public information;
3. citizen engagement;
4. employee recognition; and
5. entertainment [13].

A significant number of studies have been done in the field of sentiment analysis on Twitter during the last decade [14, 15, 16]. Many recent studies focused on novel uses of social media platforms in transportation. Hoang et al. examined real-time commuter feedback to capture bus-related micro events such as an accident, missing the bus, and skipped stops [17]. In a similar approach, Wojtowicz and Wallace assessed the role of social media in gaining a better understanding of non-routine events in transportation management [18]. Liau and Tan used a mixture of statistical techniques to segment customers' opinions toward low-cost airlines in Malaysia. They showed that the findings could help airlines with better operational decision-making [8]. Collins et al. evaluated commuter satisfaction using sentiment classification on 557 tweets on Chicago's rail transit system. They suggested that using sentiment analysis has several advantages over traditional surveys including:

- low cost of data collection;
- real-time data collection; and
- data much more focused on user-specific needs [19].

Kaufmanand Moss suggested a "co-monitoring" approach by combining analysis of real-time social media feeds and staff reports. The findings showed improved responsiveness as well as better partnerships between agencies and customers. The approach also saves cost and time com-pared with traditional methods [20]. Schwitzer studies how social media portrayed public transit and examined the way that communities might be influenced. The findings suggested planners and transit agencies have a stronger commitment to the use of social media [21].

The second group of works focused on topic modeling. This approach is gaining increasing attention in text-mining research. Papadimitriou et al. introduced latent semantic indexing based on spectral analysis of the term-document matrix to retrieve information in a corpus [22]. Blei et al. explained the basic idea of topic modeling with the probabilistic model called latent Dirichlet allocation (LDA) method to extract an underlying set of topics [23]. Koltsova and Koltcov proposed an interval semi-supervised latent Dirichlet allocation (ISLDA) approach to improve the LDA approach. They used a dataset of Live Journal posts written by $2,000$ top Russian bloggers from January to April 2013. Their findings suggested that the ISLDA is a preferable method over LDA. Also, they presented the term frequency–inverse document frequency (TF-IDF) coherence metric for evaluating topic quality [24]. Sotiropoulos et al. combined topic modeling and sentiment analysis algorithms with a semantically aware clustering procedure to identify the distribution of overall public sentiment toward two U.S. telecommunication firms [25]. Steinskog et al. extended the LDA approach by applying several pooling techniques to aggregate similar tweets. Their method alleviates the low performance of topic modeling on tweets because of the short text format of tweets. Their findings established that aggregating similar tweets increased the topic coherence [26]. There have been several studies on the application of social media in transportation planning; for further readings, the review by Rashidi et al. provides a comprehensive overview of most recent case studies [27].

### 2.4. Methodology

We address the research gap from two different perspectives:

1. sentiment analysis of dockless bike sharing-related tweets; and

2. topic modeling.

The former focuses on the classification of tweets where several classifiers are trained on publicly available datasets. Finally, the trained model is tested on the actual dataset to identify positive and negative tweets concerning dockless bike sharing. To evaluate the classification results, we also developed an unsupervised clustering model to check if the resulting clusters are in line with the classes found by the classifier. In the second part, we use statistical modeling techniques to detect trending topics and events in a corpus of tweets.

#### 2.4.1. Data Preprocessing

The analysis covered two and a half months (from May 1, 2018, to June 30, 2018) and included $32,802$ tweets in English using Twitter API. Tweets with the hashtag (#) of "dockless," "bike sharing," and "bike share" are chosen as the sample in this analysis. The hashtag (#) is useful in the categorization of tweets, and also helps simplify the process of searching for tweets. Retweets (tweets that start with "RT") are treated as duplicates and are removed from both sets of data. Tweets are then cleaned by removing punctuation, special characters, digits, emoticons (such as emojis) and uniform resource locators (URLs) so that the dataset contains only words.

Tokenization is the process of breaking up a sequence of strings into pieces called tokens. Tokenization aims to explore the words in a sentence and identify meaningful keywords. Punctuation was removed in the process of tokenization. Tokens can be made up of characters, numeric or alphanumeric. Following this, stop-words are removed from the tweets. Stop-words are words from the non-linguistic view that do not carry information. Prepositions (such as "from," "to," "after," etc.), articles (such as "a," "an," and "the") and pronouns (such as "I," "you," "she," "he," etc.) are treated as stop-words in our work. Eliminating stop-words helps to improve text processing performance. Next, word stemming is executed. Word stemming is a process of transforming words into their roots. The stem, root, is often called "lexeme" by linguistics and is the smallest unit of a word; for example, "stemming," "stemmed," and "stems" have the same root word: "stem." Lastly, capital letters are converted to lower case. Tweets are then converted into a corpus. A corpus is a large and structured set of texts. The two sub-sections below discuss the two main techniques that are used to summarize the public opinions of dockless bike sharing on the corpus of tweets.

### 2.4.2. Feature Representation

Texts have many distinct properties. There are several ways to extract the features within a text and present it in proper format for the classifier. TF-IDF method is a statistical measure for evaluating how important a word is to a document in a corpus. The importance increases proportional to the frequency of an appearance of a word in a document and is offset by the frequency of the word in the corpus. This method converts textual data into a numerical vector which is later used in the classifier.

Unigrams, bigrams, and n-gram models, with their frequency counts, are considered as features. N-grams are all combinations of adjacent words or letters of length n in a continuous sequence from a given text or speech. It is widely used in text mining and natural language processing (NLP) to develop the features for supervised machine learning techniques.

There are two primary techniques for sentiment analysis for the Twitter data:

1. machine learning; and

2. lexicon-based.

In the former approach, the model classifies public opinions into positive, neutral, or negative classes. The sentiment polarity of a tweet is determined by comparing all the opinion words in the tweet against the subjective words in the dictionary and aggregating these words to give a final opinion for each feature. The lexicon method assumes that the contextual sentiment orientation is the sum of the sentiment orientation of each word or phrase in the text. Thus it uses a sentiment dictionary with opinion words and matches them with the data to determine polarity. Sentiment scores are assigned to the opinion words to describe how positive, negative, or objective the words contained in the dictionary are. This study focuses on the machine learning approach and considers three commonly used classifiers, as described in literature review:

1. naive Bayes;

2. logistic regression; and

3. support vector machines (SVM).

#### 2.4.2.1. Naive Bayes

Naive Bayes is a probabilistic classifier based on Bayes' theorem. The model is simple, fast, reliable, and accurate in many NLP applications. In the context of text classification (positive or negative classes), the occurrence of document $d_i$ depends on only two mutually exclusive events, $c$ and not $c(\bar{c})$. Therefore, the probability that a document $d_i$ belongs to a class $c$ is calculated by the Bayes' theorem as where $p(c|d_i)$ is the posterior probability arrived at by calculating $z_{ic}$.

$$p(c|d_i) = \frac{p(d_i|c)p(c)}{p(d_i)} = \frac{p(d_i|c)p(c)}{p(d_i|c)p(c) + p(d_i|\bar{c})p(\bar{c})} = \frac{\frac{p(d_i|c)}{p(d_i|\bar{c})}p(c)}{\frac{p(d_i|c)}{p(d_i|\bar{c})}p(c) + p(\bar{c})} \tag{2.1}$$

$$z_{ic} = \log \frac{p(d_i|c)}{p(d_i|\bar{c})} \tag{2.2}$$

$$p(c|d_i) = \frac{e^{z_{ic}}p(c)}{e^{z_{ic}}p(c) + p(\bar{c})} \tag{2.3}$$

#### 2.4.2.2. Logistic Regression

Logistic regression classification is based on the logistic function method. The logistic function is also called the sigmoid function. It is an S-shaped curve that can take any real-valued number and map it into a value between $0$ and $1$, but never exactly at those limits. Whereas naive Bayes is a generative model, meaning that the model calculates the joint probability distribution $p(x, y)$, logistics regression is discriminative in that the model learns the conditional probability distribution $p(x, y)$.

#### 2.4.2.3. Support Vector Machines (SVM)

SVM is another popular classification technique. SVM analyze the data, define the decision boundaries and use the kernels for computation which are performed in the input space. The input data are two sets of vectors each of size m. Then each data point, which is represented as a vector, is classified into a class. The model then finds a margin between the two classes that are far from any document. The distance defines the margin of the classifier. Maximizing the margin reduces indecisive decisions. Aramaki et al. addressed the problem of detecting influenza epidemics by applying text-mining techniques to tweets. Their findings show SVM, logistics regression, and naive Bayes have the highest performance among other classifiers in their application [28].

## 2.5. Results and Findings

There are $18,999$ unique users represented among the $32,802$ tweets. Of those users, $16,688$ have enabled their locations at the account level. These locations do not necessarily represent the locations from which the users posted the tweets because they do not frequently change the locations. At the tweet

Table 2.1. Tweets by location.

| Location | Percentage of tweets with location |
|---|---|
| The U.S. | 78.9% |
| The U.K. | 4.8% |
| Canada | 4.2% |
| Australia | 1.6% |
| India | 1.4% |
| China | 0.8% |
| Others (38 countries) | 8.3% |

Table 2.2. U.S. tweets by location share.

| City | Share | dockless system available |
|---|---|---|
| Washington, D.C. | 14.6% | Yes |
| San Francisco, CA | 8.3% | Yes |
| Seattle, WA | 5.2% | Yes |

level, there are two parameters, "place" and "coordinates," associated with the locations that the user is tweeting from. The parameter "place" describes a larger area whereas "coordinates" defines an exact point location. Out of $32,802$ tweets, only 61 tweets have "coordinates" (almost $0.1\%$), and $1,178$ tweets have "place" ($3.6\%$). As shown in Table 2.1, the majority of tweets with locations are from the United States. Table 2.2 shows that the location share of more than $5\%$ of $929$ tweets (78.9%) originated in the United States. After removing retweets, $16,431$ tweets remained in the database. They contained $9,853$ "User Mentions," $16,832$ "URLs," and $72$ "Emojis" of which there were $50$ positives and $22$ negatives. There are $210,979$ words that contain $13,402$ unique words. Figure 2.1 illustrates the high-frequency words used in the collection of tweets before preprocessing. It verifies the dominant theme of collected tweets which is mainly dockless bike sharing. The average length of the tweets is $115.38$ and $87.13$ characters respectively, before and after cleaning.

### 2.5.1. Sentiment Analysis

Since the collected tweets do not have any labels (positive or negative), so it is impossible to develop a supervised model. Hence, we took advantage of a Stanford Twitter Sentiment Corpus [15] to develop a supervised model. Then the trained model was applied to our dataset to classify tweets as negative or positive. This dataset has $1,600,000$ training tweets of which $800,000$ are positive and $800,000$ are negative. The model is then used to classify the collected tweets on dockless bikes. For further validation, we also did the clustering to check if the supervised and unsupervised method generate the same results. The dataset is


Figure 2.1. Word cloud illustration of tweets.

split into three subsets: training, validation, and testing with $60\%$, $20\%$, and $20\%$ of the tweets, respectively. This method is preferred over k-fold cross validation because of the size of the datasets.

We started the sentiment analysis by training the three supervised machine learning algorithms (naive Bayes, logistics regression, support vector) on the Stanford dataset. Table 2.3 describes the summary of accuracy scores for the trained models with the $10,000$ unigram feature for TF-IDF. It shows that the logistics regression model has the highest performance within a reasonable amount of time. Therefore, we chose it for the final assessment.

We used a varying number of n-gram features for TF-IDF to improve the performance of the model. Figure 2.2 illustrates the accuracy of the models with an increasing number of features given n-gram. As shown in Figure 2.2, 3-gram TF-IDF achieves the highest accuracy among other cases. The performance in relation to CPU time is an indication of how each of these models could be practical when the size of tweet corpus increases dramatically. However, this issue is not critical as GPU processing facilitates heavy processing. Because the performance of the models is almost equal, we selected the model with the least CPU time for further analysis. The logistics regression model with a $100,000$ 3-gram feature was chosen for the sentiment analysis. We tested the model on the corpus of tweets and $73.8\%$, $26.2\%$ are classified as positive and negative sentiment, respectively.

Table 2.3. Models' accuracy and CPU time by the number of features.

| Multi-col-row | Accuracy | | Time (seconds) | |
|---|---|---|---|---|
| | 10,000 features | 100,000 features | 10,000 features | 100,000 features |
| Logistics regression | 78.18% | 78.48% | 32.17 | 36.28 |
| Linear SVC with L1 regularization | 78.04% | 77.84% | 459.41 | 448.61 |
| Multinomial Naive Bayes | 76.60% | 76.62% | 14.87 | 13.92 |

Figure 2.3 shows the sentiment probability distribution. Tweets are labeled either positive or negative if their probability of given class is bigger than $0.5$. Therefore, the higher the probability, the higher the confidence the tweets are labeled correctly. In total, tweets with positive sentiments have higher probability whereas negative tweets have, on average, lower probability. In other words, the model labels the tweets as positive with higher confidence than negative ones. Manual validation of $32,000$ tweets requires too many resources. To facilitate the evaluation of supervised models (the classification results), we used K-means which have been applied in a wide range of applications to cluster unlabeled tweets [29, 30]. The resultant groups comprised of $68.4\%$, and $31.6\%$ of the tweets. This is in line with the classification results in which one group has almost $70\%$ of the tweets. In the next step, we compare the original tweet with the results of the classification and clustering sections.

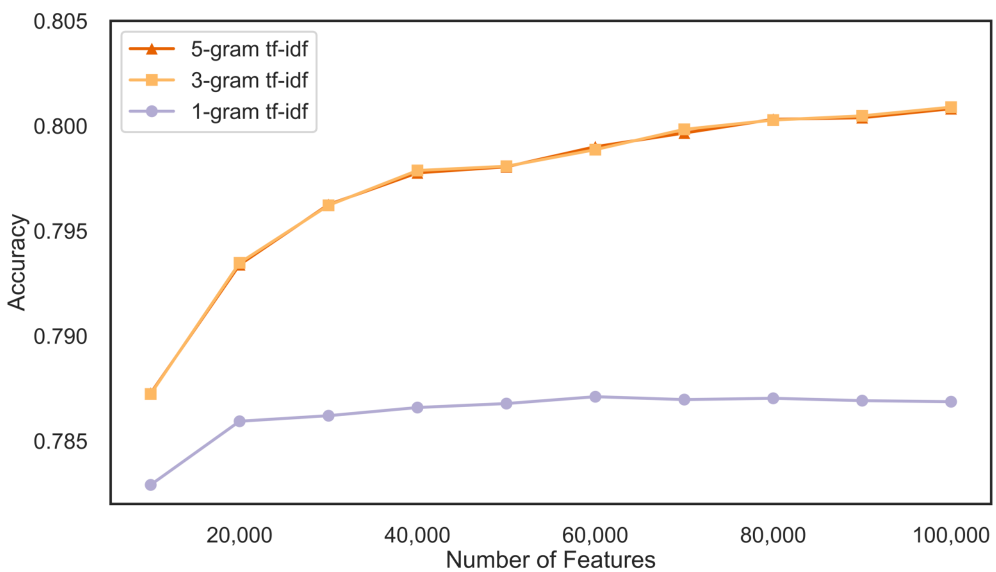


Figure 2.2. Accuracy of n-gram TF-IDF and number of features.

As shown in Table 2.4, further investigation into classification and clustering shows some interesting results. Of the labels, clustering and classification methods agree in only $61\%$ of the cases. In some

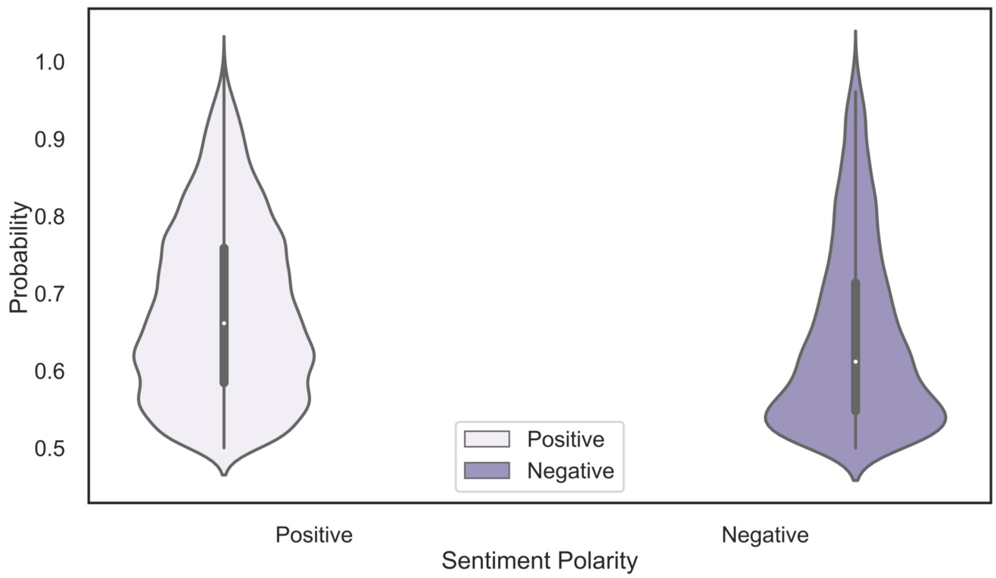


Figure 2.3. Sentiment polarity distribution of tweets.

instances, neither technique generates the same label. This could be due to several reasons. One might be K-means clustering with Euclidean distance are prone to high-dimensional data as we have in this case for TF-IDF. This could be because of the model error (row 2, 5, 6, and 7). However, in some cases, the tweets have a different sentence structure, for example, a question in which polarity is neither positive or negative. In some cases, the user shows agreement on a URL which contains a negative opinion toward dockless bike sharing. Because were moved URLs during data preprocessing, the ultimate sentiment label of the tweet is positive

### 2.5.2. Topic Modeling

Many people post messages about various issues on Twitter every day. The investigation of the relationship between the underlying topics of different authors' messages could yield interesting results about peoples' perceptions. One could, for example, find out the gender differences in the perception of individuals toward risks associated with genetically modified organisms [9]. In another case, one could identify the similarities and differences among politicians on various topics [26]. Twitter has an abundance of messages, and the enormous number of tweets posted every second makes Twitter suitable for such tasks. However, detecting topics in tweets can be challenging because of the informal language used and because tweets usually are less coherent than traditional documents. The community has also spawned user-generated meta tags, like hashtags and mentions, which have analytical value for opinion mining. This paper describes a system aimed at discovering trending topics and events in a corpus of tweets, as well as exploring the topics of different Twitter users and how they relate to each other. Utilizing Twitter metadata mitigates the dis-advantages tweets typically have when using standard topic modeling methods. User information, as well as hashtag co-occurrences, can give much insight into topics that are currently trending. Topic models are types of statistical methods used to represent abstract and latent topics in document collections. These probabilistic models usually present topics as multinomial distributions over words, assuming that each document in a collection can be explained as a mixture of topics. In this paper, we use the LDA algorithm to discover underlying topics in the collection of tweets. LDA is a generative probabilistic model of a corpus and has many applications in collaborative filtering, content-based image retrieval, and bioinformatics. Many studies described the mathematical formulation of the method in detail [23].

We applied the LDA algorithm to our dataset and the results are presented in Table 2.5. We listed seven important topics, each of which has five sub-topics. Regardless of the word "bike," or "city," as they show the overall theme of the tweets, there are some interesting sets of words about dockless bike sharing. On one hand, there are general words such as cheap, flooding, scooter, side-walk, and threat. On the other hand, there are proper names such as Uber, Lyft, Motivate, and Sharelock as well as city names including Chicago and Seattle.

In the case of sub-topic "sidewalk," users perceived sidewalk negatively, and only a minority of this cluster have positive sentiments. With the number of tweets giving negative feedback on bikes blocking sidewalks, it verifies the news about the negative externalities of the dockless programs. Hence city officials should plan proactively to mitigate this problem. For example, some practices reduce these negatives: painting areas on the ground to show proper parking places, hiring maintenance workers to take care of abandoned bikes, calling on people to behave appropriately, or enabling a system of credit checks based on user behaviors (by reviewing reported photos). If the user credit is lower than a certain threshold, they will need to pay higher prices to rent a bike.

One exciting sub-topic is "scooter" which is not the focus of this study. However, in some cases, people mention dockless scooters with dockless bike sharing. The dockless electric scooter is the second emerging trend after dockless bike share. Scooter sharing first emerged in Santa Monica, CA, in September 2017 by scooter share startup Bird. Later, Lime, Spin, and others deployed the technology in San Diego, CA; San Francisco, CA; San Jose, CA; Washington D.C.; and Austin, TX. The companies are expanding quickly and are rapidly gaining popularity. This information provides an idea for bike sharing management on how efficient, prompt replies will resolve most of the problems discussed in each cluster above.

Table 2.4. Tweet sentiment polarity from classification and clustering.

| Classification (supervised method) | Clustering (unsupervised method) | Original tweet |
|---|---|---|
| Positive | Positive | "If Portland's bikeshare sees a big increase from going dockless in May (and I think it will), it will be a turning point for dockless bikeshare. Urbanists may finally stop clinging to the dock technology that makes trips longer :)" |
| Negative | Positive | "The 'public' (at the moment) seems too irresponsible for dockless. In interim, what abt a lightweight infrastructure play - A wardrobe-sized automatic/robotic locker for bldg entries. High density stacking, dispenses scooters when needed." |
| Negative | Negative | "Considerations for dockless bikes: What happens to bikes left as sidewalk obstacles? What happens when bikes left against trees or other tender plantings/property? Who is going to work to prevent dockless bikes from becoming sidwalk "trash" in a sense?" |
| Negative | Negative | "@quick13 Saw dockless bikes recently in Charlotte. Bikes scattered all over." |
| Positive | Negative | "@NiceRideMN Having just been to Seattle, one thing that caught my eye about the docklessbikes there is the wild abandon with which the bikes get abandoned in weird places." |
| Positive | Negative | "So cool! @GridBikes rebalances its docks/stations using cargo bikes! #bikeshare" |
| Negative | Positive | "#Bellevue's dockless #bike share pilot will only allow #e-bikes A few months after #Seattle'sbike share pilot technically ended, Bellevue is rolling out their own pilot program governing #dockless bike share—and there are a few changes. Do you have..." |

Table 2.5. Topics and subsequent subtopics extracted by the LDA approach.

| Topic 1 | Topic 2 | Topic 3 | Topic 4 | Topic 5 | Topic 6 | Topic 7 |
|---|---|---|---|---|---|---|
| bike | Lyft | city | get | bike | bike | system |
| hire | Uber | scooter | new | car | scooter | Sharelock |
| world | buy | bike | program | city | Austin | bike |
| cheap | Motivate | company | Chicago | scooter | people | station |
| flooding | big | new | pilot | sidewalk | Seattle | thread |

### 2.6. Conclusions and Implications

bike sharing companies, city authorities, and public stakeholders are continually refining insights into public needs, preferences, experiences, and opinions on their services. First, by studying Twitter-user sentiments toward dockless bike sharing, we found more positive than negative polarity across the different algorithms. Hence, users, in general, are satisfied with this emerging technology. Second, there are seven main groups of topics that are discussed by Twitter users. Although these seven topics are commonly discussed, this research can enable a deeper understanding of the sentiment underlying those seven topics.

People use Twitter as a platform to express their emotion and the problems they faced in using any mode of transportation. Despite the difference between positive and negative polarity, bike sharing companies should put more effort into improving the overall image of the dockless program, as people are prone to post any disturbance on Twitter. The blockade of bikes on the side-walk plays a crucial role. bike share companies have to come up with a practical solution, or the current popular opinion will erode to a primarily negative perception of dockless bike sharing.

Opinion mining, as a derivative of NLP, is emerging in many fields such as customer analytics, health care, politics, and others. As in traditional survey methods, tweets may not reflect true opinions on a subject area. However, since social networks have become popular over the last decade, the application of NLP to social media posts could be seen as a new approach to opinion mining. There are two things which make the social network analysis unique:

1. people, regardless of the potential social bots, discuss subjects deliberately without being concerned that their tweets might be analyzed. However, the survey techniques might be prone to bias due to social desirability and impression management; and

2. in traditional methods, participants are asked specific questions whereas topic modeling gives the flexibility to social network analysis for extracting various subjects without a set of predefined questions.

Transportation science has the potential to use this opinion mining approach because of the interactions between people as customers and several transportation entities. However, there are several concerns regarding the opinion mining of social networks:

- intuitively, people usually do not post something about transportation services unless something happens. This event-triggered behavior might produce more negative views about transportation services.

Contrary to this expectation, in this study we found more positive sentiment. Because bike sharing is a new phenomenon, this could be a possible reason for our findings. Further analysis might be helpful to understand the real impact of dockless systems after such systems have been fully established;

- one of the prominent issues with Twitter sentiment analysis is the quality of the tweets. Some online posts are retweets without any additional words, emojis, or tags. These types of tweets are removed early in the preprocessing step since they could not reveal users opinion explicitly. On the other hand, some tweets do not have correct grammatical formats and may be missing com-mas, periods, or correct tense. These lead to false classifications. Advanced preprocessing techniques could alleviate this problem;

- online platforms are open to everyone. This level of flexibility could have an impact on someone's opinion toward a subject. Not only the tweets, but also the user profile might influence the perception of others toward the user's tweets. Unlike traditional surveys, in which other participants are anonymous, online platforms to some extent have exposure to others;

- a reliable source is a critical subject in the social network. Social network analysis is prone to social bots. Social bots are to some extent autonomous agents with the intention of influencing users on social media. Social bot detection is complicated, requiring separate research which is not the scope of this study. In addition to bots, the frequency of a user's tweets might also increase the bias of this approach;

- from a transportation planning perspective, every transportation study is to some extent limited to a geographic area. One might be broader than another. Social network analysis including Twitter, Facebook, and others fails to address this issue adequately. Because of users' privacy preferences, locations are either missing or not well represented. Users may choose, at their discretion, to enable/disable location tracking on their devices. For example, someone may post a tweet about an event without adding a location to the tweet - which happens in most cases - or someone may get off a plane and tweet an opinion (complaint) at their destination about the delay at the origin. These are just examples that call the social network analysis into question. As described earlier, less than $4\%$ of tweets can be linked to specific locations. Moreover, a longer observation period does not guarantee getting better geo-tagged data. For example, for a user on a location-enabled device used to tweet negative comments on a topic, longer observation would likely lead to the user having more negative tweets related to a specific location. On the other hand, a user without a geo-enabled device does not have any impact on location-related tweets.

In conclusion, bike sharing companies, stakeholders, and city officials would have a better opportunity to understand public opinion and address the issues of concern, and in turn attract more riders or customers, generate more profits, and provide more mobility options, if these problems are addressed. This study provides insight into public sentiment toward emerging dockless bike share programs. Structured interviews or surveys can be affected by response bias including the demeanor of the surveyor, the phrasing of questions, the way an experiment is conducted, and the desires of participants to act as a good experimental subject and to provide socially desirable responses [31, 32]. Hence, opinion mining on social media data could be a complement to traditional methods, as users voluntarily give their opinions on the subjects of interest when compared with surveys conducted where no interview is involved. Also, this study gives a clearer picture of dockless programs with regard to topics which are broadly discussed in social networks. Officials can identify valuable users by making proactive decisions based on the predictions of users' future behaviors. If bike share companies can analyze customer tweet data in real time by classifying customers' feedback, it will help management facilitate strategic, operational activities. Opinion mining of social networks in transportation would be beneficial for identifying new and emerging phenomena in a community. However, a well-designed survey questionnaire and in-depth analysis is undoubtedly required for any planning and decision-making.

There are several avenues for future research. First, clustering could be a reliable method for sentiment categorization in case there are no labels available. Second, photos, videos, and URLs in tweets are a great source of information to improve the sentiment analysis model. Future works should focus on developing methods to consider the relationship between textual inputs and other forms of data. Third, hashtags and emojis are usually excluded in the data preprocessing step. Future studies may investigate if hashtags and emojis have any contribution to consumer sentiments. Finally, in early 2020, the COVID-19 crisis forced some shared micromobility companies like Bird or Lime to stop operations in many markets resulting in several layoffs. Rider safety and virus containment cited as primary reasons for such decisions. Future research could investigate if there is any change in public sentiment toward public transportation and shared micromobility services. Despite these limitations, we believe that our study has contributed to this subject, especially with regard to consumer behavior toward dockless bike sharing.

# 3. MAXIMUM CLOSENESS CENTRALITY K-CLUBS: A STUDY OF DOCKLESS BIKE SHARING

## 3.1. Abstract

In this work, we investigate a new paradigm for dockless bike sharing. Recently, it has become essential to accommodate connected and free-floating bicycles in modern bike sharing operations. This change comes with an increase in the coordination cost, as bicycles are no longer checked in and out from bike sharing stations that are fully equipped to handle the volume of requests; instead, bicycles can be checked in and out from virtually anywhere. In this paper, we propose a new framework for combining traditional bike stations with locations that can serve as free-floating bike sharing stations. The framework we propose here focuses on identifying highly centralized $k$-clubs (i.e., connected subgraphs of restricted diameter). The restricted diameter reduces coordination costs as dockless bicycles can only be found in specific locations. In addition, we use closeness centrality as this metric allows for quick access to dockless bike sharing while, at the same time, optimizing the reach of service to bikers/customers. For the proposed problem, we first derive its computational complexity and show that it is $\mathcal{NP}$-hard (by reduction from the 3-SATISFIABILITY problem), and then provide an integer programming formulation. Due to its computational complexity, the problem cannot be solved exactly in a large-scale setting, as is such of an urban area. Hence, we provide a greedy heuristic approach that is shown to run in reasonable computational time. We also provide the presentation and analysis of a case study in two cities of the state of North Dakota: Casselton and Fargo. Our work concludes with the cost-benefit analysis of both models (docked vs. dockless) to suggest the potential advantages of the proposed model[1].

## 3.2. Introduction

bike sharing systems (BSSs) have become a prominent mode of transportation around the world, especially in urban areas. BSSs bring a number of advantages to existing transportation networks. Among them, we note the increased personal mobility, reduced transportation costs, reduced traffic congestion, decrease in use of and dependence in fossil fuel, increase in public transit visibility, enhancement of down-

[1]The material in this chapter was co-authored by Ali Rahim Taleqani, Chrysafis Vogiatzis, and Jill Hough. Ali Rahim Taleqani also drafted and revised all versions of this chapter. This chapter appears in Journal of Advanced Transportation (Ali Rahim Taleqani, Chrysafis Vogiatzis, and Jill Hough, 2020)

town areas along with the economic development that follows, health benefits, and increase in environmental awareness [33, 34, 35].

Since their introduction in Europe in the 1960s, BSSs have undergone a series of developments. The most recent of these developments is referred to as the *dockless* or *free-floating* BSS. In a dockless BSS, residents that are interested in using a bicycle can check out and in bicycles throughout an urban area using nothing more than their smartphones. The bicycles are equipped with a geographic positioning system (GPS), thus enabling users to locate the nearest available bicycle and to unlock it with the use of an app. Riders are then allowed to drop off (check back in) the bicycle anywhere within a geographic area (referred to as the geo-fenced area). Within that area, bicycles are allowed to be parked legally. The trip ends as soon as the checked out bicycle is parked and securely locked anywhere in the geo-fenced area.

As is obvious from the description, dockless or free-floating bicycles offer enhanced convenience and improved accessibility, which in turn translates to increased personal mobility, compared to conventional bike sharing. The enhanced convenience stems from the fact that users no longer have to wait for a parking spot to become available in a bicycle dock so as to return their bicycle after the trip (especially in heavily trafficked areas). However, as with many other technologies, dockless BSSs also present new challenges. The one we deal with in this work is the fact that bicycles can now be left unattended in improper positions [36, 37]. Piles of bicycles from bike sharing companies are increasingly becoming a public nuisance and they make the aesthetic environment of an urban area less enjoyable with clogged sidewalks that are rendered no longer walkable. Hence, a dockless BSS development without proper control and regulation mechanisms can result in blocked rights of way, increase in trash, vandalism, and other negatives.

The framework we develop here aims to address this issue of control while at the same time advocating for dockless bike sharing. We propose a framework that will both (i) allow users the increased benefits of a dockless system (easy and fast access to bicycles, reduced parking space needs) and (ii) reduce the coordination costs for controlling the sprawl of the dockless bike sharing operations by restricting the size of the geo-fenced area.

Continuing with our motivation, a critical component to the success of every BSS is that users can check out a bicycle within convenient walking distances [38]. This simple, yet powerful, principle guides BSSs to offer the right number of bikes in the right locations at the right time so as to accommodate daily commuting demands. In other words, station location and density are key factors in any BSS [39]. In the traditional dock-based system, BSS operators are expected to rebalance bike inventory between different stations so as to meet (asymmetric) demands. As an example of such rebalancing operations, Chiariotti et al. (2018) proposed a dynamic model to address the fluctuations in demands of a BSS in New York City [40]. In a similar note, Wang et al. (2018) applied a data-driven approach for defining a safe rebalancing range and provide rebalancing operators with the next targeted station and the number of bikes to move [41]. Rebalancing bicycle inventory imposes extra costs associated with human and physical capital on bike sharing programs. These costs can grow to be considerable if the system is large. Moreover, the rebalancing problem is even more pronounced in dockless BSSs because of unrestrained parking locations [42]. Finding the right locations for stations and rebalancing are correlative problems. According to the Department of Transportation, a dockless program should be initiated where demand is highest and designated bike parking areas, referred to as bike hubs, should be used to maintain some order. This policy would help mitigate the hodgepodge of problems that can result from adopting a dockless system [43].

As our framework will optimize the reach of dockless bike sharing operations while also restricting the size of the system, our model will also alleviate some of the problems involved with rebalancing. To further elaborate on our model, we offer Figure 3.1. On the right, we present a conventional dock-based BSS. The transportation network is presented with nodes and edges (representing streets), with the bicycle docks being noted with blue rectangular nodes: observe that docks are not necessarily located in nodes only, but can also be located along the edges of the network. On the other hand, the figure on the right shows our proposed framework. We now allow for a geo-fenced area (represented by the shaded area) where users can check out and in bicycles from anywhere. This allows for more people to have fast access to bicycles and reduces the need for docks within that area. Due to that, these docks could be moved to other areas, further than the geo-fenced area, to enable bike sharing use to other residents. In addition to that, the area where bicycles can be dropped off anywhere is significantly decreased, making it easier for operators to find and collect bicycles so as to rebalance their inventory. Last, we note here that the shaded area of the network on the left forms a 2-club (i.e., a subgraph of diameter equal to 2).

We can summarize our contributions in the following three components:

- first, we use the $k$-club concept, combined with closeness centrality, so as to identify candidate locations that could be geo-fenced. We also allow for a weight at each node of the network: this modification enhances the speed of the $k$-club formation through the heuristic algorithm devised;

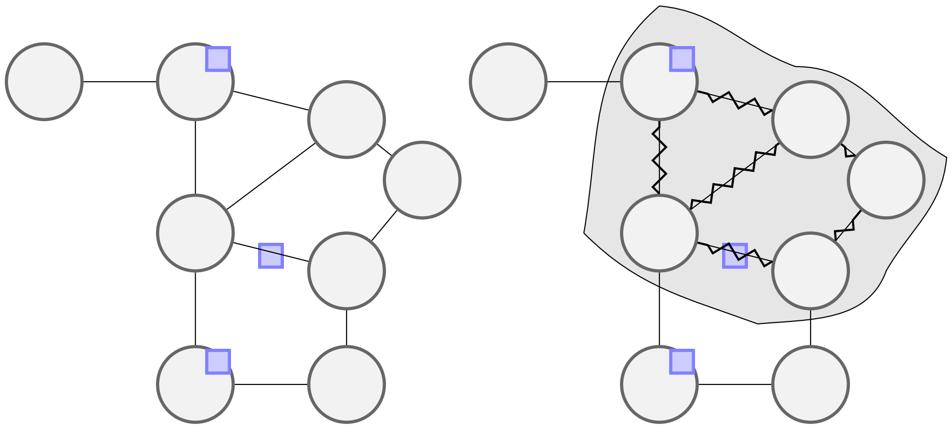

Figure 3.1. Graph without k-club (left), with k-club (right).

- then, we turn our attention to a real-world application. We present an experimental study on the cities of Fargo and Casselton. In the study, we analyze the exact optimization model and the heuristic devised and compare them in computational time and solution obtained. In each $k$-club obtained for varying values of $k$, riders (commuters) are able to reach to any other neighboring sites within a fixed distance (controlled by $k$), implying that the virtual locations provide better accessibility to demand points;

- last but not least, we present potential strategies for operators to further manage the inventory by applying incentives and making bicycle collection and rebalancing more cost-effective.

The remainder of the paper is organized as follows: the next section reviews related literature on BSS design and discusses how those relate to the objectives of this work. Then, we provide the necessary mathematical background, provide the definitions of all notation used, and derive the computational complexity of the problem studied. The next section illustrates the mathematical formulation that can be solved using a commercial optimization solver and also proposes a greedy heuristic to solve it. In the following section, we discuss two computational experiments that reveal our findings in two real-world transportation networks: namely the smaller city of Casselton, ND, and the larger city of Fargo, ND. However, due to the size of the network in Fargo, we only test and present the results of the heuristic approach. The last section of the paper is devoted to our conclusions and a brief overview of future plans.

### 3.3. Literature Review

There is a plethora of studies on bike sharing systems. These studies generally fall into three major areas:

1. general quantitative analysis;
2. facility location problems; and
3. redistribution problems.

The first body of literature focuses on the quantitative analysis of existing BSSs, analyzing their characteristics, and examining empirical evidence of usage patterns in cities including Dublin [44], Beijing [45], Montreal [46], Brisbane [47], Helsinki [48], Paris [49], Switzerland [50], and New York [51]. Nair et al. (2012) examine several aspects of the Velib BSS in Paris, France [49]. Their findings show that integrating transit and BSS can yield higher utilization. Bachand-Marleau et al. (2012) surveyed residents of Montréal, Quebec, in Canada to determine the factors leading to use BSS as well as the frequency of use [46]. Campbell and Brakewood (2017) quantify the impact that BSSs have on bus ridership in New York City [51]. They conclude that either bike sharing members substitute bike sharing for bus trips or the implemented BSS led to travel behavior changes of non-members. Audikana et al. (2017) studied the impact of a BSS in a small city (less than $100,000$ residents) in Switzerland [50]. They suggested that BSS network density along with the developed partnerships play a critical role in its success.

The second stream of literature focuses on the strategic design of BSS where the ultimate goal is to find the locations, capacity, and coverage areas of BSSs [52]. These studies try to determine the number and location of stations, fleet size, and network structure of the underlying BSSs. They consider various objectives, including the maximization of demand coverage, the minimization of transportation cost, and the minimization of the overall cost. Lin et al. (2013) address the strategic design problem by formulating it as a hub location inventory model [53]. In their work, they consider both total costs (travel cost of users, bike inventory costs, facility costs) and service level (bicycle lanes) in their model. The authors then propose a heuristic method to find high-quality solutions. In a similar study, Lin and Yang (2011) propose a nonlinear integer method to determine the optimal location, bike lanes, and routes [54]. Their model assumes a penalty for uncovered demand but does not consider relocation (rebalancing) of bikes. Martinez et al. (2012) present a mixed integer linear program to maximize the net revenue by simultaneously optimizing the locations of stations, the fleet size, and bike relocation activities for a regular operation day [55]. Nair and Miller-Hooks (2016) formulate an equilibrium network design model to address the same objective as the previous study [56]. They propose a metaheuristic solution approach to overcome the intractability of the exact solution for real-life, large-scale networks. In another study Reijsbergen (2016) identifies alternative locations with the aid of spatial data and simulation techniques: more specifically, a data-driven approach to determine how attractive city areas are for station placement is presented [57]. The literature offers other methodologies, that are not based on facility location models, to define the location of the stations. Garcia-Palomares et al. (2012) develop a GIS-based model to calculate the spatial distribution of the potential demand for trips and find the locations of bike stations by using the location-allocation modeling approach [58].

Finally, a third group of the literature is associated with the relocation of bicycles in a BSS. The problem arises from demand imbalance leading to accumulation of bicycles at some stations (and consequently, limited bicycle availability in other stations). Vogel and Mettfeld (2010) apply a system dynamic method to model the effect of dynamic repositioning on the service level [59]. Shu et al. (2013) develop a stochastic network flow model with proportionality constraints to determine bike flow in a bike sharing network. They also present a numerical analysis on the Singapore BSS and find that period distribution is the most effective for system performance [60]. Forma et al. (2015) develop a 3-step heuristic and mixed integer linear programming model for repositioning [61]. The first step involves clustering the stations based on geographic location and inventory levels using a heuristic method. In the second and third steps, they employ a mixed integer linear program to find the best routes for repositioning vehicles. Alvarez-Valdes et al. (2016) address the static repositioning problem using simulation techniques in two stages [62]. In the first stage, they estimate the levels of unsatisfied demand for a set of stations in a given period. In the second stage, they use the estimation as an input to their redistribution algorithm. Schuijbroek et al. (2017) combine service level requirements and vehicle routes to rebalance the inventory [63]. They propose a "cluster-first route-second" heuristic considering the service level feasibility and approximate routing costs simultaneously. Yan et al. (2017) develop four planning models for leisure-based BSSs given deterministic and stochastic demands [64]. They apply non-linear time-space network models to integrate bike repositioning and vehicle routing with user dissatisfaction estimations. In a recent study, Celebi et al. (2018) propose a hybrid approach jointly considering location decisions and capacity allocation [65]. Their goal is to find the optimal configurations of a BSS by combining set-covering and queuing models to determine service levels.

One of the gaps in the current state-of-the-art is that most focus only on either user accessibility or rebalancing strategies to manage supply and demand within an urban area. As described in the Introduction section, our contribution is to fill exactly that gap and propose a framework that allows for both high accessibility for the users and reliable and cost effective rebalancing and coordination for BSS operators. Our proposed model relies on the definition of a $k$-club from graph theory, whose definition and related literature is offered in the next paragraphs.

Given a simple undirected graph, a $k$-club is a subset of vertices inducing a subgraph of diameter at most $k$. These structures represent cohesive subgroups in social network analysis with common applications in network-based data mining and clustering. Several authors have discussed mathematical formulations for identifying $k$-clubs of maximum cardinality, as well as various methods to locate $k$-clubs within a network [66, 67, 68]. In addition to using $k$-clubs, our work also focuses on the centrality of a group of a specific structure. Group centrality, introduced by Everett and Borgatti (1999), aims to identify groups or classes of high centrality [69]. Centrality measures the aim to characterize the importance of an element in a network. They typically fall into three main classes [70], referred to as degree (i.e., the number of connections of a specific element in the network), closeness (i.e., how close an element is to every other element in the network), and betweenness centrality (i.e., how important an element is in the communications between any two other elements in the network, assuming all such communications take place using the shortest path between the elements).

More recently, researchers have focused on highest betweenness groups [71]. Finally, another extension of identifying highly centralized groups has to do with the added restriction that the group induces a subgraph "motif", such as being a complete subgraph/clique [72, 73], or inducing a star [74].

In this paper, we propose an integer programming formulation and a heuristic algorithm to find the most centralized $k$-club in a transportation network based on closeness centrality. The resultant $k$-club consists of a set of nodes in which the maximum traversing distance is $k$ hops (by definition), and the total weighted by population distance to a node in the $k$-club is minimized (as it will be the $k$-club with maximum closeness centrality). Based on this result, a BSS operator could then enable the area covered by the $k$-club as the geo-fenced area where dockless bike sharing is allowed and satisfy the following objectives:

1. maximize demand coverage (as the area obtained is the most centralized, with respect to closeness centrality);

2. minimize distances traversed for rebalancing operations (as the geo-fenced area is of restricted diameter); and

3. offer a large, convenient geographical area for checking in/out the available bikes without need for physical stations. As the success of a BSS heavily depends on the network of bike paths and bike stations in the community, this is an important objective facilitated by our framework.

Most of the previous work that addresses physical bike station location problems illustrates problems including station capacity decisions and demand predictions, among others. To the best of our knowledge, this paper is the first to suggest a solution to problems that have arisen from the emergence of dockless bike sharing systems with the aid of a $k$-club. The ultimate goal is to locate potential hubs in a city, referred to as $k$-clubs, by geo-fencing a suitably small area of a city.

### 3.4. Methodology

Let $G(V, E)$ be an undirected network, with $V$ symbolizing the vertices (intersections of the transportation/biking network) and $E$ the edges (streets in the transportation/biking network). Every node is assumed to be assigned a non-negative parameter, $w_i \geq 0$, referred to as the *weight* at this specific location. This weight parameter can be used to capture different aspects of the problem at hand, depending on the application. For example, the weight of a node could capture socio-economic attributes like population, points of interests in the vicinity, number of jobs, etc. Another possible way to model and use the weight parameter is through the interactions between different pairs of origin and destination, like traffic flows (outgoing traffic from an origin node, incoming traffic to a destination node, or simply a summation of outgoing and incoming traffic to a specific node). In either way, we assume a distinct, non-negative number explaining the level of attraction for that node.

We say that $(i, j) \in E$ if there exists an edge starting from node $i$ and ending in node $j$, in which case we write that $a_{ij} = 1$. We also denote with $N(i) = \{j \in V : a_{ij} = 1\}$ the open neighborhood of node $i$. We write that the diameter of graph $G$ is $D$ if the maximum shortest path distance between two nodes in the graph is $D$. Clearly, all pairs of nodes in the graph will be located at a distance $\ell$ from one another with $0 \leq \ell \leq D$. Let $d_{ij}$ be the distance between two nodes $i$ and $j$, and $d_{Sj} = \min_{i \in S} \{d_{ij}\}$ as the distance of a node $j$ to a set of nodes $S$. Then, for any set of nodes $S \subseteq V$, we define a function $f : V \mapsto \mathbb{R}$, as

$$f(S) = \sum_{i \in V} w_i d_{Si}. \tag{3.1}$$

Last, we use $\mathcal{P}^k$ to denote all paths of length less than or equal to $k$. Similarly, $\mathcal{P}_{ij}^k$ is the set of all paths of length at most $k$ connecting two nodes $i$ and $j$ ($i \neq j$). Clearly, we have that $\mathcal{P}^k = \bigcup_{i,j \in V: i \neq j} \mathcal{P}_{ij}^k$.

The decision version of the problem we are trying to solve is provided in Definition 2. Before that, we need to provide the definition of a $k$-club.

**Definition 1.** A set of nodes $S \subseteq V$ is said to form a $k$-club if the subgraph induced by $S$, $G[S]$, has a diameter $D \leq k$.

**Definition 2.** Given a graph $G(V, E)$ with a nonnegative weight $w : V \mapsto \mathbb{R}$ assigned to every node, an integer number $k$, and a real number $\ell$, does there exist a $k$-club $S \subseteq V$ such that $f(S) \leq \ell$?

Detecting a $k$-club of maximum cardinality is a well-known $\mathcal{NP}$-hard problem [75, 76]. Hence, it is expected that our problem, as described in Definition 2 will also be shown to be $\mathcal{NP}$-complete, rendering the optimization version $\mathcal{NP}$-hard. This is exactly what we show in Theorem 1. Before we do that, we define 3-SATISFIABILITY (3SAT), a famous $\mathcal{NP}$-complete problem.

**Definition 3** (3SAT)**.** Given $m > 2$ clauses $C_1, C_2, \ldots, C_m$ and $n$ literals and their complements $x_1, x_2, \ldots, x_n$ and $\overline{x}_1, \overline{x}_2, \ldots, \overline{x}_n$, does there exist an assignment such that a formula $C_1 \wedge C_2 \wedge \ldots \wedge C_m$ in conjunctive normal form is true, when every clause consists of exactly 3 literals?

**Theorem 1.** The decision version of our problem, as described in Definition 2, is $\mathcal{NP}$-complete.

*Proof.* The problem can be shown to be in $\mathcal{NP}$, as both verifying that a subset $S$ forms a $k$-club and that $f(S) \leq \ell$ can be done in polynomial time.

Now, consider an instance of 3SAT with $m$ clauses on $n$ literals. We will reduce it to a version of our problem using the following gadget/transformation. First, create two nodes for every literal and its complement ($V_\ell$); we connect every node by a chain of $k-1$ nodes ($V_{\ell \times \ell}$) to every other node, but its complement (this forms edge set $E_\ell$). Moreover, create one node for every clause ($V_c$); connect each node in $V_c$ by a chain of $M-1$ nodes ($V_{c \times \ell}$) to the literals that the corresponding clause consists of ($E_c$), where $M >> k$. Finally, assume that all nodes in $V_c$ have a weight of 1, while all other nodes in $V \setminus V_c$ have a weight of 0. We will show that the 3SAT instance has a feasible assignment if and only if the constructed graph $G(V, E)$ with $V = V_\ell \cup V_{\ell \times \ell} \cup V_c \cup V_{c \times \ell}$ and $E = E_\ell \cup E_c$ has a $k$-club $S \subseteq V$ such that $f(S) \leq m \cdot M$. The gadget is also shown in Figure 3.2.

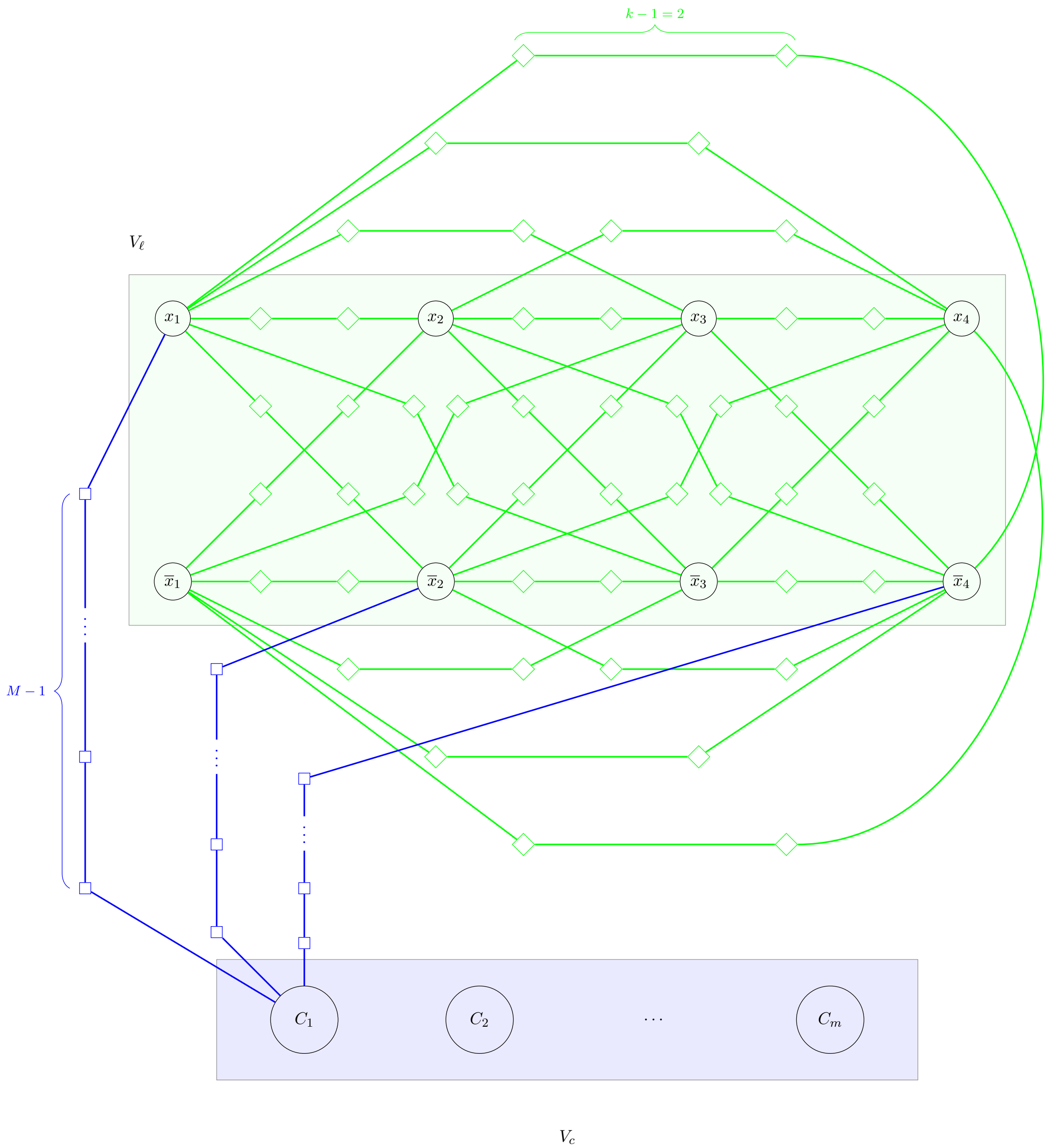


Figure 3.2. An example of the gadget used to reduce an instance of 3SAT to our problem to an instance of our problem with $k = 3$. For simplicity, we only show one clause, $C_1 = x_1 \vee \overline{x}_2 \vee \overline{x}_4$. The other clauses would be similarly connected to the nodes-literals of $V_\ell$ through a chain of $M - 1$ nodes.

Assume that the 3SAT instance has a feasible assignment $A$. Then, it is easy to see that by construction, the nodes corresponding to the literals in $A$ form a $k$-club (let them be $S$). Moreover, $A$ satisfies all clauses, hence there exists at least one node in $S$ that is at a distance of $M$ from each node in $V_c$. Hence, we have that $f(S) \leq m \cdot M$.

For the other direction of the proof, assume there exists a $k$-club $S \subseteq V$ such that $f(S) \leq m \cdot M$; yet there exists no feasible assignment of literals to satisfy the 3SAT instance. We distinguish between four cases:

1. $S$ consists of exactly one node $u_\ell \in V_\ell$ and nodes in $V_{\ell\times\ell}$ in as many as all $2n-1$ chains connecting them to all other literals (but its complement);
2. $S$ consists of exactly one node $u_c \in V_c$ and nodes in $V_{c\times\ell}$ in as many as 3 chains connecting $u_c$ to the literals clause $c$ contains;
3. $S$ consists of only nodes in $V_{c\times\ell}$ in exactly one chain connecting a literal-node $u_\ell \in V_\ell$ to a clause-node $u_c \in V_c$; and
4. $S$ consists of several nodes in $V_\ell$, along with the nodes in $V_{\ell\times\ell}$ in all chains necessary to connect them within $k$ hops.

**Case 1** Let $u_\ell \in V_\ell$ be the literal-node in $S$. From the nodes in the chains connecting $u_\ell$ to the other literals (but the node corresponding to its complement), one chain can have at most $\overline{k} \leq k-1$ nodes in $S$ and the remaining chains can have at most $k-\overline{k}$, where $1 \leq \overline{k} \leq \lceil k/2 \rceil$. Now, at best, this literal can satisfy at most $m-1$ clauses (since by assumption there exists no satisfiable assignment) whereas the literal that satisfies the remaining clause is located within a distance of $k-\overline{k}$ from $S$. Hence, we have:

$$f(S) \geq (m-1)\cdot M + M + \overline{k} = m \cdot M + \overline{k} > m \cdot M. \tag{3.2}$$

This contradicts the assumption that $S$ is a $k$-club with $f(S) \leq m \cdot M$.

**Case 2** Let $u_c \in V_c$ be the clause-node in $S$. Since we have a 3SAT instance, $u_c$ has exactly 3 chains around it, and contains at most $\overline{k} \leq k$ nodes from one chain with the remaining chains having at most $k-\overline{k}$ nodes in $S$. The three literal-nodes connected through the chains to clause-node $u_c$ can satisfy at most $m-2$

other clauses (apart from $c$). Hence, at best, we have:

$$\begin{aligned}
f(S) &\geq (m-2) \cdot \big(M - \overline{k} + M\big) + \big(M - \overline{k} + k + M\big) \geq \\
&\geq (m-2) \cdot (M - k + M) + (M - k + k + M) = \\
&= (m-2) \cdot 2 \cdot M + 2 \cdot M = && (M >> k) \\
&= 2\,(m-1)\,M. && (3.3)
\end{aligned}$$

By assumption, though, we have that $f(S) \leq m \cdot M$, which, combined with inequality (3.3), leads to:

$$2\,(m-1)\,M \leq m \cdot M \implies m \leq 2, \tag{3.4}$$

which is a contradiction.

**Case 3** A similar contradiction to Case 2 is obtained when $k$-club $S$ consists only of nodes in $V_{c\times\ell}$. Let the $k$-club be at a distance of $\overline{k}$ from the clause-node $u_c$ and at a distance of $M-k-\overline{k}$ from the literal-node $u_\ell$ of that chain. We then have one clause at a distance of $\overline{k}$, at most $m-2$ clauses (as, otherwise, literal $\ell$ satisfies all clauses, a contradiction) at a distance of $M-k-\overline{k}+M$, and at least 1 clause at a distance of, at best, $M-k-\overline{k}+k+M$, leading to:

$$\begin{aligned}
f(S) &\geq \overline{k} + (m-2) \cdot \big(M - k - \overline{k} + M\big) + \big(M - k - \overline{k} + k + M\big) = \\
&= (m-2) \cdot \big(2 \cdot M - k - \overline{k}\big) + 2 \cdot M = \\
&= (m-2) \cdot 2 \cdot M + 2 \cdot M = && (M >> k) \\
&= 2\,(m-1)\,M. && (3.5)
\end{aligned}$$

This leads to the same contradiction as in Case 2.

**Case 4** Finally, in the last case, the $k$-club $S$ is built so as to contain a series of nodes corresponding to literals. At best, those literals satisfy $m-1$ clauses (as, again, the 3SAT instance is assumed to be without a solution). Hence, we have that:

$$f(S) \geq (m-1) \cdot M + (M+k) = m \cdot M + k > m \cdot M. \tag{3.6}$$

This last contradiction finishes the proof.

□

### 3.4.1. Mathematical Formulation

In this section, we present our mathematical formulation and a greedy heuristic algorithm to solve larger scale instances. We also present some computational results on generated and real-life instances for smaller $k$-clubs ($k = 2, 3$).

We begin this section with the definition of our variables. We will use two sets of binary variables, defined as follows.

$$x_i^\ell = \begin{cases} 1, & \text{if node } i \in V \text{ is at a distance of } 0 \leq \ell \leq D \text{ from the } k\text{-club,} \\ 0, & \text{otherwise.} \end{cases}$$

$$y_p = \begin{cases} 1, & \text{if path } p \in \mathcal{P}^k \text{ is within the } k\text{-club,} \\ 0, & \text{otherwise.} \end{cases}$$

We can now proceed to describe the mathematical formulation, shown in (3.7). It is based on the maximum $k$-club chain formulation presented in [76]. Newer formulations for identifying $k$-clubs (as in, e.g., [77]) can also be employed, but are not explored here.

$$
\begin{aligned}
\min\ & \sum_{i\in V}\sum_{\ell=0}^{D} \ell \cdot w_i \cdot x_i^\ell && && (3.7a)\\
s.t.\ & x_i^0 \leq y_p, && \forall i \in p, \forall p \in \mathcal{P}^k, && (3.7b)\\
& \sum_{\ell=0}^{D} x_i^\ell = 1, && \forall i \in V, && (3.7c)\\
& x_i^{\ell+1} \leq \sum_{j\in N(i)} x_j^\ell, && \forall i \in V, 0 \leq \ell \leq D-1, && (3.7d)\\
& x_i^0 + x_j^0 \leq 1 + \sum_{p\in\mathcal{P}_{ij}^k} y_p, && \forall i,j \in V : i \neq j, && (3.7e)\\
& x_i^\ell \in \{0,1\}, && \forall i \in V, 0 \leq \ell \leq D, && (3.7f)\\
& y_p \in \{0,1\}, && \forall p \in \mathcal{P}^k. && (3.7g)
\end{aligned}
$$

The objective function in (3.7a) aims to minimize the total weighted distance every node outside the $k$-club needs to traverse until it accesses a node in the $k$-club. The constraint family in (3.7b) restricts that a path can only be within the $k$-club if every node that belongs to it belongs in the $k$-club. Constraints (3.7c) enforce that every node in the graph is at a distance $0 \leq d \leq D$ from a node in the $k$-club. The following constraints, shown in (3.7d), recursively enforce that a node can be at a distance of $\ell + 1$ from the $k$-club if it is neighboring a node that is located at a distance of $\ell$ itself. The constraint family in (3.7e) restricts that two nodes can not both belong in the $k$-club unless there exists at least one path connecting them within $k$ hops or less that is in the $k$-club. Finally, the binary nature of all variables involved is enforced with (3.7f)–(3.7g).

### 3.4.2. Greedy Heuristic

The above formulation is difficult to solve, as the underlying problem was shown to be $\mathcal{NP}$-hard (with a decision version being $\mathcal{NP}$-complete per Theorem 1). Hence, along with solving the formulation using a commercial solver, we also devise a practical heuristic. In our case, we opted for a greedy heuristic that always chooses to increase the $k$-club at hand by choosing a node with a maximum weight-to-distance ratio: that is, if a node is located near many nodes with big weights, it is more prone to being selected. This approach is shown in Algorithm 1.

**Algorithm 1:** Greedy Central $k$-club.

function Greedy_Central_$k$-club ($k$);
**Input :** A graph $G(V, E)$, weights $w : V \mapsto \mathbb{R}$
**Output:** A $k$-club
$\mathcal{I} \leftarrow V$;
$S \leftarrow \emptyset$;
**while** $|\mathcal{I}| > 0$ **do**
  **forall** $i \in \mathcal{I}$ **do**
    **forall** $j \in V$ **do**
      $d_j = \min\limits_{k \in S \cup \{i\}} d_{jk}$
    **end**
    $r_i = \sum\limits_{j \in V} \frac{w_j}{2^{d_j}}$;
  **end**
  $\bar{i} \leftarrow \arg\max\limits_{i \in \mathcal{I}} \{r_i\}$;
  $S \leftarrow S \cup \{\hat{i}\}$;
  $\mathcal{I} \leftarrow \{j \in N(S) : d_{ij} \leq k, \forall i \in S\}$;
**end**
**return** $S$

The backbone of the heuristic method is the spatial interaction model known as the gravity model (as it is similar to Newton's law of gravity). Its basic formula is as follows:

$$T_{ij} = \frac{w_i * w_j}{2^{d_j}}, \tag{3.8}$$

where $w_i$ and $w_j$ are the weight parameters (or, importance) of the origin and destination locations and $d_{ij}$ is (as defined earlier) the distance between the origin $i$ and destination $j$. In this work, we slightly change the interaction term in the numerator given in (3.8). Starting from some origin $i$, we are searching all adjacent (nearby) locations $j \in N(i)$ so as to add it to the $k$-club being built. Since the term $w_i$ is the same for all considered locations $j$ (as $(i, j) \in E$), we drop it from consideration and hence are left with a ratio of the importance of candidate location $j$ (given in the weight parameter $w_j$) versus the distance.

The algorithm is initialized with all nodes in the nodeset $V$ being in the candidate list, $\mathcal{I}$, and the starting $k$-club, $S$, is empty. Then, for every node in the candidate list, we "add" it in $S$ and calculate the shortest paths $d_j$ from every node $j$ to any node inside $S$. Then, the ratio becomes the summation of fractions $\frac{w_j}{2^{d_j}}$. The node with maximum ratio is indeed added in $S$, and the candidate list is updated with only neighboring nodes that satisfy the $k$-club criterion. A pictorial example, and its calculations are provided in Example 15 below.

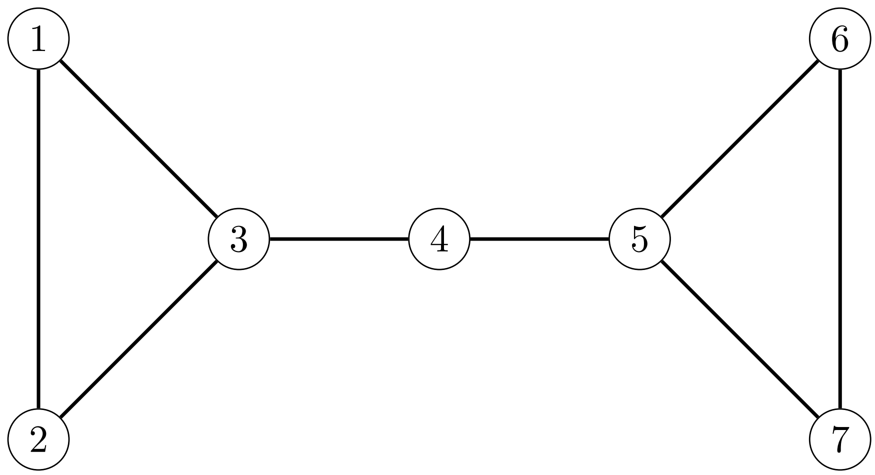

Figure 3.3. An example of how our greedy approach of Algorithm 1 works.

Assume that we have the graph of Figure 3.3 with weights $w_1 = w_2 = w_6 = w_7 = 5$, $w_3 = w_4 = w_5 = 10$, and we are looking for a 2-club. Initially, $\mathcal{I}$ contains all 7 nodes and $S$ is empty.

Starting from node 1, we see that it is located at a distance of 0 from itself, a distance of 1 from nodes 2 and 3, a distance of 2 from node 4, a distance of 3 from node 5, and a distance of 4 from nodes 6 and 7. Hence, we have that $r_1 = 5/2^0 + 5/2^1 + 10/2^1 + 10/2^2 + 10/2^3 + 5/2^4 + 5/2^4 = 16.875$. In the example, it is easy to see that exactly the same is true for nodes 2, 6, and 7.

Similarly, for nodes 3 and 5, we have $r_3 = r_5 = 10/2^0 + 10/2^1 + 5/2^1 + 5/2^1 + 10/2^2 + 5/2^3 + 5/2^3 = 23.75$. Finally, for node 4, we have that $r_4 = 10/2^0 + 10/2^1 + 10/2^1 + 5/2^2 + 5/2^2 + 5/2^2 + 5/2^2 = 25$. Hence, we update $S$ to include 4 ($S = \{4\}$) and the candidate list to include all nodes in the open neighborhood of $S$, such that their distance to 4 is less than or equal to $k = 2$. ($\mathcal{I} = \{3, 5\}$). We are now ready to start the second iteration.

For node 3 we now have the following distances from $j$ to $S \cup \{3\}$: nodes 1 and 2 are located one hop away, nodes 3 and 4 are zero hops away, node 5 is also one hop away, and nodes 6 and 7 are two hops away. Hence, we have that $r_3 = 10/2^0 + 10/2^0 + 5/2^1 + 5/2^1 + 10/2^1 + 5/2^2 + 5/2^2 = 32.5$. The key realization here is that the distances are no longer between the candidate node and every other node in the graph, but instead between $S$ including the candidate node and every other node in the graph. We also note that node 5 will have exactly the same ratio, by construction of the example. Let us add node 3 to $S$ (hence, $S = \{3, 4\}$), and $\mathcal{I} = \{1, 2, 5\}$.

For the third iteration, we have: $r_1 = r_2 = 5/2^0 + 5/2^1 + 10/2^0 + 10/2^0 + 10/2^1 + 5/2^2 + 5/2^2 = 35$ and $r_5 = 10/2^0 + 10/2^0 + 10/2^0 + 5/2^1 + 5/2^1 + 5/2^1 + 5/2^1 = 40$. Hence, 5 is added leading to $S = \{3, 4, 5\}$. Now, observe that $N(S) = \{1, 2, 6, 7\}$, but adding any of those nodes leads to a distance of 3 hops within $S$: hence, $\mathcal{I} \leftarrow \emptyset$, and Algorithm 1 terminates with $S = \{3, 4, 5\}$.

**3.4.3. Computational Results**

The developed algorithm and optimization model were implemented in Python and all numerical experiments were conducted on a Lenovo laptop with an Intel 2.50 GHz quad-core processor and 8 GB of RAM. To diversify the experiments and fully explore the behavior of the proposed algorithm as well as the optimization approach, two different sets of instances were considered. The first set of instances consists of Watts-Strogatz small-world graphs with a varying number of nodes, edges, and diameter (stylized as $G_1$–$G_6$). The second group are three cities (Sioux Falls, Eastern Massachusetts/EMA, and Berlin) from a networks repository for transportation research [78]. In Table 3.1, we present the computational times as well as information for each network (such as the number of nodes, the number of edges, and the diameter).

Table 3.1. Computational times under heuristic and optimization approach for $k = 2, 3$.

| | | | | 2-club | | 3-club | |
|---|---|---|---|---|---|---|---|
| Graph | $\|V\|$ | $\|E\|$ | $D$ | heuristic | optimization | heuristic | optimization |
| G1 | 16 | 32 | 3 | 0.12 | 0.15 | 0.80 | 0.27 |
| Sioux Falls | 24 | 38 | 6 | 0.17 | 0.29 | 0.27 | 0.29 |
| G2 | 32 | 64 | 6 | 0.18 | 0.54 | 0.55 | 0.98 |
| G3 | 64 | 128 | 6 | 0.65 | 1.95 | 1.11 | 13.24 |
| EMA | 74 | 129 | 9 | 0.87 | 2.85 | 1.84 | 8.77 |
| G4 | 128 | 256 | 6 | 3.34 | 12.05 | 4.98 | 70.76 |
| G5 | 256 | 512 | 7 | 21.67 | 80.77 | 28.69 | 561.62 |
| Berlin | 397 | 644 | 29 | 88.05 | 87623.10 | 106.61 | 73580.04 |
| G6 | 512 | 1024 | 8 | 161.28 | 473.49 | 243.18 | 2866.31 |

Although the computational time expectedly grows for both the commercial solver and the heuristic as the number of nodes increases, the growth rate is much slower for the heuristic algorithm. This is verified by Table 3.1 for identifying highly central 2 and 3-clubs. Note that, with the exception of the Berlin graph, the heuristic approach shows a speedup that is on average 3 and 7 times faster than the exact optimization model for $k = 2$ and $k = 3$, respectively. The case of the Berlin network is very important. In this transportation network, the exact optimization fails to find a solution within reasonable computational time, and instead spends hours trying to prove optimality. This happens because the diameter of the graph is big, and the number of binary variables in model (3.7) becomes prohibitively large.

## 3.5. Case Study

In this section, we investigate two case studies from the state of North Dakota, in the cities of Fargo and Casselton. Case studies and real-world visualization are necessary to put the problem in its related context and understand its implications. However, due to the computational complexity of our problem, reaching a solution within reasonable computing time is challenging. Hence, the exact optimization model of (3.7) was only solved on the (smaller) city of Casselton, whereas in the (larger) city of Fargo we only present the results of the heuristic (as in Algorithm 1).

### 3.5.1. Data Description

Casselton is a city in the state of North Dakota, with a population of $2,329$ in the 2010 census. To the best our knowledge, there is no bike sharing program planned for deployment in the near future. Figure 3.4 illustrates the overall geography of the city and the population distribution in proportionally graduated circles.

The network for the city of Casselton was built with TIGER/Line® road data and block population with ArcGIS 5.0. All roads were converted to sets of vertices and edges representing intersections and road segments, respectively. There are $|V| = 400$ vertices and $|E| = 523$ edges in the resulting graph. The block population polygons are turned to point features for weighing the graph vertices. According to a National Association of City Transportation Officials (NACT) report [38], to achieve an increase in ridership as well as in overall system utility, bike sharing kiosks should be located no more than $1,000$ feet apart from one another. Therefore, every single vertex has the potential to become a dockless bike station within $1,000$ feet. Then, each vertex is weighted based on the closeness to the population points.

For the city of Fargo, due to its size, only the greedy heuristic of Algorithm 1 was put to the test. The population in Fargo is 105,545. At the moment, a bike sharing system is in place, with 11 stations in the locations shown in Figure 3.5 with a triangle. The same figure also presents the geography of the city and the population in proportional circles. The network for the city of Fargo is obtained in the same way as the one for Casselton. The final graph contains $|V| = 2989$ vertices and $|E| = 4302$ edges, which is indeed large-scale for the exact optimization solver.

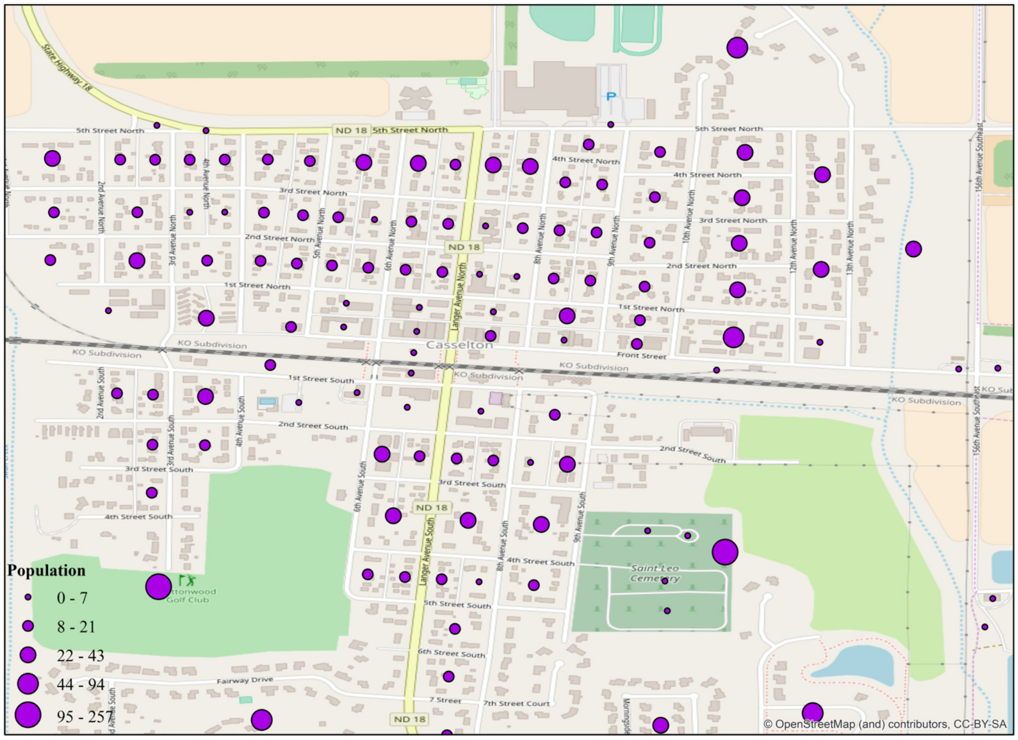


Figure 3.4. The geography, transportation network, and population of Casselton, ND.

All codes for solving the problem, both for the exact optimization model and the greedy heuristic were coded in Python. For solving the optimization model, we used Gurobi 7.5 [79]. We are now ready to present our findings in the next section.

### 3.5.2. Results and Findings

We investigate three measures obtained by both the heuristic and the exact optimization:

1. number of nodes selected in the $k$-club (cardinality);

2. population located in the selected nodes (immediate access); and

3. distance-weighted cost from all nodes to the $k$-club (general accessibility).

The number of nodes in the $k$-club represent the desirable, potentially geo-fenced, sites where a rider could check in/out a bike. The population measure represents the number of the residents within the $k$-club: they are the ones with immediate access to a location with bicycles. Finally, the distance-weighted

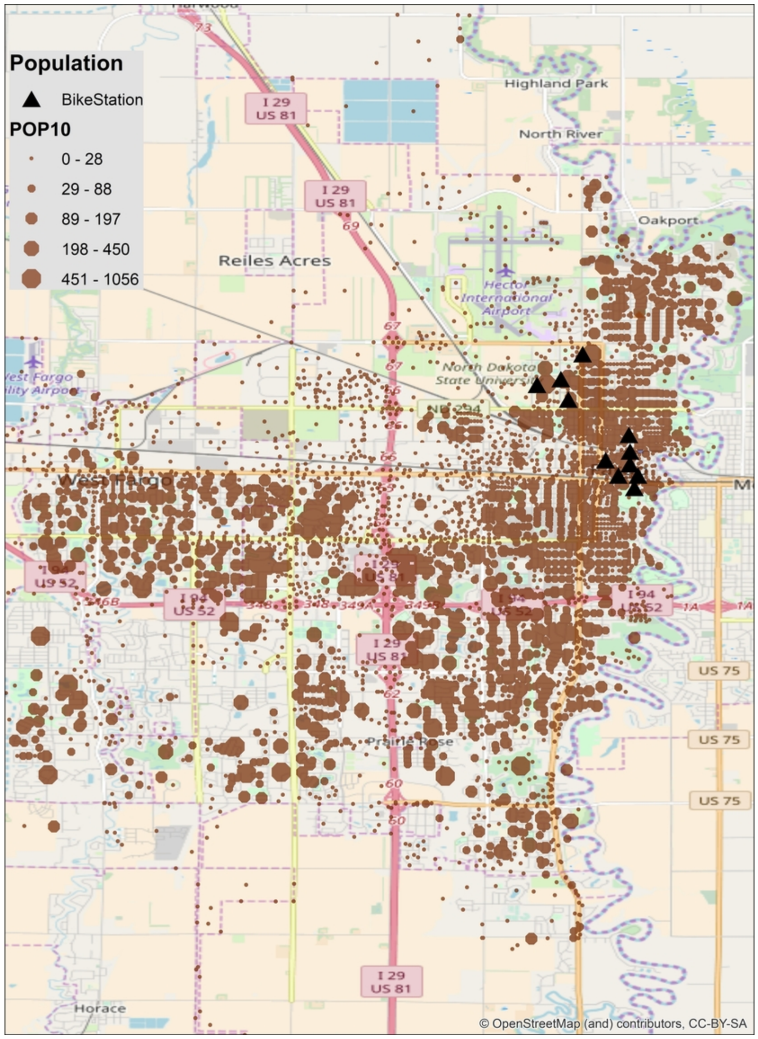


Figure 3.5. The geography, transportation network, and population of Fargo, ND. The locations of the existing eleven bike stations in Fargo are shown with a triangle in the map.

cost describes the total distance a commuter (from any location in the network) should walk to reach some node in the $k$-club to get access to a bike. Therefore, as was also shown in our optimization model, lower distance-weighted costs are preferable.

Table 3.2 summarizes the results for $k \in \{2, \ldots, 9\}$ in Casselton. The population represents the number of residents living in the $k$-club. The distance-weighted cost is the actual objective function of our optimization model. Finally, time shows the computational time required to solve the problem.

Starting from the population, in the case of exactly optimizing the formulation, it is consistently smaller than the population covered by the heuristic approach. On the other hand, distance-weighted cost represents the distance that the residents living outside the $k$-club must travel to access to designated geo-

Table 3.2. Numerical results for Casselton.

| | Cardinality | | Population | | Distance-weighted cost | | Time (seconds) | |
|---|---|---|---|---|---|---|---|---|
| $k$ | Alg. 1 | Opt. | Alg. 1 | Opt. | Alg. 1 | Opt. | Alg. 1 | Opt. |
| 2 | 5 | 5 | 187.31 | 20.04 | 27483.24 | 18387.73 | 11.42 | 705.22 |
| 3 | 7 | 6 | 226.48 | 23.02 | 27392.87 | 17403.04 | 12.09 | 929.13 |
| 4 | 10 | 12 | 214.33 | 61.04 | 23414.88 | 16116.41 | 15.88 | 1214.09 |
| 5 | 9 | 13 | 234.28 | 64.64 | 25384.43 | 15134.70 | 16.36 | 6267.38 |
| 6 | 14 | 23 | 282.25 | 128.90 | 23327.35 | 13886.10 | 18.42 | 10478.39 |
| 7 | 18 | 25 | 293.57 | 134.33 | 19752.91 | 12907.99 | 18.69 | 35906.98 |
| 8 | 21 | 38 | 310.87 | 222.63 | 18156.45 | 11723.64 | 19.84 | 59136.11 |
| 9 | 16 | 42 | 297.33 | 266.99 | 18638.10 | 11679.81 | 20.98 | 104211.74 |

fenced areas. The optimization model expectedly offers better results than the heuristic for all $k$-clubs obtained. Finally, when looking at the computational time, it becomes clear that even in a small city like Casselton, the exact optimization approach is prohibitively expensive, with $k = 7$ taking a little less than 10 hours, and $k = 9$ requiring more than 24 hours of computation before it terminates upon reporting a suboptimal solution and an optimality gap of 56.8%. The heuristic though is significantly and consistently faster, with a small uptick in computational time linear with the value of $k$ as it increases.

Figures 3.6 and 3.7 present the solutions within the city, and show the sets of nodes selected. Both the heuristic and the optimization approaches suggest groups of vertices located nearby – seeing as the resulting set of nodes forms a $k$-club. However, the heuristic approach starts with the most populated points in the city, and expands the set of nodes around that same point as the diameter of the set ($k$) increases. On the other hand, the optimization model is more dynamic, as it tries to minimize the overall distance-weighted cost.

We note that the heuristic is also inconsistent, as there are cases (see, e.g., $k = 4$ vs. $k = 5$) where a solution worsens as far as the distance-weighted cost is concerned as $k$ increases. This happens because the heuristic of Algorithm 1 myopically chooses the "best" candidate node to add so long as it respects the $k$-club diameter requirement. Because of this, the population immediately covered is bigger in the solution from the heuristic as opposed to the optimization model. We note though that this is not necessarily good, as it might result in locations where a high number of residents have immediate access to dockless bike sharing, but other residents have to travel very far to access it.

In the case of Fargo, as shown in Figure 3.8, we only applied the heuristic algorithm to validate our model, as optimizing for the values of $k$ that would be meaningful resulted in running out of memory.

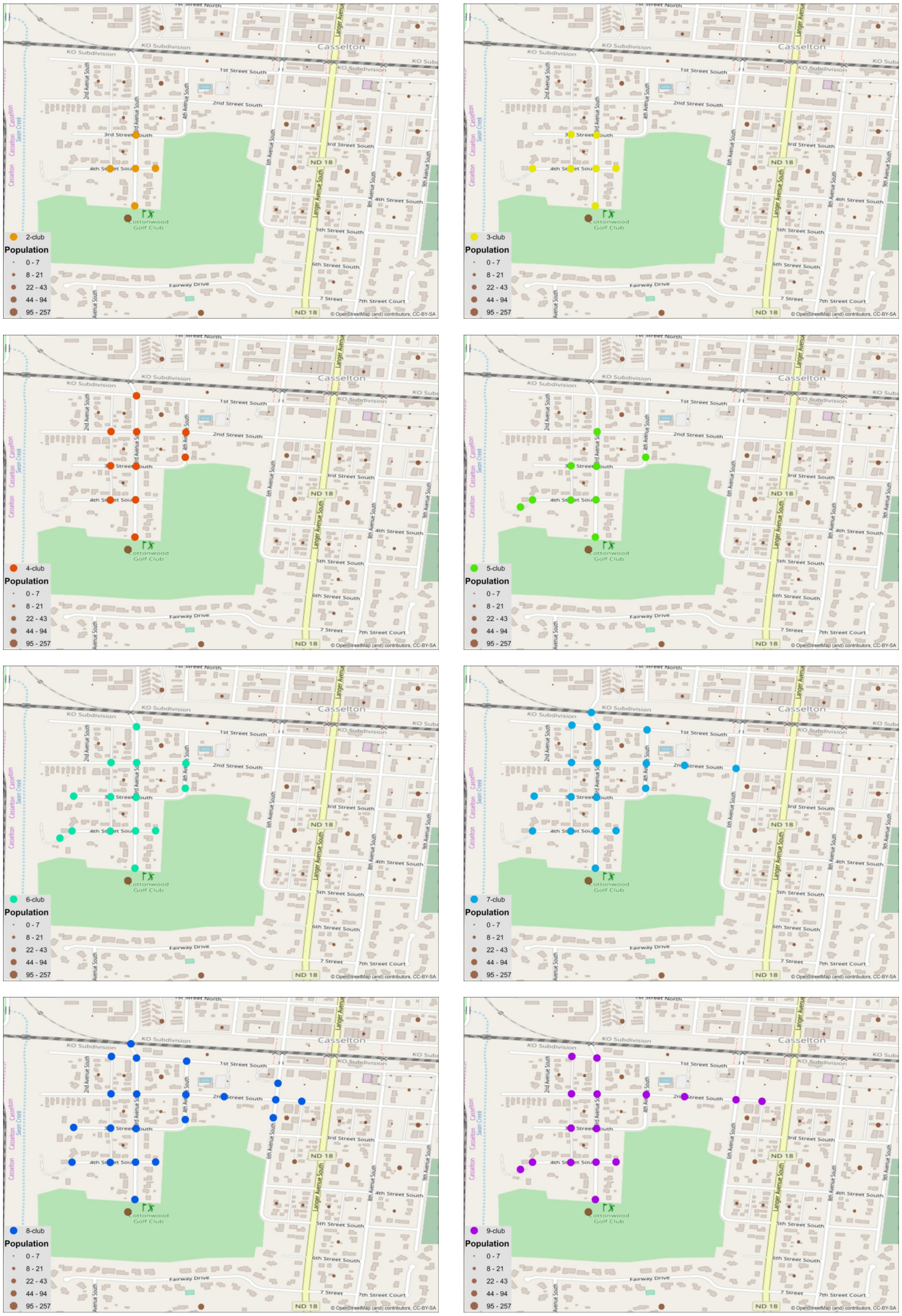


Figure 3.6. The $k$-clubs obtained from the heuristic for the city of Casselton. The results are for $k = 2, \ldots, 9$ starting from the top left (for $k = 2$) and ending in the bottom right (for $k = 9$).


Figure 3.7. The $k$-clubs obtained from optimizing the model for the city of Casselton. The results are for $k = 2, \ldots, 9$ starting from the top left (for $k = 2$) and ending in the bottom right (for $k = 9$).

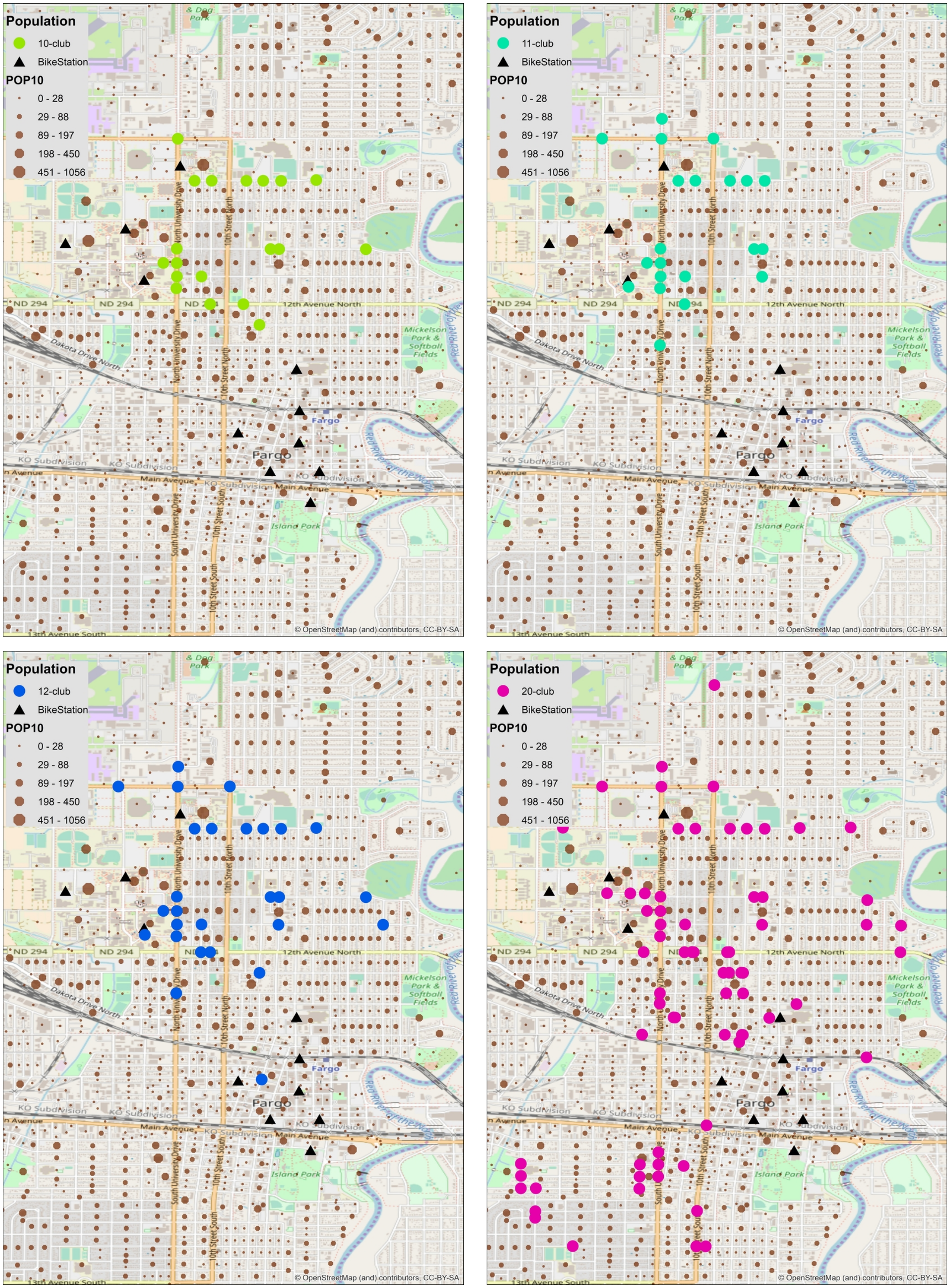


Figure 3.8. The $k$-clubs obtained from the heuristic for the city of Fargo. The results are for $k = 10, 11, 12$, and 20 in the top left, top right, bottom left, and bottom right, respectively.

Figure 3.8 illustrates the $k$-club heuristic solutions for Fargo, for $k \in \{10, 11, 12, 20\}$. The potential sites were located in a highly populated area next to the university campus. The existing 11 bike stations already in operation in Fargo are only blocks away from the suggested the 10-club. Table 3.3 summarizes the numerical results. It is intuitive that due to the fact Fargo has a larger overall population per block, the corresponding numbers in the table are much larger than the ones for Casselton.

Table 3.3. Numerical results from the heuristic model for Fargo.

| $k$ | Cardinality | Population | Distance-weighted cost |
|---|---|---|---|
| 10 | 19 | 3695.08 | 1913079.15 |
| 11 | 21 | 3984.72 | 1807995.05 |
| 12 | 27 | 4808.93 | 1798887.10 |
| 20 | 71 | 9966.84 | 1513123.94 |

### 3.5.3. Cost Benefit Analysis

Equipment, installation, and maintenance are three significant costs involved in implementing a bike sharing program. The main drawback to physical bike station systems (known as kiosk system) is their high acquisition and operating costs. Stations are costly including tens of thousands of dollars to manufacture and install along with several thousand dollars to acquire customized bikes. Moreover, kiosk systems mandate constant bike rebalancing. This happens because every bike needs to be returned to a kiosk: if the kiosk is full, the riders must find another location with available spots, resulting in higher operational cost and a decrease in customer satisfaction.

The cost of each bike is estimated at $1,234 [80]. Assuming a cost of $1,000 on average for each bike, the cost for a typical kiosk with 11 docks will range from $29,000 to $34,000, excluding operating costs. Figure 3.9 shows the relationship between the cost and number of docks. These figures are even higher at the planning stage ($55,000 per station) [81]. The optimal number of docks is another critical factor in a bike sharing program. Increasing the number of docks leads to higher costs, and a pile up of bikes in one location, which consequently results in higher re-balancing cost. At the same time, it leads to higher customer satisfaction. The dockless option would at least avoid initial capital investment and pave the way to introduce bike sharing programs to cities, without sacrificing customer satisfaction with the program.

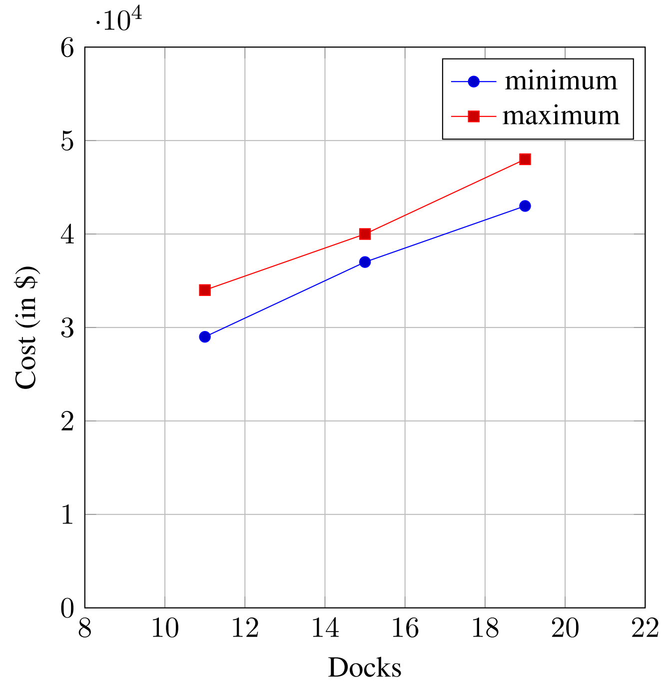


Figure 3.9. Equipment and installation cost vs. number of stations.

### 3.6. Conclusions and Implications

In this work, we discussed a new paradigm for selecting where a dockless (geo-fenced) bike sharing system should be enabled within an urban area. This paradigm tries to solve the disadvantage of kiosk-based bike sharing programs such as high equipment costs and costs associated with customer dissatisfaction due to lack of bikes/docks at the desired location. Also, the proposed model offers a better solution to existing dockless problems.

We modeled our problem as one of detecting a connected set of nodes of restricted diameter (that is, where any two nodes are reachable within $k$ hops using nodes inside the set), or a $k$-club. The goal was to find a $k$-club of maximum closeness, so as to make sure that all other nodes in the transportation network are close enough to the bike sharing locations. We showed that, as expected, the problem is $\mathcal{NP}$-hard, and provided an integer programming formulation to solve it. We also propose a greedy heuristic, which is computationally inexpensive. As $k$ increases for the obtained $k$-club, we should expect the coordination costs to increase along side as greater values of $k$ will imply larger geo-fenced areas. From a practical perspective, BSS operators would have to trade off the size of the geo-fenced area (the larger, the more easily accessible and more convenient to users) to the rebalancing costs (the smaller, the more easily coordinated and cheaper for BSS operators).

We also used our methods to study the resulting setup in two cities of the state of North Dakota, Casselton (of smaller population) and Fargo (of bigger population). The potential cost savings in the dockless approach could decrease initial capital investments for introducing a bike sharing program in a city. It also leads to an increase in the number of the virtual docks (capacity) without blocking streets or pedestrian walkways. One might say that dockless bike sharing brings chaos to cities, due to the freedom of allowing bike check in/out anywhere in a geo-fenced area. That is why our approach could mitigate the described situation and leverage this dockless alternative, by only enabling some areas with this capability. The model at the moment is built based on the population as the only location weight.

Future directions for our work include the following. First, we could investigate the identification of multiple $k$-clubs of varying sizes within a city. This would allow BSS operators to have multiple smaller geo-fenced areas or fewer larger geo-fenced areas to cover all bike sharing demands. As a second direction, we should consider more ways to build the weight parameter in our framework. For example, we plan to investigate how $k$-club formation and how the geo-fenced areas change as we consider city points of interest, distance to nearby transit points, and origin-destination demands throughout the day, among others. Next, another future avenue for our research would be to investigate more closely the interactions between different operators (e.g., dockless bike sharing and scooter sharing, or dockless bike sharing and public transit) with respect to different geo-fenced areas.

# 4. RISK PERCEPTION OF BICYCLE/SCOOTER RIDERS RISKY BEHAVIORS

## 4.1. Abstract

Bicycle and scooter use entails high safety and health risks. News stories have described the reckless behavior of users across the United States with the emergence of micromobility options. This paper investigates risky behaviors associated with bicycle and scooter riding adults residing in the United States. Two separate surveys were administered through the Qualtrics® platform. Participants were asked to rate the severity and frequency of 20 risky behaviors of riders on five Likert scales. The risk matrix is built based on the magnitude and frequency of each risk and ordered logistic regression is applied to identify significant factors. Regression analysis revealed that age and income are significant factors shared between both survey groups. Level of education and living in urban areas are two statistically significant factors explaining the different risky behaviors with bicycles or scooters. In general, the survey results show that participants perceive that there is a low risk associated with reckless behaviors. It may imply that they are exposed to fewer incidents, or the media exaggerates the news about the incidents. Further research on other aspects of risk, such as network geometry and safety education, would help better understand the underlying factors. The findings offer insight for developing new enforcement policies and safety education programs to enhance scooter/bicycle sharing programs and provide a safe environment for all road users. The authors sincerely acknowledge the funding provided by the Small Urban, Rural Tribal Center On Mobility which receives funds from the U.S. Department of Transportation's University Transportation Centers (UTC) Program.

## 4.2. Introduction

Shared micromobility options such as bike and scooter sharing are increasingly becoming an accessible mode of transportation in many cities and towns across the United States [82, 80, 83] . More than 207 million trips have been made on shared bikes and e-scooters since 2010. Riders took 84 million trips in 2018, more than double the number from the previous year, mostly because of the emergence of electric scooters [84].

City officials have a great interest in promoting sustainable shared micromobility modes for health and environmental reasons [85]. However, the behavior of riders is a recurring theme in the public debate surrounding the growth of this mode of shared micromobility. In many of these discussions, the majority of riders are characterized as displaying lawless attitudes, which are the root of individual and public safety concerns associated with riding those vehicles. Many of these safety issues are related to unsafe riding behaviors, including, but not limited to, operating under the influence of alcohol, maneuvering recklessly, overloading with multiple riders, speeding, etc. The danger of risky riding to pedestrians and other road users is amplified in urban areas (with high population densities) and nighttime [86, 87].

There were 783 cyclist deaths in 2017 in the United States, which accounted for 2.1 percent of all traffic fatalities during the year. Seventy-five percent of pedal-cyclists who died in motor vehicle crashes in 2017 were killed in accidents in urban areas [88]. The difference in data collection makes it challenging to compare statistics among different types of vehicles. For bike sharing, the average collision rate was 4.33 crashes per year among operators with more than 1,000 vehicles and fewer for operators with smaller fleets [89]. There is no reliable and consistent data available for electric scooters. For example, a pilot study attempted to measure the safety impacts of scooter sharing by reviewing reported scooter incidents in Multnomah County, Portland. On average, emergency room visits increased from less than one per week before the pilot to approximately ten per week during the pilot period. However, exact numbers are difficult to quantify because of missing data related to other types of scooters, such as mopeds and non-motorized standing scooters. Of the entire sample of scooter-related emergency visits, 83 percent did not involve another mode compared to 13.6 percent involving a motor vehicle and 2.8 percent involving a pedestrian. Only one collision (0.6%) was reported involving two scooters. These statistics are difficult to be validated because the trips were diverted from other modes such as automobiles, buses, or rail and increased the risk of the individual riders [90]. Heesch et al. (2011) analyzed cycling accidents and found that regular cyclists were involved in a relatively high number of traffic crashes in Queensland, Australia [91]. While most of the accidents were not serious, the number of crashes was more elevated in Australia than in European and Asian nations. In Australia, approximately one in 40 road crash deaths were cyclists [92]. Because cyclists comprised only 2% of national fatalities and injuries while making less than 1% of all trips, the perception that cycling is dangerous is not unfounded [88].

Safety involving bike and scooter riders and other road users has been a prevalent topic in research and media reports in most countries. Some research described the conflict between motorists and cyclists as “road wars” [93, 94, 95]. Salmon et al. suggested that such conflict is a long-standing problem, and cyclists are 12 times more at risk of death than car drivers [96]. MacMillan et al. researched the media’s reporting of cyclist fatalities in London, suggesting media coverage and the way they portray various aspects of road behaviors was a relatively vital method of preventing reckless behaviors or promoting appropriate ones. The number of fatalities covered by The Evening Standard, one of London's major newspapers, has increased tenfold while the number of trips has only doubled [97]. Some research suggested that the fear of cars and possible accidents are barriers to cycling [97, 47, 98]. The number of negative representations of cyclists is almost double the positive ones: the former includes words like "irresponsible lawbreakers," "pariahs," and "dangerous to others" and the latter "brave," "harmless," and "healthy" [99]. In similar research, Bogdanowicz recognized the negative language toward cycling by media, which characterized it as a transport mode for "eccentrics" or "tree huggers" [100]. Skinner and Rosen noted that the negative language and hostile attitude toward cyclists are much more noticeable where cycling is rare [101].

The majority of existing studies have been conducted to help engineers and city planners design and improve roads and intersections. Researchers asked cyclists to rate their general risk perception of a route through a set of videos, surveys, test courses, and simulations. Each examined several network geometry and operation-specific factors related to the safety perceptions of cyclists [102, 103, 104, 105, 106, 107, 108]. There has been no systematic study on the interactions between cyclists and other road users, particularly in the United States. This implies the need for a more comprehensive understanding of what happens when drivers and bicyclists interact.

Another group of studies investigated the reasons behind risky behaviors and addressed possible determinants among different people. Reyna and Farley [109] tried to answer why adolescents may seek out risky situations. They found that adolescents, despite conventional wisdom, generally overestimate risks. Indeed, after the age of 14, there might be no difference between teens and adults concerning the perception of risk [110]. Feenstra et al. [111] conducted a survey-based study to investigate the risky cycling behaviors of adolescents from 13 to 18 years old before developing safety education programs. They found that adolescents are capable of identifying themselves as risk-takers or not. They suggested shifting the focus of education programs from risk perception to decreasing risky attitudes in traffic and promoting a sense of responsibility. Shope and Bingham [112] provided a list of possible determinants to explain why young drivers engage in more risky behavior than adult drivers such as characteristics of the behavior, abilities, developmental factors, behavioral factors, among others.

Despite studies investigating the safety aspects of shared mobility schemes, there is little understanding of the perceived risk of reckless behavior from other road users' perspectives. For example, researchers and practitioners have little empirical knowledge about how certain aspects of risk rank compared to others, nor do they know much about road users' fear of different types of reckless bike/scooter use. To address this gap, this study explores multiple aspects of perceived risk associated with reckless behaviors of bike and scooter riders. This paper aims to investigate the overall risk perception of the risky behaviors related to bicycle and scooter riders from a general perspective, including perceptions of users and no-users. We are also interested in examining the difference in the level of the risks associated with those using bicycles versus scooters. We can summarize our contributions in the following three areas:

- first, we address risky behaviors related to two popular sustainable modes of transportation in the United States: scooter and bicycle;
- second, we develop a risk matrix to provide better insight into the magnitude and frequency of each risk. We use the ordered logistic regression to analyze the demographics and general cycling behaviors and ascertain if there are any significant underlying determinants; and
- finally, we focus on the risk perceived by the general audience (users and non-users) to determine the overall perception. We also provide feedback received from open-ended comments about real risks in individual experiences during interactions with scooters or bicycles.

The next two sections of this paper depict the methods, the data, and the findings, respectively, followed by the ordered logistics regression analysis section. Discussion of the variables in the statistical models and the limitations of this study are provided. Finally, the conclusion is provided.

### 4.3. Methodology

In short, this study is composed of three parts. First, we explain the survey that was conducted in March 2019 to capture the risk perception of risky behaviors of riders. Then we build the risk matrix from the survey results for further analysis. Finally, we do a statistical analysis of the risk to identify significant factors associated with each level of risk.

#### 4.3.1. Survey

We designed two separate cross-sectional online self-completed surveys for each vehicle type (bikes and scooters) to identify and assess reckless and risky behaviors of bike and scooter riders. Each survey has three sections:

1. socio-demographics;

2. risky behavior characterization; and
3. general riding behavior.

For risk analysis, the severity and frequency of each factor are asked using a Likert scale. Study respondents were first asked to complete a screening question to determine eligibility: respondents had to be 18 years old and U.S. residents). Eligible respondents then were asked to provide informed consent to complete the survey. Those who provided the informed consent then completed a 15-question survey about the significance and frequency of various types of risky behaviors associated with riders, overall cycling behaviors, and demographics. Participants who did not consent were not allowed to continue to the second part. The North Dakota State University Institutional Review Board (IRB) approved the questionnaires.

**4.3.1.1. Sample Size and Recruitment**

Respondents were recruited from March 1, 2019, to March 15, 2019, by Qualtrics® panel that consisted of the following criteria: residents of the United States in different geographic areas (rural, suburban, and urban areas) and ages 18+. The inclusion/exclusion criteria were the same for both surveys. We used the Qualtrics® platform for participant recruitment because Qualtrics® panel is demographically and politically representative [113]. Qualtrics® checks every IP address and uses a sophisticated digital fingerprinting technology to exclude duplication and ensure validity. Participants completed the survey from their own devices. Upon survey completion, a unique code was used to redeem an incentive. Qualtrics® was paid at a rate of $6 per subject, but the actual payment amount from Qualtrics® to respondents is unknown to us. As is typical in web-based survey research, we employed multiple attention checks and quality screens in our surveys. Attention checks confirmed that web-based survey respondents were reading questions carefully and thoroughly. The Qualtrics® panel suggested using the median time to complete the survey as the cut-off point to determine whether respondents rushed through the survey, so we applied this criterion to the survey as a part of the quality screening.

According to 2010 United States Census, the United States has a population of 308,745,538, of which 76% (234,646,609) are age 18 years and older [114]. For very large population size, we can use the following formula to calculate the sample size for the surveys.

$$\text{Sample Size} = \frac{\text{Z-score}^2 \times \text{StdDev} \times (1 - \text{StdDev})}{\text{Margin of Error}^2} \tag{4.1}$$

Considering a 90% confidence level, a 5% margin of error, and a 0.5 standard of deviation, the expected sample size is 270.

#### 4.3.2. Risk Matrix

We used a risk assessment matrix to conduct a subjective risk assessment in our model. The basis for the risk matrix is the definition of risk as a combination of the severity of the consequences occurring in a particular accident scenario and its frequency. To build the risk matrix, we first categorize and scale the severity and frequency as well as the output risk index. The categorization of the severity and frequency depends on the type of activity or the specifics of the processes involved. We categorized the frequency and the severity into five groups. This provided the basis for constructing the plane matrix with 25 cells, each representing a specific risk category. The relationships between all inputs and outputs for a standard risk matrix are suggested by the U.S. National Institute of Standards and Technology, as shown in Table 4.1 [115].

Table 4.1. Assessment scale.

| **Likelihood** | **Level of Severity (Impact)** | | | | |
|---|---|---|---|---|---|
| | **Very Low** | **Low** | **Moderate** | **High** | **Very High** |
| **Very High** | Very Low | Low | Moderate | High | Very High |
| **High** | Very Low | Low | Moderate | High | Very High |
| **Moderate** | Very Low | Low | Moderate | Moderate | High |
| **Low** | Very Low | Low | Low | Low | Moderate |
| **Very Low** | Very Low | Very Low | Very Low | Low | Low |

The application of the risk matrix is simple. After assessing the severity and frequency categories, the risk category as one out of five groups (very low, low, moderate, high, very high) is specified using the risk matrix. This is the basis for further risk control measures in the next section. We are interested in identifying any relationship between the level of risk (as represented in the risk matrix) of each risky behavior listed in Table 4.2 and explanatory variables such as sex, age, income, among others listed in Table 4.3.

#### 4.3.3. Perceived Risk Model

Because the level of the risk, as dependent variable, is ordinal (more than two categories and the value of each group have a meaningful sequential order), we use the ordered logistic regression method, also known as the proportional odds model, to investigate the determinants that influence ordinary road users' risk perceptions of the various risky behaviors in the United States, as listed in Table 4.2. Initially, eight explanatory variables were included as categorical variables in the perceived risk model, as described in Table 4.3.

Table 4.2. Dependent variables.

| Dependent Variables | Description |
|---|---|
| **Y1** | Ignoring traffic signals |
| **Y2** | Riding a scooter/bicycle while under the influence |
| **Y3** | Riding at night without lights on |
| **Y4** | Distracted riding, including, but not limited to, talking or texting on phones, eating or drinking, or other distracting activities |
| **Y5** | Ignoring stop signs |
| **Y6** | Not yielding to pedestrians |
| **Y7** | Speeding |
| **Y8** | Swerving (riding in a zigzag) |
| **Y9** | Riding on sidewalks |
| **Y10** | Riding against traffic on the roadway |
| **Y11** | Riding the wrong way on a one-way street |
| **Y12** | Stoppie – braking too quickly resulting in a skid or the rear tire lifting up |
| **Y13** | Wheelie – riding a scooter/bicycle with the front wheel raised off the ground |
| **Y14** | Jumping off a curb |
| **Y15** | Passing too closely on either side of vehicles on the road |
| **Y16** | Tailgating - riding too closely behind another vehicle |
| **Y17** | Riding without helmets |
| **Y18** | Riding with under-inflated tires |
| **Y19** | Yelling, or making angry gestures at motorists, cyclists, scooter riders, or pedestrians |
| **Y20** | Riding with a passenger |

Table 4.3. Independent variables.

| Independent Categorical Variables | Description | Levels (reference case marked with asterisk) |
|---|---|---|
| **X1** | Age | *18-24, 25-34, 35-44, 45-54, 55-64, 65-74, 75 or older |
| **X2** | Sex | *Female, Male |
| **X3** | Marital Status | *Divorced, Married, Separated, Single, Widowed |
| **X4** | Education | *Associate degree, Bachelor's degree, Graduate degree, High school diploma or equivalent, Less than a high school diploma, Some college, No degree |
| **X5** | Employment | *Disabled, Employed full-time (40+ hours a week), Employed part-time (less than 40 hours a week), Retired, Self-employed, Unemployed (currently looking for work), Unemployed (not currently looking for work) |
| **x6** | Income | *below $10k, $10k-$25k, $25k-$50k, $50k-$75k, $75k-$100k, $100k-$125k, $125k-$150k |
| **X7** | Race | *Asian American, Black or African-American, Middle eastern American (middle east, north Africa and the Arab world), Native American or Alaska native, Native Hawaiian or other pacific islander, Some other race, White American |
| **X8** | Region | *Rural, Suburban, Urban |

## 4.4. Results and Findings

### 4.4.1. Descriptive Analysis

Considering the initial estimate of 270, after 15 days of recruiting, 749 responses were collected, of which 659 (329 of scooters (S) and 330 of bicycles (B)) are valid responses and eligible for use in the analysis. A quantitative design allowed for information collected from a large number and enabled a comparison between groups, behaviors, and outcomes. Also, some qualitative analysis was possible as a result of one open-ended comment box within the questionnaire.

### 4.4.2. Data Description

The percentage of white respondents (72% S and 85% B) is approximately representative of the U.S. population (72% white). Black (15% S and 9% B) and Asian (4% S and 3% B) are the next two major groups of respondents. The respondents were more than twice as likely to hold a high school diploma, followed by some college.

Responses came from all over the United States with minimal to no responses from a couple of states, as seen in Figure 4.1, including Hawaii and Alaska (not pictured). Using Census regional designations, there were 145 responses from suburban areas (44%), 115 responses from urban areas (35%), and 69 responses from rural areas (21%) areas for scooter survey. The bicycle survey generated 180, 79, and 71 responses from suburban, urban, and rural areas, respectively. Most respondents were in full-time employment (38% S, and 25% B). Interestingly, from the bicycle survey, the second major group with 25% of respondents is retired, while 15% of respondents in the scooter survey have a part-time job.

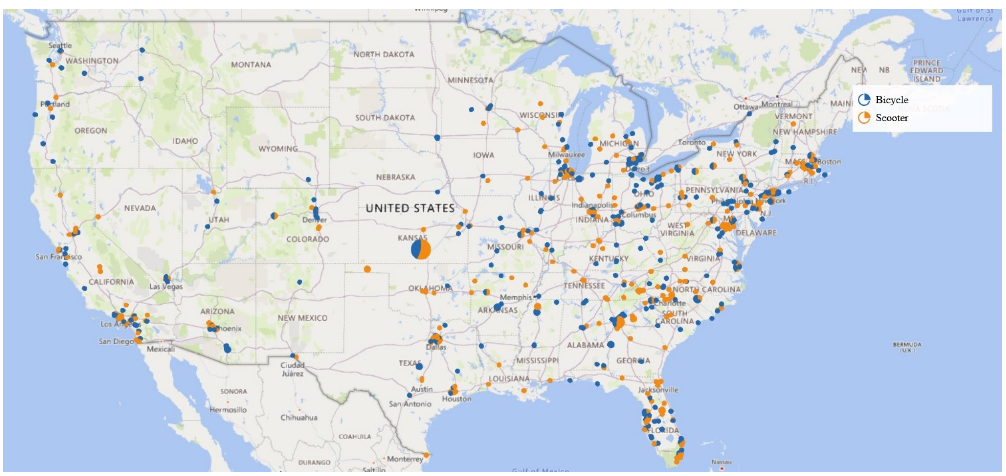


Figure 4.1. Distribution of respondents by vehicle type.

Most respondents indicated they were married (47% S, 50% B). A total of 37% in the scooter survey and 30% in the bicycle survey indicated they were single. Most respondents earn $25k-$50k (over 23% in each survey) annually in both surveys. Respondents in the second major group in the scooter survey make $10k-$25k while in the bicycle survey, most respondents make $50k-$75k annually.

Dill and McNeil adopted a topology developed by the City of Portland to describe the cycling behaviors of adults. It includes four categories: “Strong and the Fearless,” “Enthused and Confident,”

Table 4.4. Riders type by region.

| Region | Category | Scooter | Bicycle |
|---|---|---|---|
| **No Way, No How:** *unwilling to ride even if high-quality (bicycle/scooter) infrastructure is available* | Rural | 22 | 28 |
| | Suburban | 32 | 51 |
| | Urban | 32 | 22 |
| | **Total** | **86** | **101** |
| **Interested but Concerned:** *willing to ride if high-quality (bicycle/scooter) infrastructure is available* | Rural | 32 | 29 |
| | Suburban | 68 | 87 |
| | Urban | 42 | 32 |
| | **Total** | **142** | **148** |
| **Enthusiastic and Confident:** *willing to ride if some (bicycle/scooter) -specific infrastructure is available* | Rural | 10 | 8 |
| | Suburban | 29 | 33 |
| | Urban | 31 | 20 |
| | **Total** | **70** | **61** |
| **Strong and Fearless:** *willing to use scooters with limited or no (bicycle/scooter) -specific infrastructure* | Rural | 5 | 6 |
| | Suburban | 16 | 9 |
| | Urban | 10 | 5 |
| | **Total** | **31** | **20** |

"Interested but Concerned," and "No Way No How" [116]. We also adopted the same approach but asked participants to describe their riding skills, as indicated in Table 4.4.

As described in Table 4.4. in all three distinct geographic areas, there is a general trend: most respondents (more than 43%) characterize themselves as "interested but concerned," while the "strong and fearless" is the least selected option. This may mainly be because of the recent deployment of bike/scooter sharing programs across the United States. The weather might be another factor that needs to be included in future research. Interestingly, under the "strong and fearless" category, there were almost twice as many respondents who considered themselves as "strong and fearless" in riding scooters as there were for bicycles. This may be because of the greater ease of getting off a scooter versus a bicycle if there is an impending crash because scooters are usually lighter and more manageable than a bicycle. Surprisingly, the number of people who are not willing to ride a bicycle, even with high-quality infrastructure, is more than one-fourth of respondents in both cases (scooter and bicycle).

An analysis of respondents' bicycle and scooter riding frequency (includes both their own bikes/scooters and sharing systems) is illustrated in Figure 4.2, which shows that 36% of respondents never rode a scooter before, while only 3% never rode a bicycle. This might be because scooters are unavailable in some regions. Except for the categories "Never" and "More than five years ago," the riding profiles of respondents in both surveys follow the same pattern.

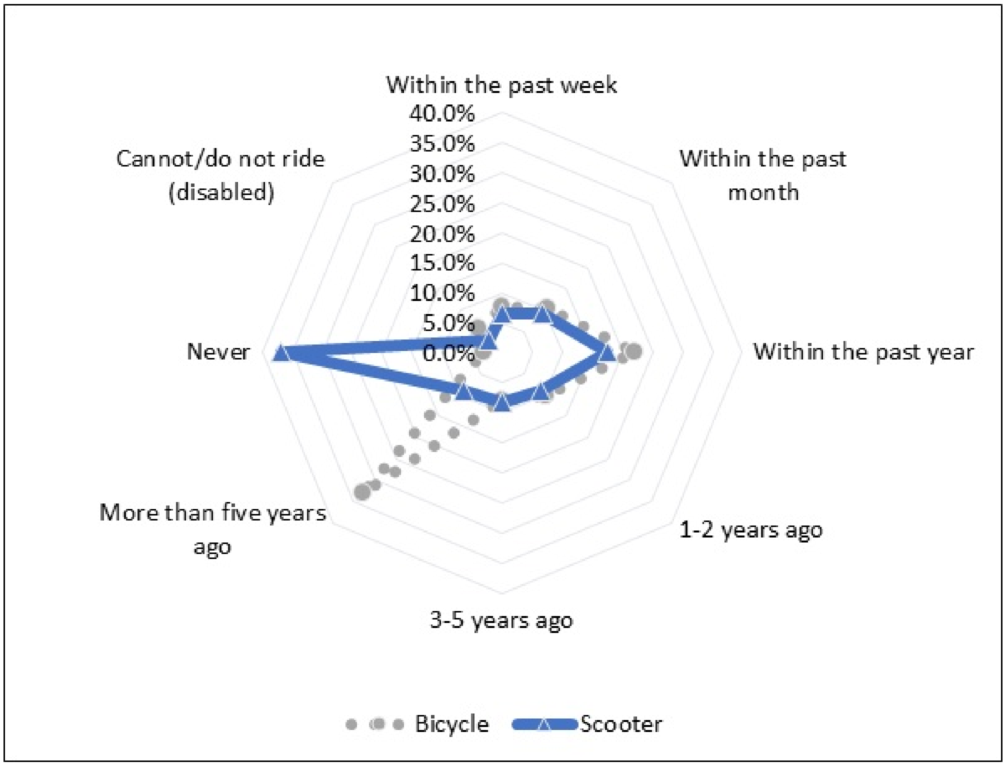


Figure 4.2. Bicycle/scooter riding profile.

Based on miles traveled, people rode slightly longer distances with scooters than with bicycles. Other than that, scooter and bicycle survey responses follow the same pattern, as presented in Tables 4.5 and 4.6. These numbers also imply that most people prefer using scooters or bicycles for short trips between 1 and 7 miles. Further data analysis was conducted to understand how far riders travel by either scooters or bicycles. On average, the most trips made by either scooters or bicycles take 16 – 30 minutes. For travel times less than 5 minutes, the percentage of people using scooters is almost 5% higher than those who rode bicycles. The flexibility offered by dockless scooters might lead to improved accessibility and shorter trip times.

On average, many respondents perceive lower risk across all risky behaviors listed in the survey. However, the distribution of responses of "low" risk is not the same across the two surveys (scooter vs. bicycle), as illustrated in Figures 4.3 and 4.4. Riding a bike with under-inflated tires and riding a scooter without a helmet are perceived as the highest risk activities than any other type of behavior. Surprisingly, "riding at night without lights on" and "distracted riding" are perceived as lower levels of risk in bicycle and scooter surveys, respectively. "Y12," "Y13," "Y14," "Y19," and "Y20" are ranked as the top least-risky

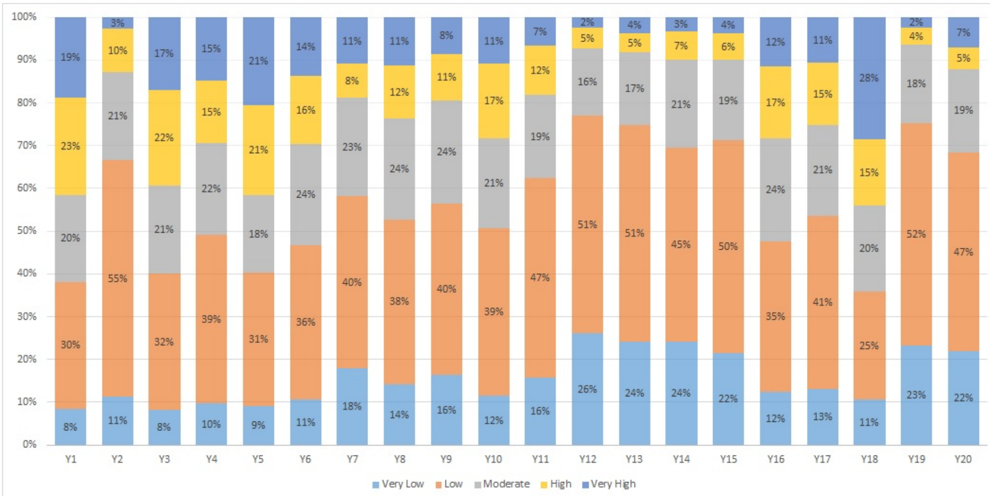


Figure 4.3. Distribution of responses by risky behaviors (bicycle).

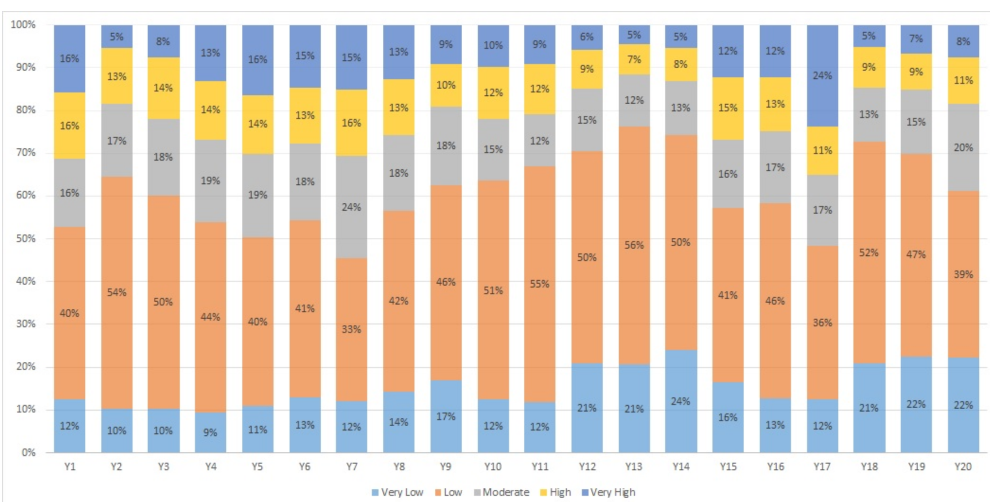


Figure 4.4. Distribution of responses by risky behaviors (scooter).

Table 4.5. Traveled time/distance per trip (Bicycle).

| | | Time (minutes) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **0-5** | **6-15** | **16-30** | **31-50** | **51-75** | **75** | **N/A** | **Total** |
| **Distance (miles)** | **1** | 1.9% | 7.7% | 3.4% | 2.4% | 0.5% | 0.0% | 1.9% | 17.8% |
| | **1-3** | 0.5% | 6.3% | 17.3% | 10.1% | 1.9% | 0.5% | 0.0% | 36.5% |
| | **4-7** | 0.0% | 1.9% | 9.1% | 8.7% | 4.3% | 1.4% | 0.0% | 25.5% |
| | **8-12** | 0.5% | 0.0% | 0.5% | 4.3% | 3.4% | 0.5% | 0.5% | 9.6% |
| | **13** | 0.0% | 0.0% | 1% | 2.4% | 1.4% | 1.9% | 0.0% | 6.7% |
| | **N/A** | 1.0% | 0.5% | 1.0% | 0.0% | 1.0% | 0.5% | 0.0% | 3.8% |
| | **Total** | 3.8% | 16.3% | 32.2% | 27.9% | 12.5% | 4.8% | 2.4% | 100.0% |

Table 4.6. Traveled time/distance per trip (Scooter).

| | | Time (minutes) | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | **0-5** | **6-15** | **16-30** | **31-50** | **51-75** | **75** | **N/A** | **Total** |
| **Distance (miles)** | **1** | 5.9% | 5.9% | 3.2% | 0.5% | 0.5% | 0.5% | 0.0% | 16.8% |
| | **1-3** | 2.2% | 8.6% | 13.5% | 4.3% | 1.6% | 0.5% | 0.5% | 31.4% |
| | **4-7** | 0.0% | 4.9% | 10.3% | 11.9% | 3.2% | 0.5% | 0.5% | 31.4% |
| | **8-12** | 0.5% | 0.5% | 2.7% | 2.7% | 2.2% | 1.6% | 0.0% | 10.3% |
| | **13** | 0.0% | 0.5% | 1.6% | 0.5% | 1.6% | 2.7% | 0.5% | 7.6% |
| | **N/A** | 0.5% | 0.5% | 0.5% | 0.0% | 1.1% | 0.0% | 0.0% | 2.7% |
| | **Total** | 9.2% | 21.1% | 31.9% | 20.0% | 10.3% | 5.9% | 1.6% | 100.0% |

behaviors by the respondents in the two surveys. This might imply that these behaviors are least physically possible to occur in the daily commute, which leads to a lower risk level.

Because the interpretation of coefficients in an ordinal logistic regression is hard to generalize, we only focused on finding statistically significant variables as described in Tables 4.7, 4.8, 4.9, and 4.10. Income and age are the top two statistically significant variables (a significance level of 0.05) for at least ten risky behaviors (dependent variables) in both surveys. Considering the age factor, the eight risky behaviors common between the two types of vehicles included Y1, Y5, Y6, Y9, Y10, Y11, Y15, Y16. Interestingly, age is not a significant factor for explaining risky actions, including "Distracted riding," Wheelie, "Riding without helmets," "Yelling," and "Riding with a passenger." From an income perspective, Y1, Y2, Y3, Y5, Y6, Y7, Y8, and Y19 are the common significant factors for both surveys.

Considering other independent variables, being married is related to "speeding" in both surveys. Further research needs to be done to identify cognitive and emotional factors that influence risk-taking among people with different marital statuses. People living in urban areas have different risk perceptions associated with vehicle types. While "Speeding" with scooters is the only risky behavior explained by the "urban" factor, it is a critical factor for reckless cycling behaviors such as distracted cycling, ignoring stop signs, ignoring traffic signals, and others. The region is the least significant value across all 20 dependent variables, implying that how people perceive risk is not dependent on location.

### 4.5. Conclusions and Implications

The emergence of shared micromobility systems like bike- and scooter sharing systems attract many commuters to use these vehicles, but they are also an inconvenience to many residents because of their reckless use. Some riders show reckless behaviors, causing a backlash against these modes of transportation. Hence, the objective of this article was to evaluate how people perceive risk associated with a set of reckless behaviors of scooter/bicycle riders. The motivation to conduct this research was the increasing number of news articles about incidents related to bikes and scooters across the United States.

The results show that using scooters as a means of transportation (36 % never rode a scooter before) is at the early stage of development in the U.S. Considering the percentage of enthusiastic and interested, there is much more room available for scooter sharing program to be expanded. Also, the percentage of people not sure about their scooters/bicycles riding skills draw the attention to more education program in this respect. The short travel time either by scooter or bicycle (16-30 minutes) requires enforcement approach should be agile and could track an individual riding behavior while using the vehicle.

In this study, eight determinants such as age, sex, income, and others were identified to explain 20 risky behaviors. In general, in both surveys, participant risk perception of each identified behavior is relatively low. This may be because the frequency of the incidents is low in the region where participants are located and are not observed by many residents. From the perceived risk model, age and income play a critical role in explaining most of the risky behaviors in both surveys. Education levels differ between the two surveys and in explaining the risky behaviors. Having a high school diploma or less made in the scooter survey scooters and having a bachelor's degree in the bicycle survey are dominant factors to determine the risky behaviors. One reason might be the availability of bike sharing programs on university campuses. However, this might change in the future as scooter sharing companies have been expanding across the United States during the past two years.

From the open-ended question, certain behaviors are not addressed in the surveys. Riding with no hands, holding onto vehicles, riding abreast instead of a single file are the respondents' major concerns. Also, respondents have observed many risky behaviors from kids, which is not the scope of this study. Not wearing a helmet is another concern that is already on the list but frequently repeated in this section. The

Table 4.7. Significant factors by response variables (Y1-Y10) from bicycle survey (95% significance level).

| **Predictor** | **Y1** | **Y2** | **Y3** | **Y4** | **Y5** | **Y6** | **Y7** | **Y8** | **Y9** | **Y10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | | | | |
| *25-34* | | | | | | | | | | |
| *35-44* | | | | | | | | | | |
| *45-54* | | | | | | | | | | * |
| *55-64* | * | | * | | * | * | | | * | * |
| *65-74* | | | | | | | | | | |
| *75 or older* | | | | | | * | * | | * | |
| **Gender** | | | | | | | | | | |
| *Male* | | | | | | | | | | |
| **Marital Status** | | | | | | | | | | |
| *Married* | | * | | | | | * | | | |
| *Separated* | | | | | | | | | | |
| *Single* | | * | | | | | | | | |
| *Widowed* | | | | | | | | | | |
| **Education** | | | | | | | | | | |
| *Bachelor degree* | | * | * | * | * | * | | | | |
| *Graduate degree* | | | | | | | | | | |
| *High school diploma or equivalent* | | | | | | | | | | |
| *Less than a high school diploma* | | | | | | | | | | |
| *Some college, no degree* | | | | | | | | | | |
| **Employment Status** | | | | | | | | | | |
| *Employed full-time* | | | | | | | | | | |
| *Employed part-time* | | | | | | | | | | |
| *Retired* | | | | | | | * | * | | |
| *Self-employed* | | | | | | | | | | |
| *Unemployed (currently looking for work)* | | | | | | | | | | |
| *Unemployed (not currently looking for work)* | | | | | | | | * | | |
| **Income** | | | | | | | | | | |
| *$10k - $25k* | | * | * | | | | * | * | | |
| *$25k - $50k* | * | | * | | * | | | | | |
| *$50k - $75k* | * | * | * | | * | * | * | * | | |
| *$75k - $100k* | | | * | | | | | | | |
| *$100k - $125k* | | | * | | * | | | | | |
| *$125k - $150k* | | | | | | | | | | |
| *Over $150k* | * | | | | | | | | | |
| **Race** | | | | | | | | | | |
| *Black or African-American* | | | | | | | | | | |
| *Middle Eastern American* | | | | | | | | | | |
| *Native American or Alaska Native* | | * | | | | | | * | * | |
| *Native Hawaiian or other Pacific islander* | | | | | | | | | | |
| *White American* | | * | * | | | | | | | * |
| *Other race* | * | * | | | | | | | * | * |
| **Region** | | | | | | | | | | |
| *Suburban* | | | | | | | | | | |
| *Urban* | * | | * | | | | | | | |

Table 4.8. Significant factors by response variables (Y11-Y20) from bicycle survey (95% significance level).

| **Predictor** | **Y11** | **Y12** | **Y13** | **Y14** | **Y15** | **Y16** | **Y17** | **Y18** | **Y19** | **Y20** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | | | | |
| *25-34* | | | | | | | | | | |
| *35-44* | | | | | | | | | | |
| *45-54* | | | | | | | | | | |
| *55-64* | * | | | | * | * | | * | | |
| *65-74* | | | | | | | | | | |
| *75 or older* | * | | | | * | | | | | |
| **Gender** | | | | | | | | | | |
| *Male* | | | | | | | | | | |
| **Marital Status** | | | | | | | | | | |
| *Married* | | | | | | | | | | |
| *Separated* | | | | | | | | | | |
| *Widowed* | | | | | | | | | | |
| **Education** | | | | | | | | | | |
| *Bachelor degree* | | | | | | * | * | * | | |
| *Graduate degree* | | | | | | | | | | |
| *High school diploma or equivalent* | | | | | | | | | | |
| *Less than a high school diploma* | | | | | | | | | | |
| *Some college, no degree* | | | | | | | | | | |
| **Employment Status** | | | | | | | | | | |
| *Employed full-time* | | | | | | | | | | |
| *Employed part-time* | | | | | | | * | | | |
| *Retired* | | | | | | | | | | |
| *Self-employed* | | | | | | | | | | |
| *Unemployed (currently looking for work)* | | | | | | | | | | |
| *Unemployed (not currently looking for work)* | | | | | | * | * | | * | |
| **Income** | | | | | | | | | | |
| *$10k - $25k* | | | | | | | | * | | |
| *$25k - $50k* | | | | | | | | * | | |
| *$50k - $75k* | | | | | | | | * | * | |
| *$75k - $100k* | | | | | | | | * | | |
| *$100k - $125k* | | * | | | | | | * | | |
| *$125k - $150k* | | | | | | | | | | |
| *Over $150k* | | | | | | | | | | |
| **Race** | | | | | | | | | | |
| *Black or African-American* | | | | | | | | | | |
| *Middle Eastern American* | | | | | | | | | | |
| *Native American or Alaska Native* | | | * | | | | | | | * |
| *Native Hawaiian or other Pacific islander* | | | | | | | | | | |
| *White American* | | | | | * | | | | | * |
| *Other race* | * | * | | | | | | | | |
| **Region** | | | | | | | | | | |
| *Suburban* | | | | | | | | | | |
| *Urban* | | | | | | | | | | |

Table 4.9. Significant factors by response variables (Y1-Y10) from scooter survey (95% significance level).

| **Predictor** | **Y1** | **Y2** | **Y3** | **Y4** | **Y5** | **Y6** | **Y7** | **Y8** | **Y9** | **Y10** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | | | | |
| *25-34* | | * | | | | | | | | |
| *35-44* | * | * | | | | * | | * | * | * |
| *45-54* | * | | | | | * | * | * | | |
| *55-64* | | | | | | | | | | |
| *65-74* | * | | | | * | | * | | | |
| *75 or older* | | | | | | | | | | |
| **Gender** | | | | | | | | | | |
| *Male* | | | | | | * | | | | * |
| **Marital Status** | | | | | | | | | | |
| *Married* | | | | | | | * | | | |
| *Separated* | * | | | | | | * | * | | |
| *Single* | | | | | | | | | | |
| *Widowed* | | | | | | | | | | |
| **Education** | | | | | | | | | | |
| *Bachelor degree* | | | | | | | | | | |
| *Graduate degree* | | | | | | | | | | |
| *High school diploma or equivalent* | | | | | | | | | | |
| *Less than a high school diploma* | | | | | | | | | | |
| *Some college, no degree* | | | | | | | | | | |
| **Employment Status** | | | | | | | | | | |
| *Employed full-time* | | | | | * | | * | * | | |
| *Employed part-time* | | | | | | | | | | |
| *Retired* | | | | | | | | | | |
| *Self-employed* | | | | | | | | * | | |
| *Unemployed (currently looking for work)* | | | | | * | | | | | |
| *Unemployed (not currently looking for work)* | | | | | | | | | | |
| **Income** | | | | | | | | | | |
| *$10k - $25k* | * | * | * | * | * | * | * | * | * | * |
| *$25k - $50k* | * | | | * | * | | | | | |
| *$50k - $75k* | * | | * | * | * | * | * | * | * | * |
| *$75k - $100k* | | | | | | | | | | |
| *$100k - $125k* | | | | | | | | | | |
| *$125k - $150k* | | | | | * | * | | | | |
| *Over $150k* | | | | | | | | | | |
| **Race** | | | | | | | | | | |
| *Black or African-American* | | | | | | | | | * | |
| *Middle Eastern American* | | | | | | | | | | |
| *Native American or Alaska Native* | | | | | | | | | | |
| *Native Hawaiian or other Pacific islander* | | | | | | | | | | |
| *White American* | | | * | | | | | | | |
| *Other race* | | | * | | | | | | | |
| **Region** | | | | | | | | | | |
| *Suburban* | | | | | | | | | | |
| *Urban* | | | | | | | * | | | |

Table 4.10. Significant factors by response variables (Y11-Y20) from scooter survey (95% significance level).

| **Predictor** | **Y11** | **Y12** | **Y13** | **Y14** | **Y15** | **Y16** | **Y17** | **Y18** | **Y19** | **Y20** |
|---|---|---|---|---|---|---|---|---|---|---|
| **Age** | | | | | | | | | | |
| *25-34* | | | | | * | | | | | |
| *35-44* | * | * | | * | * | * | | | | |
| *45-54* | | * | | * | * | * | | | | |
| *55-64* | | | | * | | | | | | |
| *65-74* | | | | | | | | | | |
| *75 or older* | | | | | | | | | | |
| **Gender** | | | | | | | | | | |
| *Male* | * | | | | * | * | | | | |
| **Marital Status** | | | | | | | | | | |
| *Married* | | | | | | | | | | |
| *Separated* | * | * | | * | * | * | | | | |
| *Single* | | | | | * | * | | | | |
| *Widowed* | | | | | | | | | | |
| **Education** | | | | | | | | | | |
| *Bachelor degree* | | | | | | | | | | |
| *Graduate degree* | | | | | | | | | | |
| *High school diploma or equivalent* | | | | | | | | | | |
| *Less than a high school diploma* | | | | | | | | | | |
| *Some college, no degree* | | | | | | | | | | |
| **Employment Status** | | | | | | | | | | |
| *Employed full-time* | | | | | | | | | | |
| *Employed part-time* | | | | | | | | | | |
| *Retired* | | | | | * | | | | | |
| *Self-employed* | | | | | | | | | | |
| *Unemployed (currently looking for work)* | | | | | | | | | | |
| *Unemployed (not currently looking for work)* | | | | | | | | | | |
| **Income** | | | | | | | | | | |
| *$10k - $25k* | * | | | * | * | * | * | | * | |
| *$25k - $50k* | | | | | * | | * | | * | |
| *$50k - $75k* | * | | | | * | | * | | * | |
| *$75k - $100k* | | | | | | | | | | |
| *$100k - $125k* | | | | | | | | | | |
| *$125k - $150k* | | | | | | | * | | * | |
| *Over $150k* | | | | | | | | | | |
| **Race** | | | | | | | | | | |
| *Black or African-American* | | * | * | | | | | | | |
| *Middle Eastern American* | | | | | | | | | | |
| *Native American or Alaska Native* | | | | | | | | | | |
| *Native Hawaiian or other Pacific islander* | | | | | | | | | | |
| *White American* | | * | * | | | | | | | |
| *Other race* | | | | | | | | | | |
| **Region** | | | | | | | | | | |
| *Suburban* | | | | | | | | | | |
| *Urban* | | | | | | | | | | |

feedback could help future research to have a comprehensive survey questionnaire or focus on a specific issue.

### 4.5.1. Limits of the Study

Though the sample size is small, the findings suggest that there are great opportunities to understand the perceived risk of road users towards scooter and bike riders. Bigger sample size and city-level survey would definitely help narrow down potential factors. The other avenue could be a longitudinal survey before, and after micromobility services are available in a region. Second, since we wanted to make the survey feasible, the number of risky behaviors listed in the surveys is limited. At the beginning of this study, little was known about the specific risky behaviors, so we tried to cover as many as possible. From the results of the open-ended question, more risky behaviors could be covered in future research. Third, this study did not systematically explore riding behaviors of an individual from current behavior models, extra question on general riding behaviors might elaborate more detail about an individual experience on use of bicycle or scooter as well as their interactions with different aspects of transportation system like network geometry, rights of way, public safety among others.

### 4.5.2. Practical Applications

#### 4.5.2.1. Contribution to Research

First, we extend the literature on risk perception by doing two separate surveys to understand the magnitude and frequency of risk from road users' perspective. We also developed the risk matrix and risk perception model to identify the significant factors explaining each risk. This work could be a starting point to identify other possible factors that cause any risky behaviors. The risk matrix is also used to build predictive models to characterize users' riskiness.

#### 4.5.2.2. Contribution to Practice

Our findings offer several insights for practitioners. This work investigates the potential risks associated with scooters and bike riders and may help city planners and system operators to set policies or appropriate enforcement to reduce any harm from rogue riders. Second, the results of the risk matrix help to quantify the penalties related to each risky behavior. The results also help officials to design educational programs to mitigate any reckless behaviors.

# REFERENCES


[1] Mobike, "How cycling changes cities." https://mobike.com/global/public/HowCyclingChangesCitiesMobike.pdf. Accessed on: 2020/03/01.

[2] NHTS, "National household travel survey." https://nhts.ornl.gov/vehicle-trips. Accessed on: 2020/03/01.

[3] R. P. Godavarthy and A. R. Taleqani, "Winter bikesharing in US: User willingness, and operator's challenges and best practices," *Sustainable Cities and Society*, vol. 30, pp. 254–262, Apr. 2017.

[4] NACTO, "Bike share in the u.s.: 2017." https://nacto.org/wp-content/uploads/2018/05/NACTO-Bike-Share-2017.pdf. Accessed on: 2018/05/10.

[5] T. Nieuwesteeg, "Dockless bikes promise the future of transportation, but litter the city of dallas." https://www.nbcnews.com/tech/innovation/dockless-bikes-promise-/future-transportation-litter-city-dallas-n866351. Accessed on: 2018/04/01.

[6] J. L. Lee, "Bike-sharing companies face an uphill ride in u.s.." https://www.reuters.com/article/us-usa-bikesharing/bike-sharing-companies-/face-an-uphill-ride-in-u-s-idUSKCN1GS0YX. Accessed on: 2018/03/05.

[7] A. Rahim Taleqani, C. Vogiatzis, and J. Hough, "Maximum closeness centrality-clubs: A study of dock-less bike sharing," *Journal of Advanced Transportation*, vol. 2020, 2020.

[8] B. Y. Liau and P. P. Tan, "Gaining customer knowledge in low cost airlines through text mining," *Industrial Management & Data Systems*, vol. 114, pp. 1344–1359, Oct. 2014.

[9] P. Dass, Y. Lu, M. Chowdhury, D. Lampl, J. Kamalanathan, and K. E. Nygard, "Gender differences in perceptions of genetically modified foods," in *proceedings of 31st International Conference on Computers and their Applications (CATA'16), Las Vegas, Nevada, USA*, 2016.

[10] K. Ravi and V. Ravi, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," *Knowledge-Based Systems*, vol. 89, pp. 14–46, Nov. 2015.

[11] D. M. Blei, “Probabilistic topic models,” *Communications of the ACM*, vol. 55, p. 77, Apr. 2012.

[12] J. S. Evans-Cowley and G. Griffin, “Microparticipation with social media for community engagement in transportation planning,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2307, pp. 90–98, Jan. 2012.

[13] S. Bregman, , and and, *Uses of Social Media in Public Transportation*. National Academies Press, May 2012.

[14] A. Pak and P. Paroubek, “Twitter as a corpus for sentiment analysis and opinion mining.,” in *LREc*, vol. 10, pp. 1320–1326, 2010.

[15] A. Go, R. Bhayani, and L. Huang, “Twitter sentiment classification using distant supervision,” *CS224N project report, Stanford*, vol. 1, no. 12, p. 2009, 2009.

[16] D. Davidov, O. Tsur, and A. Rappoport, “Enhanced sentiment learning using twitter hashtags and smileys,” in *Proceedings of the 23rd international conference on computational linguistics: posters*, pp. 241–249, Association for Computational Linguistics, 2010.

[17] T. Hoang, P. H. Cher, P. K. Prasetyo, and E.-P. Lim, “Crowdsensing and analyzing micro-event tweets for public transportation insights,” in *2016 IEEE International Conference on Big Data (Big Data)*, IEEE, Dec. 2016.

[18] J. Wojtowicz and W. A. Wallace, “Use of social media by transportation agencies for traffic management,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2551, pp. 82–89, Jan. 2016.

[19] C. Collins, S. Hasan, and S. Ukkusuri, “A novel transit rider satisfaction metric: Rider sentiments measured from online social media data,” *Journal of Public Transportation*, vol. 16, pp. 21–45, June 2013.

[20] S. M. Kaufman and M. L. Moss, “Co-monitoring for transit management: Using web-based rider input for transit management,” tech. rep., 2014.

[21] L. Schweitzer, “Planning and social media: A case study of public transit and stigma on twitter,” *Journal of the American Planning Association*, vol. 80, pp. 218–238, July 2014.

[22] C. H. Papadimitriou, P. Raghavan, H. Tamaki, and S. Vempala, “Latent semantic indexing: A probabilistic analysis,” *Journal of Computer and System Sciences*, vol. 61, pp. 217–235, Oct. 2000.

[23] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *Journal of machine Learning research*, vol. 3, no. Jan, pp. 993–1022, 2003.

[24] O. Koltsova and S. Koltcov, “Mapping the public agenda with topic modeling: The case of the russian livejournal,” *Policy & Internet*, vol. 5, pp. 207–227, June 2013.

[25] D. N. Sotiropoulos, C. D. Kounavis, P. Kourouthanassis, and G. M. Giaglis, “What drives social sentiment? an entropic measure-based clustering approach towards identifying factors that influence social sentiment polarity,” in *IISA 2014, The 5th International Conference on Information, Intelligence, Systems and Applications*, IEEE, July 2014.

[26] A. Steinskog, J. Therkelsen, and B. Gambäck, “Twitter topic modeling by tweet aggregation,” in *Proceedings of the 21st nordic conference on computational linguistics*, pp. 77–86, 2017.

[27] T. H. Rashidi, A. Abbasi, M. Maghrebi, S. Hasan, and T. S. Waller, “Exploring the capacity of social media data for modelling travel behaviour: Opportunities and challenges,” *Transportation Research Part C: Emerging Technologies*, vol. 75, pp. 197–211, Feb. 2017.

[28] E. Aramaki, S. Maskawa, and M. Morita, “Twitter catches the flu: detecting influenza epidemics using twitter,” in *Proceedings of the conference on empirical methods in natural language processing*, pp. 1568–1576, Association for Computational Linguistics, 2011.

[29] E. Jain and S. K. Jain, “Using mahout for clustering similar twitter users: Performance evaluation of k-means and its comparison with fuzzy k-means,” in *2014 International Conference on Computer and Communication Technology (ICCCT)*, IEEE, Sept. 2014.

[30] K. Kireyev, L. Palen, and K. Anderson, “Applications of topics models to analysis of disaster-related twitter data,” in *NIPS workshop on applications for topic models: text and beyond*, vol. 1, Canada: Whistler, 2009.

[31] G. Kalton and H. Schuman, “The effect of the question on survey responses: A review,” *Journal of the Royal Statistical Society. Series A (General)*, vol. 145, no. 1, p. 42, 1982.

[32] A. J. Nederhof, "Methods of coping with social desirability bias: A review," *European Journal of Social Psychology*, vol. 15, pp. 263–280, July 1985.

[33] P. DeMaio and J. Gifford, "Will smart bikes succeed as public transportation in the united states?," *Journal of Public Transportation*, vol. 7, no. 2, p. 1, 2004.

[34] S. A. Shaheen, A. P. Cohen, and E. W. Martin, "Public bikesharing in north america: early operator understanding and emerging trends," *Transportation research record*, vol. 2387, no. 1, pp. 83–92, 2013.

[35] F. F. Shahraki, A. P. Yazdanpanah, E. E. Regentova, and V. Muthukumar, "Bicycle detection using hog, hsc and mlbp," in *International Symposium on Visual Computing*, pp. 554–562, Springer, 2015.

[36] M. Harris, "The bike share war is shaking up seattle like nowhere else." `https://www.wired.com/story/the-bike-share-war-is-shaking-up-seattle-like-nowhere/-else/`. Accessed on: 2018/06/27.

[37] D. Garrick, "San diego considering crackdown on dockless bikes, including fees, new rules - the san diego union-tribune." `http://www.sandiegouniontribune.com/news/politics/sd-me-dockless-bikes-20180620-story.html`. Accessed on: 2018/06/27.

[38] National Association of City Transportation Officials, "Bike share ridership increases when stations are closer together." `https://nacto.org/wp-content/uploads/2015/09/NACTO_Walkable-Station-Spacing-Is-Key-For-Bike-Share_Sc.pdf`. "Accessed on: 2018/04/23.

[39] L. Conrow, A. T. Murray, and H. A. Fischer, "An optimization approach for equitable bicycle share station siting," *Journal of Transport Geography*, vol. 69, no. April, pp. 163–170, 2018.

[40] F. Chiariotti, C. Pielli, A. Zanella, and M. Zorzi, "A dynamic approach to rebalancing bike-sharing systems," *Sensors (Switzerland)*, vol. 18, no. 2, pp. 1–22, 2018.

[41] S. Wang, T. He, D. Zhang, Y. Shu, Y. Liu, Y. Gu, C. Liu, H. Lee, and S. H. Son, "BRAVO : Improving the Rebalancing Operation in Bike Sharing with Rebalancing Range Prediction," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 2, no. 1, pp. 1–22, 2018.

[42] L. Pan, Q. Cai, Z. Fang, P. Tang, and L. Huang, "A deep reinforcement learning framework for rebalancing dockless bike sharing systems," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1393–1400, 2019.

[43] M. Nickelsburg, "Dockless bike sharing is expanding to bellevue, applying lessons from seattle – geekwire." https://www.geekwire.com/2018/dockless-bike-sharing-expanding-bellevue-applying-lessons-seattle/. Accessed on: 2018/03/30.

[44] H. Murphy, "Dublin bikes: An investigation in the context of multimodal transport," *MSc Sustainable Development, Dublin Institute of Technology: Dublin, Ireland*, 2010.

[45] Z. Liu, X. Jia, and W. Cheng, "Solving the Last Mile Problem: Ensure the Success of Public Bicycle System in Beijing," *Procedia - Social and Behavioral Sciences*, vol. 43, pp. 73–78, 2012.

[46] J. Bachand-Marleau, B. H. Y. Lee, and A. M. El-Geneidy, "Better understanding of factors influencing likelihood of using shared bicycle systems and frequency of use," *Transportation Research Record*, vol. 2314, no. 1, pp. 66–71, 2012.

[47] E. Fishman, S. Washington, and N. Haworth, "Barriers and facilitators to public bicycle scheme use: A qualitative approach," *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 15, no. 6, pp. 686–698, 2012.

[48] S. Jäppinen, T. Toivonen, and M. Salonen, "Modelling the potential effect of shared bicycles on public transport travel times in greater helsinki: An open data approach," *Applied Geography*, vol. 43, pp. 13–24, 2013.

[49] R. Nair, E. Miller-Hooks, R. C. Hampshire, and A. Bušić, "Large-Scale Vehicle Sharing Systems: Analysis of Vélib'," *International Journal of Sustainable Transportation*, vol. 7, no. 1, pp. 85–106, 2012.

[50] A. Audikana, E. Ravalet, V. Baranger, and V. Kaufmann, "Implementing bikesharing systems in small cities: Evidence from the Swiss experience," *Transport Policy*, vol. 55, no. November 2016, pp. 18–28, 2017.

[51] K. B. Campbell and C. Brakewood, “Sharing riders: How bikesharing impacts bus ridership in New York City,” *Transportation Research Part A: Policy and Practice*, vol. 100, pp. 264–282, 2017.

[52] B. Boyaci, K. G. Zografos, and N. Geroliminis, “An optimization framework for the development of efficient one-way car-sharing systems,” *European Journal of Operational Research*, vol. 240, no. 3, pp. 718–733, 2015.

[53] J. R. Lin, T. H. Yang, and Y. C. Chang, “A hub location inventory model for bicycle sharing system design: Formulation and solution,” *Computers and Industrial Engineering*, vol. 65, no. 1, pp. 77–86, 2013.

[54] J. R. Lin and T. H. Yang Ta-Hui, “Strategic design of public bicycle sharing systems with service level constraints,” *Transportation Research Part E: Logistics and Transportation Review*, vol. 47, no. 2, pp. 284–294, 2011.

[55] L. M. Martinez, L. Caetano, T. Eiró, and F. Cruz, “An optimisation algorithm to establish the location of stations of a mixed fleet biking system: an application to the city of lisbon,” *Procedia-Social and Behavioral Sciences*, vol. 54, pp. 513–524, 2012.

[56] R. Nair and E. Miller-Hooks, “Equilibrium design of bicycle sharing systems: the case of Washington D.C.,” *EURO Journal on Transportation and Logistics*, vol. 5, no. 3, pp. 321–344, 2016.

[57] D. Reijsbergen, “Probabilistic Modelling of Station Locations in Bicycle-Sharing Systems,” in *Federation of International Conferences on Software Technologies: Applications and Foundations*, pp. 83–97, Springer, 2016.

[58] J. C. García-palomares, J. Gutiérrez, and M. Latorre, “Optimizing the location of stations in bike-sharing programs : A GIS approach,” *Applied Geography*, vol. 35, no. 1-2, pp. 235–246, 2012.

[59] P. Vogel and D. C. Mattfeld, “Modeling of repositioning activities in bike-sharing systems,” in *World conference on transport research (WCTR)*, 2010.

[60] J. Shu, M. C. Chou, Q. Liu, C.-P. Teo, and I.-L. Wang, “Models for Effective Deployment and Redistribution of Bicycles Within Public Bicycle-Sharing Systems,” *Operations Research*, vol. 61, no. 6, pp. 1346–1359, 2013.

[61] I. A. Forma, T. Raviv, and M. Tzur, "A 3-step math heuristic for the static repositioning problem in bike-sharing systems," *Transportation research part B: methodological*, vol. 71, pp. 230–247, 2015.

[62] R. Alvarez-Valdes, J. M. Belenguer, E. Benavent, J. D. Bermudez, F. Muñoz, E. Vercher, and F. Verdejo, "Optimizing the level of service quality of a bike-sharing system," *Omega (United Kingdom)*, vol. 62, pp. 163–175, 2016.

[63] J. Schuijbroek, R. C. Hampshire, and W.-J. Van Hoeve, "Inventory rebalancing and vehicle routing in bike sharing systems," *European Journal of Operational Research*, vol. 257, no. 3, pp. 992–1004, 2017.

[64] S. Yan, J. R. Lin, Y. C. Chen, and F. R. Xie, "Rental bike location and allocation under stochastic demands," *Computers and Industrial Engineering*, vol. 107, pp. 1–11, 2017.

[65] D. Çelebi, A. Yörüsün, and H. Işık, "Bicycle sharing system design with capacity allocations," *Transportation Research Part B: Methodological*, vol. 114, pp. 86–98, 2018.

[66] J.-M. Bourjolly, G. Laporte, and G. Pesant, "An exact algorithm for the maximum k-club problem in an undirected graph," *European Journal of Operational Research*, vol. 138, no. 1, pp. 21–28, 2002.

[67] F. M. Pajouh and B. Balasundaram, "On inclusionwise maximal and maximum cardinality k-clubs in graphs," *Discrete Optimization*, vol. 9, no. 2, pp. 84–97, 2012.

[68] F. D. Carvalho and M. T. Almeida, "Upper bounds and heuristics for the 2-club problem," *European Journal of Operational Research*, vol. 210, no. 3, pp. 489–494, 2011.

[69] M. G. Everett and S. P. Borgatti, "The centrality of groups and classes," *The Journal of mathematical sociology*, vol. 23, no. 3, pp. 181–201, 1999.

[70] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.

[71] A. Veremyev, O. A. Prokopyev, and E. L. Pasiliao, "Finding groups with maximum betweenness centrality," *Optimization Methods and Software*, vol. 32, no. 2, pp. 369–399, 2017.

[72] C. Vogiatzis, A. Veremyev, E. L. Pasiliao, and P. M. Pardalos, "An integer programming approach for finding the most and the least central cliques," *Optimization Letters*, vol. 9, no. 4, pp. 615–633, 2015.

[73] M. Rysz, F. M. Pajouh, and E. L. Pasiliao, “Finding clique clusters with the highest betweenness centrality,” *European Journal of Operational Research*, vol. 271, no. 1, pp. 155–164, 2018.

[74] C. Vogiatzis and M. C. Camur, “Identification of essential proteins using induced stars in protein–protein interaction networks,” *INFORMS Journal on Computing*, 2019.

[75] B. Balasundaram, S. Butenko, and S. Trukhanov, “Novel approaches for analyzing biological networks,” *Journal of Combinatorial Optimization*, vol. 10, no. 1, pp. 23–39, 2005.

[76] J.-M. Bourjolly, G. Laporte, and G. Pesant, “An exact algorithm for the maximum k-club problem in an undirected graph,” *European Journal of Operational Research*, vol. 138, no. 1, pp. 21 – 28, 2002.

[77] A. Buchanan and H. Salemi, “Parsimonious formulations of low-diameter clusters,” *Optimization Online Eprints*, vol. 3, pp. 14–21, 2017.

[78] Transportation Networks for Research Core Team, “Transportation networks for research.” `https://github.com/bstabler/TransportationNetworks`. Accessed on: 2018/07/10.

[79] Gurobi Optimization, “Gurobi optimizer reference manual,” 2016.

[80] R. Godavarthy, J. Mattson, and A. R. Taleqani, “Evaluation study of the bike share program in fargo, north dakota,” tech. rep., SURLC, 2017.

[81] F. &. Peers, “Bike share implementation for los angeles county regional plan.” `https://clkrep.lacity.org/onlinedocs/2015/15-0985_misc_f_08-20-15.pdf`. Accessed on: 2018/04/23.

[82] A. R. Taleqani, J. Hough, and K. E. Nygard, “Public opinion on dockless bike sharing: A machine learning approach,” *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2673, pp. 195–204, Apr. 2019.

[83] K. Anderson-Hall, B. Bordenkircher, R. O’Neil, and S. C. Scott, “Governing micro-mobility: A nationwide assessment of electric scooter regulations,” tech. rep., 2019.

[84] N. A. of City Transportation Officials, “Nacto shared micromobility in 2018.” `https://nacto.org/shared-micromobility-2018/`. Accessed on: 2019/06/02.

[85] M. Winters, K. Teschke, M. Grant, E. M. Setton, and M. Brauer, “How far out of the way will we travel? Built environment influences on route selection for bicycle and car travel,” *Transportation Research Record*, vol. 2190, no. 1, pp. 1–10, 2010.

[86] “E-scooters: a transport ’tsunami’ flooding cities worldwide.” https://phys.org/news/2019-07-e-scooters-tsunami-cities-worldwide.html. Accessed on: 2019/07/01.

[87] G. Gardner, “Nashville mayor proposed e-scooter ban; then council changed the rules.” https://www.forbes.com/sites/greggardner/2019/09/16/nashville-mayor-/proposed-e-scooter-ban-then-council-changed-the-rules/#3ae347ecef1a. Accessed on: 2019/10/01.

[88] N. H. T. S. Administration, “Traffic safety facts 2017 a compilation of motor vehicle crash data.” https://crashstats.nhtsa.dot.gov. Accessed on: 2019/07/01.

[89] S. A. Shaheen, E. W. Martin, A. P. Cohen, N. D. Chan, and M. Pogodzinski, “Public Bikesharing in North America During a Period of Rapid Expansion: Understanding Business Models, Industry Trends & User Impacts, MTI Report 12-29,” 2014.

[90] E. Martin, A. Cohen, J. L. Botha, and S. Shaheen, “Bikesharing and bicycle safety,” tech. rep., 2016.

[91] K. C. Heesch, J. Garrard, and S. Sahlqvist, “Incidence, severity and correlates of bicycling injuries in a sample of cyclists in queensland, australia,” *Accident Analysis & Prevention*, vol. 43, pp. 2085–2092, Nov. 2011.

[92] J. Garrard, S. Greaves, and A. Ellison, “Cycling injuries in Australia: road safety’s blind spot?,” *Journal of the Australasian College of Road Safety*, vol. 21, no. 3, p. 37, 2010.

[93] S. O’Brien, “Curbing the road wars between motorists and cyclists.” https://www.heraldsun.com.au/news/opinion/susie-obrien/curbing-the-road-wars-between-motorists-and-cyclists/news-story/130c70df71368c79730ce3463634f556. Accessed on: 2020/01/07.

[94] Z. Williams, “Bikes vs cars: why it’s war between cyclists and drivers on city streets.” https://www.theguardian.com/film/2015/may/03/bikes/

-vs-car-film-war-cyclists-drivers-fredrik-gertten-interview. Accessed on: 2020/01/07.

[95] A. Tapp, S. Rundle-Thiele, R. Anibaldi, S. Warren, and A. Beardmore, “Road wars? The role of language in perceptions of bikes and cars sharing the road: Possible implications for social marketing interventions,” in *Australia and New Zealand Marketing Academy*, Australian & New Zealand Marketing Academy (ANZMAC), 2014.

[96] P. M. Salmon, M. G. Lenné, G. H. Walker, and A. Filtness, “Investigating the factors influencing cyclist awareness and behaviour: an on-road study of cyclist situation awareness,” *Journal of the Australasian College of Road Safety*, vol. 24, no. 4, p. 7, 2013.

[97] A. Macmillan, A. Roberts, J. Woodcock, R. Aldred, and A. Goodman, “Trends in local newspaper reporting of london cyclist fatalities 1992-2012: the role of the media in shaping the systems dynamics of cycling,” *Accident Analysis & Prevention*, vol. 86, pp. 137–145, Jan. 2016.

[98] E. S. Chataway, S. Kaplan, T. A. S. Nielsen, and C. G. Prato, “Safety perceptions and reported behavior related to cycling in mixed traffic: A comparison between brisbane and copenhagen,” *Transportation Research Part F: Traffic Psychology and Behaviour*, vol. 23, pp. 32–43, Mar. 2014.

[99] C. Rissel, C. Bonfiglioli, A. Emilsen, and B. J. Smith, “Representations of cycling in metropolitan newspapers - changes over time and differences between sydney and melbourne, australia,” *BMC Public Health*, vol. 10, June 2010.

[100] T. Bogdanowicz, “Cycling and the media,” *Intermedia-London*, vol. 32, pp. 21–22, 2004.

[101] D. Skinner and P. Rosen, “Hell is other cyclists: rethinking transport and identity,” in *Cycling and society*, pp. 99–112, Routledge, 2016.

[102] M. Møller and T. Hels, “Cyclists’ perception of risk in roundabouts,” *Accident Analysis & Prevention*, vol. 40, pp. 1055–1062, May 2008.

[103] J. Parkin, M. Wardman, and M. Page, “Models of perceived cycling risk and route acceptability,” *Accident Analysis & Prevention*, vol. 39, pp. 364–371, Mar. 2007.

[104] M. S. Klobucar and J. D. Fricker, "Network evaluation tool to improve real and perceived bicycle safety," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2031, pp. 25–33, Jan. 2007.

[105] B. W. Landis, V. R. Vattikuti, R. M. Ottenberg, T. A. Petritsch, M. Guttenplan, and L. B. Crider, "Intersection level of service for the bicycle through movement," *Pedestrians and Bicycles 2003: Safety and Human Performance*, no. 1828, pp. 101–106, 2003.

[106] L. Leden, P. Gårder, and U. Pulkkinen, "An expert judgment model applied to estimating the safety effect of a bicycle facility," *Accident Analysis & Prevention*, vol. 32, pp. 589–599, July 2000.

[107] R. G. Hughes and D. L. Harkey, "Cyclists' perception of risk in a virtual environment: effects of lane conditions, traffic speed, and traffic volume," in *Proceedings of the Conference on Traffic Congestion and Traffic Safety in the 21st Century*, pp. 132–138, ASCE, 1997.

[108] R. B. Noland and H. Kunreuther, "Short-run and long-run policies for increasing bicycle transportation for daily commuter trips," *Transport Policy*, vol. 2, pp. 67–79, Jan. 1995.

[109] V. F. Reyna and F. Farley, "Risk and rationality in adolescent decision making," *Psychological Science in the Public Interest*, vol. 7, pp. 1–44, Sept. 2006.

[110] R. E. DAHL, "Adolescent brain development: A period of vulnerabilities and opportunities. keynote address," *Annals of the New York Academy of Sciences*, vol. 1021, pp. 1–22, June 2004.

[111] H. Feenstra, R. A. Ruiter, and G. Kok, "Social-cognitive correlates of risky adolescent cycling behavior," *BMC Public Health*, vol. 10, July 2010.

[112] J. T. Shope and C. R. Bingham, "Teen driving: motor-vehicle crashes and factors that contribute," *American journal of preventive medicine*, vol. 35, no. 3, pp. S261–S271, 2008.

[113] T. C. Boas, D. P. Christenson, and D. M. Glick, "Recruiting large online samples in the United States and India: Facebook, mechanical turk, and qualtrics," *Political Science Research and Methods*, pp. 1–19, 2018.

[114] D. L. Poston, Jr. and L. F. Bouvier, *Age and Sex Composition*, p. 228–264. Cambridge University Press, 2010.

[115] R. S. Ross, “Guide for conducting risk assessments,” tech. rep., 2012.

[116] J. Dill and N. McNeil, “Four types of cyclists? examination of typology for better understanding of bicycling behavior and potential,” *Transportation Research Record*, vol. 2387, no. 1, pp. 129–138, 2013.