A High Performing and Scalable Model for   
Computing and Visualizing Urban Transit Accessibility

**Yuxuan Cui**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[cuiyuuxuan@gmail.com](mailto:cuiyuuxuan@gmail.com)

**Luka Vukovic**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[luka.vuko@outlook.com](mailto:luka.vuko@outlook.com)

**Rain Shen**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[rain.ya1213@gmail.com](mailto:rain.ya1213@gmail.com)

**Graham Paul Kerford**  
Computer Science and Statistics  
Faculty of Science  
University of British Columbia

[graham.kerford@gmail.com](mailto:graham.kerford@gmail.com)

June 25th, 2021

**ABSTRACT**

Background:

Problem Statement:  
Brief hypothesis/solution statement:  
Brief methodology:  
Results:  
Future work:

**1 INTRODUCTION**

Transportation network analysis is crucial to urban planners, researchers, and policy makers to ensure accessibility remains is a key feature of the urban landscape. This requires both a realistic and a scalable measure for accessibility, which can impact the planning and development of roads and city districts, while also helping analyze a population’s access to healthcare, schools, grocery stores, and other amenities. On a fundamental level, having proper accessibility measures allows urban resource distribution to be more equitable, optimal, and based on informed decisions.

Despite the importance of such measures, much of past accessibility research only considers a limited set of travel modes. These include driving, biking, and walking, but overlook the importance of public transit as a primary mode of travel. (Liu & Zhu, 2004) This may be for a few reasons such as a lack of standardized public transit data, or simply a lack of routing engines that can efficiently model complex journeys through transit networks.

Measuring transit accessibility is crucial and will be more than ever for a few major reasons. First, being in the midst of a Third Industrial Revolution, much focus will be diverted to urban and regional planning. (Roberts, 2015) Sharing economies and sustainability driven planning will undoubtedly increase our dependence on public transit and electric vehicles, although it is uncertain what fraction of society will own such technology within the next few decades. Instead, one might find access to electric vehicles through autonomous ridesharing systems, which will more than likely be useful for filling holes in current public transit networks (this *may require a reference*). Not to mention, car dependence generally has poor outcomes on physical health, psychological health, and the environment through traffic congestion, substantial loss of time, physical inactivity, and accidents. (Martin *et al.*, 2014; Royal Society for Public Health, 2016; Sallis *et al.*, 2004; EMBARQ, 2013) As more and more people lessen their reliance on full car ownership, (*reference*) the importance of developing and improving public transit systems becomes vital for supporting urban growth.

To better understand the scale of public transportation in Canada, prior to the COVID-19 pandemic, 31.4% of Canadians regularly used sustainable transportation, where public transit comprised almost 40% of those cases. When considering Canada’s largest three metropolitan areas, as much as 40.4% of the population used sustainable transportation, of which 55% was public transit. (Statistics Canada Census, 2016) To say the least, public transit will remain an imperative area for urban developers to focus on.

Secondly, we must recognize that certain population segments do not readily have access to private transportation. These populations tend to be more vulnerable whether they be marginalized, elderly, or youth, and are likely to have larger reliance on public transit relative to other social segments (need a refence). Therefore, when considering society in its entirety, modeling transit accessibility becomes paramount in urban planning.

Thus, to address the lack of standardized methods for obtaining transit accessibility information, this project responds with a first iteration methodology for simple, scalable, and high performing network travel time computation, city block accessibility scoring, and finally visualization. A case study of Vancouver was performed measuring access to cultural amenities such as museums, libraries, art galleries, and theatres, to serve as an initial proof of concept.

**2 BACKGROUND AND MOTIVATION**

**2.1 Literature Review**

**2.3 Research Questions**

**3 METHODOLOGY**

In this section, we detail how the research problem was approached computationally. We explain what data a scalable model would require, and how transit accessibility scores are computed and visualized from the data in an efficient manner.

**3.1 Data**

Several types of data were used for computing and visualizing transit accessibility measures. Data sources are summarized in their respective tables for computation and visualization.

|  |  |  |
| --- | --- | --- |
| *Data* | *Features* | *Source* |
| General Transit Feed Specification (GTFS) | All transit network data (stop coordinates, stop times, bus routes, etc.) | TransLink Open API1 |
| Dissemination Blocks (Origins) | Unique block ID, latitude, and longitude of the city block’s centroid | Census of Population2 |
| Amenities (Destinations) | Unique destination ID, latitude, and longitude | The Open Database of Cultural and Art Facilities (ODCAF)3 |
| OpenStreetMap (OSM) | Urban street network data | openstreetmap.org |

**Table 1**. Data required for many-to-many point travel time matrix computation. 1TransLink Open API (May 2021). 2Statistics Canada (2016). 3Statistics Canada (2020)

|  |  |  |
| --- | --- | --- |
| *Data* | *Features* | *Source* |
| Geospatial Shape file (.shp) | Dissemination block unique ID, longitude/latitude polygon data | Census Cartographic Boundary File1 |

**Table 2**. Data additionally required for accessibility measure visualization. 1Statistics Canada (2016).

**3.2 Computing Travel Time Matrices**

To evaluate transit accessibility across an urban landscape, many-to-many point travel times need to be computed into a travel time matrix. This requires a street network (Open Street Maps), a transit network (GTFS), origin-destination coordinates, and a routing engine.

Popular open-source routing engines include Open Trip Planner 1 (OTP1), Open Trip Planner 2 (OTP2), Conveyal’s R5 (Rapid Realistic Routing on Real-world and Reimagined networks), and GraphHopper. OTP1 and OTP2 are generally focused on passenger facing journey planning. OTP1 has analysis functionality but performs slow relative to other engines, using a generalized cost A\* algorithm. OTP2 is better optimized with a Multi-criteria range-RAPTOR algorithm but does not support one-to-many point routing analysis. Most importantly however, OTP optimizes routing on generalized cost instead of on minimizing travel time. For example, OTP may opt for a single long bus ride over one with a few transfers that yields a shorter travel time. In reality, transit users typically aim to optimize time.

R5 supports and is optimized for time-window trip planning which better reflects how people use the transportation system. R5, being implemented in Java, was also intended for analysis applications being magnitudes faster and less memory intensive than engines that are similar to OTP1, particularly for one-to-many point routing and travel time matrix generation. (OTP website) As such, travel times were computed using the open source r5r library, an R implementation of Conveyal’s R5 routing engine, with multimodal transit networks built from the GTFS and OpenStreetMap data detailed in 3.1. (Pereira, 2021) The final travel time matrix was an aggregation of 36 trips, departing every hour from 7:00am to 7:00pm with a 30-minute departure window on a weekday, a Saturday, and a Sunday (12 x 3). This allowed us to average travel times across changing bus schedules throughout the week.

**3.3 Measuring Transit Accessibility**

Transit accessibility to cultural amenities was measured in three fashions. The first was via isochrones, which indicates the shortest time from a given point to the nearest amenity. This was the most interpretable of all three measures.

The second measure involved scoring each city block using the mean transit time to the nearest amenity and the standard deviation of that mean time. The score was computed by taking the inverse of mean transit time plus two standard deviations of that time. This is essentially the worst-case scenario transit time scaled from 0 to 1, where the shortest transit times yield scores closer to 1, while the longest transit times yield scores closer to zero.

Since scores were not very interpretable, particularly because the scores were positively skewed, the percentile of the scores were computed as a third measure. This uniformly distributes the scores allowing for more interpretability than raw scores alone.

**3.4 Amenity Accessibility Weights**

***-Google Place API***

Given a predefined geolocation data, Google Place API (Application Programming Interface) returns information of the point of interest via HTTP requests over the internet. [ site 1] There are five main categories of requests available in the current version: Place Search, Place Details, Place Photos, Place Autocomplete and Query Autocomplete. The availability of request allows developers to build applications in a more flexible way based on the demands. In this project, Place Search and Place Details were used to generate the weights data. The name of each amenity was used as input of Place Search Request, which identifies the search targets. To obtain a more precise weights data for each amenity, geolocation points were used to reduce the location bias. The output of Google Search Request returns a list of unique place identifiers for target places, which can be used as the input of  Place Details Request; Place Details Request requires two parameters, *API\_ key* and *place \_id*, once the parameters are filled, it returns more detailed information for example, rating, number of reviews, operational hours and days  based on a specific place which is determined by *ipd*.  The table below shows the required and optional parameters for Place Search Request and Place Details Request.

|  |  |
| --- | --- |
| **Place Search Request** | **Place Details Request** |
| ***Required Parameters*** | |
| *Key* - Personal API keys  *Input* - Name the point of interest  *Input\_type* - Textquery | *Key* - Personal API keys  Place\_id - Place id |
| ***Optional Parameters*** | |
| *Location\_bias -* Point {lat,lng} | *Field-*{'name', 'formatted\_address', 'rating','user\_ratings\_total','opening\_hours/weekday\_text'} |

***-Data method***

*Vancouver\_facility* dataset is a subset of Open Database of Cultural and Art Facilities (ODCAF), filtered by art facilities in the Greater Vancouver area. The chart 1 shows the overall percentage of missing values grouped by features. Using names and geolocation points from *Vancouver\_facility* dataset, 94% of amenity can be found using Google Place API, in which 83% have total reviews and rating data, for business operational hours and days only 62% of data were available.

Chart, bar chart

Description automatically generated

The main focus was on 4 types of art facilities in the Metro Vancouver area, Gallery, Library, Museum and Theater. The table 2 shows the percentage of missing values grouped by type. Among 4 types of arts facilities, *Theatre* typehas the highest missing value percentage on Open days/hours , more than half of the information was not available using Google place API. With further investigation on reasons behind the missing values, it leads to 2 conclusions. First, Covid restriction, due to BC health regularisations, majority of theaters were closed or with reduced operational hours, thus some business hour information were simply labelled “closed” or missing in Google Place. Second, Open Days/Hour may not be static, some theaters are only open on special occasions or events. Thus, additional data may be needed.  However, if additional data are not available, it is not optimal but reasonable to replace NAs using the column mean for each type of amenity, based on the assumptions of that same type having similar opening hours and days.

Table 2. Missing Values in %

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Type | API Name | Open Days | Open Hours | Rating | Total Review |  |
| Gallery | 1.01 | 28.28 | 30.30 | 11.11 | 11.11 |  |
| Library | 1.16 | 25.58 | 25.58 | 19.76 | 19.76 |  |
| Museum | 4.35 | 29.35 | 29.35 | 6.52 | 6.52 |  |
| Theater | 4.00 | 61.33 | 62.67 | 8.00 | 8.00 |  |

Once NA values were filled, it is important to normalize data, so that the data are compatible across the table. The formula that was used for normalizing data has shown below.

Formula: XXXXX

**3.5 Efficiency Model**

The efficiency score was calculated by taking the difference between the Accessibility and the Needs of each block. The needs consist of three parameters, the traffic intensity, population and the amenity densities, all of which are normalized.

The traffic intensity is based on traffic surveys taken by the British Columbia Government between 1997 and 2015 \*\*\* <https://www.th.gov.bc.ca/trafficdata/legacy.html>

\*\*\*. These surveys measure the average number of vehicles that pass over it in 15 min intervals over an 8 min period. Since not every survey site is used each year, only the data from the most recent year was used in the analysis. Due to the sparsity of the surveys and the traffic intensity score was derived by taking the mean vehicle count within a 5 km radius of each block and normalizing the data.

The amenity density is normalized value of all amenities within a 1 km walking distance of the center of each block assuming an average walking speed of 3.6 km/h.

The Need was measured as the mean of the normalized traffic intensity, population, and amenity density for each block. The efficiency score was calculated by taking the percentile of both the Accessibility and the Need and normalizing the difference between the two.

**3.6 Visualization**

To visualize transit network accessibility, R’s Leaflet library was used to generate choropleth maps from city block shape file. Visualizations were deployed on an R Shiny dashboard, which was preferred over other due to its sleekness and ease of use. The dashboard was inspired from the Washington Post, ‘Washington: A world apart’ and comprised of four tabs allowing users to switch between different types of accessibility scores, points of interest, and how many nearest points of interest to consider (1, 2, 3, …, n).

Originally, the dashboard was rendering choropleth maps on the fly each time a new selection was selected. This however was inefficient and slow. To address this Leaflet choropleth maps were exported as html files before being called by the dashboard in a matter of seconds, significantly improving performance by over 10x. This however led to significant complications in the creation of the dashboard.

Due to the scores and properties of the census blocks being built into the html files, to switch between score sets the dashboard had to select a new html file by editing the file path. This also led to the removal of graphs summarize the data as no active interaction between the html files and the dashboard.

For each new parameter, a new html file had to be created. In total 32 html files were created for the score sets. To limit the number of parameters, only 4 parameters were used when filtering the score sets: ‘Amenity Type’, ‘Weight’, and the number of amenities ‘Nearest n Amenities’. Four htmls were required for the isochrones as they were based on the time to the nearest one amenity rather than a calculated score. The Kepler animation allowed for the inclusion of more parameters than the score sets as the html incorporates a filter requiring only a single html file. However, as the html file incorporates a multitude of maps, the size of the file would grow exponentially for each new parameter introduced. Due to a file size limit of 500 MB in Kepler and to improve performance of the dashboard the number of parameters was limited to the Amenity “Type”, “time”, and the “day” of the week.

The dashboard was originally published using Shinyapps.io cloud. Due to the file size limits the dashboard was published as open sourced on the r shiny server.

**4 RESULTS**

**5 DISCUSSION**

**5.1.6 Efficiency Model**

To determine how well transit is allocated throughout the Greater Vancouver Area, an efficiency score was created by comparing the needs of each block in comparison to their accessibility score. By comparing the accessibility and the needs of each block, blocks where the needs exceed the accessibility, a lack accessibility, and are segregated from blocks with a surplus in accessibility.

The needs of each block are based on three parameters, the population, traffic intensity and the amenity density. The population reflects the accessibility requirements of the block. Traffic intensity restricts a person access to a block, and the amenity density is number of destinations in proximity to each block.

The basic equation for the efficiency score is summarized as:

------------

The efficiency model can be summarized using by taking the absolute value of the efficiency score, providing a comparable difference between scores ----, or by taking its square highlighting the extremes of the data.

The method used to calculate the efficiency score consist of taking the percentile of the Accessibility and of the Needs and normalizing the difference between both percentiles. By taking the percentile of the Accessibility and Needs, both scoresets have the same distribution. This differentiates between efficiency scores with different Accessibility and Needs, placing efficiency scores with equal Accessibility and Needs in the center of the distribution. This method emphasises both extremes of the spectrum differentiating areas of high and low efficiency.

A divergent color scheme of red, white, and blue was used to further emphasize the extremes of the scoresets. To de-emphasize blocks with an eff. This color scheme consists of 5 separate colors, white, light red, dark red, light blue, and dark blue. Color represents the following percentiles.

|  |  |
| --- | --- |
| *Color* | *Efficiency Percentile* |
| Dark Blue | 30-70 |
| Light Blue |  |
| White |  |
| Light Red |  |
| Dark Red |  |

**5.2 Limitations/Assumptions**

**5.2.1 Optimal Travel Times**

The travels calculated from R5 were based on the optimal travel times provided in the GTFS dataset. These travel times do not include increase in travel times due to high traffic volumes which vary significantly throughout the day changing the overall accessibility. Seasonal variations in the number and preferred mode of transport also effect the travels times, reflecting the overall levels of congestion in a city. social variations observed in municipality. To optimize the travels times, it would require access to past bus trip information. Since most transit buses are outfitted with a gps and can be tracked in real time, access to these databases would allow for the calculation of the actual travel time for specific times of day, day of the week and seasons based on the actual times it took for the busses to complete their route.

**5.2.2 Sparse Traffic Data**

The traffic correction used for calculating the need of the efficiency score was calculated based on the normalized mean of all vehicle counts done at specific monitory sites across the city of Vancouver. This is arbitrary since the location and year that the monitoring was being done are not uniform. As such there are variations in the traffic counts. The counts were also

**5.2.3 Kepler Rendering Speed (still issue?)**

**5.2.4 Employment Location & Efficiency Model**

**6 CONCLUSIONS**

**ACKNOWLEDGMENT**

**REFERENCES**

EMBARQ. 2013. Saving Lives with Sustainable Transport. World Resources Institute

Liu, S., Zhu, X., 2004. An integrated GIS approach to accessibility analysis. Trans. GIS 8, 45–62.

Martin, A., Y. Goryakin and M. Suhrcke. 2014. “Does active commuting improve psychological wellbeing? Longitudinal evidence from eighteen waves of the British Household Panel Survey.” Preventative Medicine. Vol. 69, p. 296–303.

Pereira, Rafael H. M., Marcus Saraiva, Daniel Herszenhut, Carlos Kaue Vieira Braga, and Matthew Wigginton Conway. 2021. “r5r: Rapid Realistic Routing on Multimodal Transport Networks with R5 in R.” Findings, March. <https://doi.org/10.32866/001c.21262>.

Roberts, Brian. (2015). The Third Industrial Revolution: Implications for Planning Cities and Regions. Working Paper Urban Frontiers.

Royal Society for Public Health, Vision, Voice and Practice. 2016. Health in a Hurry: The Impact of Rush Hour Commuting on Our Health and Wellbeing

Sallis, J.F., Frank, L.D., Saelens, B.E., Kraft, M.K., 2004. Active transportation and physical activity: opportunities for collaboration on transportation and public health research. Transport Res. A – Pol. 38, 249–268.,

Statistics Canada. 2017. “Census of Population, 2016: Commuters using sustainable transportation in census metropolitan areas”. Minister of Industry. Catalogue no. 98-200-X2016029. ISBN 978-0-660-23763-3.

Washington Post. 2013. “Washington: A world apart.” <https://www.washingtonpost.com/sf/local/2013/11/09/washington-a-world-apart/>