Equitable Public Bus Network Optimization for Social Good: A Case Study of Singapore

David Tedjopurnomo
s3702199@student.rmit.edu.au
RMIT University
Melbourne, Victoria, Australia

Zhifeng Bao*
zhifeng.bao@rmit.edu.au
RMIT University
Melbourne, Victoria, Australia

Farhana Choudhury
farhana.choudhury@unimelb.edu.au
The University of Melbourne
Parkville, Victoria, Australia

Hui Luo
huiluosunshine@gmail.com
RMIT University
Melbourne, Victoria, Australia

A. K. Qin
kqin@swin.edu.au
Swinburne University of Technology
Hawthorn, Victoria, Australia

ABSTRACT

Public bus transport is a major backbone of many cities' socioeconomic activities. As such, the topic of public bus network optimization has received substantial attention in Geographic Information System (GIS) research. Unfortunately, most of the current literature are focused on improving only the efficiency of the bus network, neglecting the important equity factors. Optimizing only the efficiency of a bus network may cause these limited public transportation resources to be shifted away from areas with disadvantaged demographics, compounding the equity problem. In this work, we make the first attempt to explore the intricacies of the equitable public bus network optimization problem by performing a case study of Singapore's public bus network. We describe the challenges in designing an equitable public bus network, tackle the fundamental problem of formulating efficiency and equity metrics, perform exploratory experiments to assess each metric's real-life impact, and analyze the challenges of the equitable bus network optimization task. For our experiments, we have curated and combined Singapore's bus network data, road network data, census area boundaries data, and demographics data into a unified dataset which we released publicly. Our objective is not only to explore this important yet relatively unexplored problem, but also to inspire more discussion and research.

CCS CONCEPTS

• **Applied computing** → **Transportation**; • **General and reference** → *Metrics*; • **Social and professional topics** → **User characteristics**.

KEYWORDS

datasets, equity, metric, public transportation, bus network

*Corresponding authors

FAccT '22, June 21–24, 2022, Seoul, Republic of Korea

ACM ISBN 978-1-4503-9352-2/22/06…$15.00
https://doi.org/10.1145/3531146.3533092

ACM Reference Format:
David Tedjopurnomo, Zhifeng Bao, Farhana Choudhury, Hui Luo, and A. K. Qin. 2022. Equitable Public Bus Network Optimization for Social Good: A Case Study of Singapore. In *2022 ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, Seoul, Republic of Korea.* ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3531146.3533092

1 INTRODUCTION

In cities with high population density and limited land resources such as Singapore, it is imperative to make the best use of this resource while taking sustainable development into account [20]. It was reported that public transportation trips occupied 63% of the peak hour trips in Singapore in 2012 [2]. The quality of a public transit route network can be evaluated in terms of a number of network parameters, such as route directness, service coverage, network efficiency, and the number of transfers required [33]. Research in bus network optimization can be divided into three tracks: route generation, in which new routes are created and utilized on top of existing bus routes or a new bus network is created from scratch [7][33][32][25][26][27]; frequency assignment, in which the bus frequency of existing bus routes are modified [31][14][7]; or a combination of both [8][16][10].

Unfortunately, due to the focus on improving bus network quality, bus network equity has largely been ignored despite its importance, as inequities in public transportation resource allocation reinforce social exclusion and limit a person's access to jobs and opportunities [23][21]. Here, equity is defined in terms of social attributes such as income and age. An equitable bus network is a network where different population demographics have access to similar levels of public transportation availability and quality. Inequities in bus networks are still common in many cities around the world. In Perth, Australia, 70% of the population share only 30% of the public transportation services, where the elderly suffer the most from inequitable public bus network [22]; whereas in Melbourne, 70% of the population share only 19%[11]. Cuthill et al. [9] analyzed the Public Transport Accessibility Level (PTAL) data from London. They found that younger demographics have better public transport accessibility while people with no educational qualifications have worse accessibility. While research that explores inequities in bus network exist, there is no current research that provides insights into the potential solution. Addressing public transport inequity, in tandem with improving or at least maintaining efficiency, is a difficult task due to several challenges.

Challenge 1. Defining Efficiency and Equity Metrics. The quality of a bus network consists of route directness, service coverage, network efficiency, and the number of transfers required. Designing a metric of bus network efficiency requires aggregating all four factors. Designing an equity metric is also a difficult task because finding the correct level of granularity, between individual-level and group-level, is hard. Additionally, an area that is considered disadvantaged in one social factor such as household income may not be disadvantaged in other social factors.

Challenge 2. Route Equity-Directness Tradeoff. Bus route equity and route directness are antitheses of each other as improving the equity of a bus route requires covering more bus stops in underrepresented areas while improving route directness requires covering fewer. Covering a larger number of bus stops may improve equity by providing more bus services to bus stops in disadvantaged areas, but at a cost of lower route directness and therefore lower efficiency (i.e., quantity over quality). On the other hand, covering a fewer number of bus stops results in a smaller coverage, but better route directness and therefore efficiency for that smaller set of disadvantaged bus stops (i.e. quality over quantity). Balancing between the two aspects is difficult because it is highly dependent on the network itself.

Challenge 3. Resource-Constrained Optimization. The large cost of operating buses and the limited number of buses complicate the bus network optimization problem by introducing complex trade-offs in both route planning and bus frequency assignment. To improve service efficiency and availability, additional bus routes may be created in underrepresented areas or existing bus routes in these areas may need to be given additional buses to reduce waiting time. However, this will divert resources away from other areas which might be underrepresented themselves.

In this paper, we perform the first work on exploring the intricacies and challenges of equitable bus network optimization. We explore this problem using the case study of Singapore. We curate our comprehensive Singapore dataset [1] and define our efficiency and equity metrics. We then perform exploratory experiments using three bus network optimization heuristics which act as investigation mechanisms to understand the problem better. Our main objective is to address Challenge 1 through our efficiency and equity metrics design, but we also explore Challenges 2 and 3, which are more general bus network optimization challenges. Our contributions are as follows:

- We curate, describe, and publicly release a comprehensive Singapore bus network dataset [1]. Due to the lack of literature, such dataset was not publicly available to the best of our knowledge. This combines bus network, bus schedule, road network, travel distance via bus network, travel distance via cars, and census demographics data of several demographics factors (more details in Section 3). Furthermore, we analyze the distribution of Singapore's advantaged and disadvantaged areas, and their bus network efficiency.
- We perform an analysis on the equity factors of bus networks. We formulate several choices of equity metrics to assess demographic-based equity and area-based equity, and show the strengths of each metric.
- We study and discuss the efficiency factors of bus networks. We formulate a novel efficiency metric that captures crucial network parameters of route directness, service coverage, network efficiency and the number of transfers, and combine them into one unified and intuitive efficiency score.
- We perform exploratory experiments using several heuristic models to show the intricacies of bus network optimization. Our experiments compare the strengths and weaknesses of the different heuristics, show how they affect the bus network's overall efficiency and equity, and how they relate to the challenges we mentioned.

2 RELATED WORK

In this section, we discuss several existing efficiency optimization work, which can be divided into three main categories: bus frequency optimization, which modifies existing bus network's timetable or headway; route generation, which generates new routes or creating a new bus network from scratch; or a combination of both.

Only bus frequency optimization. Yang et al. [31] perform bus network optimization by modifying the headway to minimize passenger and operator cost. The passenger cost consists of the waiting time cost, riding time cost, and alighting time cost. The operator cost consists of the fixed operational cost such as bus maintenance, and variable costs such as fuel consumption. The authors use a parallel genetic algorithm model. Fonseca et al. [14] minimize transfer costs and operational costs of a bus network by modifying bus timetables. They formulate the problem as a mixed integer programming problem and use a matheuristic approach. Mo et al. [18] perform bus schedule optimization to maximize the number of passengers that can be served within the waiting time threshold. They propose two variants of the problem, one without the constraint of limited vehicles and one with this constraint, and formulate a partition-based greedy method and a progressive partition-based greedy method to approximate these NP-hard problems. Finally, Banerjee and Smilowitz [5] addressed the equitable school bus scheduling problem (SBSP) that considers both school bell times and route schedules in order to minimze the number of buses required. They extend a time-indexed integer programming model with a minimax model in order to equitably reduce the disutilities caused by changing school start times.

Only route generation. Bowerman et al. [7] propose a school bus routing problem, which aims to find a set of bus routes that ensure students located in different areas have access to the bus service from their residence to school, as a multi-objective mathematical model. Zhao and Ubaka [33] perform route generation that seeks to minimize the number of transfers and optimize route directness while maximizing the service coverage. They propose a greedy search method and a fast hill-climb search method. Yang et al. [32] present a parallel ant colony algorithm to maximize the direct traveler density based on the demand for an entire bus network. An empty network is built initially and then routes are added to satisfy the goal until all users are loaded to the network. Szeto and Wu [25] study a bus network design problem in a suburban residential area of Hong Kong to improve the existing bus services by reducing the number of transfers and the total travel time of the users. They use a hybrid genetic algorithm to solve this problem. Wang and Qu [26] study a bus route design problem on a suburb in Gold Coast, Australia that has a low population density. This work is limited to finding one bus route with the minimal total length. The problem is proved as NP-hard and a dynamic programming approach is developed. Wang et al. [27] perform a multi-objective bus route optimization, balancing between fulfilling commuting demand and improving the bus network connectivity. For the commuting demand, the authors use real taxi datasets in order to measure the travel demand for each road in the network. For the bus network connectivity, the authors use graph natural connectivity. They propose an expansion-based greedy algorithm to address the problem.

Both route generation and frequency optimization. Chu [8] perform the simultaneous task of bus route generation and frequency optimization by using a mixed-integer programming model and branch-and-price-and-cut algorithm. However, this work uses two very small bus networks; the Mandl network [17], which is a minuscule network consisting of only 15 nodes and 21 edges, and a slightly larger network consisting of 26 nodes and 84 edges. Liang et al. [16] perform both route generation and frequency assignment in order to design a bus transit network around the existing metro network to balance the ridership between metro and bus. They test their method on the Beijing second ring public transit network, which consists of 106 nodes and 1,706 edges. Darwish et al. [10] apply deep reinforcement learning to optimize the bus network for the benefit of the riders and the transport operators. They aim to balance customer satisfaction and minimizing the capital and operating expenditure. However, their work also uses the small Mandl network. Finally, Bertsimas et al. [6] performed school bus routing optimization by considering bus stop assignment, bus routing and bus routes scheduling. Using an algorithm called Biobjective Routing Decomposition (BiRD), the authors are able to outperform existing state-of-the-art routing methods while maintaining equitable school bell times assignment.

Other equity research. In geospatial-related research, equity has been considered in topics such as ride-sharing [28][19][29]. Equity aspects have also been considered in a wide-range of non-geospatial fields such as equitable machine learning [3] [13] [15] and language models [4] [12].

As we can see, none of the present work consider the equity factors with regards to sensitive demographic attributes such as age and income. In addition, their efficiency metric does not intuitively capture the notion of efficiency that translates to real application. Works that define monetary cost for passengers based on waiting and traveling time, and public bus operators based on fuel cost and driver salaries are specific for a certain location because factors such as fuel cost and salaries vary depending on the country, or even city. Furthermore, defining the waiting cost and traveling cost for riders is difficult and not intuitive; with most of the aforementioned literatures using predefined values that are not clearly explained. Finally, as we will explain further in Section 4.1, works that use the bus transfer cost do not fully capture the frequency and severity of the transfer. Our research seeks to fill this gap of equitable bus network optimization research by designing several equity metrics for different applications and also designing an intuitive and comprehensive efficiency metric.

3 DATASET

We aim to formulate an efficiency metric that combines route directness, service coverage, network efficiency and the number of transfers, and several equity metrics that show the bus service level discrepancy between different areas. To fulfill these goals, we first aggregate several publicly available Singapore datasets. To capture the route directness factor, we use the bus route data and the road network data to find the travel distances between bus stops when using public bus transport and cars respectively, and then compare the two distances. To capture the service coverage factor, we use the bus stops coordinates data and mapmatch it to the road network. For the network efficiency factor, we utilize the bus schedules data to find the average waiting time for each route and use that in our calculation. For the number of transfers factor, we process the bus routes data to find the number of transfers required for every pair of bus stops. For the equity metric, we use the Singapore planning area census boundaries data and combine it with the census data to obtain the demographic profiles for each area. We aggregate the bus stops and bus routes data [1], road network data [2], census boundaries data [3], and the census demographics data [4] into a unified dataset and release it publicly [1].

3.1 Dataset Preprocessing

Bus network. The Singapore road network data is retrieved from OpenStreetMap. This road network is a directed graph where the nodes are road junctions and the edges are the road segments connecting them. Since the bus stops and the road network data are not from the same source, we first map-match the bus stops' coordinates to the road network. Next, to easily retrieve the road distance between bus stops, we treat the junctions and bus stops as vertices in the road network and the roads connecting them as edges. To do that, we split roads containing bus stops by treating each bus stop as a vertex, and connect that vertex to adjacent vertices to create edges. With this transformation, we combine the Singapore road network data and the bus stops data into one road graph where the road junctions and bus stops are the vertices, and the roads between them are weighted edges; the weight denotes the edge length in meters.

Bus routes. We perform a simplification of the bus routes by assuming that all of them run in both directions, where both directions cover the exact same route but in reverse. For bus routes that naturally contain more than one direction, we only keep the first direction. In addition, we only keep regular bus routes. Specifically, we remove weekend-only routes, and special routes such as midnight routes and express routes that complement existing full bus routes. Bus routes in Singapore often have different headway for different periods in a day. To further simplify the problem, for each bus route, we take the average of the headway for all periods and

[1] https://www.kaggle.com/gowthamvarma/singapore-bus-data-land-transport-authority and https://www.transitlink.com.sg/

[2] retrieved using the Python osmnx library

[3] https://data.gov.sg/dataset/planning-area-census2010

[4] https://www.tablebuilder.singstat.gov.sg/

assign that as the one waiting time for that route. For clarification, headway is defined as the duration between subsequent buses in a bus route.

Distance matrices. Using the created road network graph, we create two distance matrices for car (or road) and bus distance, respectively. For the car distance matrix, we simply calculate the shortest road distance between every pair of bus stops using the A* algorithm. For the bus distance matrix, we use the sequence of bus transfers that will result in the shortest distance between every pair of bus stops. Rather than using the road distance to get the distance between bus stops, we use the bus travel distance data from the transitlink website instead, since these bus routes might not use the shortest road distance to connect the bus stops. In addition, we also incorporate the waiting time and bus transfer penalty to the bus travel distance. The calculation of the overall bus travel distance is available in Section 4.1.

Census boundaries. The division of Singapore into smaller areas are done on three levels: 5 regions, 55 planning areas, and more than 300 subzones. Since the region level granularity is too coarse and the subzone is too fine, we use the planning area level. Out of the 55 planning areas, 5 do not contain bus stops, so we remove them. Several of these areas are also missing the demographic information, in which case they are not used for the equity calculations, but are still used for the efficiency calculations. For each census area, we assign it a list of bus stops belonging to that area.

Demographics data. Our experiments use three sensitive attributes: age, qualification, and income. For each attribute, we divide them into the advantaged and disadvantaged group. For the age attribute, we assign the demographics of 65 years old and above into the disadvantaged group, same as [11]. For the qualification attribute, the Singapore dataset contains eight groups of maximum qualification attained: NoQualification, Primary, LowerSecondary, Secondary, Post-Secondary(Non-Tertiary), Polytechnic, University, and ProfessionalQualificationAndOtherDiploma. Similar to [9], we assign the NoQualification group into the disadvantaged group, but we also add the Primary group as well. The remaining are assigned into the advantaged group. Finally, for the income attribute, we assign demographics with income of less than $2,000 per month into the disadvantaged group and vice versa.

3.2 Dataset Statistics

After the aforementioned preprocessing and simplification steps, the Singapore bus network consists of 5,021 bus stops. Out of those bus stops, 81 bus stops are considered "terminal" bus stops. Terminal bus stops are bus stops that are used as the start or ending bus stop of at least one bus route. They are typically bus terminals, which contain a large bus parking lot. The Singapore bus network has 361 bus routes. The minimum, average, and maximum number of bus stops visited are 2, 40, and 105 respectively. The minimum, average, and maximum route length are 1.1, 19, and 73.5 kilometers respectively. Out of the 50 remaining planning areas, 14 areas are missing the age information while 27 are missing the income and qualification information.

4 EVALUATION METRIC

In this work, we seek to assess a bus network's performance based on both the efficiency and equity metric.

4.1 Efficiency Metric

In our work, we define efficiency by comparing bus travel distance with the car (or road) travel distance. This is because the car travel distance is the shortest possible distance between two bus stops in the road network. We also take into account the convenience of the travel into the efficiency calculation. To quantify travel convenience, we use bus transfers because riders prefer to transfer as few times as possible [24]. Bus transfer has been used as a metric of travel efficiency before, with two different types of implementations used: count-based [33][10], which counts the number of transfers required and assigns a penalty value accordingly, and unit-based [16][8], which defines a unit such as dollar cost that penalizes the severity of the transfers. Each of these approaches has its own weakness. The count-based approach treats every transfer of the same count equally, but does not take into account the different severity (e.g. waiting time) for each transfer. The unit-based approach addresses this problem, but it does not impose penalties for additional transfers. In addition, all of these previous approaches impose a limit on the transfers; usually allowing a maximum of two transfers per trip. This approach is too restrictive, as there are pairs of bus stops that cannot be connected within two transfers. Our approach takes into account both the number of transfers and the severity of the transfer, while also not imposing an upper limit on the number of transfers. We will provide some important formal definitions below.

Definition 4.1 (Stop-to-stop Bus Travel Distance). Here, we define two concepts: bus route and bus path. A bus route refers to a sequence of bus stops that a bus visits sequentially in one trip. This can either be government-issued routes or new routes generated by a model. A bus path is a sequence of different bus routes that connect a pair of bus stops, where passengers must transfer between these bus routes. For a pair of bus stops s_x and s_y, stop-to-stop bus travel distance is defined as the shortest bus path between the two stops when strictly following the existing bus routes.

Definition 4.2 (Stop-to-stop Car Travel Distance). The stop-to-stop car travel distance is the length of the shortest path between two bus stops when following the road network. The shortest path is retrieved using algorithms such as A* and the route length is based on summing the edge length of that path.

Definition 4.3 (Transfer Penalty). Given a bus path consisting of N bus routes (connected by $N-1$ transfers), the calculation of the transfer penalty is given below:

$$p = \sum_{n}^{N} w\_dist_n * 2^{n-1} \tag{1}$$

$$w\_dist_n = \frac{w\_time_n}{60} \times 15 \tag{2}$$

For every bus route n in the bus path, we first transform the waiting time w_time_n into waiting distance w_dist_n. Given a waiting time, waiting distance refers to the number of kilometers a bus will travel during the duration of that waiting time using the standard 15 km/h speed based on transitlink [5]. We do this so that we can directly combine the waiting time (in minutes) and travel distance (in kilometers) into a single value. We can then calculate the total transfer penalty p as the sum of each of the N bus routes' waiting distance multiplied by a factor. This factor increases after every subsequent transfer; here we double the factor after every transfer. This encourages fewer transfers while also shortening passengers' waiting; both affecting their convenience. Finally, to get the final travel distance between a pair of bus stops, we sum the actual travel distance and this transfer penalty for every bus route in the path.

Definition 4.4 (Stop-to-Area Efficiency). The stop-to-area efficiency is calculated based on the comparison between the stop-to-area bus and car travel distance. The stop-to-area bus travel distance is calculated by averaging the stop-to-stop bus distance between the source stop and all bus stops in the destination area, where we add the transfer penalty to every distance. Similarly, the stop-to-area car travel distance averages all stop-to-stop car travel distance.

Definition 4.5 (Area-to-Area Efficiency). The area-to-area efficiency captures the bus network's performance using a coarser-grained scale of census areas instead of bus stops by averaging the stop-to-area efficiency for all bus stops in the source area. Then, by averaging the area-to-area efficiency for a source area to all of its destination areas, we can get the overall efficiency of that source area, denoted as $\overline{\mathbb{C}_I}$ for area I. Furthermore, averaging the efficiency of all source areas will result in the overall bus network efficiency.

As mentioned in Section 1, the quality of a public transit route network can be evaluated through route directness, service coverage, network efficiency, and the number of transfers required. Our efficiency metric covers all aspects:

- **Route directness.** We assign larger efficiency scores to trips that are more direct. To do this, we compared bus travel distance with the car travel distance; the latter of which is the gold-standard of route directness.
- **Service coverage.** By aggregating the efficiency metric to area-level, we capture service coverage in two ways. Firstly, the area-level metric averages the coverage from all bus stops in that area, ensuring the efficiency score of an area is not dominated by a single well-connected bus stop and rewarding areas where most of the bus stops are well-connected. Secondly, the connectivity of an area is based on its average connectivity to all other areas; no matter whether or not that area is urban or rural, or advantaged or disadvantaged.
- **Network efficiency.** We incorporate bus network efficiency through the aforementioned route directness calculation. In addition, we also incorporate the waiting time into account; meaning that routes with shorter waiting time will have a greater efficiency score.
- **Number of transfers required.** By considering the transfer penalty, our efficiency metric takes into account the number of transfers required to connect any pair of bus stops and assigns the penalty accordingly. We also take into account the severity of each transfers.

[5] https://www.transitlink.com.sg/

4.2 Equity Metrics

The concept of equity in bus network optimization with respect to sensitive attributes has not been formally defined before. Thus, in this section, we make use of the efficiency metric to formulate several equity metrics with different use cases. Before that, we formally define the advantaged and disadvantaged areas.

Definition 4.6 (Advantaged and Disadvantaged Areas). We assign every census area into either the set of advantaged areas $\mathbb{I}^+$ or the set of disadvantaged areas $\mathbb{I}^-$. The division is dependent on the sensitive attributes outlined in Section 3.1. We use the overall (i.e countrywide) percentage of the disadvantaged population. If the percentage of an area's disadvantaged population is over the overall percentage, then that area is counted as disadvantaged.

Definition 4.7 (Population-scaled Equity (PEQ)). Here, we take into account the different demographic profiles for each area. We use the population density. Each advantaged (disadvantaged) area $I^+ \in \mathbb{I}^+$ ($I^- \in \mathbb{I}^-$) has their population size I^+_p (I^-_p) and there is also the total population size for all advantaged (disadvantaged) areas $\mathbb{I}^+_p$ ($\mathbb{I}^-_p$). With these values defined, below is the calculation for the overall efficiency score:

$$PEQ = 1 - \left| \left(\sum_{I^+ \in \mathbb{I}^+} \frac{I^+_p}{\mathbb{I}^+_p} \cdot \overline{\mathbb{C}_{I^+}} \right) - \left(\sum_{I^- \in \mathbb{I}^-} \frac{I^-_p}{\mathbb{I}^-_p} \cdot \overline{\mathbb{C}_{I^-}} \right) \right| \tag{3}$$

For each advantaged or disadvantaged area, we take its overall efficiency, and weigh it based on the fraction of the advantaged or disadvantaged population; giving areas with bigger population greater influence. The left (right) hand side of the equation calculates the overall efficiency of the advantaged (disadvantaged) areas. Thus, this equation measures equity based on the discrepancy of the bus efficiency. *PEQ* is in the scale of (0,1]; the higher the value, the better. Since this metric scales every advantaged-disadvantaged areas by the population size, it treats every person equally. However, at the same time, it marginalizes areas with relatively fewer people.

Definition 4.8 (Area-level Equity (AEQ)). The area-level equity calculates the average connectivity of the advantaged and disadvantaged areas, and then calculates the discrepancy between the two. This is calculated as below:

$$AEQ = 1 - \left| \left(\sum_{I^+ \in \mathbb{I}^+} \frac{\overline{\mathbb{C}_{I^+}}}{|\mathbb{I}^+|} \right) - \left(\sum_{I^- \in \mathbb{I}^-} \frac{\overline{\mathbb{C}_{I^+}}}{|\mathbb{I}^-|} \right) \right| \tag{4}$$

This metric is the opposite of the population-scaled equity above, as this metric treats every area equally, but does not treat every person equally.

Definition 4.9 (Maximum Difference (MD)). The previous two metrics calculate equity based on the discrepancy between advantaged and disadvantaged areas. The weakness of these metrics is that it does not capture the inequity where an area with very poor efficiency and an area with very good efficiency are in the same group (either advantaged or disadvantaged). Thus, we define the maximum difference metric as below:

$$MD = 1 - \left(\max_{\forall I \in \mathbb{I}} \overline{\mathbb{C}_I} - \min_{\forall J \in \mathbb{I}} \overline{\mathbb{C}_J} \right) \tag{5}$$

This metric finds the area with the best and worst mean efficiency, then calculates the difference. We subtract 1 with this value to get a metric where a larger value is better; making it align with the previous metrics. This metric ignores the concept of advantaged and disadvantaged areas and simply compares the best and worst performing areas.

Definition 4.10 (Standard Deviation (SD)). Similar to the maximum difference metric above, this metric also ignores the concept of advantaged and disadvantaged areas. However, the maximum difference metric only takes into account the efficiency difference of two areas. For this metric, we use the standard deviation of all areas:

$$SD = 1 - \sqrt{\frac{\sum_{I \in \mathbb{I}} (\overline{\mathbb{C}_I} - EF)}{|\mathbb{I}|}} \tag{6}$$

Here, EF is the mean bus efficiency of all areas. Similar to the maximum difference metric, we subtract 1 with the standard deviation to get a metric where a larger value is better.

Definition 4.11 (Residual Difference (RD)). Residual difference has been used in equity-related applications in [30]. The authors use this metric in regression tasks. Residual difference in regression tasks measures the extent of overestimation or underestimation discrepancy for advantaged or disadvantaged groups. A positive (negative) value means that the prediction for the advantaged group is generally overestimated (underestimated) compared to the disadvantaged group. A value close to 0 means that both demographics receive the same level of over or underestimation.

In the context of bus network optimization, residual difference can be used to compare two bus networks; a generated bus network, and the existing bus network which we will refer to as the **ground truth** network. We use this metric to evaluate how equitable the bus optimization algorithms are in terms of how they improve or degrade the efficiency of advantaged or disadvantaged areas over the ground truth. The algorithm is defined as follows:

$$RD = 1 - \left| \left(\sum_{I^+ \in \mathbb{I}^+} \frac{\overline{\hat{\mathbb{C}}_{I^+}} - \overline{\mathbb{C}_{I^+}}}{|\mathbb{I}^+|} \right) - \left(\sum_{I^- \in \mathbb{I}^-} \frac{\overline{\hat{\mathbb{C}}_{I^-}} - \overline{\mathbb{C}_{I^-}}}{|\mathbb{I}^-|} \right) \right| \tag{7}$$

Where $\overline{\hat{\mathbb{C}}_{I^+}}$ $\left(\overline{\hat{\mathbb{C}}_{I^-}}\right)$ denotes the overall efficiency of an advantaged (disadvantaged) area in the generated bus network, while $\overline{\mathbb{C}_{I^+}}$ $\left(\overline{\mathbb{C}_{I^-}}\right)$ denotes the overall efficiency of an advantaged (disadvantaged) area in the ground truth network. We also subtract 1 with this value to get a metric where a larger value is better. A value of 1 means that the efficiency of advantaged and disadvantaged areas are equally improved or degraded, i.e. they are treated equally by the algorithm.

In summary, here are our five equity metrics as below:

- Population-scaled Equity (PEQ) compares the bus service discrepancy between advantaged and disadvantaged population.
- Area-level Equity (AEQ) compares the bus service discrepancy between advantaged and disadvantaged areas.
- Maximum Difference (MD) compares the difference between the best and worst-served area.
- Standard Deviation (SD) compares the overall service level of all areas.
- Residual Difference (RD) compares the way bus network optimization algorithms treat advantaged and disadvantaged areas.

5 DATASET ANALYSIS

We visualize the efficiency distribution of Singapore's areas and whether or not they are advantaged or disadvantaged based on our curated dataset [1]. There are three sensitive attributes: age, income, and qualification. However, due to space constraints, we only provide the equity figure for the age attribute.

The advantaged/disadvantaged area distribution is shown in Figure 1. Singapore contains 55 planning areas. However, not all census data are available for all areas. In particular, 14 areas are missing the age data, and 27 areas are missing the qualification and income data. The missing areas are usually non-residential; the southwestern part of Singapore, such as Tuas and Pioneer, are industrial areas. The five patches of missing areas in the center (from top to bottom, left to right) are: Lim Chu Kang, which is an agricultural and military area; Tengah, which is a planned residential area; Central Water Catchment, which consists of Singapore's water reservoirs; Paya Lebar, which is mostly occupied by an air base; and Marina South and Straits View, which are commercial and tourism areas. Since these areas do not have a lot of residences, census data for these areas are not available. The area efficiency distribution is visualized in Figure 2. The easternmost area of Singapore is Changi, which contains Singapore's main airport. As such, it has many bus stops connecting it to different parts of the country. Some areas west of Changi are also high-efficiency areas because of their proximity to the airport. On the contrary, the southwestern part of Singapore consists of the industrial areas Tuas and Pioneer, which tend to not be as well-connected, resulting in lower efficiency scores.

In Figure 3, we visualize the five most and least efficient areas in Singapore. Two of the most efficient areas, Changi and Bedok have efficiency values of 0.5727 and 0.5212. In Figure 3, Changi is the easternmost area and Bedok is its western neighbor. Their bus routes' coverage are visualized in Figures 5 and 6 respectively. These two areas are highly efficient due to their large countrywide coverage. These two areas are crucial; Changi houses Singapore's main airport while Bedok is the most populous planning area in Singapore, with 289,740 residents according to the census data. Conversely, Singapore's two least efficient areas are Straits View and Marina South, with efficiency values of 0.3935 and 0.4039 respectively. In Figure 3, they are the two small areas near the south-central. Since they only have one bus route, route 400 that only provides direct connection between the two areas, we do not visualize the bus route coverage. Due to this low coverage, reaching any other areas in Singapore require bus transfers, resulting in low efficiency. Here we also visualize the third-worst efficient area, Tuas (efficiency score: 0.4218), in Figure 4. We can see that the bus coverage is confined to a small portion of Singapore's eastern area. Similar to Straits View and Marina South, connections to other parts of Singapore from Tuas require bus transfers. The three least efficient areas we

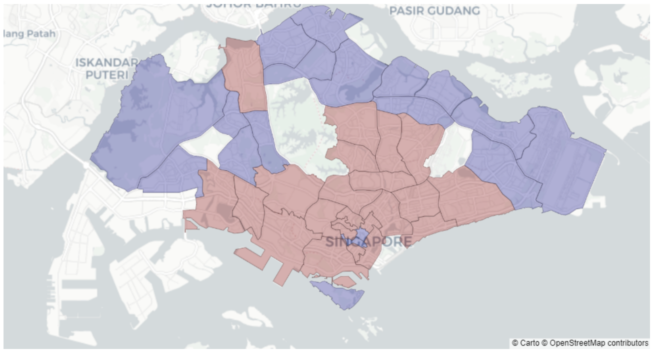

Figure 1: Choropleth of the areas for the 'age' attribute. Red = disadvantaged. Blue = advantaged.

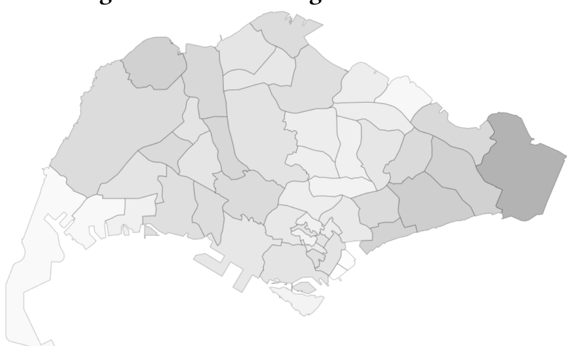

Figure 2: Choropleth of the areas' efficiency. A darker shade denotes better efficiency.

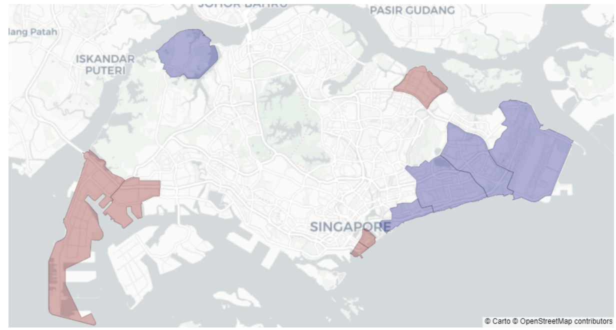

Figure 3: The top-5 (blue) and bottom-5 (red) areas in terms of efficiency.

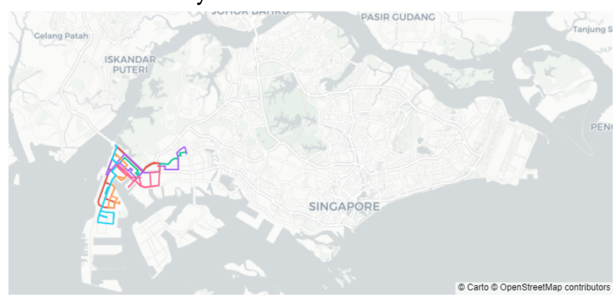

Figure 4: Coverage of the 7 bus routes of the low-efficiency Tuas area

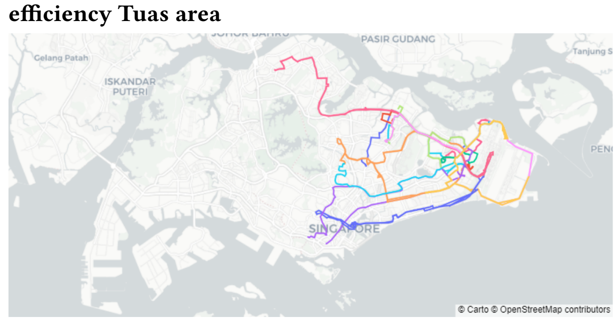

Figure 5: Coverage of the 20 routes of the high-efficiency Changi area

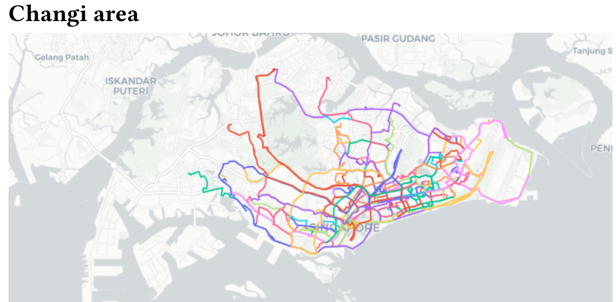

Figure 6: Coverage of the 73 routes of the high-efficiency Bedok area

discussed are not residential areas. Thus, they receive low bus transport resources. Furthermore, Straits View and Marina South are not directly connected to any bus interchanges while Tuas is connected to only a few. More bus routes that connect these areas to nearby large bus interchanges will help greatly in improving efficiency without creating long routes.

6 EXPERIMENTS

In this section, we describe our experimental setup and the analyses of our results. Our experiments are intended to be exploratory. We use heuristic methods to perform bus network optimization, analyze their performance in terms of efficiency and equity, and explore the methods further through case studies.

6.1 Experiment Setup

6.1.1 Dataset Preparation. We use several simple models, which we will describe in detail in Section 6.1.3, to perform route generation and compare the equity and efficiency with the original bus network, which we refer to as the **ground truth** network. The purpose of this experiment is to show and explain the difficulty of the equitable bus network design, and provide insights on the different methods' strengths and weaknesses for this task and how they can be improved. In order to compare our methods against the ground truth network, we perform a reduction on the bus network. We randomly remove 25% of the bus routes, i.e., 90 out of the 361 Singapore bus routes. We call this reduced network as the **initial** network. We then feed this network to our models in order to generate 90 replacement routes. We then compare the efficiency and equity of these new networks with the ground truth, and analyse the different methods' strengths and weaknesses. This novel experiment seeks to compare human and machine-generated routes. One can view the initial network as a real bus network consisting of 271 routes, and the ground truth network as an updated network with 90 new human-generated routes. Our models will also generate a new set of 90 routes, which can then be compared with the ground truth. Finally, we obtain a set of terminal stops, a concept which we defined in Section 3.2, for later use.

6.1.2 Restrictions. In order to ensure that the created bus networks are realistic, we impose several restrictions.

Restriction on routes. All created bus routes must be under 75 kilometers, slightly longer than the longest Singapore bus route of 73.5 kilometer. All routes are assumed to be symmetrical, where for each route it is implied that the exact same route in the opposite direction with the same bus frequency exists. In addition, all bus routes operate on the same schedule starting from 6AM to 12AM and must start and end at a terminal stop. Finally, since the majority of Singapore's routes are non-looping routes, we also only generate non-looping routes.

Bus resource restriction. In order to produce realistic routes and waiting times, we impose a restriction on the number of buses needed to serve the bus routes. Specifically, the minimum number of buses needed to serve the new bus network must not exceed the ground truth's minimum number of buses.

Network cost restriction. We also impose an operational cost limit for the solution. The operational cost captures the cost of running the bus fleet to serve the whole network, e.g. fuel costs. For better disambiguation between this and the fleet cost above, we use the term network cost instead. The network cost is formulated as follows:

$$G = \sum_{n \in N} \left\lceil \frac{1080}{w\_time_n} \right\rceil * 2 * n.len \tag{8}$$

We find the number of daily trips for a route, multiply it by 2 to take into account both directions, and multiply that by the length of the route to get the total kilometers traveled for one route and then sum for all routes. The number of daily trips for a route is calculated by dividing the number of minutes in a daily operation–1080 minutes or 18 hours from 6AM to 12AM–by the waiting time for the route. The network cost captures the idea of variable cost as used in the work of Yang et al. [31], but we did not convert our cost into monetary value; we directly use the travel distance itself. Both this restriction and the bus resource restriction are related to Challenge 3.

6.1.3 Baseline Models. We compare the ground truth network with networks created from three baseline models, which we will explain below. Due to space limitations, we do not provide detailed descriptions of our model.

Direct. The direct method is predicated on the idea that by directly connecting two bus stops using the shortest road distance, we can maximize the connectivity between them. Thus, for each route, we find the current worst-connected pair of bus stops (i.e., with the worst stop-to-stop efficiency) and directly connect them. Afterwards, for both bus stops, we find their nearest terminal stop and form a direct connection, creating a bus route with four main bus stops as its skeleton. Any bus stops that the route passes through between these four stops will also be added to the bus route.

Gradual. Unlike the direct method that directly connects two bus stops, this model builds the bus route by gradually adding bus stops closer and closer to the destination bus stop. It takes a pair of terminal stops and from the starting stop, it adds the closest bus stop in the direction of the destination until the destination bus stop itself is reached.

Gradual-Equity. The gradual method above is concerned only with reaching the destination. This modification of the gradual method takes into account the equity score, allowing it to take detours and add bus stops not in the direct path to the destination if the estimated equity improvement justifies adding that bus stop.

6.2 Experiment Results

We report the efficiency and equity scores of all models in Table 1. We use the five equity scores we described in Section 4.2: Population-scaled Equity (PEQ), Area-level Equity (AEQ), Maximum Difference (MD), Standard Deviation (SD), and Residual Difference (RD). Due to space constraints, we only display the values for the "age" attribute. In our experiments, we find that if we disallow our methods' network cost to exceed the ground truth, our methods are likely to assign far lower bus frequencies. Thus, we allow our methods to exceed the ground truth network cost by less than 1%. We show the analyses of our results below and discuss some case studies.

6.2.1 Numerical Analysis: Efficiency. The baseline models' efficiency ranked from worst to best are Gradual-Equity, Gradual, and Direct. From our observation, we find that the efficiency score is negatively correlated with the number of bus stops in the route. The Gradual-Equity, Gradual, and Direct methods produce bus routes with 44, 36, and 20 bus stops on average respectively. This shows that routes that are more direct to their destination tend to have better efficiency despite the fewer number of bus stops to improve. Efficiency improvements for a pair of bus stops only happen when a new route provides a shorter travel than existing routes. The larger bus stop coverage for the Gradual-Equity and Gradual method is inconsequential because of the most part, existing routes are more efficient. While the number of bus stops has a correlation with the efficiency, there is no correlation between the route length in kilometers with the efficiency. We find that the Gradual-Equity, Gradual, and Direct methods route lengths are 33, 23.7, and 27 kilometers respectively. This experiment relates to Challenge 2 we mentioned in Section 1.

6.2.2 Numerical Analysis: Equity. We show the efficiency results for the "Age" attribute only in Table 1 due to space constraints, although we observe similar results for all the other attributes. Note that the sensitive attributes' results are only applicable for PEQ, AEQ, and RD since MD and SD consider all areas equally without the advantaged/disadvantaged categorization. Additionally, we do not have the RD score for the Ground Truth since RD compares the improvements/degradation compared to the ground truth. We find several observations:

Table 1: Experiment results for all models

Model	# Routes	Efficiency	Equity (Age)					Cost	
			PEQ	AEQ	MD	SD	RD	Network	Fleet
Ground truth	361	0.4714	0.9967	0.9930	0.8208	0.9689	-	1,344,359	5,328
Initial	271	0.4004	0.9921	0.9863	0.8367	0.9696	0.9933	1,011,441	4,012
Direct	361	0.4338	0.9866	0.9827	0.8515	0.9716	0.9898	1,321,586	5,236
Gradual	361	0.4224	0.9939	0.9880	0.8698	0.9736	0.9950	1,345,429	5,284
Gradual-Equity	361	0.4166	0.9933	0.9898	0.8675	0.9722	0.9969	1,349,921	5,310

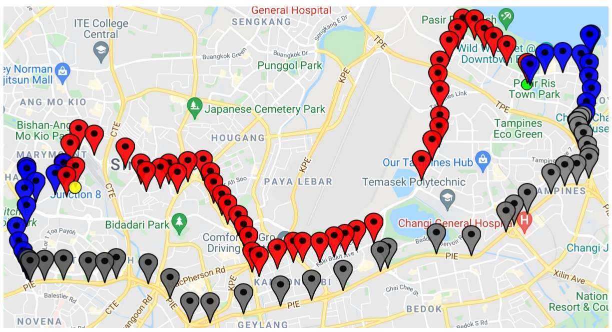

Figure 7: Comparison between one ground truth (red) and direct route (blue and gray); both starting on the yellow and ending on the green circle. The grey points represent the road connecting the two sections of the blue bus routes. This road has no bus stop.

- Due to the demographic spread and Singapore's extensive bus coverage, public transport in Singapore is equitable, with most efficiency scores nearing the gold standard of 1.
- Compared to the Ground Truth, our baseline methods are able to improve upon the MD and SD scores; which measure the discrepancy among all areas' efficiency scores. The Singapore bus network is designed by the government to prioritize some hotspot areas such as Changi Airport whereas our methods are impartial to the context of different areas' purpose, resulting in a better equity.
- The Direct method's equity is overall lower than the other two baselines. As we will explain in detail in the Direct method case study below, the reason is that the efficiency improvements are only applied to a set of bus stops within a small set of areas.
- Gradual-Equity does not show noticeable improvements over Gradual. As we will discuss in greater detail in the Gradual-Equity case study below, while Gradual-Equity covers more bus stops that are relatively disadvantaged, the resulting bus route is not efficient enough to improve the service quality of these stops.

6.2.3 Case Study: Direct Method. Figure 7 compares one ground truth route (red) with one direct route (blue); starting from the yellow dot to the green dot. The direct route is split into two. We visualize the road junctions passed by the direct route in shades of gray to show the full route and we find that this route passes through the Pan Island Expressway (PIE). Since the direct method uses the shortest road distance, it often passes expressways that do not have any bus stops. Thus, this route improves the efficiency of only a few bus stop. This is related to Challenge 2. Additionally, the covered bus stops are concentrated near the start and the end of the route. As a result, bus routes created using the Direct method will greatly boost the efficiency of the areas near the origin and destination bus stops only. The relatively low equity score is caused by this lopsided coverage.

6.2.4 Case Study: Gradual Method. We compare one ground truth route with one route created using the direct method in Figures 8 and 9, respectively. The gradual method manages to avoid the bias towards expressways as the direct method. However, since it is still intended to reach the destination bus stop as soon as possible, it did not manage to detect nearby bus stops that may need coverage. We can see that in the ground truth route, the bus stop on the top left is included. This bus stop is a bus stop nearby a tourist spot; the Sungei Buloh Wetland Reserve. A weakness of automated methods is that they are unable to utilize these specific contextual information.

6.2.5 Case Study: Gradual-Equity Method. One route comparison between the Gradual and Gradual-Equity is displayed in Figures 10 and 11 respectively. We display the routes using lines instead of points for better clarity. Including the equity factors results in longer routes, as the method will often detour from the shortest route in order to add bus stops that are less covered. However, in our observation, we find that on metropolitan areas with lots of street blocks, this method creates routes with lots of inefficient detours and 180 degree turns. This can be seen in the area highlighted in red. This is the main reason why the Gradual-Equity method has the lowest efficiency out of the baselines. The lower efficiency means that most of the covered bus stops do not receive any boost in their efficiency score. Thus, the equity score do not change much compared to the Gradual method.

6.2.6 Benefits and Challenges for Practical Applications. Equitable public bus network optimization provides better service for disadvantaged groups, leading to better social equity thanks to easier access to jobs and opportunities. Our aim in this work is to explore the option of using automated methods. While we apply our methods only on Singapore, our methods are applicable for other cities as well as long as the road network graph data, bus network data, bus schedules data, and demographics data are available. Simply applying these methods would not bring similar improvements to different cities, however, as Singapore has its own idiosyncracies with regards to their bus networks. One example is the terminal

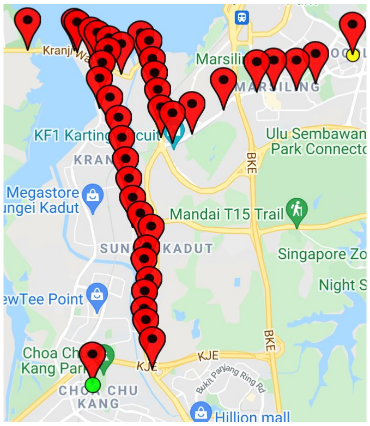

Figure 8: One ground truth route starting from the yellow and ending on the green point

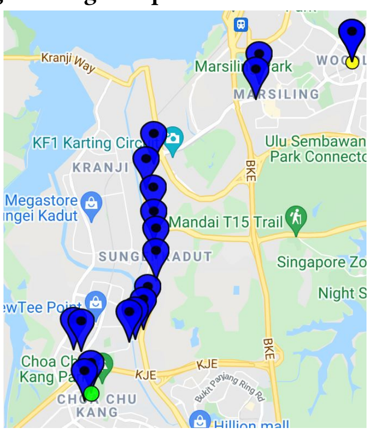

Figure 9: One route created using the gradual method starting from the yellow and ending on the green point

stops. In Singapore, many bus routes start and end at designated bus terminals that houses spacious bus parking spots. This minimizes deadheading, but it requires large space consumption; sometimes on metropolitan areas. This approach may not be applicable in other countries, in which case the concept of terminal stops may not be applicable. This is one amongst many differences that makes applying a one-size-fits-all solution challenging.

7 CONCLUSION

In this work, we introduce the difficulty of tackling equitable bus network optimization. As our work is the first to introduce equitability in bus network optimisation with regards to sensitive demographic attributes, we curate, describe, and publicly release a comprehensive dataset for the case study of Singapore that combines bus network, road network, and demographics data. We then propose an improved bus efficiency metric and several equity metrics. Finally, using these metrics, we perform several exploratory analyses to explain the difficulty of the problem. With our dataset and the metrics we propose, we endeavor to inspire the community

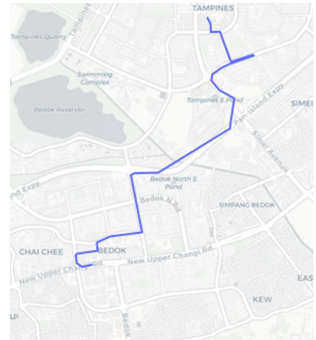

Figure 10: A route created by the Gradual method

Figure 11: A route created by the Gradual-Equity method

to explore this crucial problem by providing a starting point for future research in this field.

ACKNOWLEDGMENTS

This research is supported in part by DP220101434 and DP200102611. ARC had no role in the design and conduct of the study; access and collection of data; analysis and interpretation of data; preparation, review, or approval of the manuscript; or the decision to submit the manuscript for publication. The authors declare no other financial interests.

REFERENCES

[1] 2022. Aggregated Singapore Bus Dataset. https://github.com/rmitbggroup/facct_2022_dataset.

[2] Land Transport Authority. 2013. Household Interview Travel Survey 2012: Public Transport Mode Share Rises to 63%. https://www.lta.gov.sg/content/ltagov/en/newsroom/2013/10/2/household-interview-travel-survey-2012-public-transport-mode-share-rises-to-63.html.

[3] Pranjal Awasthi, Alex Beutel, Matthäus Kleindessner, Jamie Morgenstern, and Xuezhi Wang. 2021. Evaluating Fairness of Machine Learning Models Under Uncertain and Incomplete Information. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 206–214.

[4] Ari Ball-Burack, Michelle Seng Ah Lee, Jennifer Cobbe, and Jatinder Singh. 2021. Differential Tweetment: Mitigating Racial Dialect Bias in Harmful Tweet Detection. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 116–128.

[5] Dipayan Banerjee and Karen Smilowitz. 2019. Incorporating equity into the school bus scheduling problem. *Transportation Research Part E: Logistics and Transportation Review* 131 (2019), 228–246.

[6] Dimitris Bertsimas, Arthur Delarue, and Sebastien Martin. 2019. Optimizing schools start time and bus routes. *Proceedings of the National Academy of Sciences* 116, 13 (2019), 5943–5948.

[7] Robert Bowerman, Brent Hall, and Paul Calamai. 1995. A multi-objective optimization approach to urban school bus routing: Formulation and solution method. *Transportation Research Part A: Policy and Practice* 29, 2 (1995), 107–123.

[8] James C. Chu. 2018. Mixed-integer programming model and branch-and-price-and-cut algorithm for urban bus network design and timetabling. *Transportation Research Part B: Methodological* 108 (2018), 188–216.

[9] Neil Cuthill, Mengqiu Cao, Yuqi Liu, Xing Gao, and Yuerong Zhang. 2019. The Association between Urban Public Transport Infrastructure and Social Equity and Spatial Accessibility within the Urban Environment: An Investigation of Tramlink in London. *Sustainability* 11, 5 (2019).

[10] Ahmed Darwish, Momen Khalil, and Karim Badawi. 2020. Optimising Public Bus Transit Networks Using Deep Reinforcement Learning. In *2020 IEEE 23rd International Conference on Intelligent Transportation Systems (ITSC)*. 1–7.

[11] Alexa Delbosc and Graham Currie. 2011. Using Lorenz curves to assess public transport equity. *Journal of Transport Geography* 19, 6 (2011), 1252 – 1259.

[12] Jwala Dhamala, Tony Sun, Varun Kumar, Satyapriya Krishna, Yada Pruksachatkun, Kai-Wei Chang, and Rahul Gupta. 2021. BOLD: Dataset and Metrics for Measuring Biases in Open-Ended Language Generation. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 862–872.

[13] Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. 2021. Bridging Machine Learning and Mechanism Design towards Algorithmic Fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 489–503.

[14] João Paiva Fonseca, Evelien van der Hurk, Roberto Roberti, and Allan Larsen. 2018. A matheuristic for transfer synchronization through integrated timetabling and vehicle scheduling. *Transportation Research Part B: Methodological* 109 (2018), 128–149.

[15] Fereshte Khani and Percy Liang. 2021. Removing Spurious Features Can Hurt Accuracy and Affect Groups Disproportionately. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. 196–205.

[16] Jinpeng Liang, Jianjun Wu, Ziyou Gao, Huijun Sun, Xin Yang, and Hong K. Lo. 2019. Bus transit network design with uncertainties on the basis of a metro network: A two-step model framework. *Transportation Research Part B: Methodological* 126 (2019), 115–138.

[17] Christoph E. Mandl. 1980. Evaluation and optimization of urban public transportation networks. *European Journal of Operational Research* 5, 6 (1980), 396–404. https://doi.org/10.1016/0377-2217(80)90126-5

[18] Songsong Mo, Zhifeng Bao, Baihua Zheng, and Zhiyong Peng. 2020. Towards an Optimal Bus Frequency Scheduling: When the Waiting Time Matters. *IEEE Transactions on Knowledge and Data Engineering* (2020), 1–1. https://doi.org/10.1109/TKDE.2020.3036573

[19] Vedant Nanda, Pan Xu, Karthik Sankararaman, John Dickerson, and Aravind Srinivasan. 2019. Balancing the Tradeoff between Profit and Fairness in Rideshare Platforms During High-Demand Hours. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society* (12 2019), 2210–2217.

[20] Aybike Ongel, Henriette Cornet, Penny Kong, Raymond Khoo, Tao Liu, and Manfred Kloeppel. 2018. Public transport service quality improvement using Universal Design standards and advanced vehicle technologies. In *2018 International Conference on Intelligent Autonomous Systems (ICoIAS)*. 211–216.

[21] Chelsey Palmateer and David M Levinson. 2017. *Justice, exclusion, and equity: An analysis of 48 US metropolitan areas*. Technical Report. Nexus Research Group.

[22] Anthony Michael Ricciardi, Jianhong(Cecilia) Xia, and Graham Currie. 2015. Exploring public transport equity between separate disadvantaged cohorts: a case study in Perth, Australia. *Journal of Transport Geography* 43 (2015), 111–122.

[23] Susan Shaheen, Corwin Bell, Adam Cohen, and Balaji Yelchuru. 2017. *Travel behavior: Shared mobility and transportation equity*. Technical Report. U.S. Department of Transportation.

[24] Richard Stern. 1996. Passenger Transfer System Review. *TCRP Synthesis of Transit Practice* (1996).

[25] Wai Yuen Szeto and Yongzhong Wu. 2011. A simultaneous bus route design and frequency setting problem for Tin Shui Wai, Hong Kong. *European Journal of Operational Research* 209, 2 (2011), 141–155.

[26] Shuaian Wang and Xiaobo Qu. 2015. Rural bus route design problem: Model development and case studies. *KSCE Journal of Civil Engineering* 19, 6 (2015), 1892–1896.

[27] Sheng Wang, Yuan Sun, Christopher Musco, and Zhifeng Bao. 2021. Public Transport Planning: When Transit Network Connectivity Meets Commuting Demand. In *Proceedings of the 2021 International Conference on Management of Data*. 1906–1919.

[28] Ouri Wolfson and Jane Lin. 2017. Fairness versus Optimality in Ridesharing. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*. 118–123.

[29] An Yan and Bill Howe. 2020. Fairness-Aware Demand Prediction for New Mobility. *Proceedings of the AAAI Conference on Artificial Intelligence* 34, 01 (Apr. 2020), 1079–1087.

[30] An Yan and Bill Howe. 2021. EquiTensors: Learning Fair Integrations of Heterogeneous Urban Data. In *Proceedings of the 2021 International Conference on Management of Data*. 2338–2347.

[31] Zhong Yang, Xueshan Sun, Baozhen Yao, Qingcheng Zeng, and Erik Jeppesen. 2011. Parallel genetic algorithm in bus route headway optimization. *Applied Soft Computing* 11 (12 2011), 5081–5091.

[32] Zhongzhen Yang, Bin Yu, and Chuntian Cheng. 2007. A parallel ant colony algorithm for bus network optimization. *Computer-Aided Civil and Infrastructure Engineering* 22, 1 (2007), 44–55.

[33] Fang Zhao and Ike Ubaka. 2004. Transit Network Optimization - Minimizing Transfers and Optimizing Route Directness. *Journal of Public Transportation* 7, 1 (03 2004), 4.