Equity Promotion in Public Transportation

Anik Pramanik[1]*, Pan Xu[1]*, Yifan Xu[2]*

[1] Department of Computer Science, New Jersey Institute of Technology, Newark, USA
[2] School of Cyber Science and Engineering, Southeast University, Nanjing, China
ap2645@njit.edu, pxu@njit.edu, xyf@seu.edu.cn

Abstract

There are many news articles reporting the obstacles confronting poverty-stricken households in access to public transits. These barriers create a great deal of inconveniences for these impoverished families and more importantly, they contribute a lot of social inequalities. A typical approach addressing the issue is to build more transport infrastructure to offer more opportunities to access the public transits especially for those deprived communities. Examples include adding more bus lines connecting needy residents to railways systems and extending existing bus lines to areas with low socioeconomic status. Recently, a new strategy is proposed, which is to harness the ubiquitous ride-hailing services to connect disadvantaged households with the nearest public transportations. Compared with the former infrastructure-based solution, the ride-hailing-based strategy enjoys a few exclusive benefits such as higher effectiveness and more flexibility.

In this paper, we propose an optimization model to study how to integrate the two approaches together for equity-promotion purposes. Specifically, we aim to design a strategy of allocating a given limited budget to different candidate programs such that the overall social equity is maximized, which is defined as the minimum covering ratio among all pre-specified protected groups of households (based on race, income, etc.). We have designed a linear-programming (LP) based rounding algorithm, which proves to achieve an optimal approximation ratio of $1 - 1/e$. Additionally, we test our algorithm against a few baselines on real data assembled by outsourcing multiple public datasets collected in the city of Chicago. Experimental results confirm our theoretical predictions and demonstrate the effectiveness of our LP-based strategy in promoting social equity, especially when the budget is insufficient.

Introduction

We consider the last-mile problem in public transportation. There are roughly 20% of households that are at or below the federal poverty line lack access to a car, and the percentage can get as high as 33% among the low-income African and Latino population (Berube, Deakin, and Raphael 2006). For these impoverished families: On the one hand, they rely on the public transportation as the only travel option for daily activities such commuting to work, shopping, *etc.*; on the other hand, many of them live relatively far away from public transits and need to walk a long distance to even the nearest bus and/or metro stops. The difficulty in access to public transport creates a great deal of trouble and inconvenience for these poverty-stricken families and contributes to many social inequalities (DeGood and Schwartz 2016). One recent example is the radically unequal access to COVID-19 vaccines in the early stage of rollout in 2021, where it had been widely reported that vaccination rate of Black People was greatly lagging behind that of White counterparts, and one of the main causes was the obstacle in access to public health providers (Blackstock and Blackstock 2021; Stolberg 2021; Abcarian 2021; Jones 2021).

The traditional way to improve the access to public transits includes creating more bus lines connecting needy residents to railway systems, extending existing bus lines to deprived communities, and increasing the frequency and service hours of transit providers, to name a few. Recently, local officials are weighing the option of integrating the popular ride-hailing services offered by Uber and Lyft to the existing toolkit to combat the last-mile problem (Jin, Kong, and Sui 2019; Kong, Zhang, and Zhao 2020). One example is to establish an auxiliary subsidized program to allocate reimbursed ride-hailing trips to needy households to help them access to the near transit stations. Compared with the traditional solutions, ride-hailing services enjoy benefits such as more flexibility and mobility (can be requested via mobile apps upon needed) and a lower cost in general. Additionally, ride-hailing-based program can help effectively target needy households, especially when they sparsely and remotely scatter over a large area where traditional means are either cost prohibitive or ineffective.

There are a few challenging issues in the last-mile problem, including how to craft eligibility to identify the set of qualified needy households, how to set subsidizing guidelines for ride-hailing services for people under different spatial and financial conditions, how to design routes and schedules of new bus lines, and how to split limited budget to different approaches. For policy-related issues, existing literatures have already proposed a few answers. For example, DeGood and Schwartz (2016) outlined eligibility spatial criteria among other financial factors as follows: walking distance to the nearest bus lines should fall between [0.25, 3.5] miles and to

---

the nearest railway station between $[0.5, 3.5]$ miles.[1] They also proposed detailed suggestions on setting tailored subsidizing policy for needy households based on the income level and the household size.

In this paper, we consider a non-policy-related question of optimizing budget allocation: Given a limited budget, a set of established qualified households, and a collection of well-defined promotion programs (*e.g.,* a set of new bus lines with full information regarding the operational cost, routes and schedules, a comprehensive subsidizing framework of ride-hailing trips), how to design a best strategy of allocating the limited fund to different proposed programs to maximize the social equity? There are several different metrics for the social equity; one example is the min coverage ratio among all pre-specified protected groups based on sensitive information like race and ethnicity, which is commonly used in fairness- and equity-related promotions (Ma, Xu, and Xu 2020; Nanda et al. 2020; Hosseini et al. 2022).

We formalize our problem, called **Equity Promotion in Public Transportation** (EPPT), as follows. Suppose we have a set I of qualified needy households, where each i has the following three kinds of information: (1) spatial attributes such as the living location, walking distances to the nearest bus lines and the nearest railway station; (2) financial factors including income level and household size, and (3) basic demographics such as race and ethnicity. Assume we are offered a total fund B for a given time window (*e.g.,* one year or one quarter). We have a collection J of candidate bus lines to open, where each bus line j is characterized as an operational cost c_j and a subset $S_j \subseteq I$ denoting the set of target households covered,[2] and where $\{S_j | j \in J\}$ can be possibly overlapping. Each target household $i \in I$ can be covered either by a candidate bus line $j \in J$ if $S_j \ni i$ or by being enrolled into the ride-hailing-based welfare program, where the latter will induce a cost c_i determined by i's spatial and financial traits. Suppose we have a collection of protected groups $\mathcal{G} = \{g\}$, where each group $g \subseteq I$ is a subset of target households sharing specific demographics (*e.g.,* Black, White, and Asian). Note that all information of $\{I, J, B, \{c_i, c_j, S_j | i \in I, j \in J\}, \mathcal{G} = \{g\}\}$ is given as part of the input. In order to better expose the technical challenges in our problem, we present two versions of a budget allocation strategy, as follows.

Deterministic version of a budget allocation strategy. For each bus line $j \in J$, let $x_j = 1$ indicate that j is to open; for each qualified household $i \in I$, let $x_i = 1$ indicate that i should be enrolled into the ride-hailing program. Thus, a deterministic budget allocation plan can be captured as a binary vector $\mathbf{x} \in \{0, 1\}^{|J| + |I|}$, and it is called *feasible* if $\sum_{j \in J} c_j x_j + \sum_{i \in I} c_i x_i \leq B$, *i.e.,* the total cost is within the given budget. For each needy household $i \in I$, let $y_i := \min(1, x_i + \sum_{j:S_j \ni i} x_j)$, which indicates if i is covered ($y_i = 1$)

or not. Observe that $y_i = 0$ (i is not covered) iff $x_i = x_j = 0$ for all j with $S_j \ni i$, which means i is not added to the ride-hailing program, either none of bus lines covering i is open. The coverage ratio of group g on $\mathbf{x}$ can be then expressed as $\sum_{i \in g} y_i / |g|$, where $|g|$ refers to the cardinality of group g. The resulting objective of the social equity can be computed as $\min_{g \in \mathcal{G}} \sum_{i \in g} y_i / |g|$.

Randomized version of a budget allocation strategy. Let $\Phi = \{\phi_k | 1 \leq k \leq K\}$ be a collection of all possible feasible deterministic strategies as described above.[3] A randomized budget allocation strategy ALG then can be captured as a distribution $\mathcal{D}$ over Φ such that ALG will sample and run a strategy from Φ following distribution $\mathcal{D}$. For each household $i \in I$, let random variable $Y_i = 1$ indicate that i is covered in ALG (and $Y_i = 0$ otherwise). For each group $g \in \mathcal{G}$, the *expected* coverage ratio under ALG then can be expressed as $\mathsf{E}[\sum_{i \in g} Y_i / |g|]$, where the expectation is taken over $\mathcal{D}$, the random choice taken by ALG. The final objective of the social equity can be computed as $\min_{g \in \mathcal{G}} \mathsf{E}[\sum_{i \in g} Y_i / |g|]$.

Deterministic vs. randomized strategies. (1) Note that any deterministic strategy can be cast as a special randomized one. This suggests that by expanding the focus from deterministic to randomized, we (the algorithm designer) can *potentially* land at a better strategy in terms of achieving a larger objective value, while how much extra value we can gain really depends on the specific problem. As Example 1 shows, some randomized strategy can far outperform the optimal deterministic on some instances of our model. (2) Randomized strategies make more sense in the context of promoting equity in public transits compared with deterministic ones. In most practical scenarios, local governments can secure only a limited budget that can cover a small portion of impoverished households. In this case, any deterministic budget-allocation strategy could benefit only a small fixed set of targets and inevitably leave far more vulnerable households unaffected, which could cause more social injustices. In contrast, randomized strategies can potentially impact a far larger deprived population. Also, they are easy to implement in practice, *e.g.,* by randomly sampling a set of targets every time and then recruiting them to the ride-hailing program, and/or by alternatively operating different bus lines based on monthly/quarterly schedules.

Preliminaries and Main Contributions

Throughout this paper, we denote $[n] = \{1, 2, \ldots, n\}$ for a generic positive integer n; we use OPT to denote both of an optimal strategy and the corresponding performance, and the same for ALG, which denotes both a generic algorithm and its performance.

Approximation ratio (AR). For NP-hard combinatorial optimization problems, a powerful framework is called *approximation algorithms*, where we aim to design an efficient algorithm (polynomial running time) with a guaranteed performance from the optimal. Consider a maximization problem

---

[1]Here upper bounds are set to ensure a high utilization of limited funds, which can benefit more people that live relatively closer to the existing public system than those far away.

[2]Here we view bus lines with the same route but different schedules and/or frequencies as distinct since they can incur different costs and cover varied-sized sets of needy households.

---

[3]Note that the number K of all possible feasible strategies is upper bounded by $2^{|I| + |J|}$.

like EPPT as studied here. Let ALG be an approximation algorithm (possibly randomized) and OPT denote an optimal algorithm with no running-time constraint and its performance. We say ALG achieves an approximation ratio of $\rho \in [0,1]$ if $\mathsf{E}[\mathsf{ALG}] \geq \rho \cdot \mathsf{OPT}$ for all possible input instances.

Connection to Budgeted Maximum Coverage Problem (BMCP). To the best of our knowledge, BMCP is the model closest to ours, which is a generalization of the classical Maximum Coverage Problem. The basic setting is as follows. We have a ground set of I and a collection $\mathcal{S} = \{S_j | j \in J\}$ of subsets of I, where each subset S_j indexed by $j \in J$ is associated with a cost $c_j > 0$. Suppose we are given a total budget B, and we aim to identify a sub-collection, denoted by $J' \subseteq J$, to maximize the total coverage $|\cup_{j \in J'} S_j|$ subject to the budget constraint, *i.e.*, $\sum_{j \in J'} c_j \leq B$. BMCP and its related variants are well studied in theoretical computer science community (Cohen and Katzir 2008; Khuller, Moss, and Naor 1999), and most of them can be solved via a greedy-based framework with an optimal approximation ratio of $1 - 1/\mathsf{e} \sim 0.632$ (Feige 1998). *Note that BMCP can be cast as a special case of our problem EPPT when it has only one single protected group and offers no any ride-hailing-based program, which suggests that EPPT admits no approx-ratio better than $1 - 1/\mathsf{e}$ unless P=NP.* The lemma below further highlights the difference between BMCP and EPPT.

Lemma 1. *For BMCP, any optimal randomized strategy can be realized by a deterministic one. In contrast, there exists some instance of EPPT where an optimal randomized strategy can strictly beat an optimal deterministic.*

The lemma above suggests that expanding the set of strategy choices from deterministic to randomized will offer no extra power for BMCP but will possibly do for EPPT. The difference is mainly due to the fact that the objective function of BMCP is linear (thus, linearity of expectation can be applied), whereas that of EPPT is nonlinear.[4]

Proof. We prove the first claim for BMCP. Consider a given instance of BMCP with a ground set of I and a given budget B. Let OPT_R be an optimal randomized strategy that is characterized by the following distribution over the collection $\Phi = \{\phi_k | k \in [K]\}$ of all feasible deterministic strategies: OPT_R will run ϕ_k with probability q_k with $\sum_{k \in [K]} q_k = 1$. For each element $i \in I$ and each deterministic strategy k, let $Y_{ik} = 1$ indicate that i is covered in ϕ_k and $Y_{ik} = 0$ otherwise. Thus, $\sum_{i \in I} Y_{ik}$ captures the coverage of strategy ϕ_k. Observe that

$$\mathsf{E}[\mathsf{OPT}_R] = \sum_{k \in [K]} \left(\sum_{i \in I} Y_{ik} \right) \cdot q_k \leq \max_{k \in [K]} \left(\sum_{i \in I} Y_{ik} \right) = \mathsf{OPT}_D,$$

where OPT_D denotes the performance of an optimal deterministic. The claim for EPPT can be seen on Example 1. □

Example 1. *Consider a toy example of EPPT, where $\mathcal{G} = \{g_1, g_2\}$ with $g_1 = \{a\}$, $g_2 = \{b\}$ and $I = \{a, b\}$. Let $J = \emptyset$*

(no bus lines to open) and covering households a and b via the ride-hailing program each incur a unit cost that is equal to the budget, i.e., $c_a = c_b = B = 1$. We can verify that for any deterministic strategy can achieve an equity being 0 since it can cover only one household in one group; thus, $\mathsf{OPT}_D = 0$. Consider such a randomized strategy that is to cover only a or b via the ride-hailing program each with probability $1/2$. Then it achieves an expected equity of $1/2$, which suggests that $\mathsf{E}[\mathsf{OPT}_R] \geq 1/2 > \mathsf{OPT}_D = 0$.

Main Contributions and Techniques

In this paper, we consider a technical issue of promoting the social equity by optimizing the budget allocation to different candidate welfare programs assisting deprived households in access to the public transits. In all technical sections, we assume WLOG that the cost associated with each program is no more than 1 with $B \geq 1$ by scaling down all costs such that $\max_{\ell \in I \cup J} c_\ell = 1$. Our main technical result is stated as follows.

Theorem 1. *[Section] There exists a linear-programming (LP) based rounding strategy (RAS) that achieves an optimal approximation ratio of $1 - 1/\mathsf{e}$ for EPPT, which uses a budget no more than B in expectation and no more than $B + 1$ for any realization.*

Remarks on Theorem 1. (1) Note that our problem EPPT captures BMCP as a special case; thus, no algorithm (including randomized) can achieve an approx-ratio better than $1 - 1/\mathsf{e}$, which suggests the optimality RAS in terms of approx-ratio. (2) As stated in Theorem 1, our strategy is feasible in budget under expectation, though no always. This is inconsequential and can be mitigated technically. Note that the total absolute overflow of the budget is 1, which represents the max cost associated with any candidate programs. By decomposing or splitting a given program into several copies (*e.g.*, replacing a bus line operating 12 hours a day by 12 bus lines sharing the same route but operating one hour only a day), we can significantly reduce the max cost incurred by any candidate welfare programs. Another note is that our input setting is positioned on a given time window (*e.g.*, one quarter or a year). Thus, we expect to run our randomized strategy repeatedly stretching over a long period covering multiple units of the time window considered here. The fact that our strategy is feasible in the budget under expectation suggests its feasibility in *the long run*, making it somewhat acceptable in practice.

We implement RAS and compare it against several natural baselines on real data that is assembled by integrating multiple real datasets from different public sources. Experimental results confirm our theoretical predictions and demonstrate the effectiveness of our LP-based randomized strategy (RAS) in promoting social equity in public transportation, especially when the budget is small. This highlights the practical value of our proposed strategy in improving social equity since the lack of fund for the public infrastructure is widely reported and is prevalent across the USA; see, *e.g.*, (APTA 2021; Mawad 2021; Freemark 2021). Furthermore, the results suggest the great complementary value brought by the ride-hailing-based program to the traditional bus-line-based

---

[4]Actually, the first part of statement of Lemma 1 can be generalized to any optimization problem where the objective function can be expressed as a linear function of decision variables.

when the budget is insufficient.

Technical challenges. The main technical challenge in our problem is partially due to the non-linear objective function of maximizing the overall social equity that is quantified as the minimum (expected) coverage ratio among all protected groups. One the one hand, as shown in Example 1, randomized strategies can be substantially more powerful than deterministic ones on the objective studied here; on the other hand, the design and analysis of a randomized algorithm are more technically challenging in general compared with that of a deterministic, requiring to craft and add appropriate random bits based on the specific input structure. Note that the introduction of ride-hailing welfare program does not add any new *technical* challenges to the problem,[5] though it could mean a lot in practice in terms of promoting the social equity; see experimental results of the impact on the equity brought by the ride-hailing program in Section 12.

One of the main technical contributions in the paper is to propose a *weighted* version of Dependent Rounding (DR), where the original version of DR was introduced by Gandhi et al. (2006) that is to round a vector in a fractional bipartite-matching polytope to an integral that is required to lie in the same polytope and satisfy a few properties. One specific feature imposed on the final rounded solution in (Gandhi et al. 2006) is that the *unweighted* sum of variables incident to every vertex should remain invariant before and after rounding. In our case, we need to ensure the *weighted* sum of all variables has a gap as small as possible between the original fractional and the final rounded integral solutions. To solve this challenge, we craft a more delicate weighted version of DR; see Algorithm 1.

Other related work. There are a few works that have considered resource allocation in online setting under different contexts; see, *e.g.,* equity promotion in vaccination (Xu and Xu 2022), ride-hailing resource allocation (Lesmana, Zhang, and Bei 2019), online resource-allocation problems with limited choices in the long-chain design (Asadpour, Wang, and Zhang 2020), and online fair division problem (Aleksandrov et al. 2015). Another research line has studied matching policy design and simulation in integrating the ride-hailing and public transits (Basu et al. 2018; Boone, Steinberger, and Wafa 2018; Shen, Zhang, and Zhao 2018; Stiglic et al. 2018; Yan, Levine, and Zhao 2019). However, these studies haven't considered the objective of promoting the overall social equity as here.

An LP-based Rounding Algorithm

Recall that the introduction of ride-hailing-based program brings no technical challenge to EPPT; see discussions in "**Technical challenges**." For the ease of exposition, we consider an input instance of EPPT with the bus-lines-based program only by treating to cover each household via ride-hailing trips as a special one-one bus line covering the specific household only. Consider an optimal randomized strat-

egy OPT.[6] For each target household $i \in I$, let y_i be the probability that i is covered in OPT. For each candidate bus line $j \in J$, let x_j be the probability that j is set to open in OPT. Consider the Linear Program (LP) below.

$$\max \quad \min_{g \in \mathcal{G}} \Big(\sum_{i \in g} y_i / |g| \Big), \tag{1}$$

$$\sum_{j \in J} c_j x_j \leq B, \tag{2}$$

$$y_i \leq \min\Big(1, \sum_{j:S_j \ni i} x_j\Big), \ \forall i \in I \tag{3}$$

$$0 \leq y_i, x_j \leq 1 \qquad i \in I, j \in J. \tag{4}$$

Lemma 2. *The optimal value of* LP (1) *is a valid upper bound for the performance of an optimal randomized strategy.*

Proof. We can verify that Objective (1) captures the exact performance of OPT (*i.e.,* the min expected coverage ratio among all protected groups). To prove our claim, it suffices to show that the strategy $\{x_j, y_i\}$ of OPT is feasible to all constraints there. For each $j \in J$ and $i \in I$, let $X_j = 1$ and $Y_i = 1$ indicate j is set to open and i is covered in OPT, respectively. Thus, $\mathsf{E}[X_j] = x_j$ and $\mathsf{E}[Y_i] = y_i$ for every j and i. Since OPT can be viewed as a certain randomization over all feasible deterministic strategies, $\sum_{j \in J} c_j X_j \leq B$. Thus, $\mathsf{E}[\sum_{j \in J} c_j X_j] \leq B$, which leads to Constraint (2). Note that for each $i \in I$, $Y_i \leq 1$ and $Y_i \leq \sum_{j:S_j \ni i} X_j$. Taking expectation on both sides, we get Constraint (3). Constraint (4) is valid since $\{x_i, y_j\}$ are all probability values. □

Based on an optimal solution to LP (1), we design a strategy, denoted by RAS (Algorithm 1), which is built on randomized dependent rounding. Let $\mathbf{x}^* = (x_j^*)$ be part of the optimal solution of LP (1). The main idea is to apply a series of rounding procedures to transform $\mathbf{x}^*$ to a random binary vector $\mathbf{X}^* \in \{0, 1\}^{|J|}$ such that the expected total cost $\sum_{j \in J} c_j \mathsf{E}[X_j^*]$ is as small as possible (ideally no more than B), while each household $i \in I$ can get covered with a probability as large as possible.

The overall picture of our rounding procedure is as follows. During each step, we identify two remaining fractional values in $\mathbf{x}^*$ if any, and "twist" the two values together following one of the two directions randomly, and in each direction, one fractional value will be rounded up and the other will be down. We carefully design the rounding procedure such that it has Properties (**P1**), (**P2**), (**P3**), and (**P4**) as outlined in Lemma 3, which is vital to the performance of the randomized strategy RAS in Algorithm 1. Generally, (**P1**) suggests that at least one fractional value in $\mathbf{x}^*$ will be rounded in each step and thus, RAS will terminate after at most $|J|$ steps.[7]; (**P2**) says that the expectation on each variable remains the same throughout the rounding process; (**P3**) implies that the total budget on the final rounded solution could get an overflow by at most 1; and (**P4**) indicates *negative* correlation over any subset of variables.

---

[5]We can actually view covering a household i by the ride-hailing program alternatively as a virtual new bus line j' with cost $c_{j'} = c_i$ and coverage $S_{j'} = \{i\}$.

[6]Throughout this paper, we use the two terms "strategy" and "algorithm" interchangeably.

[7]More precisely, RAS will stop after at most $\lceil |J|/2 \rceil$ steps.

Algorithm 1: A Randomized Allocation Strategy (RAS) for Equity Promotion in Public Transportation.

Input: An input instance of Equity Promotion in Public Transportation (EPPT): $\{I, J, B, \{c_j, S_j | j \in J\}, \mathcal{G} = \{g\}\}$.

Output: A *randomized* budget-allocation strategy denoted by a binary vector $\mathbf{X} \in \{0, 1\}^{|J|}$ with $X_j = 1$ indicating to open bus line $j \in J$ and $X_j = 0$ otherwise.

1 Solve LP (1) and let $\mathbf{x}^* = \{x_j^* | j \in J\}$ be part of the optimal solution. Set $\mathbf{X} = \mathbf{x}^*$, *i.e.*, $X_j = x_j^*$ for all $j \in J$.
2 **while** *1* **do**
3   **if** *there exists two fractional values in* $\mathbf{X}$*, say,* $0 < X_p, X_q < 1$*,* **then**
4   compute $\alpha := \max\{\epsilon > 0 : X_p + \epsilon \le 1, X_q - c_p \cdot \epsilon/c_q \ge 0\}$; $\beta := \max\{\epsilon > 0 : X_p - \epsilon \ge 0, X_q + c_p \cdot \epsilon/c_q \le 1\}$;
5   with probability $\beta/(\alpha + \beta)$, update $X_p \leftarrow X_p + \alpha, X_q \leftarrow X_q - c_p \cdot \alpha/c_q$;
6   with probability $\alpha/(\alpha + \beta)$, update $X_p \leftarrow X_p - \beta, X_q \leftarrow X_q + c_p \cdot \beta/c_q$.
7   **else if** *there exists one single fractional value* $0 < X_j < 1$ *in* $\mathbf{X}$*,* **then**
8   with probability X_j, set $X_j \leftarrow 1$; and with probability $1 - X_j$, set $X_j \leftarrow 0$.
9   **else**
10   **Break**.
11   **end**
12 **end**

Lemma 3. RAS *in Algorithm 1 satisfies the following properties in each rounding step: (**P1**) At least one fractional value is rounded to either 0 or 1; (**P2**) The expectation on each value keeps invariant after rounding; (**P3**) The total cost remains the same if two fractional values are involved and will get increased by at most 1 if only one value gets involved; (**P4**) For any subset* $\mathcal{S} \subseteq J$, $\mathsf{E}[\prod_{j \in \mathcal{S}}(1 - X_j)]$ *will never get increased.*

We present the proof of (**P1**), (**P2**), and (**P3**) here and leave that of (**P4**) to Appendix.

Proof. Consider two cases. **Case 1**: There exists two fractional values in $\mathbf{X}$, say $0 < X_p, X_q < 1$. Following the definition of α and β, we see that the two updating procedures, $X_p \leftarrow X_p + \alpha, X_q \leftarrow X_q - c_p \cdot \alpha/c_q$ and $X_p \leftarrow X_p - \beta$, $X_q \leftarrow X_q + c_p \cdot \beta/c_q$, both will end up with at least one integral value on either X_p or X_q. Let X_p' and X_q' be the rounded value in the end. We see

$$\mathsf{E}[X_p'|X_p, X_q] = (X_p + \alpha) \cdot \frac{\beta}{\alpha + \beta} + (X_p - \beta) \cdot \frac{\alpha}{\alpha + \beta} = X_p;$$

$$\mathsf{E}[X_q'|X_p, X_q] = \left(X_q - \frac{c_p\alpha}{c_q}\right) \cdot \frac{\beta}{\alpha + \beta} + \left(X_q + \frac{c_p\beta}{c_q}\right) \cdot \frac{\alpha}{\alpha + \beta}$$
$$= X_q.$$

Thus, we conclude that the expectations of the two values both remain unchanged after rounding. Observe that for each updating procedure, the total cost remains the same:

$$(X_p + \alpha) \cdot c_p + (X_q - \frac{c_p\alpha}{c_q}) \cdot c_q = X_p \cdot c_p + X_q \cdot c_q;$$

$$(X_p - \beta) \cdot c_p + (X_q + \frac{c_p\beta}{c_q}) \cdot c_q = X_p \cdot c_p + X_q \cdot c_q.$$

Now consider **Case 2**: There exists one single fractional values in $\mathbf{X}$, say $0 < X_j < 1$. Let X_j' be the value after being rounded. According to our procedure, we see that $X_j' \in \{0, 1\}$. Furthermore, $\mathsf{E}[X_j'|X_j] = X_j \cdot 1 = X_j$. Note that in this case the total cost will get increased by at most $c_j \le 1$

(recalled that we assume WLOG that $c_j \le 1$ for all $j \in J$). $\qquad\square$

We are ready to prove the main Theorem 1. For ease of exposition, we restate Theorem 1 in the lemma below.

Lemma 4. *For the randomized strategy* RAS*, (1) it uses a total budget no more than B in expectation; (2) it uses a total budget no more than $B + 1$ for any realization; and (3) it achieves an approx-ratio at least of $1 - 1/\mathsf{e}$.*

Proof. Let $\{x_j^*, y_i^*\}$ be an optimal solution to LP (1). We prove Claim (1) first. Let $\mathbf{X}^* = (X_j^*)$ the final rounded binary vector output by RAS. Property (**P2**) in Lemma 3 suggests that $\mathsf{E}[X_j^*] = x_j^*$ for every $j \in J$ since we initialize $\mathbf{X} = \mathbf{x}^* = (x_j^*)$. Thus, the expected total cost on $\mathbf{X}^*$ should satisfy $\mathsf{E}[\sum_{j \in J} c_j X_j^*] = \sum_{j \in J} c_j x_j^* \le B$, where the last inequality is due to Constraint (2). For Claim (2): Note that the case when there is only one fractional value involved in the rounding could happen at most once in RAS, which implies that the total cost can get inflated by at most 1.

Now we show Claim (3). For each $i \in I$, let $Y_i^* = 1$ indicate that i is covered in the final strategy $\mathbf{X}^*$ and $Y_i^* = 0$ otherwise. Let $\mathcal{S}_i = \{j : S_j \ni i\}$, which denotes the set of indices of bus lines that cover i. We can verify that $Y_i^* = 1 - \prod_{j \in \mathcal{S}_i}(1 - X_j^*)$, and $\mathsf{E}[\prod_{j \in \mathcal{S}_i}(1 - X_j^*)] \le \prod_{j \in \mathcal{S}_i}(1 - x_j^*)$ (by repeatedly applying (**P4**) in Lemma 3). Thus,

$$\mathsf{E}[Y_i^*] = 1 - \mathsf{E}\left[\prod_{j \in \mathcal{S}_i}(1 - X_j^*)\right]$$
$$\ge 1 - \prod_{j:j \in \mathcal{S}_i}(1 - x_j^*) \quad ((\mathbf{P4}) \text{ in Lemma 3})$$
$$\ge 1 - \prod_{j:j \in \mathcal{S}_i} \mathsf{e}^{-x_j^*} = 1 - \mathsf{e}^{-\sum_{j \in \mathcal{S}_i} x_j^*}. \qquad (5)$$

Note that by Constraint (3), $y_i^* \le \min\left(1, \sum_{j \in \mathcal{S}_i} x_j^*\right)$. Consider these two cases. **Case (1)**: $\sum_{j \in \mathcal{S}_i} x_j^* \ge 1$. Then

$$\mathsf{E}[Y_i^*] \ge 1 - \mathsf{e}^{-\sum_{j \in \mathcal{S}_i} x_j^*} \ge 1 - 1/\mathsf{e} \ge (1 - 1/\mathsf{e}) \cdot y_i^*,$$
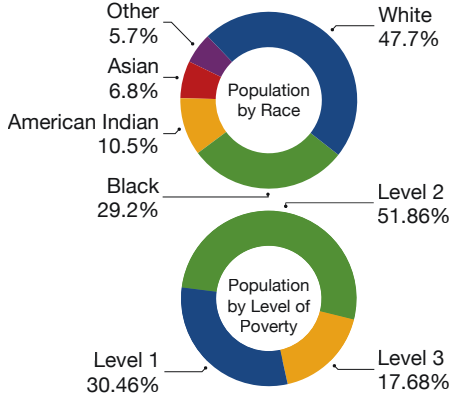
11894

Figure 1: The household distributions used in our experiments. Top: the distribution of households by races, *i.e.,* White, Black, American Indian, Asian, and some other ethic background. Bottom: the distribution of households by poverty levels.

which is due to Inequality (5) and the fact $y_i^* \leq 1$. **Case (2)**: $\sum_{j \in \mathcal{S}_i} x_j^* \leq 1$. In this case, $y_i^* \leq \sum_{j \in \mathcal{S}_i} x_j^* \leq 1$. Observe that

$$\mathsf{E}[Y_i^*] \geq 1 - \mathsf{e}^{-\sum_{j \in \mathcal{S}_i} x_j^*} \geq (1 - 1/\mathsf{e}) \cdot \left(\sum_{j \in \mathcal{S}_i} x_j^* \right) \geq (1 - 1/\mathsf{e}) \cdot y_i^*,$$

where the second inequality follows from the fact that function $(1 - \mathsf{e}^{-x})/x$ is decreasing over $x \in (0, 1]$. Summarizing the two cases above, we conclude that $\mathsf{E}[Y_i^*] \geq (1 - 1/\mathsf{e}) \cdot y_i^*$. Note that the (expected) performance of RAS satisfies

$$\mathsf{E}[\mathsf{RAS}] = \min_{g \in \mathcal{G}} \mathsf{E}\left[\sum_{i \in g} Y_i^* / |g| \right] \geq (1 - 1/\mathsf{e}) \cdot \min_{g \in \mathcal{G}} \sum_{i \in g} y_i^* / |g|$$

$$= (1 - 1/\mathsf{e}) \cdot \mathsf{LP}(1) \geq (1 - 1/\mathsf{e}) \cdot \mathsf{OPT},$$

where LP (1) and OPT denote the optimal value of LP (1) and the performance of an optimal randomized strategy, and where the last inequality is due to Lemma 2. □

Experimental Results

Our experiments involve quite a few datasets that were collected in the city of Chicago. For each dataset, we cite it as a reference pointing to the public link. Additionally, we offer more details in Appendix about how we use them in our experiments. All experiments are conducted on a PC with 2GHz Quad-Core Intel Core i7 processor and 8GB main memory.

Experiment Setup

Data preprocessing. We focus on needy households that are far away from the public transits, that is between 0.5 and 3.5 miles away form the nearest train stations and between 0.25 and 3.5 miles away from the nearest bus stations (De-Good and Schwartz 2016). We identify a total number of 17,875 target households based on data from the Chicago Metropolitan Agency for Planning (CMAP 2019) and the public transportation data by the Chicago Data Portal (CDP

2022). For each needy household, we set a poverty level following the guideline based on the income level estimation information provided by the Department of Health and Human Services (ASPE 2022). We separate all target households into three groups: the first is at or above 200% level of the poverty line (Group Level 1), the second group is between 185% and 200% of the line (Group Level 2), while the third is at or below 175% (Group Level 3). A recent survey showed that the average cost for a ride-hailing trip is about $25.37 (Salas 2019), and accordingly, we set the subsidy in the ride-hailing program as $10, $15, and $20 per ride for Group Level 1, 2, and 3, respectively. We assign each household a random race following the distribution recorded by the official US Census in Chicago, 2022 (Review 2022); see Figure 1 (Top).

The set of candidate bus routes is generated as follows. We cluster all needy households with the travel distance no more than 0.25 miles and then set a bus stop right in the center. After filtering out those impractical stations that are more than 3.5 miles away from any others, we have 649 candidate bus stops in total. Recall that each candidate bus line is for connecting residents with the nearest public transits. Thus, each route should end at either a metro station or a bus stop. Following this principle, we create 20 candidate bus routes with proper length, that is with 10-18 stops and no more than 0.75 miles between any two stops. For each bus route, we set the operating expense per vehicle revenue hour to $140 based on information offered by the Chicago Transit Authority (Carter 2019). We create two schedules for each bus route, one with the full-day and the other with the half-day working schedule.[8]

Algorithms. Suppose we have a total budget B and a collection J of candidate bus lines (which could potentially include those one-one virtual bus lines representing to cover a household via the ride-hailing program in case the latter is considered). In addition to the LP-based algorithm proposed in this paper (RAS), we implement two baselines as follows. (a) Greedy: Always select the bus line $j \in J$ that maximizes the marginal gain of equity based on the current budget allocation (break ties arbitrarily), and repeat this procedure until the budget B is exhausted or all target households are covered; (b) Uniform: Select a bus line from J uniformly at random once at a time, and repeat this procedure until the budget B is exhausted or all target households are covered; As for randomized algorithms of RAS and Uniform, we run each 1000 times and take the average as the final performance. We compare the performance of each algorithm against the optimal value to the benchmark LP (1) and compute that ratio as the final approximation ratio achieved. We also compute the 95% confidence interval for each algorithm's performance to show the potential robustness.

Computational complexity of RAS. The running time of RAS consists of two parts: Solving LP (1) and the rounding procedure. Theoretically, the running time to solve LP (1) can be as low as $O^*(N^{2+1/6} \log(N/\delta))$ (Cohen, Lee, and Song

---

[8]We assume that the half-day working schedule can serve only half of the households on the route.
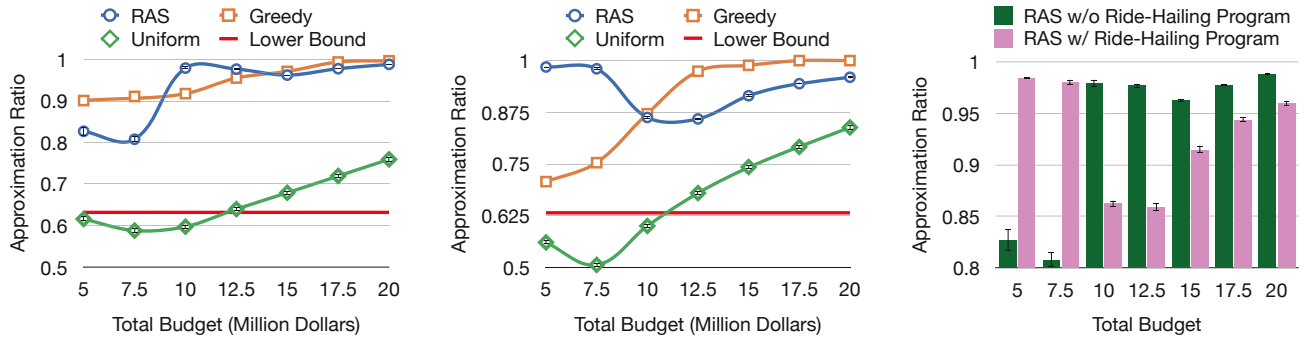
(a) Approximation ratios with the bus-line-based program only.

(b) Approximation ratios when the ride-hailing-based program added.

(c) Comparison between before and after adding the ride-hailing-based program.

Figure 2: Experimental results on real datasets collected in Chicago: The total number of budgets B takes values from $\{5, 7.5, ..., 20\}$ (million dollars).

2021), where δ is the relative accuracy and $N = |I| + |J|$ with I and J being the total number of households and bus lines, respectively. As for the rounding procedure, RAS is guaranteed to eliminate at least one fractional value from the vector $\mathbf{X}$ of size $|J|$ each round and thus, it takes $O(|J|)$ time. Therefore, theoretically, the dominant part of the running time is to solve the benchmark LP (1).

Results and Discussions

We run our experiments in two cases: the first is with bus-lines-based program only; the second is to add ride-hailing program and consider the two programs together. In our experiments, we aim to make a plan for one quarter with a total of budget of $B \in \{5, 7.5, 10, \ldots, 20\}$ (million dollars).

Figure 2a shows results in the first case when the bus-lines-based program is the only choice. The approximation ratios of RAS always stay above the theoretical lower bound of $1 - 1/\mathsf{e} \sim 0.632$ (plotted in red line) as shown in Theorem 1. As the total budget increases, all three algorithms' approximation ratios are increasing and approaching 1. This can be seen as follows: the more budget we have, the more capability each algorithm has to cover as many households as possible in each protected group, which leads to a higher equity as a result. Greedy can outperform RAS at some larger budgets since in that case, there is less need in optimizing the budget allocation among different approaches as RAS is designed for. Note that Greedy does not have any theoretical guarantees, which can perform arbitrary bad on some worst-scenario instance (since Greedy is deterministic, and as Example 1 shows, any deterministic achieves an approximation ratio of zero).

Figure 2b shows the results in the second case when both bus-lines-based and ride-hailing-based programs are considered. Our algorithm RAS dominates the other two baselines when the total budget is capped by 10 million dollars. This highlights the superiority of the LP-based rounding algorithm when budget is limited, which is commonly observed in reality. As reported in (KHARA 2022), "the needed investment for public transit in U.S. is a far cry from the current numbers that have finally been passed."

Figure 2c demonstrates the high potentials of the ride-hailing program in promoting equity among protected groups when budget is relatively small. In the case of scarcity of public funds, the approach of covering needy households via ride-hailing trips prove to be far more efficient than the traditional one through opening new bus lines. This is due to the fact that the ride-hailing-based approach enjoys more flexibility and a higher utilization rate of budget when the funds are insufficient. When budget becomes more abundant, it seems a better choice to consider the traditional approach only since the advantage of ride-hailing trips diminishes.

Conclusion

In this paper, we propose a theoretical model to promote equity in public transportation by optimizing budget allocation over different approaches of improving the access to public transits for needy households. We design an LP-based rounding algorithm and prove that it achieves an optimal $1 - 1/\mathsf{e}$ approximation ratio. Additionally, we test our algorithm against a few natural baselines on real datasets, and experimental results confirm our theoretical predictions and highlight the effectiveness in promoting the social equity among households with low socioeconomic status.

Our work opens a few new directions. The first is to introduce more constraints to the model reflecting practical restrictions in the real world to make our model one step closer to reality. One example is to consider adding area-based capacity to the ride-hailing program, which is due to the limited availability of ride-hailing drivers in that area. We expect any newly added constraints could bring significant technical challenges in algorithmic design and analysis. The second is to parallelize the current algorithm such that it can be implemented and deployed more efficiently.

Acknowledgments

References

Abcarian, R. 2021. Column: The answer to equitable distribution of COVID-19 vaccines is hiding in plain sight. https://tinyurl.com/2p8tcskt. Accessed: 2022-08-12.

Aleksandrov, M. D.; Aziz, H.; Gaspers, S.; and Walsh, T. 2015. Online fair division: Analysing a food bank problem. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

APTA. 2021. Public Transportation Facts by APTA. https://tinyurl.com/yz82rhmw. Accessed: 2022-08-12.

Asadpour, A.; Wang, X.; and Zhang, J. 2020. Online Resource Allocation with Limited Flexibility. *Management Science*, 66(2): 642–666.

ASPE. 2022. HHS Poverty Guidelines for 2022. https://tinyurl.com/2p9mcnxy. Accessed: 2023-03-20.

Basu, R.; Araldo, A.; Akkinepally, A. P.; Biran, B. H. N.; Basak, K.; Seshadri, R.; Deshmukh, N. M.; Kumar, N.; Azevedo, C. L.; and Ben-Akiva, M. E. 2018. Automated Mobility-on-Demand vs. Mass Transit: A Multi-Modal Activity-Driven Agent-Based Simulation Approach. *Transportation Research Record*, 2672: 608 – 618.

Berube, A.; Deakin, E.; and Raphael, S. 2006. Socioeconomic differences in household automobile ownership rates: Implications for evacuation policy. In *Program on housing and urban policy conference paper series, University of California, Berkeley, Working Paper*. University of California, Berkeley.

Blackstock, U.; and Blackstock, O. 2021. White Americans are being vaccinated at higher rates than Black Americans. Such inequity cannot stand. https://tinyurl.com/4f5xrhsu. Accessed: 2022-08-12.

Boone, S.; Steinberger, S.; and Wafa, Z. 2018. Mobility as a Service: Examining Transit Agency-Ride-Hailing Partnerships. In *Transportation Research Board 97th Annual Meeting*.

Carter, D. R. 2019. President's 2019 Budget Recommendations. https://tinyurl.com/mvkwznf6. Accessed: 2023-03-20.

CDP. 2022. Transportation of Chicago Data Portal. https://tinyurl.com/3smckwtp. Accessed: 2023-03-20.

CMAP. 2019. My Daily Travel Survey, 2018-2019: Public Data. https://tinyurl.com/2cvwk4p5. Accessed: 2023-03-20.

Cohen, M. B.; Lee, Y. T.; and Song, Z. 2021. Solving linear programs in the current matrix multiplication time. *Journal of the ACM (JACM)*, 68(1): 1–39.

Cohen, R.; and Katzir, L. 2008. The generalized maximum coverage problem. *Information Processing Letters*, 108(1): 15–22.

DeGood, K.; and Schwartz, A. 2016. Can New Transportation Technologies Improve Equity and Access to Opportunity? https://tinyurl.com/34purvc2. Accessed: 2022-07-11.

Feige, U. 1998. A threshold of ln n for approximating set cover. *Journal of the ACM (JACM)*, 45(4): 634–652.

Freemark, Y. 2021. US Public Transit Has Struggled to Retain Riders over the Past Half Century. Reversing This Trend Could Advance Equity and Sustainability. https://tinyurl.com/yc6h4ntm. Accessed: 2022-08-12.

Gandhi, R.; Khuller, S.; Parthasarathy, S.; and Srinivasan, A. 2006. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM (JACM)*, 53(3): 324–360.

Hosseini, H.; Huang, Z.; Igarashi, A.; and Shah, N. 2022. Class Fairness in Online Matching. *arXiv preprint arXiv:2203.03751*.

Jin, S. T.; Kong, H.; and Sui, D. Z. 2019. Uber, public transit, and urban transportation equity: A case study in new york city. *The Professional Geographer*, 71(2): 315–330.

Jones, G. A. 2021. Opinion: How we're working toward racial equity in distributing coronavirus vaccines in D.C. https://tinyurl.com/yjcxfw3f. Accessed: 2022-08-12.

KHARA, I. 2022. A massive step but is it enough? US Infrastructure Bill invests over $100 billion in public transport. https://tinyurl.com/bddkm45w. Accessed: 2022-07-25.

Khuller, S.; Moss, A.; and Naor, J. S. 1999. The budgeted maximum coverage problem. *Information processing letters*, 70(1): 39–45.

Kong, H.; Zhang, X.; and Zhao, J. 2020. How does ridesourcing substitute for public transit? A geospatial perspective in Chengdu, China. *Journal of Transport Geography*, 86: 102769.

Lesmana, N. S.; Zhang, X.; and Bei, X. 2019. Balancing efficiency and fairness in on-demand ridesourcing. In *Advances in Neural Information Processing Systems*, 5310–5320.

Ma, W.; Xu, P.; and Xu, Y. 2020. Group-level Fairness Maximization in Online Bipartite Matching. *arXiv preprint arXiv:2011.13908*.

Mawad, T. F. 2021. The Complex 50-Year Collapse of U.S. Public Transit. https://tinyurl.com/mr33xr32. Accessed: 2022-08-12.

Nanda, V.; Xu, P.; Sankararaman, K. A.; Dickerson, J.; and Srinivasan, A. 2020. Balancing the tradeoff between profit and fairness in rideshare platforms during high-demand hours. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, 2210–2217.

Review, W. P. 2022. Chicago, Illinois Population 2022. https://tinyurl.com/2p8ah52c. Accessed: 2023-03-20.

Salas, E. B. 2019. https://tinyurl.com/2p8853rc. Accessed: 2022-07-20.

Shen, Y.; Zhang, H.; and Zhao, J. 2018. Integrating shared autonomous vehicle in public transportation system: A supply-side simulation of the first-mile service in Singapore. *Transportation Research Part A: Policy and Practice*, 113: 125–136.

Stiglic, M.; Agatz, N.; Savelsbergh, M.; and Gradisar, M. 2018. Enhancing urban mobility: Integrating ride-sharing and public transit. *Computers & Operations Research*, 90: 12–21.

Stolberg, S. G. 2021. As Biden Pushes for Racial Equity in Vaccination, Data Is Lagging. https://tinyurl.com/5n8hhd3k. Accessed: 2022-08-12.

Xu, P.; and Xu, Y. 2022. Equity Promotion in Online Resource Allocation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, 9962–9970.

Yan, X.; Levine, J.; and Zhao, X. 2019. Integrating ridesourcing services with public transit: An evaluation of traveler responses combining revealed and stated preference data. *Transportation Research Part C: Emerging Technologies*, 105: 683–696.