Vancouver School Optimization

Chris Kuharchuk, Kyle Tsia, Edwin Zhu

Department of Mathematics, Simon Fraser University,
Surrey, BC, Canada

1 Introduction

1.1 Background

Vancouver has undergone extreme population growth in the last 20 years [17]. Many city services have struggled to keep up with this increase in demand [10]. The Vancouver school system is a prime example of this struggle. Many of the existing schools are showing clear signs of aging, and were not built with today's needs in mind. This is particularly apparent in higher density, and typically lower income, areas of the city [8]. The number of students per classroom has increased, especially in these areas [8]. Larger class sizes have been linked to decreased academic performance, which is further compounded among economically disadvantaged students [5, 13]. Structural changes to the Vancouver school system must be made in order to provide all students with the same standard of education.

1.2 Scope of Problem

In determining the scope of our problem, we made use of data from the City of Vancouver's Open Data Portal. This provided data on the population, number of schools, and local area boundaries within Vancouver. We decided that using the entire data set would be unmanageable, so we needed some way to narrow it down.
The data has Vancouver divided into 22 different areas. These divisions were common across all of the data sets which we used. School catchment zones are used to determine what school a child in a given area will attend [15]. These zones roughly correlate with the aforementioned divisions [14].

We decided to focus on only one of these areas. We began by looking for differences between them, in the hopes of finding an indication that one would be in particular need of additional resources.

First we looked at the total population, as well the population of children, aged 0 to 19, within each area.

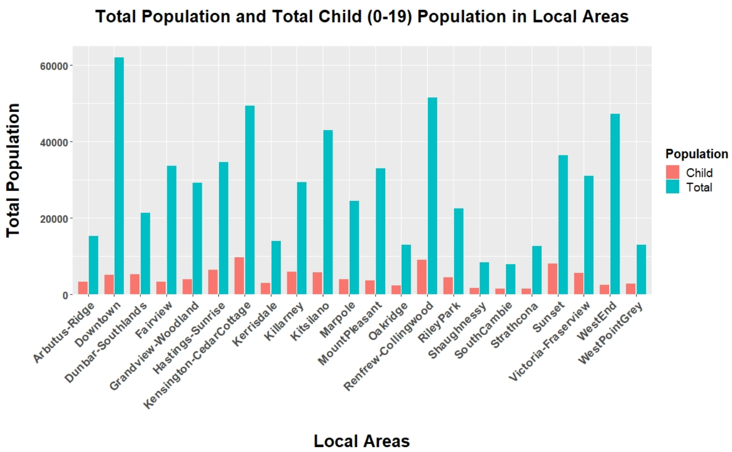

Figure 1: **Plot of Total Population and Total Child Population in Local Areas [8]**

This was a good start, but did not provide enough information on its own. Greater population does not necessarily indicate a greater need, as the area could already have a sufficient number of schools. For example, Renfrew-Collingwood area has a child population of 9120, but also has the most schools of any area at 21 [7, 8].

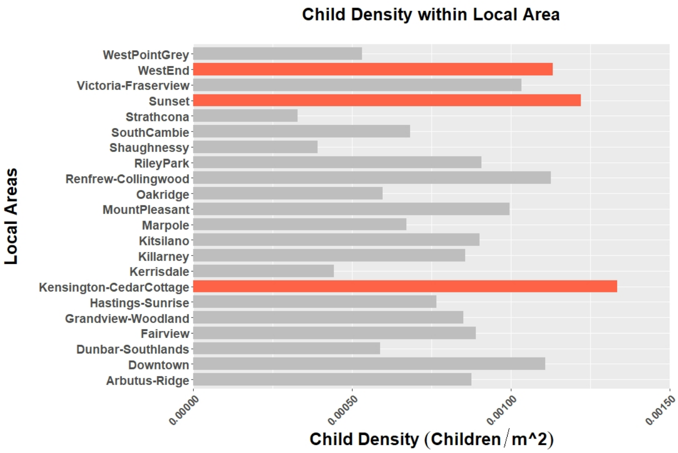

Figure 2: **"Child Density" per Local Area [7]**

Next we considered the child density across the different areas. Child density was determined by dividing the population of children by the total area of each area. The three areas with the highest child density were Sunset, Kensington-Cedar Cottage, and the West End [7].

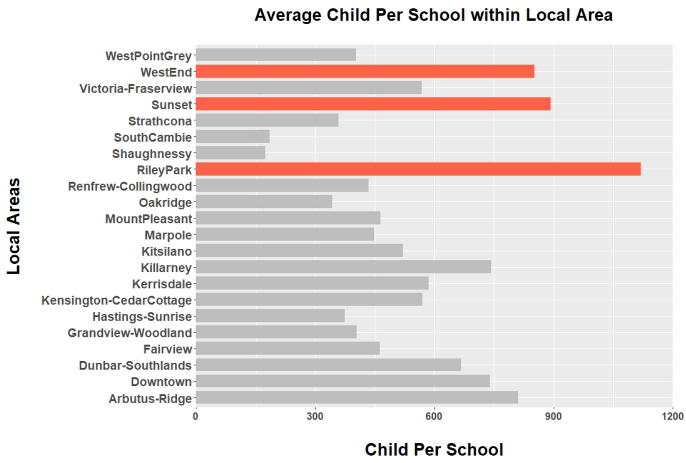

Figure 3: **Average Child Per School in Local Areas [7]**

The average number of children per school was the final statistic we considered. This was determined by dividing the population of children by the number of schools in the area. We identified the three areas with the highest average: Riley Park, Sunset, and the West End [8]. Each of these were significantly above the average of 551 across all areas [7, 8].

Both Sunset and West End appeared in the top three for child density as well as average children per school [7, 8]. This indicated that they may both be in need of additional resources. Sunset has a child population of 8,030, a child density of 0.00122 children/m2, and an average of 892 children per school [7, 8]. This puts it above average in all three categories [7, 8]. Additionally, there seems to be little investment into Sunset schools in recent years, where the newest school is Pierre Elliott Trudeau Elementary School, which was founded in 2002 [14]. Multiple schools in the area have had only minimal renovations since being founded in the early 20th century [4, 16]. For these reasons we decided to focus on Sunset.

1.3 Problem Limitations

Our goal is to reduce the number of children per school, and there are two ways in which this can be achieved. Either we can reduce the population of children, or we increase the number of schools. We have chosen to explore the possibility of adding additional schools within the area. The Vancouver School Board's budget for 2020-2021 is $510.9 million. Historically, 13% of the annual budget is put into operation, maintenance, and transportation expenses, including the construction of new facilities (Vancouver School Board, 2020). This means that there is approximately $66 million available for new school construction. Given an approximate cost of $28 million to build an average school in Vancouver 2020, and assuming that all of this $66 million can be used to construct new schools, the Vancouver School Board budget allows for the construction of at most two new schools [1, 16].

2 Modelling

Assuming that two new schools could be built within the Sunset area, we next had to decide where they should be built [1, 16](Altus, 2020). Our model was intended to determine the best possible location for a new school, given the location of existing schools in the area. This led to the issue of defining what an optimal location was in this context.

One option was to minimize the average distance someone would need to travel from within the area, to the closest school also within the area. This would have been a reasonable choice. Of course some people would not need to travel very far at all, while some people would be a more considerable distance away. This would always be an issue, regardless of where the schools were built.

In the end we decided to model this as a vertex restricted p-center problem [3]. Such a model minimizes the maximum distance between all locations in the area, and the respective location's nearest school. In this form, some locations could end up being somewhat further from a school, but none will end up extremely far away. This solution would minimize the worst-case scenario across the area.

2.1 Data

The Open Data Portal also contains data on all public intersections within Vancouver. We filtered this data to contain only intersections within the Sunset area, ignoring all intersections south of SE Marine Drive. Past this point there are no residential streets, as shown in Figure 4 [9]. This left us with a total of 275 intersections, which we will henceforth index from 1 to 275.

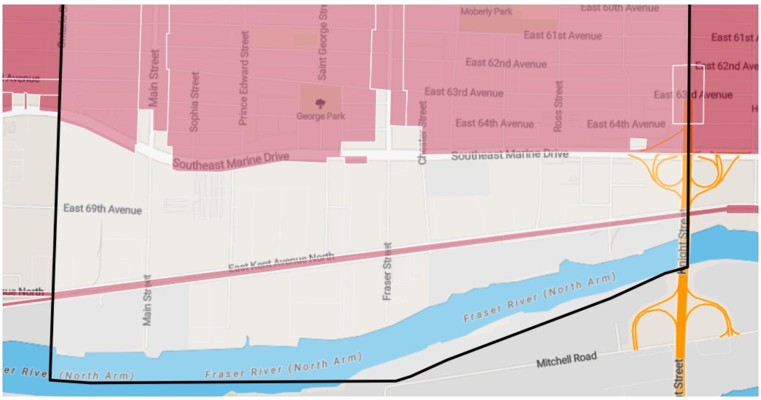

Figure 4: **Zoning of Southern Sunset Area; Red is residential zoning, Black line is Sunset area boundary [9]**

We needed a way to determine the distance between intersections. We decided that Euclidean distance was a poor choice of distance metric for within a city. An alternative would have been calculating the Manhattan distance between pairs of intersections, but this still would not take into account the actual layout of the streets. Instead, we wanted to calculate distance in a way more representative of how people travel. The differences between these metrics can be seen in Figure 5 below.

We represented the reduced intersection data as a graph by determining adjacency between intersections. Edge weights were assigned by calculating the geographical distance between adjacent pairs of intersections. These distances were calculated using the Vincenty inverse formula for ellipsoids method. This method takes the latitude and longitude of each point, and calculates the distance between them in meters using an ellipsoidal representation of Earth [12]. We chose this method as it provides a higher level of accuracy than those that assume a spherical Earth [12].

With the data represented in this way, we could apply Dijkstra's shortest path algorithm in order to determine the shortest path between all pairs of intersections. We stored these values within a matrix D, where d_{ij} the shortest path between intersection i and intersection j.

Figure 5: **Differences between distance metrics**

2.2 Assumptions

One assumption we made was that the children of one area would be restricted to only the schools of their catchment. This is reasonable, as by default children will attend a school within their catchment [15]. Although cross boundary applications are possible, priority first goes to those within their catchment [15]. This allowed us to focus solely on the Sunset area's schools and intersections. We ignored schools in other areas, even though schools just outside the area's boundaries could be closer in some circumstances.

2.3 Decision Variables

Our decision variables for this problem are as follows:

$$X_i = \begin{cases} 1, & \text{if a school is located at intersection i} \\ 0, & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1, & \text{if intersection i is serviced by a school at intersection j} \\ 0, & \text{otherwise} \end{cases}$$

$$Z = \text{maximum distance between an intersection and the school it is serviced by}$$

The most important of these is our set of X variables. These are what will primarily be manipulated in order to reach a solution. The values of Y and Z will be dependent on the state of the X variables, as will be seen in the Constraints section presented below.

2.4 Objective Function

Our objective function is as follows:

$$Minimize\ Z$$

On the surface this objective appears to be very simple. However, based on the construction of our constraints all aspects of this problem will be represented by the variable Z. This will be detailed below.

2.5 Base Constraints

Our objective function will be minimized subject to the constraints listed below. These will work together in order to give Z meaning.

$$\sum_{j=1}^{275} Y_{ij} = 1 \text{ for all i = 1...275} \tag{1}$$

$$Y_{ij} \leq X_j \text{ for all i,j = 1...275} \tag{2}$$

$$Z \geq d_{ij} Y_{ij} \text{ for all i,j = 1...275} \tag{3}$$

$$\sum_{i=1}^{275} X_i = p \tag{4}$$

$$X_i = 1 \text{ for } i \in \{\text{existing schools}\} \tag{5}$$

$$Y_{ij} \leq 1 \text{ for all i,j = 1...275} \tag{6}$$

$$Y_{ij} \geq 0 \text{ for all i,j = 1...275} \tag{7}$$

$$Y_{ij} \in \text{ integers for all i,j = 1...275} \tag{8}$$

$$X_i \leq 1 \text{ integers for all i = 1...275} \tag{9}$$

$$X_i \geq 0 \text{ integers for all i = 1...275} \tag{10}$$

$$X_i \in \text{ integers for all i = 1...275} \tag{11}$$

These constraints will affect our solution as follows:

1. $\sum_{j=1}^{275} Y_{ij} = 1$ for all i = 1...275

   The purpose of this constraint is to ensure that each intersection is serviced by only one school. This will be the school that is closest to a given intersection.

2. $Y_{ij} \leq X_j$ for all i,j = 1...275

   By constraining each Y_{ij} in this way we guarantee proper construction of our matrix Y. It is possible for intersection i to be serviced by a school at intersection j if and only if a school at intersection j exists.

3. $Z \geq d_{ij} Y_{ij}$ for all i,j = 1...275

   Each $d_{ij} Y_{ij}$ represents the distance from intersection i to a school at intersection j. It follows from our second constraint that $d_{ij} Y_{ij}$ equals zero if no school exists at intersection j. This forces Z to take the maximum value from the set of all $d_{ij} Y_{ij}$.

4. $\sum_{i=1}^{275} X_i = p$

This constraint allows us to limit the number of schools which can be placed. Without this the optimal solution would be to place a school at every intersection.

5. $X_i = 1$ for $i \in \{$existing schools$\}$

   The fifth constraint maintains the location of existing schools in our solution. We only want to add new schools. Removing or relocating existing schools is not part of our plan.

6. $Y_{ij} \leq 1$ for all i,j = 1...275
7. $Y_{ij} \geq 0$ for all i,j = 1...275
8. $Y_{ij} \in$ integers for all i,j = 1...275

   Together, constraints 6,7, and 8 make Y_{ij} a binary variable.

9. $X_i \leq 1$ integers for all i = 1...275
10. $X_i \geq 0$ integers for all i = 1...275
11. $X_i \in$ integers for all i = 1...275

    Together, constraints 9,10, and 11 make X_i a binary variable.

2.6 Heuristic Approximation

Solving this as a linear program requires a total of 75,901 decision variables. There are 275 for the X_i's, 75,625 (275^2) for the Y_{ij}'s, and one for Z. Additionally we require a minimum of 151,526 constraints, with one additional constraint for each existing school in the area. This means that the constraint matrix would contain approximately 11.5 billion elements. Each of these elements would require four bytes of memory, whether they are stored in R as integers or floating point values [11]. Simply creating this matrix would require a minimum of 46 Gigabytes of memory. This is a fairly conservative estimate which simplifies the creation of a matrix greatly. However, these numbers are intended to showcase some of the difficulties in working with this type of problem, even when using relatively small amounts of data.

Ignoring these limitations, we would still run into other computational issues. The p-center problem is proven NP-hard in less complex circumstances than what we are facing [3]. Namely, we do not have unit edge lengths, and many of our vertices have degrees of greater than three.

The difficulty of finding solutions for this problem is known. As a result, many heuristic approaches have been proposed [2, 3]. We will also be following a heuristic approach in an attempt to find a solution to our problem.

We will be using the following algorithm:

1. Create a set S, which contains the indices of all existing schools
2. While we want to add another school
   (a) For $i = 1...275$, where intersection i does not currently have a school
      i. For $j = 1...275$
         A. Find the distance between intersection j and all schools, including a school at intersection i
         B. The minimum of these distances is the distance from intersection j to its closest school
         C. Record this as the distance from intersection j to its closest school
      ii. Put closest distances into descending order
      iii. If the first element of closest distance is less than the previous best recorded, set i to be the best location currently found
   (b) Add the best location found to the set S of existing schools

Based on the existing schools, additional schools are placed in a greedy manner. This algorithm is not guaranteed to find the optimal result, as opposed to some other methods which are available. The shortcomings of this algorithm will be explored further within the Limitations section found below.

3 Results and Findings

3.1 Solution

A total of nine Sunset schools are listed in the City of Vancouver's data on schools [8]. However, some schools share the same physical space [8]. For example, the John Henderson StrongStart Centre is located within John Henderson Elementary School [8]. When taking this into account we are left with a total of six physical school locations. In this initial layout there is a maximum distance of 1.18 km between all intersections and their closest schools [6, 8].

The first school we added, in the upper right quadrant of Figure 6, reduced this to a maximum distance of 1.02 km, and the second school, in the upper left quadrant of Figure 6, further reduced this to 1.00 km [8] [6].

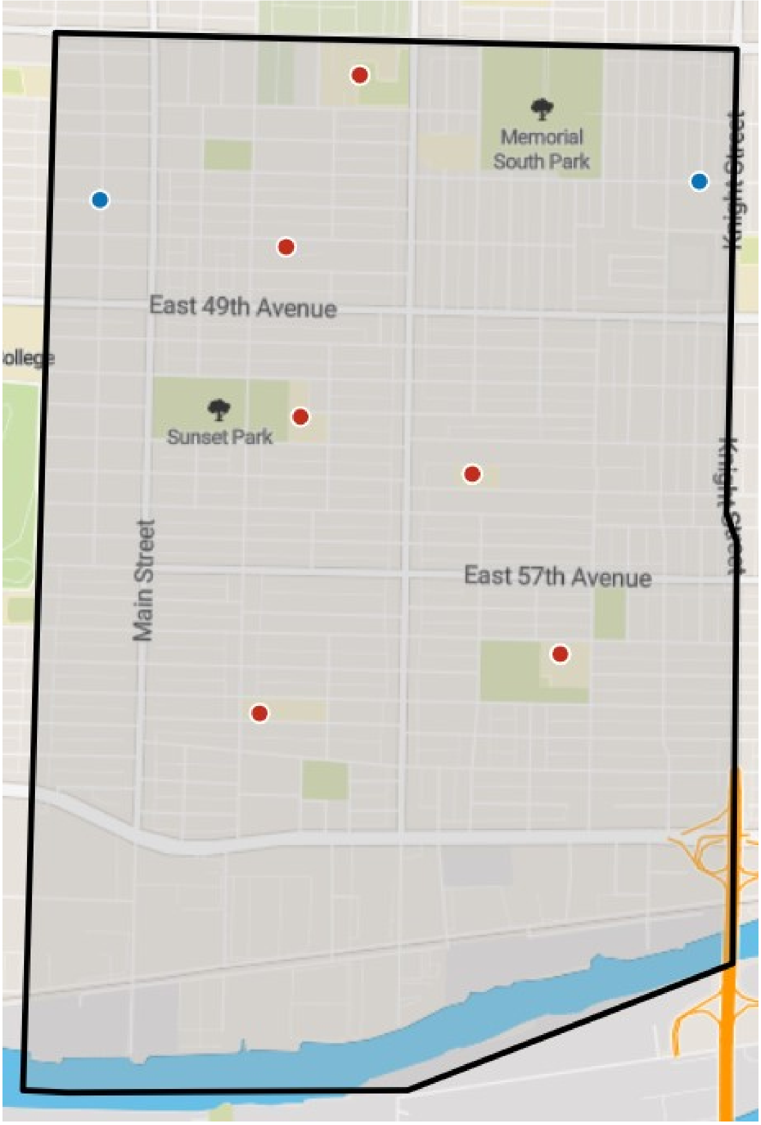

Figure 6: **Proposed placement of two new schools in the Sunset area;Red points are currently existing schools, Blue points are our proposed additions [15]**

This solution does show an improvement from the existing layout. The first school placed significantly reduced the maximum distance. After that we saw some diminishing returns.

The placement of the schools was a bit of a surprise. Optimal locations for both schools were found to be quite close to the area boundary, whereas we expected them to be a little more central. However, to verify the functionality of our algorithm we attempted to place one school in the region, while ignoring all existing schools. As expected, this resulted in a placement directly in the center of the area.

3.2 Result Limitations

There are multiple sources of error within the construction of this problem. Each of these could affect the validity of our solution to some degree. Below we will cover the main sources which we have identified.

3.2.1 Incomplete and Inaccurate Data

Accurate data regarding the layout of the Sunset region is required for a valid solution. The data that we had available was incomplete. Some intersections within the area did not appear within the original data set. We did not add any additional intersections to fill these gaps.

Additionally, the data on intersections did not properly correspond to the data available for streets. This meant that we needed to manually record the links between each intersection. It is possible that some level of human error was introduced at this stage.

Some error was also introduced when we calculated the distances between these intersections. Estimates can be unreliable when calculating short distances using latitude and longitude [12]. This is due to the assumptions which must be made about the shape of the Earth [12]. However, this should not affect our solution too much, as all points are impacted.

3.2.2 Problem Premise

We simplified this problem in order to make it more manageable. Most notably was our decision to use intersections as an analogue for prospective school locations. In order to create our decision variables it was necessary to discretize the region. Data for intersections was available, and seemed to offer a reasonable approximation.

The problem was further simplified by focusing on the Sunset area in isolation. It is likely that some residents of the Sunset area may attend a school elsewhere. It is equally likely that some non-residents of the Sunset area may attend a Sunset school. Our model ignores the fact that the closest school to some residents may not actually be located within Sunset.

Additionally, we chose to treat each school as equivalent to one another. This assumes that each school offers the same programs, has the same capacity, and teaches all ages of students.

3.2.3 Heuristic Method

The algorithm we used for estimating a solution has some clear shortcomings. A linear program would be able to place all schools simultaneously. The iterative nature of our solution means that it can only make the optimal choice for one placement at a time. However, when placing multiple schools this results in a lack of synergy between individual placements. A placement may not be optimal once you account for the placement of all those to follow.

3.3 Extensions

We did not use the real catchment boundaries because we did not have the exact data on them. However, this model could be made more realistic by using the same boundaries that the Vancouver School Board uses.

Another possible extension would be to change our assumption that all children will remain within their catchment. We could then use the entire available set of data rather than focusing on only one area. This would be more computationally intensive, however, may yield a more accurate representation of the areas that need more attention. This would also help in cases where the closest schools for a location are ignored, because they are outside of the catchment. Additional constraints would likely be necessary to manage the varying child density across different areas.

3.4 Final Thoughts

From the start, we knew that we wanted to solve a local problem. We explored the large amount of data available from the City of Vancouver and looked for a problem which we could solve. Our original idea was to allocate fire halls throughout the entirety of Vancouver. The technical limitations, as discussed above, prevented us from pursuing this. However, this laid the foundation to this optimization on school placement.

It became clear that our approach to this problem was not exclusive to schools. These methods can be applied in any case where a limited number of facilities are being placed throughout an area. City infrastructure is particularly well suited to optimizations in this manner. Examples include hospitals, post offices, fire halls, mailboxes, and anything else which should be reasonably accessible to all residents.

References

[1] Altus Group. 2020 Canadian cost guide, 2020.

[2] Anil Aswani. Vertex P-Center Problem. University of Berkeley, Berkeley, CA, 2019.

[3] Hatice Calik. EXACT SOLUTION METHODOLOGIES FOR THE P-CENTER PROBLEM UNDER SINGLE AND MULTIPLE ALLOCATION STRATEGIES. Doctoral Thesis, Bilkent University, Ankara, Turkey, 2013.

[4] Catholic Independent Schools Vancouver Archdiocese. St. Andrews Elementary, 2020.

[5] Matthew M. Chingos and Grover J. "Russ" Whitehurst. Class Size: What Research Says and What it Means for State Policy, 2011.

[6] City of Vancouver. Street Intersections, 2012.

[7] City of Vancouver. Census Local Area Profiles 2016, 2018.

[8] City of Vancouver. Schools, 2019.

[9] City of Vancouver. Zoning Districts and Labels, 2020.

[10] Iglika Ivanova. Reality Check on the Size of BC's Public Sector, 2013.

[11] P Murrell. Data Types and Data Structures. University of Auckland, Auckland, NZ, 2013.

[12] Mario Pineda-Krch. Great-circle distance calculations in R, 2010.

[13] The Alberta Teachers' Association. What is the effect of class size on student learning, 2020.

[14] Vancouver School Board. Pierre Elliott Trudeau, 2018.

[15] Vancouver School Board. Cross Boundary Registration, 2019.

[16] Vancouver School Board. John Oliver; About us, 2020.

[17] World Population Review. Vancouver Population 2020, 2020.