Trade the System Efficiency for the Income Equality of Drivers in Rideshare

Yifan Xu[1] and **Pan Xu**[2]
[1]Key Lab of CNII, MOE, Southeast University, Nanjing, China
[2]Department of Computer Science, New Jersey Institute of Technology, Newark, USA
xyf@seu.edu.cn, pxu@njit.edu

Abstract

Several scientific studies have reported the existence of the income gap among rideshare drivers based on demographic factors such as gender, age, race, etc. In this paper, we study the income inequality among rideshare drivers due to discriminative cancellations from riders, and the tradeoff between the income inequality (called fairness objective) with the system efficiency (called profit objective). We proposed an online bipartite-matching model where riders are assumed to arrive sequentially following a distribution known in advance. The highlight of our model is the concept of *acceptance rate* between any pair of driver-rider types, where types are defined based on demographic factors. Specially, we assume each rider can accept or cancel the driver assigned to her, each occurs with a certain probability which reflects the acceptance degree from the rider type towards the driver type. We construct a bi-objective linear program as a valid benchmark and propose two LP-based parameterized online algorithms. Rigorous online competitive ratio analysis is offered to demonstrate the flexibility and efficiency of our online algorithms in balancing the two conflicting goals, promotions of fairness and profit. Experimental results on a real-world dataset are provided as well, which confirm our theoretical predictions.

1 Introduction

Rideshares such as Uber and Lyft have received significant attention among research communities of computer science, operations research, and business, to name a few. One main research topic is the matching policy design of pairing drivers and riders, see, *e.g.,* [Danassis *et al.*, 2019; Ashlagi *et al.*, 2019; Bei and Zhang, 2018; Dickerson *et al.*, 2018; Zhao *et al.*, 2019]. Most of the current work focuses on either the promotion of system efficiency or that of users' satisfaction or both.

In this paper, we study the fairness among rideshare drivers. There are several reports showing the earning gap among drivers based on their demographic factors such as age, gender and race, see, *e.g.,* [Cook *et al.*, 2018; Rosenblat *et al.*, 2016]. In particular, [Hinchliffe, 2017] has reported that "Black Uber and Lyft drivers earned $13.96 an hour compared to the $16.08 average for all other drivers" and "Women drivers reported earning an average of $14.26 per hour, compared to $16.61 for men". The wage gap among drivers from different demographic groups is partially due to the discriminative cancellations from riders, which can be well spotted especially during off-peak hours when the number of riders is comparable or even less than that of drivers. Note that in rideshares like Uber and Lyft, after a driver accepts a rider: (1) all sensitive information of the driver such as name and photo will be accessible to the rider and (2) riders can cancel the driver for the first two minutes free of charge [Dough, 2019]. This makes the discriminative cancellations from riders technically possible and economically worry-free.

We aim to address the income disparity among drivers due to discriminative cancellations from riders and its tradeoff with system efficiency. Note that the two goals, promoting the group-level income equality among drivers and the system efficiency, are somewhat conflicting. Consider the off-peak hours for example, when riders are kinds of scarce resources. To maximize the system efficiency, rideshares like Uber should please riders by assigning them to their "favorite" drivers. This can effectively reduce any possible cancellations from riders and thus, minimize the risk of driving away riders to other rivals like Lyft. This measure, however, will offer those drivers "popular" among riders much more chances of getting orders than others and as a result, hurt the group-level income equality greatly.

In this paper, we propose two parameterized matching policies, which can smoothly tradeoff the above two goals with provable performances. We adopt the online-matching based model to capture the dynamics in rideshare, as commonly used before [Dickerson *et al.*, 2018; Zhao *et al.*, 2019]. Assume a bipartite graph $G = (U, V, E)$ where U and V represent the sets of *types* of offline drivers and online requests, respectively. Each driver type represents a specific demographic group (defined by gender, age, race, *etc.*) with a given location, while each request type represents a specific demographic group with a given starting and ending location. There is an edge $f = (u, v)$ if the driver (of type) u is capable of serving the request (of type) v[1] (*e.g.,* the distance

[1]For simplicity, we refer to a driver of type u and a request of

between them is below a given threshold). The online phase consists of T rounds and in each round, a request $v \in V$ arrives dynamically. Upon its arrival an *immediate and irrevocable* decision is required: either reject v or assign it to a neighboring driver in U. We assume each u has a matching capacity of $B_u \in \mathbb{Z}^+$, which captures the number of drivers belonging to the type u. Additionally, we have the following key assumptions in the model.

Arrivals of online requests. We consider a finite time horizon T (known to the algorithm). For each time-step or round $t \in [T] \doteq \{1, 2, \ldots, T\}$, a request of type v will be sampled (or v arrives) from a known distribution $\{q_v\}$ such that $\sum_{v \in V} q_v = 1$. Note that the sampling process is independent and identical across the online T rounds. For each v, let $r_v = T \cdot q_v$, which is called the *arrival rate* of request v with $\sum_{v \in V} r_v = T$. Our arrival assumption is commonly called the *known identical independent distributions* (KIID). This is mainly inspired from the fact that we can often learn the arrival distribution from historical logs [Yao *et al.*, 2018; Li *et al.*, 2018]. KIID is widely used in many practical applications of online matching markets including rideshare and crowdsourcing [Zhao *et al.*, 2019; Dickerson *et al.*, 2018; Singer and Mittal, 2013; Singla and Krause, 2013].

Edge existence probabilities. Each edge $f = (u, v)$ is associated with an existence probability $p_f \in (0, 1]$, which captures the statistical acceptance rate of a request of type v toward a driver of type u. The random process goes as follows. Once we assign u to v, we observe an immediate random outcome of the existence, which is present (*i.e.,* v accepts u) with probability p_f and not (v cancels u) otherwise. We assume that (1) the randomness associated with the edge existence is independent across all edges; (2) the values $\{p_f\}$ are given as part of the input. The first assumption is motivated by individual choice and the second from the fact that historical logs can be used to compute such statistics with high precision.

Patience of requests. Each request v is associated with patience $\Delta_v \in \mathbb{Z}^+$, which captures an upper bound of unsuccessful assignments the request v can tolerate before leaving the platform. Under patience constraints, we can dispatch each request v to at most Δ_v different drives. Observe that we cannot broadcast v to a set of at most Δ_v different drives simultaneously. Instead, we should assign v to at most Δ_v distinct drives (maybe of the same type though) *in a sequential manner* until either v accepts one or v leaves the system after running out of patience. We refer to this as the *online probing process* (OPP). Note that OPP starts immediately after a request v arrives if v not rejected by the algorithm, and ends within one single round before the next request arrives.

We say an assignment $f = (u, v)$ is *successful* if u is assigned to v, and v *accepts* u which occurs with probability p_f. Assume that the platform will gain a profit w_f from a successful assignment $f = (u, v)$ (we call a match then). For a given policy ALG, let $\mathcal{M}$ be the set of (possibly random) successful assignments; we interchangeably use the term *matching* to denote this set $\mathcal{M}$. Inspired by the work of [Nanda *et al.*, 2019; Lesmana *et al.*, 2019], we define two objectives, namely *profit* and *fairness*, which capture the system efficiency and group-level income equality among drivers, respectively.

Profit: The expected total profit over all matches obtained by the platform, which is defined as $\mathbb{E}[\sum_{f \in \mathcal{M}} w_f]$.

Fairness: Let $\mathcal{M}_u$ be the set of edges in $\mathcal{M}$ incident to u. Define the fairness achieved by ALG over all driver types as $\min_{u \in U} \frac{\mathbb{E}[|\mathcal{M}_u|]}{B_u}$.

1.1 Preliminaries and Main Contributions

Competitive ratio. The competitive ratio is a commonly-used metric to evaluate the performance of online algorithms. Consider an online maximization problem for example. Let $\mathrm{ALG}(\mathcal{I}) = \mathbb{E}_{I \sim \mathcal{I}}[\mathrm{ALG}(I)]$ denote the expected performance of ALG on an input $\mathcal{I}$, where the expectation is taken over the random arrival sequence I. Let $\mathrm{OPT}(\mathcal{I}) = \mathbb{E}[\mathrm{OPT}(I)]$ denote the expected *offline optimal*, where $\mathrm{OPT}(I)$ refers to the optimal value after we observe the full arrival sequence I. Then, competitive ratio is defined as $\min_{\mathcal{I}} \frac{\mathrm{ALG}(\mathcal{I})}{\mathrm{OPT}(\mathcal{I})}$. It is a common technique to use an LP to upper bound the $\mathrm{OPT}(\mathcal{I})$ (called the benchmark LP) and hence get a valid lower bound on the target competitive ratio. In our paper, we conduct online competitive ratio analysis on both objectives.

Main Contributions

Our contributions can be summarized in the following three aspects. First, we propose a new online-matching based model to address the income inequality among drivers from different demographic groups and its trade-off with the system efficiency in rideshare. Second, we present a robust theoretical analysis for our model. We first construct a bi-objective linear program (LP-(1) and LP-(2)), which is proved to offer valid upper bounds for the respective maximum profit and fairness in the offline optimal. Then, we propose LP-based parameterized online algorithms WarmUp and AttenAlg with provable performances on both objectives. We say an online algorithm achieves an (α, β)-competitive ratio if it achieves competitive ratios α and β on the profit and fairness against benchmarks LP-(1) and LP-(2), respectively. Results in Theorems 2 and 3 suggest that AttenAlg can achieve a nearly optimal ratio on each single objective either fairness or profit, though there is some space of improvement left for the summation of both ratios.

Theorem 1. WarmUp(α, β) *achieves a competitive ratio at least* $\left(\alpha \cdot \frac{1-1/e}{2}, \beta \cdot \frac{1-1/e}{2}\right)$ *simultaneously on the profit and fairness for any* $\alpha, \beta > 0$ *with* $\alpha + \beta \leq 1$.

Theorem 2. AttenAlg(α, β) *achieves a competitive ratio at least* $\left(\alpha \cdot \frac{e-1}{e+1}, \beta \cdot \frac{e-1}{e+1}\right) \sim (0.46 \cdot \alpha, 0.46 \cdot \beta)$ *simultaneously on the profit and fairness for any* $\alpha, \beta > 0$ *with* $\alpha + \beta \leq 1$.

Theorem 3. *No algorithm can achieve an* (α, β)*-competitive ratio simultaneously on the profit and fairness with* $\alpha + \beta > 1$ *or* $\alpha > 0.51$ *or* $\beta > 0.51$ *using* LP-(1) *and* LP-(2) *as benchmarks.*

Last, we test our model and algorithms on a real dataset collected from a large on-demand taxi dispatching platform.

type v directly as a driver u and request v when the context is clear.

Experimental results confirm our theoretical predictions and demonstrate the flexibility of our algorithms in tradeoffing the two conflicting objectives and their efficiency compared to natural heuristics.

2 Related Work

Here is a few recent work addressing the fairness issue in rideshares. [Sühr *et al.*, 2019] proposed two notions of amortized fairness for fair distribution of income among drivers, one is related to absolute income equality, while the other is averaged income equality over active time. [Lesmana *et al.*, 2019] considered nearly the same two objectives as proposed in this paper. Note that both of the aforementioned work considered an essential offline setting in the way that all arrivals of online requests are known in advance by considering a short time window. Additionally, both ignore the potential cancellations from riders, and assume each rider will accept the assigned driver surely (*i.e.,* all $p_f = 1$). [Nanda *et al.*, 2019] studied an interesting "dual" setting to us. They focused on the peak hours and examined the fairness on the rider side due to discriminative cancellations from drivers.

Our model technically belongs to a more general optimization paradigm, called *Multi-Objective Optimization*. Here are a few theoretical work which studied the design of approximation or online algorithms to achieve a bi-criterion approximation and/or online competitive ratios, see, *e.g.,* [Ravi *et al.*, 1993; Korula *et al.*, 2013; Aggarwal *et al.*, 2014; Esfandiari *et al.*, 2016]. The work of [Brubach *et al.*, 2020; Fata *et al.*, 2019] have the closest setting to us: each edge has an independent existence probability and each vertex from the offline and/or online side has a patience constraint on it. However, all investigated one single objective: maximization of the total profit over all matched edges.

3 Valid Benchmarks for Profit and Fairness

We first present our benchmark LPs and then an LP-based parameterized algorithm. For each edge $f = (u, v)$, let x_f be the expected number of probes on edge f (*i.e.,* assignments of v to u but not necessarily matches) in the offline optimal. For each u (v), let E_u (E_v) be the set of neighboring edges incident to u (v). Consider the following bi-objective LP.

$$\max \ \sum_f w_f x_f p_f \tag{1}$$

$$\max \min_{u \in U} \ \frac{\sum_{f \in E_u} x_f p_f}{B_u} \tag{2}$$

$$\text{s.t.} \sum_{f \in E_u} x_f p_f \le B_u \quad \forall u \in U \tag{3}$$

$$\sum_{f \in E_v} x_f \le \Delta_v \cdot r_v \quad \forall v \in V \tag{4}$$

$$\sum_{f \in E_v} x_f p_f \le r_v \quad \forall v \in V \tag{5}$$

$$0 \le x_f \le r_v \quad \forall f \in E_v \tag{6}$$

Let LP-(1) and LP-(2) denote the two LPs with the respective objectives (1) and (2), each with Constraints (3), (4), (5), (6). Note that we can rewrite Objective (2) as a linear one like $\max \eta$ with additional linear constraints as $\eta \le \frac{\sum_{f \in E_u} x_f p_f}{B_u}$ for all $u \in U$. For presentation convenience, we keep the current compact version. The validity of LP-(1) and LP-(2) as benchmarks for our two objectives can be seen in the following lemma.

Lemma 1. LP-(1) *and* LP-(2) *are valid benchmarks for the two respective objectives, profit and fairness. In other words, the optimal values to* LP-(1) *and* LP-(2) *are valid upper bounds for the expected profit and fairness achieved by the offline optimal, respectively.*

Proof. We can verify that objective functions (1) and (2) each captures the exact expected profit and fairness achieved by the offline optimal by the linearity of expectation. To prove the validity of the benchmark for each objective, it suffices to show the feasibility of all constraints for any given offline optimal. Recall that for each edge f, x_f denotes the expected number of probes on f (*i.e.,* assignments of u to v but not necessarily matches) in the offline optimal. Constraint (3) is valid since each driver u has a matching capacity of B_u. Note that the expected arrivals of v during the whole online phase is r_v and v can be probed at most Δ_v times upon each online arrival. Thus, the expected number of total probes and matches over all edges incident to v should be no more than $r_v \Delta_v$ and r_v, respectively. This rationalizes Constraints (4) and (5). The last constraint is valid, since for each edge, the expected number of probes should be no more than that of arrivals. Therefore, we justify the feasibility of all constraints for any given offline optimal. □

4 LP-based Parameterized Algorithms

The following lemma suggests that for any online algorithm ALG, the worst-case scenario (*i.e.,* the instance on which ALG achieves the lowest competitive ratio) arrives when each driver type has a unit matching capacity. The proof is deferred to the full version [2].

Lemma 2. *Let* ALG *be an online algorithm achieving an* (α, β)*-competitive ratio on instances with unit matching capacity (*i.e., *all* B_u*=1). We can twist* ALG *to* ALG$'$ *such that* ALG$'$ *achieves at least an* (α, β)*-competitive ratio on instances with general integral matching capacities.*

From Lemma 2, we assume *unit capacity for all driver types* throughout this paper w.l.o.g. In the following, we will present a warm-up algorithm (WarmUp) and then another refined algorithm (AttenAlg), which can be viewed as a polished version of WarmUp with simulation-based attenuation techniques. The main idea of AttenAlg is primarily inspired by the work [Brubach *et al.*, 2020]. Both WarmUp and AttenAlg invoke the following dependent rounding techniques (denoted by GKPS) introduced by [Gandhi *et al.*, 2006]. For simplicity, we state a simplified version of GKPS tailored to star graphs which suffices in our paper.

Recall that E_v is the set of edges incident to v in the compatible graph G. GKPS is such a dependent rounding technique that takes as input a fractional vector $\mathbf{z} = \{z_f, f \in E_v, z_f \in [0, 1]\}$ on E_v, and output a random binary vector $\mathbf{Z} = \{Z_f, f \in E_v\}$, which satisfied the following properties. (1) **Marginal distribution**: $\mathbb{E}[Z_f] = z_f$ for all $f \in E_v$; (2) **Degree preservation**: $\Pr[\sum_{f \in E_v} Z_f \le \sum_{f \in E_v} z_f] = 1$;

[2] https://tinyurl.com/yd9ucv8p

Algorithm 1 Sub-Routine $\mathsf{SR}(\mathbf{z})$

```
Apply GKPS to the fractional vector z and let Z be the
  random binary vector output.
Choose a random permutation π over E_v.
Follow the order π to process each f = (u, v) ∈ E_v until
  v is matched:
    if Z_f = 1 and u is available then
        Probe the edge f (i.e., assign v to u).
    else
        Skip to the next one.
```

(3) **Negative correlation**: For any pair of edges $f, f' \in E_v$, $\mathbb{E}[Z_f = 1 | Z_{f'} = 1] \le z_f$.

Throughout this section, we assume (1) $\mathbf{x}^* = \{x_f^*\}$ and $\mathbf{y}^* = \{y_f^*\}$ are optimal solutions to LP-(1) and LP-(2) respectively; (2) (α, β) are two given parameters with $0 \le \alpha, \beta \le 1, \alpha + \beta \le 1$; (3) $B_u = 1$ for all u from Lemma 2; (4) $\mathbf{x}^v = \{x_f^*/r_v, f \in E_v\}$ and $\mathbf{y}^v = \{y_f^*/r_v, f \in E_v\}$, which are scaled solutions from $\mathbf{x}^*$ and $\mathbf{y}^*$ respectively restricted on E_v. Note that from Constraints (4), (5) and (6), we have that $\mathbf{x}^v$ and $\mathbf{y}^v$ are two fractional solutions on E_v and each has a total sum at most Δ_v.

4.1 The First Algorithm $\mathsf{WarmUp}(\alpha, \beta)$

Let an online vertex v arrive at t. Our job is to probe at most Δ_v edges in E_v until v is matched. Let $\mathbf{z}$ be a given fractional solution on E_v. $\mathsf{WarmUp}(\alpha, \beta)$ invokes the following procedures (denoted by $\mathsf{SR}(\mathbf{z})$) as a subroutine during each online round: it first selects a set $\mathcal{S}_v$ of at most Δ_v edges from E_v in a random way guided by a given fractional vector $\mathbf{z}$ on E_v and then follows a random order to process all edges in $\mathcal{S}_v$ one by one. The details of SR are stated in Algorithm 1.

Based on SR, the main idea of $\mathsf{WarmUp}(\alpha, \beta)$ is as simple as follows: each round when an online vertex v arrives, it invokes $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ with probabilities α and β respectively. Recall that $\mathbf{x}^v$ and $\mathbf{y}^v$ are the scaled optimal solutions to LP-(1) and LP-(2) restricted to E_v, each has a total sum at most Δ_v. Thus, when we run $\mathsf{SR}(\mathbf{x}^v)$ or $\mathsf{SR}(\mathbf{y}^v)$ after v arrives online, we will probe at most Δ_v edges incident to v since the final rounded binary vector has at most Δ_v ones due to Property of Degree Preservation in the dependent rounding. The details of $\mathsf{WarmUp}(\alpha, \beta)$ are as follows.

Algorithm 2 $\mathsf{WarmUp}(\alpha, \beta)$

```
Let v arrive at time t.
With probability α, run SR(x^v).
With probability β, run SR(y^v).
With probability 1 − α − β, reject v.
```

We conduct an edge-by-edge analysis. It would suffice to show that each f is probed with probability at least $x_f^* \cdot \alpha \cdot (1 - 1/e)/2$ and $y_f^* \cdot \beta \cdot (1 - 1/e)/2$ in $\mathsf{WarmUp}(\alpha, \beta)$. Then by linearity of expectation, we can get Theorem 1. Focus on a given u and a time $t \in [T]$. Let $\mathsf{SF}_{u,t}$ be the event that u is *available* at (the beginning of) t.

Lemma 3. *For any given u and $t \in [T]$, we have* $\Pr[\mathsf{SF}_{u,t}] \ge \left(1 - \frac{1}{T}\right)^{t-1}$.

Proof. Recall that we assume w.l.o.g. that each B_u=1 due to Lemma 2. For each given $\ell < t$ and $f = (u, v) \in E_u$, let $X_{f,\ell}$ indicate if v arrives at time t; $Y_{f,\ell}$ indicate if f is probed during round ℓ; $Z_{f,\ell}$ indicate if f is present when probed. Note that in each subroutine of $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ after v arrives, f will be probed only when the final rounded vector has the entry one on f. Thus we claim that $\mathbb{E}[Y_{f,\ell}] \le \alpha x_f^*/r_v + \beta y_f^*/r_v$ due to Property of Marginal Distribution in dependent rounding and statements of $\mathsf{WarmUp}(\alpha, \beta)$. Thus,

$$
\begin{aligned}
&\Pr[\mathsf{SF}_{u,t}] = \prod_{\ell<t} \Pr\left[\sum_{f \in E_u} X_{f,\ell} Y_{f,\ell} Z_{f,\ell} = 0\right] \\
&= \prod_{\ell<t} \left(1 - \Pr\left[\sum_{f \in E_u} X_{f,\ell} Y_{f,\ell} Z_{f,\ell} \ge 1\right]\right) \\
&= \prod_{\ell<t} \left(1 - \sum_{f \in E_u} \frac{r_v}{T}\left(\alpha \frac{x_f^*}{r_v} + \beta \frac{y_f^*}{r_v}\right) p_f\right) \\
&= \prod_{\ell<t} \left(1 - \frac{1}{T} \sum_{f \in E_u} \left(\alpha x_f^* p_f + \beta y_f^* p_f\right)\right) \\
&\ge \left(1 - \frac{1}{T}\right)^{t-1}
\end{aligned}
$$

□

Now assume $\mathsf{SF}_{u,t}$ occurs (*i.e.,* u is available at t). Consider a given $f = (u, v)$ and let $\mathbf{1}_{f,t}$ indicate f is probed during round t in $\mathsf{WarmUp}(\alpha, \beta)$. Notice that $\mathbf{1}_{f,t}$ occurs if (1) v arrives at time t and (2) f is probed either in $\mathsf{SR}(\mathbf{x}^v)$ or $\mathsf{SR}(\mathbf{y}^v)$.

Lemma 4. $\Pr[\mathbf{1}_{f,t} | \mathsf{SF}_{u,t}] \ge \frac{\alpha x_f^*}{2T}, \Pr[\mathbf{1}_{f,t} | \mathsf{SF}_{u,t}] \ge \frac{\beta y_f^*}{2T}$.

Proof. We focus on the first inequality and try to show that f is probed at t in $\mathsf{SR}(\mathbf{x}^v)$ with probability at least $\frac{\alpha x_f^*}{2T}$ (including the probability of its online arrival). Observe that events v arrives at time t and $\mathsf{WarmUp}(\alpha, \beta)$ runs the subroutine $\mathbf{x}^v$ both happen with probability $\frac{\alpha r_v}{T}$. Let $\mathbf{X}^v$ be the rounded binary vector from $\mathbf{x}^v$ and we use X_f^v to denote its entry on f. Let $E_{v,\neg f}$ be the set of edges in E_v excluding $f = (u, v)$. For each $f' \in E_{v,\neg f}$, let $Y_{f'}$ indicate if f' falls before f in the random order π and $Z_{f'}$ indicate if f' is present when probed. Thus we have

$$
\Pr[\mathbf{1}_{f,t} | \mathsf{SF}_{u,t}] \tag{7}
$$

$$
\ge \frac{\alpha r_v}{T} \Pr[X_f^v = 1] \Pr\left[\sum_{f' \in E_{v,\neg f}} X_{f'}^v Y_{f'} Z_{f'} = 0 | X_f^v = 1\right] \tag{8}
$$

$$
= \frac{\alpha r_v}{T} \frac{x_f^*}{r_v} \left(1 - \Pr\left[\sum_{f' \in E_{v,\neg f}} X_{f'}^v Y_{f'} Z_{f'} \ge 1 | X_f^v = 1\right]\right) \tag{9}
$$

$$
\ge \frac{\alpha x_f^*}{T} \left(1 - \mathbb{E}\left[\sum_{f' \in E_{v,\neg f}} X_{f'}^v Y_{f'} Z_{f'} | X_f^v = 1\right]\right) \tag{10}
$$

$$
\ge \frac{\alpha x_f^*}{T} \left(1 - \sum_{f' \in E_{v,\neg f}} \mathbb{E}\left[X_{f'}^v Y_{f'} Z_{f'} | X_f^v = 1\right]\right) \tag{11}
$$

$$
\ge \frac{\alpha x_f^*}{T} \left(1 - \sum_{f' \in E_{v,\neg f}} \frac{x_{f'}^*}{r_v} \frac{p_f}{2}\right) \tag{12}
$$

$$
\ge \frac{\alpha x_f^*}{T} \frac{1}{2}. \tag{13}
$$

Inequality (10) follows from Markov's inequality. Inequality (12) is due to these two observations: (1) $\mathbb{E}[X_{f'}^v | X_f^v =$

$1] \leq x^*_f/r_v$ due to negative correlation in dependent rounding and (2) $\mathbb{E}[Y_{f'}] = 1/2$, $\mathbb{E}[Z_{f'}] = p_f$. Inequality (13) follows from the fact $\sum_{f' \in E_v} x^*_{f'} p_{f'} \leq r_v$ due to Constraint (5). Following a similar analysis, we can prove the second part. □

Now we have all ingredients to prove the main Theorem 1.

Proof. Consider a given $f = (u, v) \in E$, let $\kappa^{\mathbf{x}}_f$ and $\kappa^{\mathbf{y}}_f$ be the expected number of *successful* probes of f in $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ respectively. Here a probe of $f = (v, u)$ is successful iff u is available when we assign v to u (but no necessarily means f is present).

$$\kappa^{\mathbf{x}}_f \geq \sum_{t=1}^{T} \Pr[\mathsf{SF}_{u,t}] \Pr[\mathbf{1}_{f,t} | \mathsf{SF}_{u,t}]$$
$$\geq \sum_{t=1}^{T} \left(1 - \frac{1}{T}\right)^{t-1} \frac{\alpha x^*_f}{2T} \sim \frac{\alpha x^*_f (1 - 1/e)}{2}$$

The last term is obtained after taking $T \to \infty$. Similarly, we can show that $\kappa^{\mathbf{y}}_f \geq \frac{\beta y^*_f (1-1/e)}{2}$.

Let $\mathsf{Profit}(\alpha, \beta)$ be the expected total profit obtained by $\mathsf{WarmUp}(\alpha, \beta)$. By linearity of expectation, we have $\mathsf{Profit}(\alpha, \beta) \geq \frac{(1-1/e)\alpha}{2} \sum_{f \in E} x^*_f p_e w_e$. From Lemma 1, we know that the expected profit in offline optimal is upper bounded by $\sum_{f \in E} x^*_f p_e w_e$. Thus we claim that $\mathsf{WarmUp}(\alpha, \beta)$ achieves a ratio at least $\alpha(1 - 1/e)/2$ on the profit. Similarly, we can argue that $\mathsf{WarmUp}(\alpha, \beta)$ achieves a ratio at least $\beta(1 - 1/e)/2$ on the fairness. □

4.2 The Second Algorithm $\mathsf{AttenAlg}(\alpha, \beta)$

Inspired by [Brubach *et al.*, 2020], we can improve at least the theoretical performance of WarmUp with attenuation techniques applied to edges and (offline) vertices. The motivation behind is very simple. Note that edges in E_v are competing for each other since we have to stop probing whenever v is matched. Thus, attenuating those edges which win the higher chance of probing over others can potentially boost the worst-case performance.

Let $\{\gamma_t, \mu_t | t \in [T]\}$ be such a series that is defined as $\gamma_1 = 1, \mu_t = 1 - \gamma_t/2, \gamma_{t+1} = \gamma_t(1 - \mu_t/T)$, and $E_{v,t}$ be the set of available edges $f = (u, v) \in E_v$ at time t (*i.e.,* u is available at t). The formal description of AttenAlg is stated in Algorithm 3. We defer the proofs of Theorems 2 and 3 to the full version [2].

Algorithm 3 $\mathsf{AttenAlg}(\alpha, \beta)$

```
1 for t = 1, 2, ..., T do
2     Apply vertex-attenuation such that each u ∈ U is
        available at t with probability equal to γ_t.
3     Let v arrive at time t.
4     With probability α,
5         Run SR(x^v). Apply edge-attenuation such that
            each edge f ∈ E_{v,t} is probed in SR(x^v) with
            probability equal to μ_t x*_f / r_v.
6     With probability β,
7         Run SR(y^v). Apply edge-attenuation such that
            each edge f ∈ E_{v,t} is probed in SR(y^v) with
            probability equal to μ_t y*_f / r_v.
8     With probability 1 − α − β, reject v.
```

5 Experiments

5.1 Data Preprocessing

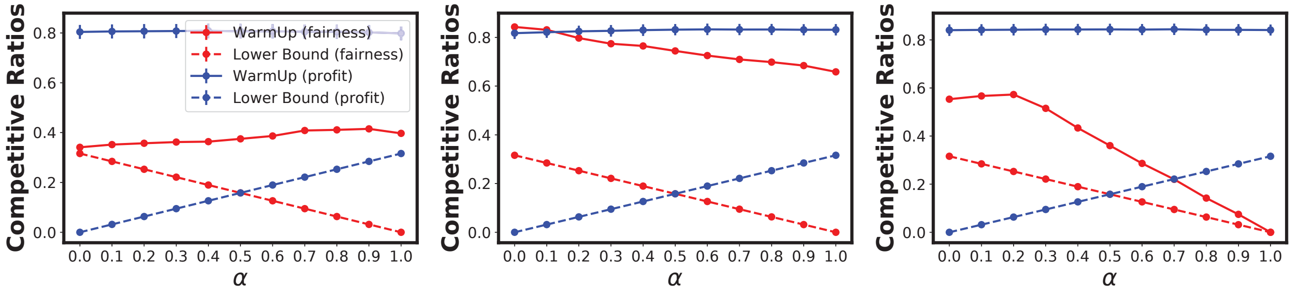
We use the New York City yellow cabs dataset[3], which is collected during the year of 2013. Although the demographics of the drivers and riders are not recorded in the original dataset, we synthesize the racial demographics for riders and drivers in a similar way to [Nanda *et al.*, 2019]. To simplify the demonstration, we consider a single demographic factor of the race only, which takes two possible options between "disadvantaged" (D) or "advantaged" (A). We set the ratio of D to A to be $1 : 2$ among riders, which roughly matches the racial demographics of NYC [Review, 2019]. Similarly, we set the ratio of D to A among drivers to be $1 : 2$ [Hall and Krueger, 2017]. The acceptance rates among the four possible driver-rider pairs (based on race status only), (A,A), (A,D), (D,A), (D,D), are set to be $0.6, 0.1, 0.1$ and 0.3, respectively. These probabilities are then scaled up by a factor η such that $p_f = \eta + (1 - \eta) \cdot p_f$. In our experiments we set $\eta = 0.5$. Note that we can apply our model straightforwardly to the case when the real-world distribution of $\{p_f\}$ values is known or can be learned. We collect records during the off-peak period when a lot of drivers are on the road while the requests are relatively lower than peak hours. On January 31, $20,701$ trips were completed in the off-peak hour (4–5 PM), compared to $35,109$ trips in the peak hour (7–8 PM). We focus on longitude and latitude ranging from $(-73, -75)$ and $(40.4, 40.95)$ respectively. We partition the area into 40×11 grids with equal size.

We construct the compatibility graph $G = (U, V, E)$ as follows. Each $u \in U$ represents a driver type which has attributes of the starting location and race. Each $v \in V$ represents a request type which has attributes of the starting location, ending location, and race. We downsample from all driver and request types such that $|U| = 57$ and $|V| = 134$. For each driver type u, we assign its capacity B_u with a random value uniformly sampled from $[1, B]$ where we vary $B \in \{10, 15, 25\}$. For each request of type v, we sample a random patience value Δ_v uniformly from $\{1, 2\}$ and a random arrival rate $r_v \sim \mathcal{N}(5, 1)$ (Normal distribution), and then set $T = \sum_{v \in V} r_v$. We add an edge $f = (u, v)$ if the Manhattan distance between starting location of request type v and the location of driver type u is not larger than 1. The profit w_f for each f is defined as the normalized trip length of the request type v such that $0 \leq w_f \leq 1$.

Algorithms. We test the $\mathsf{WarmUp}(\alpha, \beta)$ with $\alpha + \beta = 1$ against two natural heuristic baselines, namely Greedy-P (short for Greedy-Profit) and Greedy-F (short for Greedy-Fairness)[4]. Suppose a request type of v arrives at time t. Recall that E_v is the set of neighboring edges incident to v (*i.e.,* the set of assignments feasible to v). Let $E'_v \subseteq E_v$ be the set of *available* assignments $f = (u, v)$ such that there

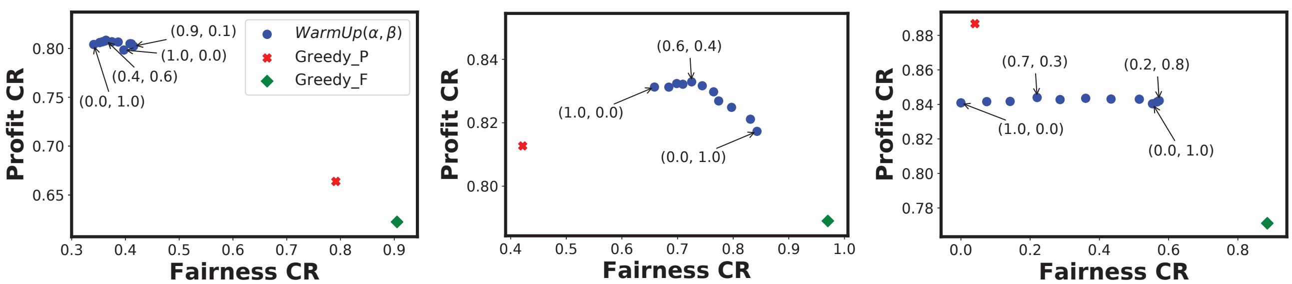
[3]http://www.andresmh.com/nyctaxitrips/

[4]A future direction is to consider a hybrid version of Greedy-P and Greedy-F, which optimizes the two objectives simultaneously.



(a) $B = 10$ (b) $B = 15$ (c) $B = 25$

Figure 1: Competitive ratios for profit and fairness with different values of α and β with $\alpha + \beta = 1$.



(a) $B = 10$ (b) $B = 15$ (c) $B = 25$

Figure 2: Performance comparisons with Greedy_P and Greedy_F.

exists at least one drive of type u at t. For Greedy-P, it will repeat greedily selecting an available assignment $f \in E'_v$ with the maximum weight $w_f p_f$ over E'_v (breaking ties arbitrarily) until either v accepts a driver or v runs out of patience. In contrast, Greedy-F will repeat greedily selecting an available $f = (u^*, v) \in E'_v$ with u^* having the least matching rate before either v accepts a driver or leaves the system. Note that we use LP-(1) and LP-(2) as the default benchmarks for profit and fairness, respectively.

5.2 Results and Discussions

Figure 1 shows the results of competitive ratios for the proposed algorithm with different values of α with ($\beta = 1 - \alpha$). We can observe that the profit and fairness competitive ratios of WarmUp always stay above the theoretical lower bounds (in dotted lines), as predicted in Theorem 1. The gaps between performances and lower bounds suggest that theoretical worst scenarios occur rarely in the real world. Note that when $B = 25$ and $\alpha = 1$ as shown in Figure 1(c), the lower bound is tight and matches the fairness performance.

Figure 2 shows the profit and fairness performances of WarmUp compared to Greedy-P and Greedy-F. Here are a few interesting observations. (1) As for profit, Greedy-P can always beat Greedy-F but not necessarily for WarmUp. The advantage of Greedy-P over WarmUp becomes more apparent when B is large and less when B is small. Note that in our experiment, the expected total number of arrivals of riders is fixed and therefore, B directly controls the degree of imbalance between drivers and riders. When B is larger, we have more available drivers compared to riders and thus, Greedy-P will outperform all the rest for profit. When B is small, however, we really need to carefully design the policy to boost profit. That's why WarmUp becomes dominant. (2) As for fairness, Greedy-F seemingly can always dominate the rest, though WarmUp shows high flexibility in the fairness performance. WarmUp shows a relatively low sensitivity toward the first parameter α for profit while high sensitivity toward the second parameter β for fairness: the latter becomes particularly obvious when B is large.

6 Conclusion

In this paper, we present a flexible approach for matching requests to drivers to balance the two conflicting goals, maximizations of income equality among all rideshare drivers and the total revenue earned by the system. Our proposed approach allows the policy designer to specify how fair and how profitable they want the system to be via two separate parameters. Extensive experimental results on the real-world dataset show that our proposed approaches not only are far above the theoretical lower bounds but also can smoothly tradeoff the two objectives between the two natural heuristics. Our work opens a few directions for future research. The most direct one is to shorten the gap between the sum of ratios of profit and fairness achieved by AttenAlg (which is 0.46). It will be interesting to give a tighter online analysis than what are presented here or offer a sharper hardness result which suggests the sum of the two ratios should be much lower than 1.

Acknowledgments

Pan Xu was partially supported by NSF CRII Award IIS-1948157. The authors would like to thank the anonymous reviewers for their helpful feedback.

References

[Aggarwal *et al.*, 2014] Gagan Aggarwal, Yang Cai, Aranyak Mehta, and George Pierrakos. Biobjective online bipartite matching. In *International Conference on Web and Internet Economics*, pages 218–231. Springer, 2014.

[Ashlagi *et al.*, 2019] Itai Ashlagi, Maximilien Burq, Chinmoy Dutta, Patrick Jaillet, Chris Sholley, and Amin Saberi. Edge weighted online windowed matching. In *ACM EC'19*, pages 729–742, 2019.

[Bei and Zhang, 2018] Xiaohui Bei and Shengyu Zhang. Algorithms for trip-vehicle assignment in ride-sharing. In *AAAI '18*, pages 3–9, 2018.

[Brubach *et al.*, 2020] Brian Brubach, Karthik Sankararaman, Aravind Srinivasan, and Pan Xu. Attenuate locally, win globally: Attenuation-based frameworks for online stochastic matching with timeouts. *Algorithmica*, 82(1):64–87, 2020.

[Cook *et al.*, 2018] Cody Cook, Rebecca Diamond, Jonathan Hall, John List, and Paul Oyer. The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Technical report, National Bureau of Economic Research, 2018.

[Danassis *et al.*, 2019] Panayiotis Danassis, Marija Sakota, Aris Filos-Ratsikas, and Boi Faltings. Putting ridesharing to the test: Efficient and scalable solutions and the power of dynamic vehicle relocation. *arXiv preprint arXiv:1912.08066*, 2019.

[Dickerson *et al.*, 2018] John Dickerson, Karthik Abinav Sankararaman, Aravind Srinivasan, and Pan Xu. Assigning tasks to workers based on historical data: Online task assignment with two-sided arrivals. In *AAMAS '18*, pages 318–326, 2018.

[Dough, 2019] Dough. Uber fees: How to avoid the cancellation fee, cleaning fee, and more. https://www.ridesharingdriver.com/uber-fees-cancellation-booking-cleaning-fees/, 2019. Accessed: 2019-12-27.

[Esfandiari *et al.*, 2016] Hossein Esfandiari, Nitish Korula, and Vahab Mirrokni. Bi-objective online matching and submodular allocations. In *Advances in Neural Information Processing Systems*, pages 2739–2747, 2016.

[Fata *et al.*, 2019] Elaheh Fata, Will Ma, and David Simchi-Levi. Multi-stage and multi-customer assortment optimization with inventory constraints. *Available at SSRN 3443109*, 2019.

[Gandhi *et al.*, 2006] Rajiv Gandhi, Samir Khuller, Srinivasan Parthasarathy, and Aravind Srinivasan. Dependent rounding and its applications to approximation algorithms. *Journal of the ACM (JACM)*, 53(3):324–360, 2006.

[Hall and Krueger, 2017] Jonathan Hall and Alan Krueger. An analysis of the labor market for uber's driver-partners in the united states. *Ilr Review*, 2017.

[Hinchliffe, 2017] Emma Hinchliffe. Yes, there's a wage gap for uber and lyft drivers based on age, gender and race. https://mashable.com/2017/01/18/uber-lyft-wage-gap-rideshare/, 2017. Accessed: 2019-12-27.

[Korula *et al.*, 2013] Nitish Korula, Vahab Mirrokni, and Morteza Zadimoghaddam. Bicriteria online matching: Maximizing weight and cardinality. In *International conference on web and internet economics*, pages 305–318. Springer, 2013.

[Lesmana *et al.*, 2019] Nixie Lesmana, Xuan Zhang, and Xiaohui Bei. Balancing efficiency and fairness in on-demand ridesourcing. In *Advances in Neural Information Processing Systems*, pages 5310–5320, 2019.

[Li *et al.*, 2018] Yaguang Li, Kun Fu, Zheng Wang, Cyrus Shahabi, Jieping Ye, and Yan Liu. Multi-task representation learning for travel time estimation. KDD '18, pages 1695–1704, 2018.

[Nanda *et al.*, 2019] Vedant Nanda, Pan Xu, Karthik Abinav Sankararaman, John P Dickerson, and Aravind Srinivasan. Balancing the tradeoff between profit and fairness in rideshare platforms during high-demand hours. *arXiv preprint arXiv:1912.08388*, 2019.

[Ravi *et al.*, 1993] Ramamoorthi Ravi, Madhav Marathe, Sekharipuram Ravi, Daniel Rosenkrantz, and Harry Hunt. Many birds with one stone: Multi-objective approximation algorithms. In *STOC '93*, pages 438–447. Citeseer, 1993.

[Review, 2019] World Population Review. New york city population. http://worldpopulationreview.com/us-cities/new-york-city-population/, 2019. Accessed: 2020-01-12.

[Rosenblat *et al.*, 2016] Alex Rosenblat, Karen Levy, Solon Barocas, and Tim Hwang. Discriminating tastes: Customer ratings as vehicles for bias. *Available at SSRN 2858946*, 2016.

[Singer and Mittal, 2013] Yaron Singer and Manas Mittal. Pricing mechanisms for crowdsourcing markets. In *WWW '13*, pages 1157–1166, 2013.

[Singla and Krause, 2013] Adish Singla and Andreas Krause. Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *WWW '13*, pages 1167–1178, 2013.

[Sühr *et al.*, 2019] Tom Sühr, Asia Biega, Meike Zehlike, Krishna Gummadi, and Abhijnan Chakraborty. Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *KDD '19*, pages 3082–3092, New York, NY, USA, 2019. ACM.

[Yao *et al.*, 2018] Huaxiu Yao, Fei Wu, Jintao Ke, Xianfeng Tang, Yitian Jia, Siyu Lu, Pinghua Gong, Jieping Ye, and Zhenhui Li. Deep multi-view spatial-temporal network for taxi demand prediction. In *AAAI '18*, pages 2588–2595, 2018.

[Zhao *et al.*, 2019] Boming Zhao, Pan Xu, Yexuan Shi, Yongxin Tong, Zimu Zhou, and Yuxiang Zeng. Preference-aware task assignment in on-demand taxi dispatching: An online stable matching approach. In *AAAI '19*, pages 2245–2252, 2019.