

## Research Article

# Maximum Closeness Centrality $k$ -Clubs: A Study of Dock-Less Bike Sharing

Ali Rahim Taleqani <sup>1</sup>, Chrysafis Vogiatzis <sup>2</sup>, and Jill Hough <sup>1</sup>

<sup>1</sup>North Dakota State University, Fargo, ND 58108, USA

<sup>2</sup>University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA

Correspondence should be addressed to Chrysafis Vogiatzis; [chrys@illinois.edu](mailto:chrys@illinois.edu)

Received 10 January 2019; Revised 20 June 2019; Accepted 9 August 2019; Published 24 January 2020

Academic Editor: Monica Menendez

Copyright © 2020 Ali Rahim Taleqani et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this work, we investigate a new paradigm for dock-less 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 dock-less bicycles can only be found in specific locations. In addition, we use closeness centrality as this metric allows for quick access to dock-less 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. 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 downtown areas along with the economic development that follows, health benefits, and increase in environmental awareness [1–3].

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 *dock-less* or *free-floating* BSS. In a dock-less 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, dock-less 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, dock-less BSSs also present new challenges. The one we deal

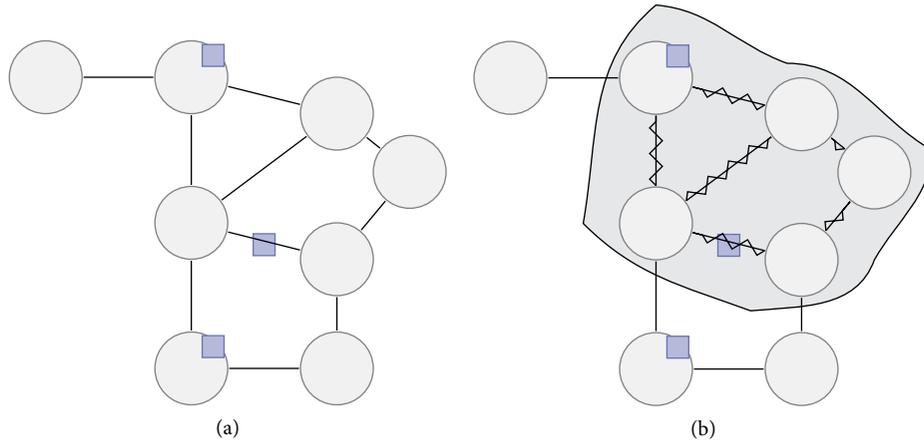


FIGURE 1: (a) Without  $k$ -club, (b) with  $k$ -club.

with in this work is the fact that bicycles can now be left unattended in improper positions [4, 5]. 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 dock-less 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 dock-less bike sharing. We propose a framework that will both (i) allow users the increased benefits of a dock-less system (easy and fast access to bicycles, reduced parking space needs) and (ii) reduce the coordination costs for controlling the sprawl of the dock-less 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 [6]. 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 [7]. 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. proposed a dynamic model to address the fluctuations in demands of a BSS in New York City [8]. In a similar note, Wang et al. 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 [9]. 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 dock-less BSSs because of unrestrained parking locations [10]. Finding the right locations for stations and rebalancing are correlative problems. According to the Department of Transportation, a dock-less 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 dock-less system [11].

As our framework will optimize the reach of dock-less 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 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:

- (i) 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.
- (ii) 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.

- (iii) 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.

## 2. Related Works

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;
- (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 [12], Beijing [13], Montreal [14], Brisbane [15], Helsinki [16], Paris [17], Switzerland [18], and New York [19]. Nair et al. examine several aspects of the Velib BSS in Paris, France [17]. Their findings show that integrating transit and BSS can yield higher utilization. Bachand-Marleau et al. surveyed residents of Montréal, Quebec, in Canada to determine the factors leading to use BSS as well as the frequency of use [14]. Campbell and Brakewood quantify the impact that BSSs have on bus ridership in New York City [19]. 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. studied the impact of a BSS in a small city (less than 100,000 residents) in Switzerland [18]. 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 [20]. 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. address the strategic design problem by formulating it as a hub location inventory model [21]. 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 propose a nonlinear integer method to determine the optimal location, bike lanes, and routes [22]. Their model assumes a penalty for uncovered demand but does not consider relocation (rebalancing) of bikes. Martinez et al. 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 [23]. Nair and Miller-Hooks formulate an equilibrium network design model to address the same objective as the previous study [24]. They propose a meta-heuristic solution approach to overcome the intractability of the exact solution for real-life, large-scale networks. In another study Reijsbergen 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 [25]. The literature offers other methodologies, that are not based on facility location models, to define the location of the stations. Garcia-Palomares et al. develop a GIS-based model to calculate the spatial distribution of the potential demand for trips and find the locations of bike stations using the location-allocation modeling approach [26].

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 apply a system dynamic method to model the effect of dynamic repositioning on the service level [27]. Shu et al. 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 [28]. Forma et al. develop a 3-step heuristic and mixed integer linear programming model for repositioning [29]. 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. address the static repositioning problem using simulation techniques in two stages [30]. 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. combine service level requirements and vehicle routes to rebalance the inventory [31]. They propose a “cluster-first route-second” heuristic considering the service level feasibility and approximate routing costs simultaneously. Yan et al. develop four planning models for leisure-based BSSs given deterministic and stochastic demands [32]. They apply nonlinear time-space network models to integrate bike repositioning and vehicle routing with user dissatisfaction estimations. In a recent study, Celebi et al. propose a hybrid approach jointly considering location decisions and capacity allocation [33]. 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 [34–36]. 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, aims to identify groups or classes of high centrality [37]. Centrality measures the aim to characterize the importance of an element in a network. They typically fall into three main classes [38], 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 [39]. 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 [40, 41], or inducing a star [42].

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 dock-less 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);
- (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 dock-less 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. Definitions and Notation

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 nonnegative 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, nonnegative 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

$$f(S) = \sum_{i \in V} w_i d_{si}. \quad (1)$$

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 [34, 43]. 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  $C_1, C_2, \dots, C_m$  and  $n$  literals and their complements  $x_1, x_2, \dots, x_n$  and  $\bar{x}_1, \bar{x}_2, \dots, \bar{x}_n$ , does there exist an assignment such that a formula  $C_1 \wedge C_2 \wedge \dots \wedge C_m$  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  $M \gg k$ . 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 2.

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  $u_c \in V_c$ .
- (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  $\bar{k} \leq k - 1$  nodes in  $S$  and the remaining chains can have at most  $k - \bar{k}$ , where  $1 \leq \bar{k} \leq \lceil k/2 \rceil$ . 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  $k - \bar{k}$  from  $S$ . Hence, we have:

$$f(S) \geq (m - 1) \cdot M + M + \bar{k} = m \cdot M + \bar{k} > m \cdot M. \quad (2)$$

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  $\bar{k} \leq k$  nodes from one chain with the remaining chains having at most  $k - \bar{k}$  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:

$$\begin{aligned} f(S) &\geq (m - 2) \cdot (M - \bar{k} + M) + (M - \bar{k} + k + M) \\ &\geq (m - 2) \cdot (M - k + M) + (M - k + k + M) \quad (M \gg k) \\ &= (m - 2) \cdot 2 \cdot M + 2 \cdot M \\ &= 2(m - 1)M. \end{aligned} \quad (3)$$

By assumption, though, we have that  $f(S) \leq m \cdot M$ , which, combined with inequality (3), leads to:

$$2(m - 1)M \leq m \cdot M \Rightarrow m \leq 2, \quad (4)$$

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  $\bar{k}$  from the clause-node  $u_c$  and at a distance of  $M - k - \bar{k}$  from the literal-node  $u_\ell$  of that chain. We then have one clause at a distance of  $\bar{k}$ , at most  $m - 2$  clauses (as, otherwise, literal  $\ell$  satisfies all clauses, a contradiction) at a distance of  $M - k - \bar{k} + M$ , and at least 1 clause at a distance of, at best,  $M - k - \bar{k} + k + M$ , leading to:

$$\begin{aligned} f(S) &\geq \bar{k} + (m - 2) \cdot (M - k - \bar{k} + M) \\ &\quad + (M - k - \bar{k} + k + M) \\ &= (m - 2) \cdot (2 \cdot M - k - \bar{k}) + 2 \cdot M \quad (M \gg k) \\ &= (m - 2) \cdot 2 \cdot M + 2 \cdot M \\ &= 2(m - 1)M. \end{aligned} \quad (5)$$

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:

$$f(S) \geq (m - 1) \cdot M + (M + k) = m \cdot M + k > m \cdot M. \quad (6)$$

This last contradiction finishes the proof.  $\square$

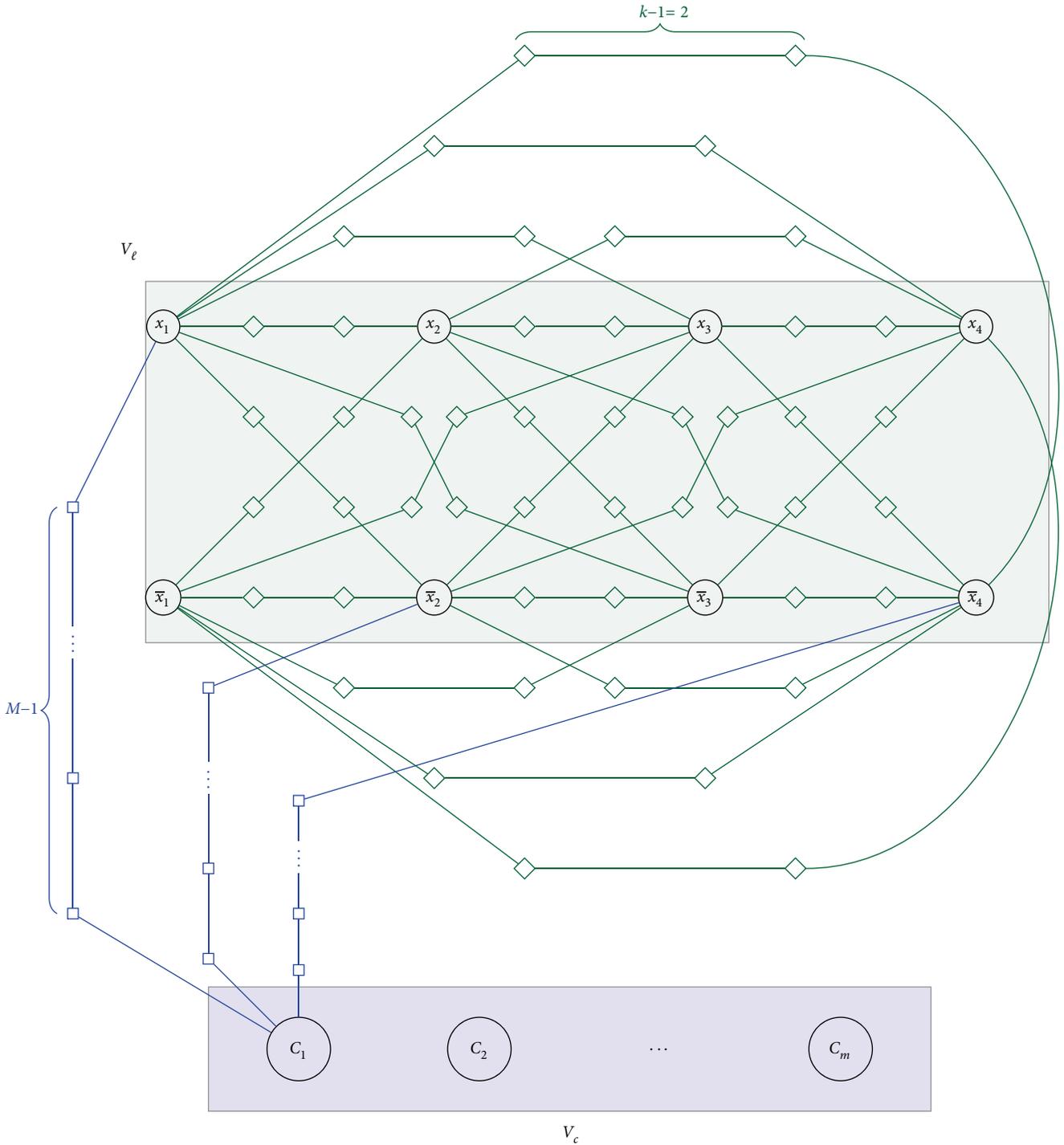


FIGURE 2: An example of the gadget used to reduce an instance of 3SAT to our problem to an instance of our problem with  $k = 3$ . For simplicity, we only show one clause,  $C_1 = x_1 \vee \bar{x}_2 \vee \bar{x}_4$ . The other clauses would be similarly connected to the nodes-literals of  $V_\ell$  through a chain of  $M - 1$  nodes.

### 4. 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 ( $k = 2, 3$ ).

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

$$\begin{aligned}
 x_i^\ell &= \begin{cases} 1, & \text{if node } i \in V \text{ is at a distance of } 0 \leq \ell \leq D \text{ from the } k\text{-club,} \\ 0, & \text{otherwise.} \end{cases} \\
 y_p &= \begin{cases} 1, & \text{if path } p \in \mathcal{P}^k \text{ is within the } k\text{-club,} \\ 0, & \text{otherwise.} \end{cases}
 \end{aligned}
 \tag{7}$$

We can now proceed to describe the mathematical formulation, shown in (7). It is based on the maximum  $k$ -club chain formulation presented in [34]. Newer formulations for identifying  $k$ -clubs (as in, e.g., [44]) can also be employed, but are not explored here.

$$\min \sum_{i \in V} \sum_{\ell=0}^D \ell \cdot w_i \cdot x_i^\ell, \quad (8a)$$

$$s.t. \ x_i^0 \leq y_p, \quad \forall i \in p, \forall p \in \mathcal{P}^k, \quad (8b)$$

$$\sum_{\ell=0}^D x_i^\ell = 1, \quad \forall i \in V, \quad (8c)$$

$$x_i^{\ell+1} \leq \sum_{j \in N(i)} x_j^\ell, \quad \forall i \in V, 0 \leq \ell \leq D-1, \quad (8d)$$

$$x_i^0 + x_j^0 \leq 1 + \sum_{p \in \mathcal{P}_{ij}^k} y_p, \quad \forall i, j \in V : i \neq j, \quad (8e)$$

$$x_i^\ell \in \{0, 1\}, \quad \forall i \in V, 0 \leq \ell \leq D, \quad (8f)$$

$$y_p \in \{0, 1\}, \quad \forall p \in \mathcal{P}^k. \quad (8g)$$

The objective function in (8a) 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 (8b) restricts that a path can only be within the  $k$ -club if every node that belongs to it belongs in the  $k$ -club. Constraints (8c) 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 (8d), 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 (8e) 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 (8f) and (8g).

**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.

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:

$$T_{ij} = \frac{w_i * w_j}{2^{d_{ij}}}, \quad (9)$$

```

1 function Greedy_Central_k-club ( $k$ );
   Input : A graph  $G(V, E)$ , weights  $w : V \mapsto \mathbb{R}$ 
   Output: A  $k$ -club
2  $\mathcal{I} \leftarrow V$ ;
3  $S \leftarrow \emptyset$ ;
4 while  $|\mathcal{I}| > 0$  do
5   forall  $i \in \mathcal{I}$  do
6     forall  $j \in V$  do
7        $d_j = \min_{k \in S \cup \{i\}} d_{jk}$ 
8     end
9      $r_i = \sum_{j \in V} \frac{w_j}{2^{d_j}}$ 
10    end
11     $\bar{i} \leftarrow \arg \max_{i \in \mathcal{I}} \{r_i\}$ ;
12     $S \leftarrow S \cup \{\bar{i}\}$ ;
13     $\mathcal{I} \leftarrow (j \in N(S) : d_{ij} \leq k, \forall i \in S)$ ;
14  end
15 return  $S$ 

```

ALGORITHM 1: Greedy Central  $k$ -Club.

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 (9). 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  $w_j/2^{d_j}$ . 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 1.

*Example 1.* Assume that we have the graph of Figure 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

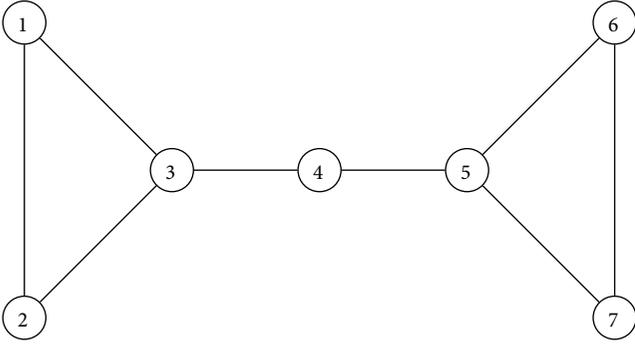


FIGURE 3: An example of how our greedy approach of Algorithm 1 works.

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  $k = 2$ . ( $\mathcal{S} = \{3, 5\}$ ). 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  $\mathcal{S} = \{1, 2, 5\}$ .

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  $S = \{3, 4, 5\}$ .

## 5. 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 2.50GHz 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 [45]. In Table 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).

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 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 (6) becomes prohibitively large.

## 6. 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 (6) 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).

**6.1. 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 of our knowledge, there is no bike-sharing program planned for deployment in the near future. Figure 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 [6], to achieve an increase in ridership as well as in overall system utility, bike-sharing kiosks should be located no more than 1000 feet apart from one another. Therefore, every single vertex has the potential to become a dock-less bike station within 1000 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 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.

All codes for solving the problem, both for the exact optimization model and the greedy heuristic were coded in

TABLE 1: Computational times under heuristic and optimization approach for  $k = 2, 3$ .

Graph	V	E	$D$	2-Club		3-Club	
				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

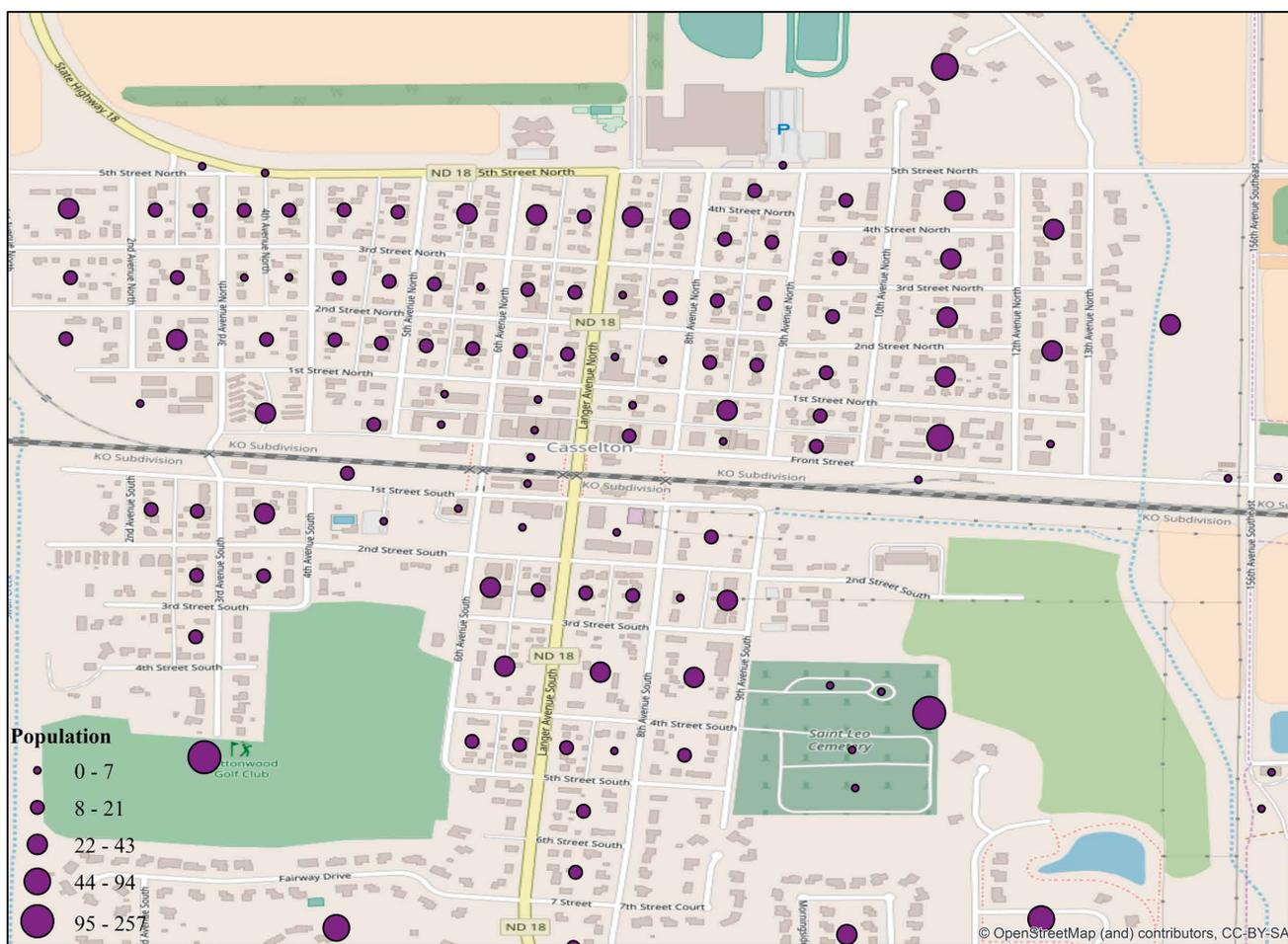


FIGURE 4: The geography, transportation network, and population of Casselton, ND.

Python. For solving the optimization model, we used Gurobi 7.5 [46]. We are now ready to present our findings in the next section.

### 7. Results

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);
- (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

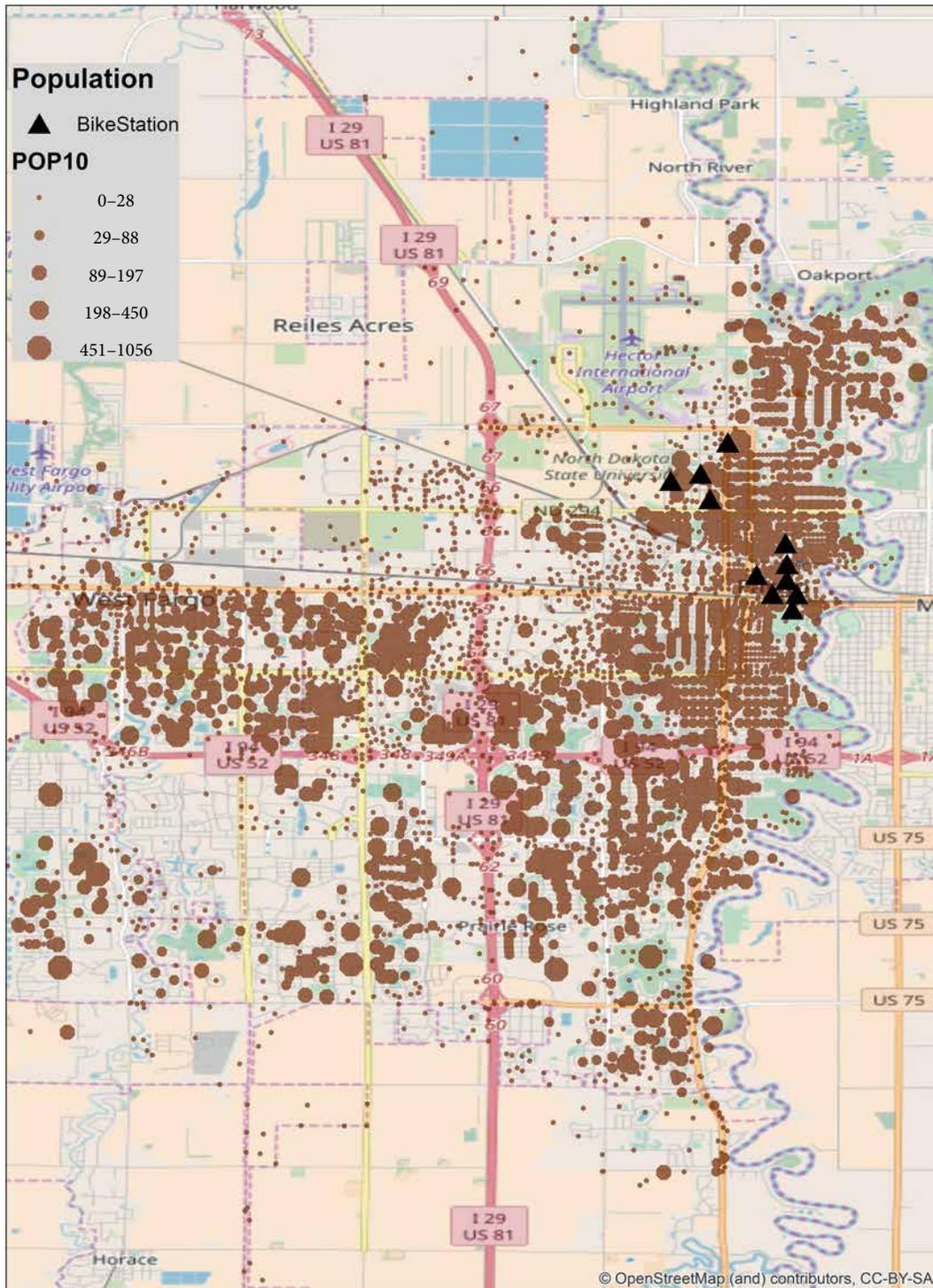


FIGURE 5: The geography, transportation network, and population of Fargo, ND. The locations of the existing eleven bike stations in Fargo are shown with a triangle in the map.

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 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 2 summarizes the results for  $k \in \{2, \dots, 9\}$  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

TABLE 2: Numerical results for Casselton.

$k$	Cardinality		Population		Distance-weighted cost		Time (seconds)	
	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

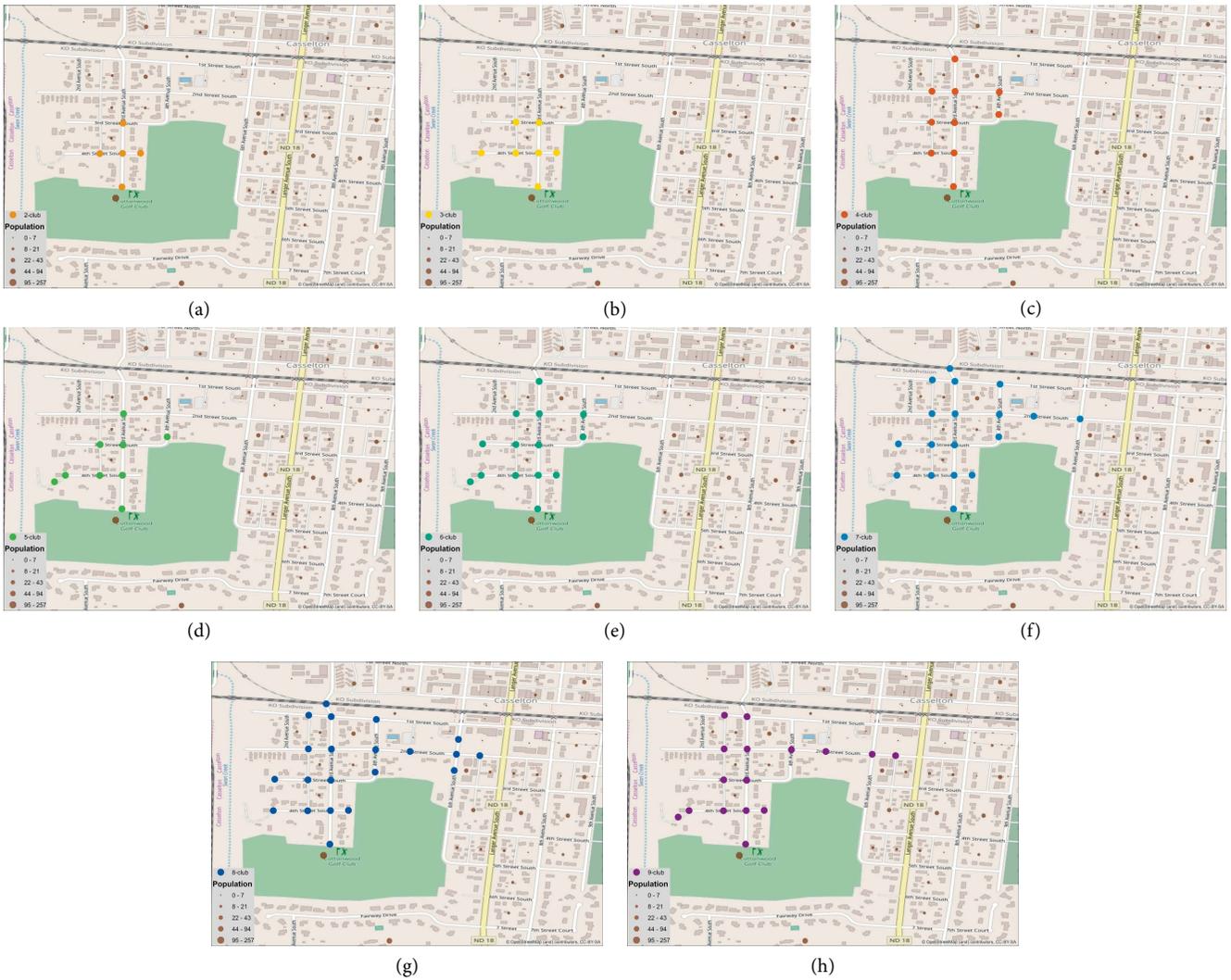


FIGURE 6: The  $k$ -clubs obtained from the heuristic for the city of Casselton. The results are for  $k = 2, \dots, 9$  starting from the top left (for  $k = 2$ ) and ending in the bottom right (for  $k = 9$ ).

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-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

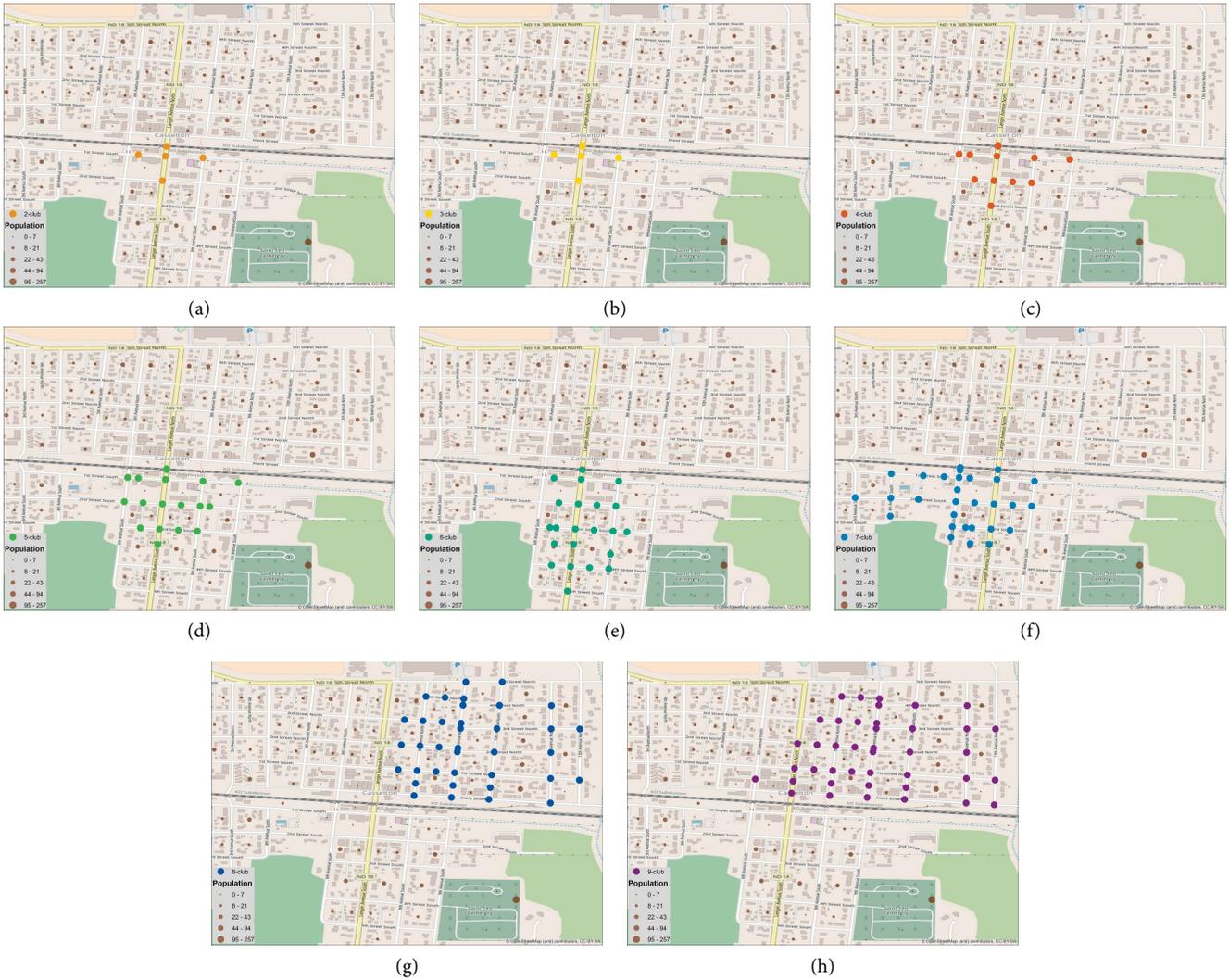


FIGURE 7: The  $k$ -clubs obtained from optimizing the model for the city of Casselton. The results are for  $k = 2, \dots, 9$  starting from the top left (for  $k = 2$ ) and ending in the bottom right (for  $k = 9$ ).

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 6 and 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 dock-less bike sharing, but other residents have to travel very far to access it.

In the case of Fargo, as shown in Figure 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 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 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.

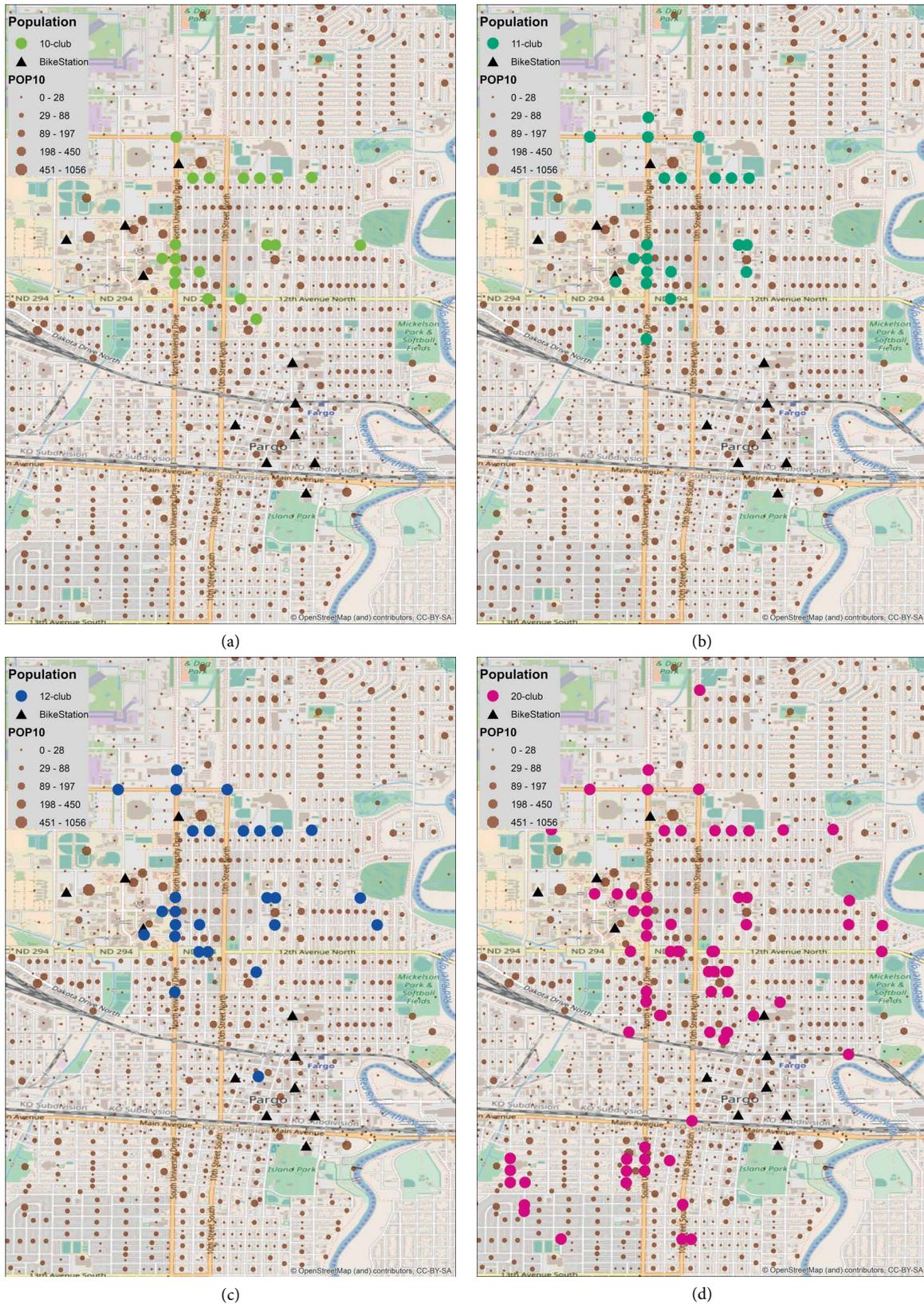


FIGURE 8: The  $k$ -clubs obtained from the heuristic for the city of Fargo. The results are for  $k=10, 11, 12,$  and  $20$  in the top left, top right, bottom left, and bottom right, respectively.

TABLE 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

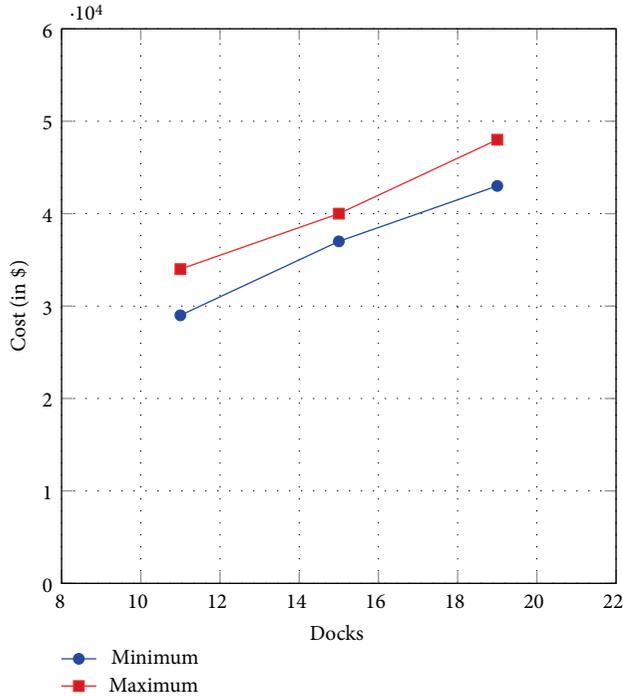


FIGURE 9: Equipment and installation cost vs. number of stations.

## 8. 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 [47]. Assuming a cost of \$1000 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 9 shows the relationship between the cost and number of docks. These figures are even higher at the planning stage (\$55,000 per station) [48]. 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 dock-less 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.

## 9. Conclusions

In this work, we discussed a new paradigm for selecting where a dock-less (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 dock-less 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 dock-less 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 dock-less 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 dock-less 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., dock-less bike sharing and scooter sharing, or dock-less bike sharing and public transit) with respect to different geo-fenced areas.

## Data Availability

The geospatial data used to support this study are made available from the North Dakota GIS Hub Data Portal at <https://gishubdata.nd.gov>. The processed data resulting in the graphs produced are available upon request by Ali-Rahim Taleqani at [ali.rahimtaleqani@ndsu.edu](mailto:ali.rahimtaleqani@ndsu.edu). Previously reported networks were also used to support this study and are available at <https://github.com/bstabler/TransportationNetworks>. The dataset is cited at a relevant place within the text as reference [45].

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## References

- [1] P. DeMaio and J. Gifford, "Will smart bikes succeed as public transportation in the united states?" *Journal of Public Transportation*, vol. 7, no. 2, pp. 1–15, 2004.
- [2] S. A. Shaheen, A. P. Cohen, and E. W. Martin, "Public bikesharing in North America: early operator understanding and emerging trends," *Journal of the Transportation Research Record*, vol. 2387, no. 1, pp. 83–92, 2013.
- [3] F. F. Shahraki, A. P. Yazdanpanah, E. E. Regentova, and M. Venkatesan, "Bicycle detection using hog, hsc and mlbp," *Advances in Visual Computing*, Springer, pp. 554–562, 2015.
- [4] D. Garrick, "San Diego considering crackdown on dockless bikes, including fees, new rules-The San Diego Union-Tribune," 2018, <http://www.sandiegouniontribune.com/news/politics/sd-me-dockless-bikes-20180620-story.html>.
- [5] M. Harris, "The bike share war is shaking up seattle like nowhere else," *Wired*, 2018, <https://www.wired.com/story/the-bike-share-war-is-shaking-up-seattle-like-nowhere-else/>.
- [6] National Association of City Transportation Officials, "Bike share ridership increases when stations are closer together," National Association of City Transportation Officials, 2015, [https://nacto.org/wp-content/uploads/2015/09/NACTO\\_Walkable-Station-Spacing-Is-Key-For-Bike-Share\\_Sc.pdf](https://nacto.org/wp-content/uploads/2015/09/NACTO_Walkable-Station-Spacing-Is-Key-For-Bike-Share_Sc.pdf).
- [7] L. Conrow, A. T. Murray, and H. A. Fischer, "An optimization approach for equitable bicycle share station siting," *Journal of Transport Geography*, vol. 69, pp. 163–170, 2018.
- [8] F. Chiariotti, C. Pielli, A. Zanella, and M. Zorzi, "A dynamic approach to rebalancing bike-sharing systems," *Sensors*, vol. 18, no. 2, pp. 1–22, 2018.
- [9] S. Wang, T. He, D. Zhang et al., "BRAVO: improving the rebalancing operation in bike sharing with rebalancing range prediction," *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 20, no. 1, pp. 1–22, 2018.
- [10] L. Pan, Q. Cai, Z. Fang, P. Tang, and L. Huang, "A deep reinforcement learning framework for rebalancing dockless bike sharing systems," *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, pp. 1393–1400, 2019.
- [11] M. Nickelsburg, "Dockless bike sharing is expanding to Bellevue, applying lessons from Seattle," *GeekWire*, 2018, <https://www.geekwire.com/2018/dockless-bike-sharing-expanding-bellevue-applying-lessons-seattle/>.
- [12] H. Murphy, "Dublin bikes: An investigation in the context of multimodal transport," Dublin Institute of Technology: MSc Sustainable Development, Dublin, Ireland, 2010.
- [13] 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.
- [14] 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: Journal of the Transportation Research Board*, vol. 2314, no. 1, pp. 66–71, 2012.
- [15] 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.
- [16] 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.
- [17] 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.
- [18] A. Audikana, E. Ravalet, V. Baranger, and V. Kaufmann, "Implementing bikesharing systems in small cities: evidence from the swiss experience," *Transport Policy*, vol. 55, pp. 18–28, 2017.
- [19] 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.
- [20] 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.
- [21] 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.
- [22] J. R. Lin and T.-H. Yang, "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.
- [23] 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.
- [24] 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.
- [25] D. Reijtsbergen, "Probabilistic modelling of station locations in bicycle-sharing systems," *Federation of International Conferences on Software Technologies: Applications and Foundations*, Springer, Vienna, Austria, pp. 83–97, 2016.

- [26] 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.
- [27] P. Vogel and D. C. Mattfeld, "Modeling of repositioning activities in bike-sharing systems," *World Conference on Transport Research (WCTR)*, 2010.
- [28] 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.
- [29] I. A. Forma, T. Raviv, and T. Michal, "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.
- [30] R. Alvarez-Valdes, J. M. Belenguer, E. Benavent et al., "Optimizing the level of service quality of a bike-sharing system," *Omega*, vol. 62, pp. 163–175, 2016.
- [31] J. Schuijbroek, R. C. Hampshire, and W.-J. Van Hoes, "Inventory rebalancing and vehicle routing in bike sharing systems," *European Journal of Operational Research*, vol. 257, no. 3, pp. 992–1004, 2017.
- [32] 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.
- [33] D. Çelebi, A. Yörüşün, and H. Isik, "Bicycle sharing system design with capacity allocations," *Transportation Research Part B: Methodological*, vol. 114, pp. 86–98, 2018.
- [34] 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.
- [35] 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.
- [36] M. F. Pajouh and B. Balasundaram, "On inclusionwise maximal and maximum cardinality  $k$ -clubs in graphs," *Discrete Optimization*, vol. 9, no. 2, pp. 84–97, 2012.
- [37] 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.
- [38] L. C. Freeman, "Centrality in social networks conceptual clarification," *Social networks*, vol. 1, no. 3, pp. 215–239, 1978.
- [39] 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.
- [40] 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.
- [41] 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. 90, no. 4, pp. 615–633, 2015.
- [42] C. Vogiatzis and M. C. Camur, "Identification of essential proteins using induced stars in protein-protein interaction networks," *INFORMS Journal on Computing*, 2019.
- [43] 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.
- [44] A. Buchanan and H. Salemi, "Parsimonious formulations of low-diameter clusters," *Optimization Online Eprints*, vol. 3, pp. 14–21, 2017.
- [45] Transportation Networks for Research Core Team, "Transportation networks for research, 2016," 2018, <https://github.com/bstabler/TransportationNetworks>.
- [46] G. Optimization, "Gurobi optimizer reference," *Manual*, 2016, <http://www.gurobi.com>.
- [47] R. Godavarthy, J. Mattson, and A. R. Taleqani, "Evaluation study of the bike share program in Fargo, North Dakota," SURLC, Fargo, 2017.
- [48] Fehr & Peers, "Regional bike share implementation plan," Fehr & Peers, Los Angeles, 2015, [http://clkrep.lacity.org/onlinedocs/2015/15-0985\\_misc\\_f\\_08-20-15.pdf](http://clkrep.lacity.org/onlinedocs/2015/15-0985_misc_f_08-20-15.pdf).