Journal of Artificial Intelligence Research 79 (2024) 569-597 Submitted 06/2023; published 02/2024

Exploring the Tradeoff Between System Profit and Income Equality Among Ride-hailing Drivers

Evan Yifan Xu XYF@SEU.EDU.CN
Southeast University, No.2 , Sipailou
Nanjing, Jiangsu 210096 CN

Pan Xu PXU@NJIT.EDU
New Jersey Institute of Technology, 323 Dr Martin Luther King Jr Blvd
Newark, NJ 07102 USA

Abstract

This paper examines the income inequality among rideshare drivers resulting from discriminatory cancellations by riders, considering the impact of demographic factors such as gender, age, and race. We investigate the tradeoff between income inequality, referred to as the fairness objective, and system efficiency, known as the profit objective. To address this issue, we propose an online bipartite-matching model that captures the sequential arrival of riders according to a known distribution. The model incorporates the notion of acceptance rates between driver-rider types, which are defined based on demographic characteristics. Specifically, we analyze the probabilities of riders accepting or canceling their assigned drivers, reflecting the level of acceptance between different rider and driver types. We construct a bi-objective linear program as a valid benchmark and propose two LP-based parameterized online algorithms. Rigorous analysis of online competitive ratios is conducted to illustrate the flexibility and efficiency of our algorithms in achieving a balance between fairness and profit. Furthermore, we present experimental results based on real-world and synthetic datasets, validating the theoretical predictions put forth in our study.

1. Introduction

Rideshare platforms such as Uber (2023), Didi (2023) and Lyft (2023) have garnered significant attention from various research communities, including computer science, operations research, and business. A key area of research in this field revolves around designing matching policies that effectively pair drivers with riders. Several studies have been conducted on this topic, see, *e.g.,* (Danassis et al., 2022; Curry et al., 2019; Ashlagi et al., 2019; Lowalekar et al., 2018; Bei & Zhang, 2018; Dickerson et al., 2018; Zhao et al., 2019; Tong et al., 2021). Most of the existing work in this domain focuses on either enhancing system efficiency or improving user satisfaction, or both.

In this paper, our focus is on examining the issue of fairness among rideshare drivers. Previous reports have highlighted the existence of an earning gap among drivers based on various demographic factors, including age, gender, and race, as discussed in studies such as (Cook et al., 2018; Rosenblat et al., 2016). For instance, Hinchliffe (2017) 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 observed among drivers from different demographic groups

Table 1: Summary of major notations.

Notation	Description
G	Bipartite graph
U	Set of types of offline drivers
V	Set of types of online requests
$f = (u, v)$	Edge indicates that u can serve v
B_u	Matching capacity of driver type u
r_v	Arrival rate of request type v
T	Total number of online rounds
p_f	Edge existence probability of edge f
Δ_v	Patience of request type v
$\mathcal{M}$	Set of successful assignments
$\mathcal{M}_u$	Set of edges in $\mathcal{M}$ incident to u
w_f	Profit of a successful assignment f
$E_u(E_v)$	Set of edges incident to $u(v)$

is partly attributed to discriminatory cancellations by riders. These discriminatory practices are more noticeable during off-peak hours when the number of riders is comparable to or even lower than the number of available drivers. It is important to note that in rideshare platforms like Uber and Lyft, once a driver accepts a rider, the rider gains access to sensitive information about the driver, such as their name and photo. Furthermore, riders have the option to cancel the driver within the first two minutes without any charges (H, 2023). This creates a situation where discriminatory cancellations by riders are technically feasible and economically risk-free.

Our objective is to address the issue of income disparity among rideshare drivers resulting from discriminatory cancellations by riders, while considering the tradeoff with system efficiency. It is important to note that these two goals, promoting income equality among drivers and maximizing system efficiency, can sometimes be conflicting. Let's consider the scenario of off-peak hours when the number of riders is relatively low. In order to maximize system efficiency, rideshare platforms like Uber aim to satisfy riders by assigning them to their preferred or "favorite" drivers. This strategy reduces the likelihood of cancellations by riders and minimizes the risk of losing riders to competing platforms like Lyft. However, implementing such a strategy would lead to certain drivers, who are favored by riders, receiving a significantly higher number of ride requests compared to others. This would result in a substantial disparity in income among drivers, negatively affecting group-level income equality. Therefore, finding the right balance between promoting income equality among drivers and optimizing system efficiency becomes crucial. It requires designing fair allocation policies that consider both the drivers' income distribution and the overall effectiveness of the rideshare system. This way, we can mitigate the negative impact of discriminative cancellations while still maintaining a satisfactory level of system efficiency.

2. Related Work

The literature examining on-demand service platforms is vast and encompasses various aspects of optimization and coordination. For instance, Ma et al. (2017) proposed a linear programming model to optimize autonomous vehicle trip chains and determine the required fleet size using AVSR networks. Taylor (2018) investigated the impact of delay sensitivity and agent independence on optimal per-service prices and wages. Bai et al. (2019) studied the coordination of endogenous demand and supply, characterizing optimal price and wage rates based on a queueing model. Bernstein et al. (2021) focused on the competition between platforms and formulated a pricing game to analyze the equilibrium prices that emerge from their interaction. Feng et al. (2021) compared the efficiency of on-demand hailing systems to traditional street-hailing systems in specific circumstances. Addressing uncertain demand and idle vehicle supply, Beirigo et al. (2022) proposed a learning-based optimization approach to approximate the marginal value of vehicles iteratively under different availability settings. Chen et al. (2022) modeled the decision-making processes of drivers and the platform's optimization problem as a Stackelberg game, conducting a counterfactual analysis to determine optimal bonus rates for various scenarios. Guo et al. (2023) focused on sustainability-oriented operational models and developed an efficient method for generating the Pareto front. These studies collectively contribute to the understanding and improvement of on-demand service platforms, offering insights into different optimization approaches, pricing strategies, competition dynamics, and sustainability considerations.

Fairness in operations is an interesting topic which has a large body of work (Bertsimas et al., 2011, 2012; Chen & Wang, 2018; Lyu et al., 2019; Cohen et al., 2021; Ma et al., 2020; Chen et al., 2023; Shao et al., 2021; Hosseini et al., 2023; Xu & Xu, 2022). 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 rideshare 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 $p_f = 1$). While Nanda et al. (2020) concentrated on peak hours when the demand for rides exceeds the supply of drivers, our focus is off-peak hours, characterized by an oversupply of drivers compared to rider demand. Our work introduces a more sophisticated sampling technique, specifically attenuation, in comparison to (Nanda et al., 2020). This advancement effectively narrows the competitive ratio gap between our proposed algorithms and the associated hardness results.

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; Grandoni et al., 2009; Korula et al., 2013; Aggarwal et al., 2014; Esfandiari et al., 2016). Aggarwal et al. (2014) studied the problem of biobjective online bipartite matching and proposed both deterministic and randomized algorithms for solving this problem. They focused on a relative simpler model and examined simpler matching algorithms such as Greedy and RANKING. Esfandiari et al. (2016) studied the bi-objective online submodular optimization and provided almost matching upper and lower bounds for allocating items to agents with two submodular value functions. But they did not consider the fairness objective in their model. The work of (Bansal et al., 2012;

Table 2: Summary of related work of ride-sharing problem.

Author(s) & Year	Research Questions	Fairness Definition	Online/ Offline	Objective	Main Mechanisms	Key Findings
(Ma et al., 2017)	AV trip chains and fleet size	N/A	Offline	Min: Cost	Linear programing	Improving mobility and sustainability
(Taylor, 2018)	Price and wage setting	N/A	Offline	Max: Profit	Pricing	Agent independence and delay sensitivity
(Bai et al., 2019)	Price and wage rates setting	N/A	Online	Max: Profit	Surge pricing	Time-based payout ratio
(Bernstein et al., 2021)	The platform competition	N/A	Online	Equilibria	Pricing game	Incentive mechanism to discourage multihoming
(Feng et al., 2021)	The matching mechanisms	N/A	Offline	Waiting time	Capped matching	Trade-offs between different mechanisms
(Beirigo et al., 2022)	Autonomous ridesharing	N/A	Online	Cumulative contribution	Dynamic programming	Service level contracts
(Chen et al., 2022)	Incentive mechanism	N/A	Offline	Capacity and profit	Stackelberg	Optimal bonus strategy
(Guo et al., 2023)	Sustainability operational	N/A	Offline	Profit and travel cost	Approximative	Generate pareto front efficiently
(Sühr et al., 2019)	Two-sided fair	Amortized fairness	Offline	Utility	Two-sided optimization	Improving income equity
(Lesmana et al., 2019)	Efficiency-Fair Tradeoff	Maxmin	Offline	Efficiency and fairness	Reassignment algorithm	Theoretical lower bound
(Nanda et al., 2020)	Rider fariness	Maxmin	Online	Profit and fairness	Sampling	Competitive ratios: theoretical lower bound
Our work	Driver fairness	Maximin	Online	Profit and fairness	Sampling and attenuation	Competitive ratios: theoretical lower bound

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. Table 2 offers a summary of related work of ride-hailing problems.

3. Main Model

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; Ma et al., 2023). 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 (*e.g.,* the distance between them is below a given threshold).[1] The online phase[2] 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 driver

1. For simplicity, we refer to a driver of type u and a request of type v directly as a driver u and request v when the context is clear.
2. The online phase represents the period when online requests arrive and are processed by the algorithm.

instances belonging to the type u.[3] Additionally, we have the following key definitions in the model.

Definition 1 (Arrivals of Online Requests). *Consider a finite time horizon T (known to the algorithm). For each 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 & Mittal, 2013; Singla & Krause, 2013; Sumita et al., 2022). We start by concentrating on a brief time frame and discretize it so that each round only has one arrival (keep in mind that we can add some dummy nodes to simulate no arrivals), and as a result, we can always find a large enough value of T to make it. In the literature on online bipartite matching under KIID (Feldman et al., 2009; Haeupler et al., 2011; Jaillet & Lu, 2013; Manshadi et al., 2012), it is customary to make the assumption that $T \to \infty$.

Definition 2 (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 u is assigned to v, an immediate random outcome of the existence can be observed, which is present (*i.e., *v accepts u) with probability p_f and not (v cancels u) otherwise. Suppose 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 requestor's individual choice and the second from the fact that historical logs can be used to compute such statistics with high precision.*

Definition 3 (Patience of Requests). *Each request v is associated with a patience value $\Delta_v \in \mathbb{Z}^+$, which represents the maximum number of unsuccessful assignments that the request v can tolerate before leaving the platform. This means that under the patience constraints, request v can be dispatched to at most Δ_v different drivers. Note that v cannot be simultaneously broadcasted to a set of Δ_v different drivers. Instead, v should be sequentially assigned to at most Δ_v distinct drivers (potentially of the same type) until either v is accepted or it leaves the system after reaching its patience limit. This process is referred to as the* Online Probing Process *(OPP). It's worth mentioning that the OPP begins immediately after a request v arrives, provided that v is not rejected, and it concludes within a 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 algorithm ALG, let $\mathcal{M}$ be the set

3. Note that during the time horizon T, an individual driver instance (as opposed to a driver type) can accept at most one online request.

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., 2020; 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.

Definition 4 (Profit Objective)**.** *The expected total profit over all matches obtained by the platform, which is defined as* $\mathbb{E}[\sum_{f\in\mathcal{M}} w_f]$.

Definition 5 (Fairness Objective)**.** *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* $\min_{u\in U} \mathbb{E}[|\mathcal{M}_u|]/B_u$ *, which can be interpreted as the group-level income equality among drivers.*

4. Preliminaries and Main Contributions

In this section, we define the notations and terminologies needed in the paper and summarize the main contributions.

4.1 Competitive Ratio and Benchmark Linear Programs (LPs)

Competitive Ratio (CR). CR is a metric commonly used to evaluate the performance of online algorithms. Consider a given (online) algorithm ALG and an offline optimal (OPT), which is also known as *clairvoyant optimal*. Note that ALG is subject to the real-time decision-making requirement, *i.e.,* ALG has to make an irrevocable decision upon every arrival of online agents (*e.g.,* riders) before observing future arrivals. In contrast, OPT is exempt from that requirement: it enjoys the privilege of observing the full arrival sequence of online agents *before* optimizing decisions. Consider a given instance of an online maximization problem as studied here, and let $\mathbb{E}[\mathsf{ALG}]$ and $\mathbb{E}[\mathsf{OPT}]$ denote the *expected* performance achieved by ALG and OPT under a given metric, respectively, where the expectation is taken over the randomness on both the random arrivals of online agents and that in ALG and OPT. We say ALG achieves a CR of at least $\rho \in [0, 1]$ if $\mathbb{E}[\mathsf{ALG}] \geq \rho \cdot \mathbb{E}[\mathsf{OPT}]$ for any possible instances. Essentially, CR captures the gap between a policy and a clairvoyant optimal due to the real-time decision-making requirement imposed on the former.

For each edge $f = (u, v)$, let x_f be the number of probes on edge f (*i.e.,* assignments of v to u but not necessarily getting matched) in an offline optimal (OPT). For each u (v), let E_u (E_v) be the set of edges incident to u (v). Consider the following bi-objective LP.[4]

$$\max \sum_{f\in E} w_f x_f p_f \tag{1}$$

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

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

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

4. Note that $\{x_f\}$ are variables for the LPs.

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

$$0 \le x_f \le B_u \cdot r_v \quad \forall f = (u, v) \in E. \tag{6}$$

Let LP-(1) and LP-(2) denote the two LPs with the respective objectives (1) and (2), each with Constraints (3), (4), (5), and (6). LP-(1) aims to maximize the profit objective defined in Definition 4, which is to maximize the expected total profit over all matches obtained by the platform. On the other hand, LP-(2) is formulated to maximize the fairness objective defined in Definition 5. Constraint (3) indicates that the number of agent type $u \in \mathcal{U}$ is finite rather than unbounded. For example, in the context of online ride-hailing, this constraint implies that the number of drivers in a given area is limited. Constraint (4) and Constraint (5) are used to bound the total number of probing times and the total number of matches, respectively. These constraints ensure that the number of probes and matches remain within certain limits. Note that we can rewrite Objective (2) as a linear one like $\max \eta$ with additional linear constraints as $\eta \le \sum_{f \in E_u} x_f p_f / B_u$ for all $u \in U$. For presentation convenience, we keep the current compact version. The validity of LP-(1) and LP-(2) as benchmarks are justified in the lemma below.

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. By leveraging the linearity of expectation, we can verify that objective functions (1) and (2) accurately capture the expected profit and fairness attained by the offline optimal solution. To demonstrate the validity of the benchmark for each objective, it is sufficient to establish the feasibility of all constraints for any given offline optimal solution. For each edge f, let us recall that x_f represents the expected number of probes on f, which corresponds to the assignments of agent u to task v without necessarily resulting in a match, in the offline optimal solution. The validity of Constraint (3) is evident as it ensures that each driver u adheres to their matching capacity of B_u. Additionally, we consider the expected behavior of task v during the entire online phase. Given that v is expected to arrive r_v times, and it can be probed at most Δ_v times upon each online arrival, it follows that the total expected number of probes and matches over all edges incident to v should not exceed $r_v \Delta_v$ and r_v, respectively. This justification supports the inclusion of Constraints (4) and (5). The last constraint guarantees that, for each edge $f = (u, v)$, the total number of probes conducted on f does not surpass B_u multiplied by the number of arrivals of task v. Consequently, the total expected number of probes on f remains below $B_u \cdot r_v$. By establishing the feasibility of all constraints for any offline optimal solution, we affirm the validity of the benchmark for both profit and fairness objectives. □

4.2 Main Contributions

In this paper, we propose two parameterized matching policies that can smoothly trade off the income inequality among drivers from different demographic groups and its trade-off with the system efficiency in rideshare. Our contributions are summarized as follows. First, we propose a new online-matching based model to address the income inequality and system efficiency in rideshare. Second, we present a robust theoretical analysis for our model. We construct a bi-objective linear program that provides valid upper bounds for the maximum profit and fairness in the offline optimal. Third, we propose two LP-based parameterized online algorithms, namely WarmUp and AttenAlg, with provable performances on both objectives; see Theorem 1 and Theorem 2 below. We say an online algorithm is (α, β)-competitive if it achieves competitive ratios of α and β on the profit and fairness against benchmarks LP-(1) and LP-(2), respectively. Lastly, we consider a special example and formally state the hardness results (see Theorem 3 below).

Theorem 1. WarmUp(α, β) *achieves a competitive ratio at least* $\left(\alpha \cdot \frac{1-1/e}{2}, \beta \cdot \frac{1-1/e}{2}\right) \sim (0.316 \cdot \alpha, 0.316 \cdot \beta)$ *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.*

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. We test our model and algorithms on a synthetic dataset and a real dataset collected from a large on-demand taxi dispatching platform. Experimental results confirm our theoretical predictions and demonstrate the flexibility of our algorithms in trading off the two conflicting objectives and their efficiency compared to natural heuristics (*e.g.,* Greedy-like algorithms).

4.3 Reduction to Unit Capacity for Every Offline Agent

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. We say an online algorithm achieves an (α, β)-competitive ratio if it achieves competitive ratios of α and β on the profit and fairness against LP-(1) and LP-(2), respectively.

Lemma 2. *Let* $\overline{\text{ALG}}$ *be an algorithm that is* (α, β)*-competitive for the special case when every offline agent has a unit matching capacity. We can twist* $\overline{\text{ALG}}$ *to* ALG *such that* ALG *is at least* (α, β)*-competitive for the general case when each offline agent is allowed to have any integer matching capacity.*

By Lemma 2, we assume w.l.o.g. *unit capacity for all offline agent types (driver types)* throughout this paper.

Proof. Consider a given instance[5] $\mathcal{I}$ with general integer matching capacities. We create another instance, denoted by $\widetilde{\mathcal{I}}$, by replacing each u with a set S_u consisting of $|S_u| = B_u$

5. The term "instance" is referred to as an input instance for an online algorithm, which includes the budget constraints B_u for all $u \in U$, the patience Δ_v and arrival rate r_v for all $v \in V$, the weight w_f and edge existence probability p_f for all edges $\{f\}$, *etc.*

identical copies of u. We claim that the optimal values of LP-(1) and LP-(2) on $\mathcal{I}$ each remain the same as those on $\widetilde{\mathcal{I}}$. Consider LP-(2) for example, and let $\{x_f\}$ be a feasible solution on instance $\mathcal{I}$. Set $\tilde{x}_{\tilde{f}=(\tilde{u},v)} = x_{f=(u,v)}/B_u$ for every $\tilde{u} \in S_u$ and $u \in U$. We can verify that $\{\tilde{x}_{\tilde{f}}\}$ is feasible to LP-(2) with respect to instance $\widetilde{\mathcal{I}}$. Furthermore, for each given $\tilde{u} \in S_u$,

$$\frac{\sum_{\tilde{f}=(\tilde{u},v)\in E_{\tilde{u}}} \tilde{x}_{\tilde{f}} \cdot p_{\tilde{f}}}{B_{\tilde{u}}} = \sum_{\tilde{f}=(\tilde{u},v)\in E_{\tilde{u}}} \tilde{x}_{\tilde{f}} \cdot p_{\tilde{f}} = \sum_{f=(u,v)\in E_u} (x_{f=(u,v)}/B_u) \cdot p_f.$$

The above equality follows from facts that (1) $E_{\tilde{u}} = E_u$ for every $\tilde{u} \in S_u$ and $u \in S_u$ and (2) $p_{\tilde{f}=(\tilde{u},v)} = p_{f=(u,v)}$ since each S_u consists of B_u identical copies of u. Therefore,

$$\min_{\tilde{u}\in\bigcup_{u\in U} S_u} \frac{\sum_{\tilde{f}=(\tilde{u},v)\in E_{\tilde{u}}} \tilde{x}_{\tilde{f}} \cdot p_{\tilde{f}}}{B_{\tilde{u}}} = \min_{u\in U} \sum_{f=(u,v)\in E_u} (x_{f=(u,v)}/B_u) \cdot p_f.$$

Thus, we claim that the optimal value of LP-(2) on $\mathcal{I}$ should be no more than that on $\widetilde{\mathcal{I}}$, denoted by LP (2)($\mathcal{I}$) $\leq$ LP (2)($\widetilde{\mathcal{I}}$).

Now, we prove the other way. Let $\{\tilde{x}_{\tilde{f}}\}$ be any *optimal* solution of LP-(2) on $\widetilde{\mathcal{I}}$. w.l.o.g. assume that $\left\{\tilde{x}_{\tilde{f}=(\tilde{u},v)} | \tilde{u} \in S_u\right\}$ all take a uniform value, say z_u, for each given $u \in U$ since otherwise we can decrease any outstanding value to make it match the rest while maintaining the optimal objective value unchanged. Note that $\{z_u | u \in U\}$ may not necessarily take the same value due to disparities over $\{p_f\}$. Consider such a solution that $x_{f=(u,v)} = B_u \cdot \tilde{x}_{\tilde{f}=(\tilde{u},v)} = B_u \cdot z_u$ for every $u \in U$. We can verify that $\{x_f\}$ is feasible to LP-(2) on $\mathcal{I}$. Furthermore,

$$\begin{aligned} \text{LP (2)}(\mathcal{I}) &\geq \min_{u\in U} \frac{\sum_{f\in E_u} x_f \cdot p_f}{B_u} = \min_{u\in U} \frac{\sum_{f\in E_u} B_u \cdot z_u \cdot p_f}{B_u} = \min_{u\in U} \sum_{f\in E_u} z_u \cdot p_f \\ &= \min_{\tilde{u}\in\bigcup_{u\in U} S_u} \sum_{\tilde{f}\in E_{\tilde{u}}} x_{\tilde{f}=(\tilde{u},v)} \cdot p_{\tilde{f}} = \text{LP (2)}(\widetilde{\mathcal{I}}). \end{aligned}$$

Thus, we conclude that LP-(2) yields the same optimal value for both $\mathcal{I}$ and $\widetilde{\mathcal{I}}$. Similarly, through similar analysis, we find that LP-(1) also achieves the same optimal value for both $\mathcal{I}$ and $\widetilde{\mathcal{I}}$.

Now, suppose we have at hand an online algorithm $\overline{\mathsf{ALG}}$ that is (α, β)-competitive for an instance $\widetilde{\mathcal{I}}$ where each offline agent (driver) has a unit matching capacity. We can twist $\overline{\mathsf{ALG}}$ to an algorithm ALG for any general instance $\mathcal{I}$ where each offline agent u has an integer capacity B_u as follows. First, convert $\mathcal{I}$ to $\widetilde{\mathcal{I}}$ by replacing each u with a set S_u of $|S_u| = B_u$ identical copies. Second, apply $\overline{\mathsf{ALG}}$ to $\widetilde{\mathcal{I}}$. We can verify that ALG is a valid algorithm on $\mathcal{I}$ since (1) each u will be matched at most B_u times since each $\tilde{u} \in S_u$ is matched at most once by $B_{\tilde{u}} = 1$ in $\overline{\mathsf{ALG}}$; and (2) each v is probed at most Δ_v times upon arrival in $\overline{\mathsf{ALG}}$, and so is v in ALG.

For each $f = (u, v)$ and $\tilde{f} = (\tilde{u}, v)$, let $Z_f = 1$ and $Z_{\tilde{f}} = 1$ indicate that f is matched (probed and present) in ALG and $\tilde{f}$ matched in $\overline{\mathsf{ALG}}$, respectively. Observe that for profit, the performance of the two algorithms, denoted by $\mathsf{ALG}_P(\mathcal{I})$ and $\overline{\mathsf{ALG}}_P(\widetilde{\mathcal{I}})$, satisfies

$$\mathsf{ALG}_P(\mathcal{I}) = \sum_{u \in U} \sum_{f \in E_u} w_f \cdot \mathbb{E}[Z_f] = \sum_{u \in U} \sum_{\tilde{u} \in S_u} \sum_{\tilde{f} \in E_{\tilde{u}}} w_{\tilde{f}} \cdot \mathbb{E}[Z_{\tilde{f}}] = \overline{\mathsf{ALG}}_P(\widetilde{\mathcal{I}}).$$

Note that LP-(1) has the same optimal value on $\mathcal{I}$ and $\widetilde{\mathcal{I}}$. Thus, we claim that ALG achieves the same competitive ratio of α as $\overline{\mathsf{ALG}}$ for profit against LP-(1). As for fairness, the performance of the two algorithms, denoted by $\mathsf{ALG}_F(\mathcal{I})$ and $\overline{\mathsf{ALG}}_F(\widetilde{\mathcal{I}})$, satisfies

$$\begin{aligned}\mathsf{ALG}_F(\mathcal{I}) &= \min_{u \in U} \frac{\sum_{f \in E_u} \mathbb{E}[Z_f]}{B_u} = \min_{u \in U} \frac{\sum_{\tilde{u} \in S_u} \sum_{\tilde{f} \in E_{\tilde{u}}} \mathbb{E}[Z_{\tilde{f}}]}{B_u} \\ &\geq \min_{u \in U} \min_{\tilde{u} \in S_u} \sum_{\tilde{f} \in E_{\tilde{u}}} \mathbb{E}[Z_{\tilde{f}}] = \min_{\tilde{u} \in \bigcup_{u \in U} S_u} \sum_{\tilde{f} \in E_{\tilde{u}}} \mathbb{E}[Z_{\tilde{f}}] = \overline{\mathsf{ALG}}_F(\widetilde{\mathcal{I}}).\end{aligned}$$

This suggests ALG is at least β-competitive for fairness against LP-(2). □

4.4 Randomized Dependent Rounding

Randomized dependent rounding techniques, denoted by GKPS, was introduced by (Gandhi et al., 2006). For simplicity, we state a simplified version of GKPS tailored to star graphs, which suffices for our problem. Recall that E_v is the set of edges incident to v in the input graph G. GKPS is such a rounding technique that takes as input a fractional vector $\mathbf{z} = (z_f : z_f \in [0,1] | f \in E_v)$ on E_v and outputs a random binary vector $\mathbf{Z} = (Z_f : Z_f \in \{0,1\} | f \in E_v)$, which satisfies the following properties.

- **Marginal Distribution**: $\mathbb{E}[Z_f] = z_f$ for all $f \in E_v$.
- **Degree Preservation**: $\Pr[\sum_{f \in E_v} Z_f \leq \sum_{f \in E_v} z_f] = 1$.
- **Negative Correlation**: For any pair of edges $f, f' \in E_v$, $\mathbb{E}[Z_f = 1 | Z_{f'} = 1] \leq z_f$.

The following notations and assumptions are used throughout this paper: (1) $\mathbf{x}^* = \{x_f^*\}$ and $\mathbf{y}^* = \{y^*\}$ are optimal solutions to LP-(1) and LP-(2), respectively; (2) $B_u = 1$ for all $u \in U$ by Lemma 2; (3) $\mathbf{x}^v = (x_f^*/r_v | f \in E_v)$ and $\mathbf{y}^v = (y_f^*/r_v | f \in E_v)$, which are scaled solutions from $\mathbf{x}^v$ and $\mathbf{y}^v$, respectively, restricted to E_v.

In the next two sections, we will present three algorithms, and each of them invokes a subroutine that applies GKPS to either $\mathbf{x}^v$ or $\mathbf{y}^v$ upon v's arrival and outputs a random binary vector $\mathbf{Z}$. From Constraints (4), (5), and (6), we observe that: (1) $\mathbf{x}^v$ and $\mathbf{y}^v$ are both fractional vectors since $0 \leq x_f^* \leq B_u \cdot r_v = r_v$ and $0 \leq y_f^* \leq B_u \cdot r_v = r_v$ for all $f \in E_v$, and (2) both vectors have a total sum of at most Δ_v. By applying the **Degree Preservation** property of GKPS, we claim that the rounded vector $\mathbf{Z}$ with probability one has at most Δ_v entries equal to one (and all the remaining entries equal to zero). The two algorithms exploit this fact by using $\mathbf{Z}$ to guide online probing, such that edge f is probed only when $Z_f = 1$. Consequently, we automatically ensure the patience constraint on v, *i.e.*, v receives no more than Δ_v probes upon its arrival. The properties of **Marginal Distribution** and **Negative Correlation** play a key role in ensuring the proper functioning of the attenuations in the second algorithm, as we need to carefully upper bound the conditional probabilities of other edges $f' \in E_v$ being rounded ($Z_{f'} = 1$) given that one edge f is rounded.

Algorithm 1: Sub-Routine $\mathsf{SR}(\mathbf{z})$: Dependent Rounding with Random Permutation

Apply GKPS to the fractional vector $\mathbf{z} = (z_f : z_f \in [0,1] | f \in E_v)$, and let $\mathbf{Z} = (Z_f : Z_f \in \{0,1\} | f \in E_v)$ be the random binary vector output.
Choose a random permutation π over E_v.
Follow the order π to process each $f = (u, v) \in E_v$:
  if $Z_f = 1$ *and u is available* **then**
    Probe edge f (*i.e.,* assign v to u).
    if *f is present (which occurs with probability p_f)* **then**
      Break.
  else
    Skip to the next one.

5. Three LP-based Algorithms and Related Competitive Analysis

In this section, we present three LP-based algorithms, and offer formal competitive analyses for the first two algorithms.

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

WarmUp incorporates SR, as shown in Algorithm 1, as a subroutine during each online round. Specifically, SR takes as part of the input a fractional vector $\mathbf{z} = (z_f : z_f \in [0,1] | f \in E_v)$, and it follows a random order to probe all edges in E_v guided by a random binary vector output by GKPS.

WarmUp takes two parameters: α and β with $0 \leq \alpha, \beta \leq 1$ and $\alpha + \beta \leq 1$. The main idea of $\mathsf{WarmUp}(\alpha, \beta)$ is as follows: In each round when an online agent v arrives, it invokes $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ with probabilities α and β, respectively. Recall that $\mathbf{x}^v = \{x^*_f / r_v, f \in E_v\}$ and $\mathbf{y}^v = \{y^*_f / r_v, f \in E_v\}$ are scaled optimal solutions of LP-(1) and LP-(2) restricted to E_v, and 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, 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 **Degree Preservation** of GKPS. $\mathsf{WarmUp}(\alpha, \beta)$ is formally in Algorithm 2.

We conduct an edge-by-edge analysis. It would suffice to show that each f is probed with probability at least $\alpha \cdot x^*_f \cdot (1 - 1/e)/2$ and $\beta \cdot y^*_f \cdot (1 - 1/e)/2$ in $\mathsf{WarmUp}(\alpha, \beta)$. Then, by linearity of expectation, we establish Theorem 1. For each $u \in U$ and $t \in [T]$, let $\mathsf{SF}_{u,t} = 1$ indicate that u is *available or safe* at (the beginning of) t and $\mathsf{SF}_{u,t} = 0$ otherwise.

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

Proof. Recall that we assume w.l.o.g. that each $B_u = 1$ due to Lemma 2. For each given $\ell < t$ and each $f = (u, v) \in E_u$, let $X_{f,\ell} = 1$ indicate that v arrives at time ℓ, $H_{f,\ell} = 1$ indicate f is probed at ℓ, and $P_f = 1$ indicate that 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 is probed only when the final rounded vector has the entry of one on f (*i.e.,* $Z_f = 1$). Therefore, we claim that

Algorithm 2: A Warm-Up Algorithm: $\mathsf{WarmUp}(\alpha, \beta)$ with $0 \leq \alpha, \beta \leq 1$ and $\alpha + \beta \leq 1$

Offline Phase:
Solve LP-(1) and LP-(2), and let $\mathbf{x}^* = (x^*_f)$ and $\mathbf{y}^* = (y^*_f)$ be optimal solutions, respectively.
For each $v \in V$, let $\mathbf{x}^v = (x^*_f / r_v : f \in E_v)$ and $\mathbf{y}^v = (y^*_f / r_v : f \in E_v)$ are scaled optimal solutions from LP-(1) and LP-(2) restricted to E_v.
Online Phase:
for $t = 1, 2, \ldots, T$ **do**
  Let an online agent v arrive at time t.
  With probability α, run $\mathsf{SR}(\mathbf{x}^v)$;
  with probability β, run $\mathsf{SR}(\mathbf{y}^v)$;
  and with probability $1 - \alpha - \beta$, reject v.

$\mathbb{E}[H_{f,\ell}] \leq \alpha x^*_f / r_v + \beta y^*_f / r_v$ due to **Marginal Distribution** of GKPS.

$$\begin{aligned}
\mathbb{E}[\mathsf{SF}_{u,t}] &= \prod_{\ell < t} \Pr\Big[\sum_{f \in E_u} X_{f,\ell} \cdot H_{f,\ell} \cdot P_f = 0\Big] = \prod_{\ell < t}\Big(1 - \Pr\Big[\sum_{f \in E_u} X_{f,\ell} \cdot H_{f,\ell} \cdot P_f \geq 1\Big]\Big) \\
&\geq \prod_{\ell < t}\Big(1 - \mathbb{E}\Big[\sum_{f \in E_u} X_{f,\ell} \cdot H_{f,\ell} \cdot P_f\Big]\Big) \quad \text{(by Markov's inequality)} \\
&\geq \prod_{\ell < t}\Big(1 - \sum_{f \in E_u} \frac{r_v}{T} \cdot \Big(\alpha \frac{x^*_f}{r_v} + \beta \frac{y^*_f}{r_v}\Big) \cdot p_f\Big) \\
&= \prod_{\ell < t}\Big(1 - \frac{1}{T}\Big(\alpha \cdot \sum_{f \in E_u} x^*_f p_f + \beta \cdot \sum_{f \in E_u} y^*_f p_f\Big)\Big) \geq \Big(1 - \frac{1}{T}\Big)^{t-1},
\end{aligned}$$

where the last inequality above follows from (1) $\sum_{f \in E_u} x^*_f p_f \leq B_u = 1$ and $\sum_{f \in E_u} y^*_f p_f \leq B_u = 1$ due to Constraint (3) in the benchmark LP and (2) $\alpha + \beta \leq 1$. $\square$

Consider a given $\tilde{u} \in U$ and a given $\tilde{f} = (\tilde{u}, v)$, and let $\chi_{\tilde{f},t} = 1$ indicate $\tilde{f}$ is probed during round t in $\mathsf{WarmUp}(\alpha, \beta)$ and $\chi_{\tilde{f},t} = 0$ otherwise.

Lemma 4. *$\mathbb{E}[\chi_{\tilde{f},t} | \mathsf{SF}_{\tilde{u},t} = 1] \geq \frac{\alpha \cdot x^*_{\tilde{f}}}{2T}$ and $\mathbb{E}[\chi_{\tilde{f},t} | \mathsf{SF}_{\tilde{u},t} = 1] \geq \frac{\beta \cdot y^*_{\tilde{f}}}{2T}$.*

Proof. We focus on the first inequality, which addresses the case when $\tilde{f} = (\tilde{u}, v)$ is probed in $\mathsf{SR}(\mathbf{x}^v)$. Observe that $\tilde{f}$ is probed in $\mathsf{SR}(\mathbf{x}^v)$ if (1) v arrives at t, which occurs with probability r_v/T; (2) WarmUp invokes $\mathsf{SR}(\mathbf{x}^v)$, which happens with probability α; (3) $\tilde{u}$ is available at t, *i.e.,* $\mathsf{SF}_{\tilde{u},t} = 1$; (4) when $\mathsf{SR}(\mathbf{x}^v)$ is invoked, $Z_{\tilde{f}} = 1$, *i.e.,* the random vector output by GKPS has value one for $\tilde{f}$, which occurs with probability $x^*_{\tilde{f}}/r_v$ by **Marginal Distribution** of GKPS; (5) $\tilde{f}$ survives the random order π to get probed before $\tilde{u}$ is matched by some other $f \in E_v$. For each edge $f \in E_v$ (including $\tilde{f}$), let $Y_f = 1$ indicate that f falls before $\tilde{f}$ in the random order π (with $Y_{\tilde{f}} = 0$ with probability one), $P_f = 1$ that f is present when probed, and Z_f be the entry on f in the random vector output by GKPS in

$\mathsf{SR}(\mathbf{x}^v)$. Note that a *sufficient* condition ensuring the occurrence of event (5) can be that $\sum_{f\neq \tilde{f}, f\in E_v} Z_f \cdot Y_f \cdot P_f = 0$, where we refer to some $f \neq \tilde{f}$ with $Z_f \cdot Y_f \cdot P_f = 1$ as that "f blocks $\tilde{f}$ with respect to the order π."[6]

$$
\begin{aligned}
\mathbb{E}[\chi_{\tilde{f},t}|\mathsf{SF}_{\tilde{u},t}=1] &\geq \frac{\alpha r_v}{T}\cdot \Pr[Z_{\tilde{f}}=1]\cdot \Pr\Big[\sum_{f\neq \tilde{f}, f\in E_v} Z_f \cdot Y_f \cdot P_f = 0 \;\Big|\; Z_{\tilde{f}}=1\Big] \\
&= \frac{\alpha r_v}{T}\cdot \frac{x^*_{\tilde{f}}}{r_v}\cdot \Big(1-\Pr\Big[\sum_{f\neq \tilde{f}, f\in E_v} Z_f \cdot Y_f \cdot P_f \geq 1 \;\Big|\; Z_{\tilde{f}}=1\Big]\Big) \\
&\geq \frac{\alpha x^*_{\tilde{f}}}{T}\cdot \Big(1-\mathbb{E}\Big[\sum_{f\neq \tilde{f}, f\in E_v} Z_f \cdot Y_f \cdot P_f \;\Big|\; Z_{\tilde{f}}=1\Big]\Big) \qquad (7)\\
&= \frac{\alpha x^*_{\tilde{f}}}{T}\cdot \Big(1-\sum_{f\neq \tilde{f}, f\in E_v}\mathbb{E}\Big[Z_f \cdot Y_f \cdot P_f \;\Big|\; Z_{\tilde{f}}=1\Big]\Big) \\
&\geq \frac{\alpha x^*_{\tilde{f}}}{T}\cdot \Big(1-\sum_{f\neq \tilde{f}, f\in E_v}\frac{x^*_f}{r_v}\cdot\frac{p_f}{2}\Big) \qquad (8)\\
&\geq \frac{\alpha x^*_{\tilde{f}}}{T}\cdot\frac{1}{2}, \qquad (9)
\end{aligned}
$$

where Inequality (7) follows from Markov's inequality; Inequality (8) is valid since (a) $\mathbb{E}[Z_f|Z_{\tilde{f}}=1] \leq \mathbb{E}[Z_f] = x^*_f/r_v$ due to **Negative Correlation** of GKPS and (b) $\mathbb{E}[Y_f]=1/2$ and $\mathbb{E}[Z_f]=p_f$ for every $f\neq\tilde{f}, f\in E_v$; Inequality (9) follows from $\sum_{f\in E_v} x^*_f p_f \leq r_v$ due to Constraint (5) of the benchmark LP. By applying similar analysis, we can get the second part of the result. ☐

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

Theorem 1. $\mathsf{WarmUp}(\alpha,\beta)$ *achieves a competitive ratio at least* $\left(\alpha\cdot\frac{1-1/e}{2}, \beta\cdot\frac{1-1/e}{2}\right)\sim(0.316\cdot\alpha, 0.316\cdot\beta)$ *simultaneously on the profit and fairness for any* $\alpha,\beta>0$ *with* $\alpha+\beta\leq 1$.

Proof of Theorem 1. For each $f=(u,v)\in E$, let $\kappa^{\mathbf{x}}_f$ and $\kappa^{\mathbf{y}}_f$ be the expected numbers of probes of f in $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$, respectively. Note that

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

where the last term is obtained by taking $T\to\infty$. Similarly, we can show that $\kappa^{\mathbf{y}}_f \geq \beta\cdot y^*_f(1-1/e)/2$. By linearity of expectation, we claim that the expected profit achieved

6. Note that the condition stated here is *sufficient but not necessary*, since some other $f=(u,v)\in E_v$ with $f\neq\tilde{f}$ may have u matched before t and, thus, cannot pose any threat to $\tilde{f}$. In the current analysis of WarmUp, we assume that every edge $f\in E_v$ other than $\tilde{f}$ has the potential to block $\tilde{f}$, which is defined as $Z_f\cdot Y_f\cdot P_f=1$. This observation is the exact motivation for our second algorithm shown in Section 5.2, which exploits a refined definition of some edge $f\neq\tilde{f}$ blocking $\tilde{f}$ and exhibits improved performance over WarmUp.

by $\mathsf{WarmUp}(\alpha, \beta)$ should be at least $(1-1/e) \cdot (\alpha/2) \cdot \sum_{f \in E} x_f^* p_f w_f$, which is further lower bounded by $(1-1/e) \cdot (\alpha/2) \cdot \mathrm{OPT}_P$ due to Lemma 1, where OPT_P denotes the performance of an offline optimal on profit. This establishes $\mathsf{WarmUp}(\alpha, \beta)$ is at least $(1-1/e) \cdot (\alpha/2)$-competitive for profit. Similarly, we can argue that $\mathsf{WarmUp}(\alpha, \beta)$ achieves the same ratio for fairness as well. □

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

Let $\{\gamma_t, \mu_t | t \in [T]\}$ be a series defined as follows,

$$\gamma_1 = 1; \mu_t = 1 - \gamma_t/2, \forall 1 \le t \le T; \gamma_{t+1} = \gamma_t(1 - \mu_t/T), \forall 1 \le t \le T-1. \tag{10}$$

Overview of $\mathsf{AttenAlg}(\alpha, \beta)$ in Algorithm 3. Inspired by the work of (Brubach et al., 2020), we present an enhanced version of WarmUp called AttenAlg, which incorporates (offline) vertex- and edge-attenuations *guided by an auxiliary series* $\{\gamma_t, \mu_t | t \in [T]\}$, as defined in Equation (10). AttenAlg consists of two phases, **Offline Phase** and **Online Phase**, similar to WarmUp. The **Offline Phase** aims to output a set of values $\{\phi_{u,t}, \psi_{f,t}, \varphi_{f,t} | u \in U, f \in E, 1 \le t \le T\}$. Note that (1) $\phi_{u,t}$ represents the precise estimate of the probability that offline agent $u \in U$ is available at (the beginning of) time $t \in [T]$; and (2) $\psi_{f,t}$ and $\varphi_{f,t}$ denote estimates of probabilities that edge $f = (u, v) \in E_v$ survives the random order π and is ready to probe in Step (5) of SR when $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ are invoked, respectively, where the two probabilities both are conditional on that v arrives at t, u is available at (the beginning of) t, and $Z_f = 1$ in the random vector output by GKPS. All estimates of $\{\phi_{u,t}, \psi_{f,t}, \varphi_{f,t} | u \in U, f \in E, t \in [T]\}$ can be obtained through Monte-Carlo simulations of **Online Phase**'s Steps up to time t.[7]

The **Online Phase** aims to achieve two goals by exploiting estimates obtained in the **Offline Phase**. (**Goal A**) The first goal is to ensure that every offline agent is available at t with a target probability of γ_t, which is achieved through (offline) vertex attenuations. Specifically, at each time step t, we independently *relabel* each available offline agent u as either available or unavailable, with respective probabilities of $\gamma_t/\phi_{u,t}$ and $1 - \gamma_t/\phi_{u,t}$. This ensures that every offline agent has a probability of γ_t of being available after the vertex attenuations. *Note that the transition of each offline agent from being available to unavailable is permanent and irreversible.* (**Goal B**) The second goal is to ensure that each edge $f = (u, v) \in E_v$ is probed with target probabilities equal to $\alpha \cdot \mu_t \cdot x_f^*/r_v$ and $\beta \cdot \mu_t \cdot y_f^*/r_v$ in $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$, respectively, and both probabilities are conditional on v arriving at time t and u being available at time t after vertex attenuations. The second goal is achieved by edge attenuations. Particularly, each edge $f \in E_v$ is probed with an extra factor of $\mu_t/\psi_{f,t}$ and $\mu_t/\varphi_{f,t}$ when f is ready to probe in Step (5) when $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$ are invoked, respectively.

Lemma 5. AttenAlg *in Algorithm 3 is valid with respect to series $\{\gamma_t, \mu_t | t \in [T]\}$ defined in* (10) *such that $\phi_{u,t} \ge \gamma_t$, $\psi_{f,t} \ge \mu_t$, and $\varphi_{f,t} \ge \mu_t$, for each $u \in U$, $t \in [T]$, and $f \in E$.*

The above lemma justifies the validity of AttenAlg. We first show how the lemma leads to Theorem 2.

7. More precisely, we can get an estimate for each target with a multiplicative error of at most $\epsilon > 0$ and a high confidence of at least $1-\delta > 0$ with a sample complexity of $O(\epsilon^{-2} \cdot \log(1/\delta))$; see detailed discussions in (Dickerson et al., 2021; Ma et al., 2023).

Algorithm 3: An LP-Based Algorithm with Attenuations: $\mathsf{AttenAlg}(\alpha, \beta)$

1 **Offline Phase**:
2 Solve LP-(1) and LP-(2), let $\mathbf{x}^*$ and $\mathbf{y}^*$ be the optimal solutions.
3 For each $v \in V$, let $\mathbf{x}^v = \{x^*_f / r_v, f \in E_v\}$ and $\mathbf{y}^v = \{y^*_f | f \in E_v\}$ be the scaled optimal solutions restricted on E_v.
4 *Initialization*: Set $\phi_{u,t} = 1$ for every $u \in U$ at $t = 1$.
5 **for** $t = 2, \ldots, T$ **do**
6 By simulating Steps from (13) to (17) of **Online Phase** for all rounds over $t' = 1, 2, \ldots, t-1$, we get a sharp estimate of $\phi_{u,t}$ for the probability of each $u \in U$ being available at (the beginning of) t *before* vertex attenuations.
7 **for** $t = 1, \ldots, T$ **do**
8 By simulating Steps from (13) to (17) of **Online Phase** for all rounds over $t' = 1, 2, \ldots, t-1$, and from Step (13) to $\mathsf{SR}(\mathbf{x}^v)$ on Step (15) *without edge attenuations* in **Online Phase** at t, we get a sharp estimate of $\psi_{f,t}$ for the conditional probability that edge $f = (u, v) \in E_v$ reaches Step (5) of $\mathsf{SR}(\mathbf{x}^v)$, *i.e.,* v is not matched when it comes to f under the random order π, which assumes that v arrives at t, $Z_f = 1$ (the random binary vector output by GKPS to $\mathbf{x}^v$), and u is available after vertex attenuations are applied at t.
9 By simulating Steps from (13) to (17) of **Online Phase** for all rounds over $t' = 1, 2, \ldots, t-1$, and from Step (13) to $\mathsf{SR}(\mathbf{y}^v)$ on Step (15) *without edge attenuations* in **Online Phase** at t, we get a sharp estimate of $\varphi_{f,t}$ for the conditional probability that edge $f = (u, v) \in E_v$ reaches Step (5) of $\mathsf{SR}(\mathbf{y}^v)$, *i.e.,* v is not matched when it comes to f under the random order π, which assumes that v arrives at t, $Z_f = 1$ (the random binary vector output by GKPS to $\mathbf{y}^v$), and u is available after vertex attenuations are applied at t.
10 **Online Phase**:
11 *Initialization*: Label all offline vertices *available* at $t = 1$.
12 **for** $t = 1, 2, \ldots, T$ **do**
13 **Vertex Attenuations**: *Independently* relabel each *available* offline agent u as available and unavailable with respective probabilities $\gamma_t / \phi_{u,t}$ and $1 - \gamma_t / \phi_{u,t}$.
14 Let an online vertex $v \in V$ arrive at time t.
15 With probability α, run $\mathsf{SR}(\mathbf{x}^v)$ with **edge attenuations**. Specifically, when each edge $f \in E_{v,t}$ is ready to probe in Step (5) of $\mathsf{SR}(\mathbf{x}^v)$, probe f with probability $\mu_t / \psi_{f,t}$, or go to Step (9) of $\mathsf{SR}(\mathbf{x}^v)$ otherwise, *i.e.,* skip to the next one with probability $1 - \mu_t / \psi_{f,t}$;
16 with probability β, run $\mathsf{SR}(\mathbf{y}^v)$ with **edge attenuations**. Specifically, when each edge $f \in E_v$ is ready to probe in Step (5) of $\mathsf{SR}(\mathbf{y}^v)$, probe f with probability $\mu_t / \varphi_{f,t}$ or go to Step (9) of $\mathsf{SR}(\mathbf{y}^v)$ otherwise;
17 with probability $1 - \alpha - \beta$, reject v.

Theorem 2. $\mathsf{AttenAlg}(\alpha, \beta)$ *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* $\alpha + \beta \leq 1$.

Proof of Theorem 2. Consider a given $f=(u,v)$. Let $\kappa_f^{\mathbf{x}}$ and $\kappa_f^{\mathbf{y}}$ be the numbers of probes of f in $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$, respectively. Following statements in **Goal A** and **Goal B**, we have

$$\kappa_f^{\mathbf{x}}=\sum_{t=1}^{T}\gamma_t\cdot\frac{r_v}{T}\cdot\frac{\alpha\cdot\mu_t\cdot x_f^*}{r_v}=(\alpha x_f^*)\cdot\sum_{t=1}^{T}\frac{\mu_t\cdot\gamma_t}{T}. \tag{11}$$

By definition of $\{\gamma_t,\mu_t\}$, we can verify that $\sum_{t=1}^{T}\mu_t\cdot\gamma_t/T=\frac{e-1}{e+1}$ when $T\to\infty$. Thus, we claim that $\kappa_f^{\mathbf{x}}=(\alpha x_f^*)\cdot(\frac{e-1}{e+1})$. By linearity of expectation, we claim that the expected profit achieved by $\mathsf{AttenAlg}(\alpha,\beta)$ should be at least $(\alpha\cdot\frac{e-1}{e+1})\cdot\sum_{f\in E}x_f^*p_fw_f$, which is further lower bounded by $(\alpha\cdot\frac{e-1}{e+1})\cdot\mathrm{OPT}_P$ due to Lemma 1, where OPT_P denotes the performance of an offline optimal on profit. This establishes $\mathsf{AttenAlg}\cdot(\alpha,\beta)$ is at least $(\alpha\cdot\frac{e-1}{e+1})$-competitive for profit. Similarly, we can argue that $\mathsf{AttenAlg}(\alpha,\beta)$ achieves the same ratio for fairness as well. $\square$

Proof of Lemma 5. We prove by induction over $t=1,2,\ldots,T$. Consider the base case $t=1$. By definition, we see that $\phi_{u,t}=\gamma_t=1$ for all $u\in U$. In the following, we show $\psi_{f,t}\geq\mu_t$ for every $f\in E$ at $t=1$. By definition, $\mu_1=1-\gamma_1/2=1/2$. Consider a given $\tilde{f}=(\tilde{u},v)$, and recall that $\psi_{\tilde{f},t=1}$ denotes the probability that $\tilde{f}$ reaches Step (5) of $\mathsf{SR}(\mathbf{x}^v)$, *i.e.,* v is not matched when it comes to $\tilde{f}$ under a random order π, which is conditional on events that v arrives at t, $Z_{\tilde{f}}=1$ (the entry on $\tilde{f}$ of the random binary vector output by GKPS to $\mathbf{x}^v$), and $\tilde{u}$ is available after vertex attenuations are applied at $t=1$. Observe that at $t=1$, every $u\in U$ is available with probability one, and we essentially apply no vertex attenuations since $\gamma_t/\phi_{u,t}=1$ (every available offline agent remains available). For each edge $f\in E_v$ (including $\tilde{f}$), let $Y_f=1$ indicate that f falls before $\tilde{f}$ in the random order π (with $Y_{\tilde{f}}=0$ with probability one), $P_f=1$ that f is present when probed, and $\mathsf{SF}_{f,t}=1$ that $f=(u,v)$ is *available or safe* at t (*i.e.,* u is available or safe) after vertex attenuations at t but before observing any online arrival at t. Thus, for $t=1$,

$$\begin{aligned}
\psi_{\tilde{f},t}&=\Pr\left[\tilde{f}\text{ reaches Step (5) of }\mathsf{SR}(\mathbf{x}^v)\;\middle|\;Z_{\tilde{f}}=1,\mathsf{SF}_{\tilde{f},t}=1\right]\\
&=1-\Pr\Big[\sum_{f\neq\tilde{f},f\in E_v}Z_f\cdot\mathsf{SF}_{f,t}\cdot Y_f\cdot P_f\geq 1\;\Big|\;Z_{\tilde{f}}=1,\mathsf{SF}_{\tilde{f},t}=1\Big] &&(12)\\
&\geq 1-\mathbb{E}\Big[\sum_{f\neq\tilde{f},f\in E_v}Z_f\cdot\mathsf{SF}_{f,t}\cdot Y_f\cdot P_f\;\Big|\;Z_{\tilde{f}}=1,\mathsf{SF}_{\tilde{f},t}=1\Big].\\
&=1-\sum_{f\neq\tilde{f},f\in E_v}\mathbb{E}\Big[Z_f\cdot\mathsf{SF}_{f,t}\cdot Y_f\cdot P_f\;\Big|\;Z_{\tilde{f}}=1,\mathsf{SF}_{\tilde{f},t}=1\Big] &&(13)\\
&=1-\sum_{f\neq\tilde{f},f\in E_v}\mathbb{E}[Z_f|Z_{\tilde{f}}=1]\cdot\mathbb{E}[\mathsf{SF}_{f,t}|\mathsf{SF}_{\tilde{f},t}=1]\cdot\mathbb{E}[Y_f]\cdot\mathbb{E}[P_f] &&(14)\\
&\geq 1-\sum_{f\neq\tilde{f},f\in E_v}\frac{x_f^*}{r_v}\cdot\frac{1}{2}\cdot p_f &&(15)\\
&\geq 1-1/2=1/2=\mu_t=\mu_1, &&(16)
\end{aligned}$$

where Equality (12) follows from that $\tilde{f}$ fails to reache Step (5) of $\mathsf{SR}(\mathbf{x}^v)$ iff there exists one other edge $f \in E_v$ with $f \neq \tilde{f}$ that blocks $\tilde{f}$ in π, *i.e.,* $f = (u, v)$ satisfies $Z_f = 1$, u is available at t, falls before $\tilde{f}$ in π, and $P_f = 1$, and thus, leading to v's matching; Equality (13) due to independence among $\{Z_f, \mathsf{SF}_{f,t}, Y_f, P_f\}$ for each given $f \in E_v$; Inequality (15) due to facts (1) $\mathbb{E}[Z_f | Z_{\tilde{f}} = 1] \leq \mathbb{E}[Z_f] = x^*_f / r_v$ following properties of **Negative Correlation** and **Marginal Distribution** of GKPS, (2) $\mathbb{E}[\mathsf{SF}_{f,t} | \mathsf{SF}_{\tilde{f},t} = 1] = 1$ (every $u \in U$ is available and no vertex attenuations are applied at $t = 1$), and (3) $\mathbb{E}[Y_f] = 1/2$ and $\mathbb{E}[P_f] = p_f$; Inequality (16) follows from Constraint (5) in the benchmark LP. Therefore, we claim that at $t = 1$, $\psi_{\tilde{f},t} \geq \mu_t$ for every $\tilde{f}$, and we can get the result of $\varphi_{\tilde{f},t} \geq \mu_t$ by applying the same analysis.

Now, we show the induction from $\tilde{t}$ to $\tilde{t}+1$. Assume that $\phi_{u,t} \geq \gamma_t$, $\psi_{f,t} \geq \mu_t$, and $\varphi_{f,t} \geq \mu_t$, for every $u \in U$, $f \in E$ and all $1 \leq t \leq \tilde{t}$. This means $\mathsf{AttenAlg}$ is valid during the first $\tilde{t}$ rounds, and thus, the two goals of vertex and edge attenuations are achieved for any $1 \leq t \leq \tilde{t}$: (1) every offline agent u has a probability of γ_t of being available after the vertex attenuations and (2) each edge $f = (u, v) \in E_v$ is probed with target probabilities equal to $\alpha \cdot \mu_t \cdot x^*_f / r_v$ and $\beta \cdot \mu_t \cdot y^*_f / r_v$ in $\mathsf{SR}(\mathbf{x}^v)$ and $\mathsf{SR}(\mathbf{y}^v)$, respectively, conditional v's arriving. We show the claim for $t = \tilde{t}+1$. Consider a given $\tilde{u} \in U$. Recall that $\phi_{\tilde{u},\tilde{t}+1}$ denotes the probability that $\tilde{u}$ is available at the beginning of $\tilde{t}+1$ before vertex attenuations. This event happens iff (1) $\tilde{u}$ survives the vertex attenuations at $\tilde{t}$, which occurs with probability equal to $\gamma_{\tilde{t}}$ by the inductive assumption, and (2) $\tilde{u}$ survives the online matching process from some arriving neighbor, *i.e.,* Steps from (14) to (17) at $\tilde{t}$ in **Online Phase**. Thus,

$$
\begin{aligned}
\phi_{\tilde{u},\tilde{t}+1} &= \gamma_{\tilde{t}} \cdot \Pr\left[\tilde{u} \text{ survives Steps from (14) to (17) at } \tilde{t} \text{ in } \mathsf{AttenAlg}\right] \\
&\geq \gamma_{\tilde{t}} \cdot \left(1 - \sum_{f \in E_{\tilde{u}}} \Pr\left[f \text{ is matched at } \tilde{t}\right]\right) \\
&= \gamma_{\tilde{t}} \cdot \left(1 - \sum_{f=(\tilde{u},v) \in E_{\tilde{u}}} \frac{r_v}{T} \cdot \Big(\frac{\alpha \mu_{\tilde{t}} \cdot x^*_f}{r_v} + \frac{\beta \mu_{\tilde{t}} \cdot y^*_f}{r_v}\Big) \cdot p_f\right) \qquad (17) \\
&= \gamma_{\tilde{t}} \cdot \left(1 - \frac{\mu_{\tilde{t}}}{T} \cdot \Big(\alpha \cdot \sum_{f \in E_{\tilde{u}}} x^*_f \cdot p_f + \beta \cdot \sum_{f \in E_{\tilde{u}}} y^*_f \cdot p_f\Big)\right) \\
&\geq \gamma_{\tilde{t}} \cdot (1 - \mu_{\tilde{t}}/T) = \gamma_{\tilde{t}+1}. \qquad (18)
\end{aligned}
$$

Equality (17) is due to inductive assumption on the goal achieved by edge attenuations during $\tilde{t}$. Inequality (18) follows from (1) $\sum_{f \in E_{\tilde{u}}} x^*_f p_f \leq B_{\tilde{u}} = 1$ and $\sum_{f \in E_{\tilde{u}}} y^*_f p_f \leq B_{\tilde{u}} = 1$ due to Constraint (3) in the benchmark LP; (2) $\alpha + \beta \leq 1$; and (3) $\gamma_{\tilde{t}} \cdot (1 - \mu_{\tilde{t}}/T) = \gamma_{\tilde{t}+1}$ by definition. Thus, we claim that $\phi_{\tilde{u},\tilde{t}+1} \geq \gamma_{\tilde{t}+1}$ for any $\tilde{u} \in U$. The results of $\psi_{f,t} \geq \mu_t$ and $\varphi_{f,t} \geq \mu_t$ for every f at $t = \tilde{t}+1$ follow from a similar analysis to the base case at $t = 1$. The only annotation to add is for Inequality (15). For any $f \in E$ and $1 \leq t \leq \tilde{t}+1$, we have

$$\mathbb{E}[\mathsf{SF}_{f,t} | \mathsf{SF}_{\tilde{f},t} = 1] \leq \mathbb{E}[\mathsf{SF}_{f,t}] = \gamma_t,$$

where the first inequality follows from Lemma 3.1 in (Brubach et al., 2020) and the second equality from inductive assumption on the goal achieved by vertex attenuations up to time

Algorithm 4: An LP-based Algorithm with Boosting: $\mathsf{Boosting}(\alpha, \beta)$

Offline Phase:
Solve LP-(1) and LP-(2), and let x^*_f and y^*_f be optimal solutions, respectively.
Online Phase:
for $t = 1, 2, \ldots, T$ **do**
  Let an online agent v arrive at time t.
  With probability α, run $\mathsf{SR}((x^*_f/r_v) \cdot \frac{\Delta_v}{\sum_{\tilde{f} \in E_{v,t}} x^*_{\tilde{f}}/r_v})$;
  with probability β, run $\mathsf{SR}((y^*_f/r_v) \cdot \frac{\Delta_v}{\sum_{\tilde{f} \in E_{v,t}} y^*_{\tilde{f}}/r_v})$;
  and with probability $1 - \alpha - \beta$, reject v.

$\tilde{t} + 1$. As a result, for any given $\tilde{f} = (\tilde{u}, v)$, we have at $t = \tilde{t} + 1$,

$$\psi_{\tilde{f}, \tilde{t}+1} \geq 1 - \sum_{f \neq \tilde{f}, f \in E_v} \frac{x^*_f}{r_v} \cdot \gamma_{\tilde{t}+1} \cdot \frac{1}{2} \cdot p_f \geq 1 - \frac{\gamma_{\tilde{t}+1}}{2} = \mu_{\tilde{t}+1},$$

where the last equality is due to the definition of the series $\{\psi_t\}$ in (10). Similarly, we can show that $\varphi_{\tilde{f}, \tilde{t}+1} \geq \mu_{\tilde{t}+1}$. Thus, we complete the induction. □

5.3 The Third Algorithm $\mathsf{Boosting}(\alpha, \beta)$

Finally, we introduce an LP-based algorithm with boosting, named $\mathsf{Boosting}(\alpha, \beta)$. In each round when an online agent v arrives, with probability α, we run $\mathsf{SR}((x^*_f/r_v) \cdot \frac{\Delta_v}{\sum_{\tilde{f} \in E_{v,t}} x^*_{\tilde{f}}/r_v})$; with probability β, we run $\mathsf{SR}((y^*_f/r_v) \cdot \frac{\Delta_v}{\sum_{\tilde{f} \in E_{v,t}} y^*_{\tilde{f}}/r_v})$; with probability $1 - \alpha - \beta$, reject v. In this way, we adjust the sum of entries in the sampling vector to ensure it does not exceed Δ_v. The rationale behind designing $\mathsf{Boosting}$ is to leverage the advantages of a greedy approach. Note that similar to the challenges encountered in (Ma et al., 2023), the competitive analysis for $\mathsf{Boosting}$ presents an exceptionally intricate problem that warrants exploration in future work. The algorithm is formally stated in Algorithm 4.

6. Hardness Results

We prove Theorem 3 in this section. Consider the below example.

Example 1. *Consider a graph which consists of n identical units, each unit $i \in [n]$ is a star graph which includes the center of v_i and two other neighbors u^a_i and u^b_i. Set $p_{i,a} = 1$ and $p_{i,b} = \epsilon$ where we use $\{i, a\}$ ($\{i, b\}$) to index the edges (u^a_i, v_i) and (u^b_i, v_i) respectively. Assume that (1) unit edge weight on all edges; (2) $T = n$ and unit arrival rate on all v_i (*i.e., *all $r_v = 1$); (3) unit matching capacity on all u (*i.e., *all B_u=1); and (4) unit patience on all v (*i.e., *all $\Delta_v = 1$).*

Let OPT-P *and* OPT-F *be the optimal LP values of* LP-(1) *and* LP-(2) *on the above example respectively. We can verify that: (1)* $\text{OPT-P} = n$*, where there is a unique optimal*

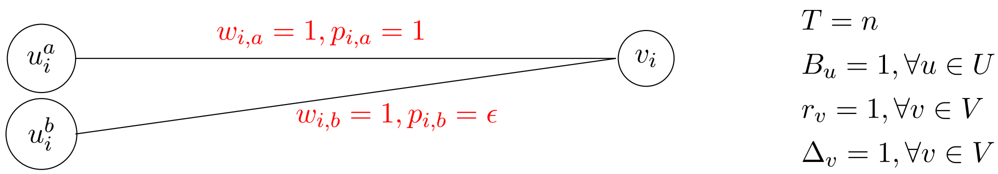


Figure 1: A toy example on which no algorithm can achieve a competitive ratio larger than $1 - 1/e$ on the profit and no algorithm can achieve competitive ratios on the profit and fairness with a sum larger than 1.

solution $x^*_{i,a} = 1$ *and* $x^*_{i,b} = 0$ *for all* $i \in [n]$*; (2)* OPT-F $= \epsilon/(1+\epsilon)$*, where there is a unique optimal solution* $y^*_{i,a} = \frac{\epsilon}{1+\epsilon}$ *and* $y^*_{i,b} = \frac{1}{1+\epsilon}$ *for all* $i \in [n]$*.*

Now based on Example 1, we prove the below lemma.

Lemma 6. *Consider Example 1 and assume* LP-(1) *and* LP-(2) *as benchmarks. We have (1) no algorithm can achieve a competitive ratio larger than* $1 - 1/e$ *on the profit; (2) no algorithm can achieve competitive ratios on the profit and fairness with a sum larger than 1.*

Proof. Consider a given online algorithm ALG, in which the expected number of probes for (u_i^a, v_i) and (u_i^b, v_i) are α_i and β_i for each $i \in [n]$, respectively. Let ALG-P and ALG-F be the profit and fairness achieved by ALG. We have that ALG-P $= \sum_{i\in[n]}(\alpha_i + \beta_i \epsilon)$, ALG-P $= \min_{i\in[n]}\Big(\alpha_i, \beta_i \epsilon\Big)$. Set $\alpha \doteq \sum_{i\in[n]} \alpha_i$ and $\beta \doteq \sum_{i\in[n]} \beta_i$. Note that (1) $\alpha + \beta \leq n$, and (2) $\alpha \leq (1 - 1/e)n$. The latter inequality is due to each $\alpha_i \leq 1 - 1/e$. Thus, the sum of competitive ratios on profit and fairness should be

$$\frac{\text{ALG-P}}{\text{OPT-P}} + \frac{\text{ALG-F}}{\text{OPT-F}} = \frac{\sum_{i\in[n]} \alpha_i + \beta_i \epsilon}{n} + \frac{\min_{i\in[n]}\Big(\alpha_i, \beta_i \epsilon\Big)}{\epsilon/(1+\epsilon)}$$
$$\leq \frac{\alpha + \epsilon\beta}{n} + \frac{\beta(1+\epsilon)}{n} = \frac{\alpha + \beta + 2\epsilon\beta}{n} \leq 1 + 2\epsilon.$$

As for profit, we see that $\frac{\text{ALG-P}}{\text{OPT-P}} = \frac{\alpha+\epsilon\beta}{n} \leq 1 - 1/e + \epsilon$. □

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.*

Proof of Theorem 3. Based on the example presented in Lemma 5 of Section 3.1 of (Fata et al., 2019), we can get a stronger version of statement (2) in Lemma 6, which states that no online algorithm can get an online ratio better than 0.51 for either the profit or fairness based on LP-(1) and LP-(2). Summarizing all analysis we prove Theorem 3. □

7. Experiments

In this section, we describe our experimental results on the synthetic dataset and a real dataset: the New York City yellow cabs dataset[8] which contains the trip histories for thousands of taxis across Manhattan, Brooklyn, and Queens.

7.1 Experiment Setup

Real Dataset. The dataset is collected during the year of 2013. Each trip record includes the (desensitized) driver's license, the pick-up and drop-off locations for the passenger, the duration and distance to complete the trip, the starting and ending time of the trip and some other information such as the number of customers. 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., 2020). 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, 2023). Similarly, we set the ratio of D to A among drivers to be 1 : 2 (FinancesOnline, 2023). 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 of 4–5 PM when a lot of drivers are on the road while the requests are relatively lower than peak hours. On January 31, 2013, $20,701$ trips were completed in the off-peak hour (from 16:00 to 17:00), compared to $35,109$ trips in the peak hour (from 19:00 to 20:00). 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. Each grid is indexed by a unique number to represent a specific pick-up and drop-off location.

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 $B \in \{10, 15, 20, 25\}$. 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 $T = \sum_{v \in V} r_v$ (here, we use the rounded value of the sampled r_v as the final arrival rate). 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 $0 \leq w_f \leq 1$.

Synthetic Dataset. We generate the bipartite graph by setting $|U| = 50$, $|V| = 50$ and $T = 500$. We randomly sample the arrival rates for each request type v, such that $\sum_{v \in V} r_v = T$. For each pair of driver type and request type (u, v), we set an edge between

8. http://www.andresmh.com/nyctaxitrips/.

them with probability 0.1. For each edge f, we choose its edge existence probability p_f uniformly at random from $[0.5, 1)$, and sample the profit w_f uniformly at random from $[0, 1]$. For each driver type u, we assign its capacity B_u with a random value uniformly sampled from $[1, B]$ where we vary $B \in \{10, 15, 20\}$. We set a uniform patience value $\Delta \in \{1, 2, 3\}$ for all v, *i.e.,* $\forall_{v \in V} \Delta_v = \Delta$.

Compared Algorithms. We first compare the performance our proposed algorithms, *i.e.,* WarmUp, AttenAlg and Boosting. Then, we test the WarmUp(α, β) with $\alpha + \beta = 1$ against two natural heuristic baselines, namely Greedy_P (short for Greedy-Profit) and Greedy_F (short for Greedy-Fairness). 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 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. We run all WarmUp(α, β) algorithms for 1000 independent trials and take the average as the expectations. We also run Greedy_P and Greedy_F for 1000 instances and take the average values as the final performance. Note that we use LP-(1) and LP-(2) as the default benchmarks for profit and fairness, respectively.

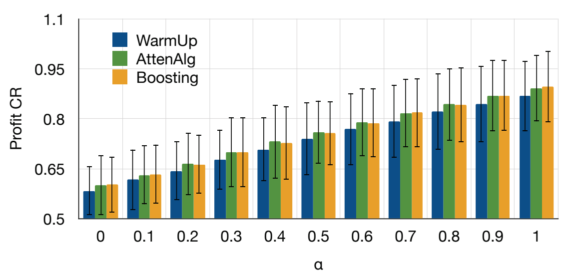


Figure 2: Competitive ratios of profit for LP-based algorithms with different values of α and β with $\alpha+\beta = 1$, while fixing $B = 10$ and $\Delta = 1$.

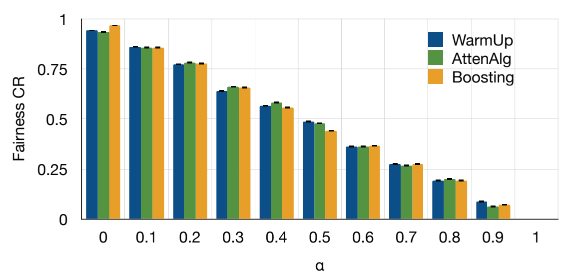
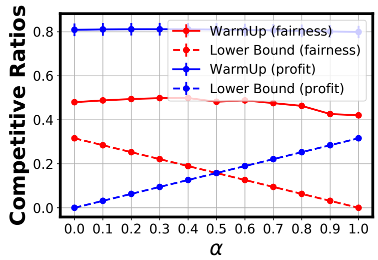


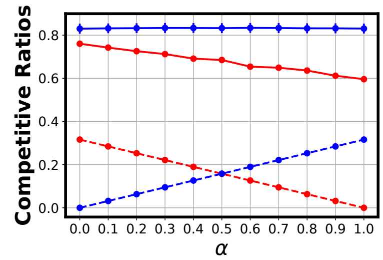
Figure 3: Competitive ratios of fairness for LP-based algorithms with different values of α and β with $\alpha+\beta = 1$, while fixing $B = 10$ and $\Delta = 1$.

7.2 Results and Discussions

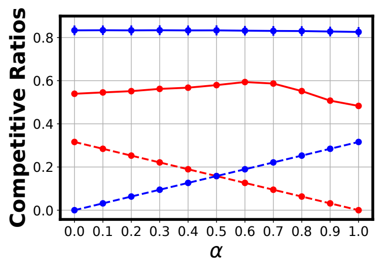
Figure 2 and Figure 3 visually demonstrate an expected trend: as the value of α increases, the profit competitive ratios (CR) of all LP-based algorithms show an upward trend, while their fairness CRs exhibit a decline. Theoretical analysis, as presented in Theorem 1 and Theorem 2, confirms this observation, stating that the lower bound of the profit CR is directly proportional to the value of α, while the lower bound of the fairness CR is proportional to $1 - \alpha$. Notably, owing to the attenuation and boosting strategies employed, AttenAlg and Boosting outperform WarmUp in terms of profit CR. However, for the sake of brevity, we have omitted the plots of AttenAlg and Boosting in the subsequent figures since there is no discernible difference in performance among these three algorithms when utilizing the same parameter settings.



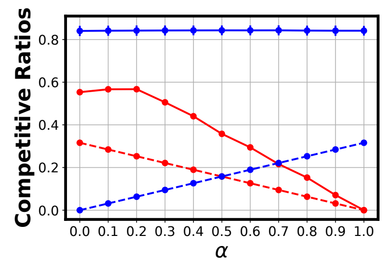
(a) $B = 10$



(b) $B = 15$

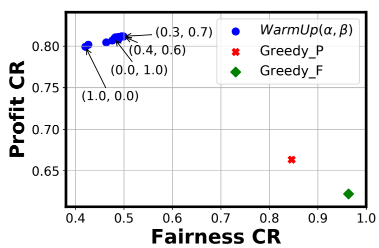


(c) $B = 20$

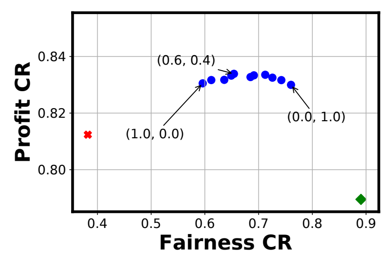


(d) $B = 25$

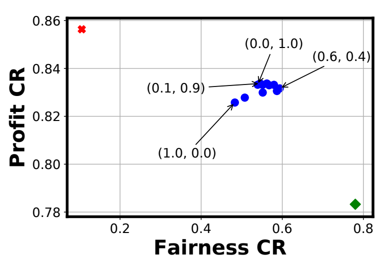
Figure 4: Real dataset: competitive ratios for profit and fairness with different values of α and β with $\alpha + \beta = 1$. (Red solid lines: fairness competitive ratios for WarmUp; Red dotted lines: lower bound of fairness competitive ratios for WarmUp; Blue solid lines: profit competitive ratios for WarmUp; Blue dotted lines: lower bound of profit competitive ratios for WarmUp)



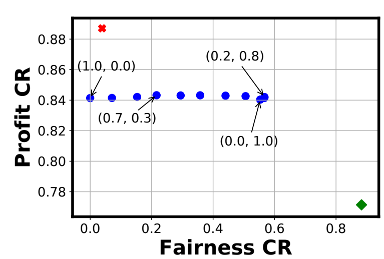
(a) $B = 10$



(b) $B = 15$

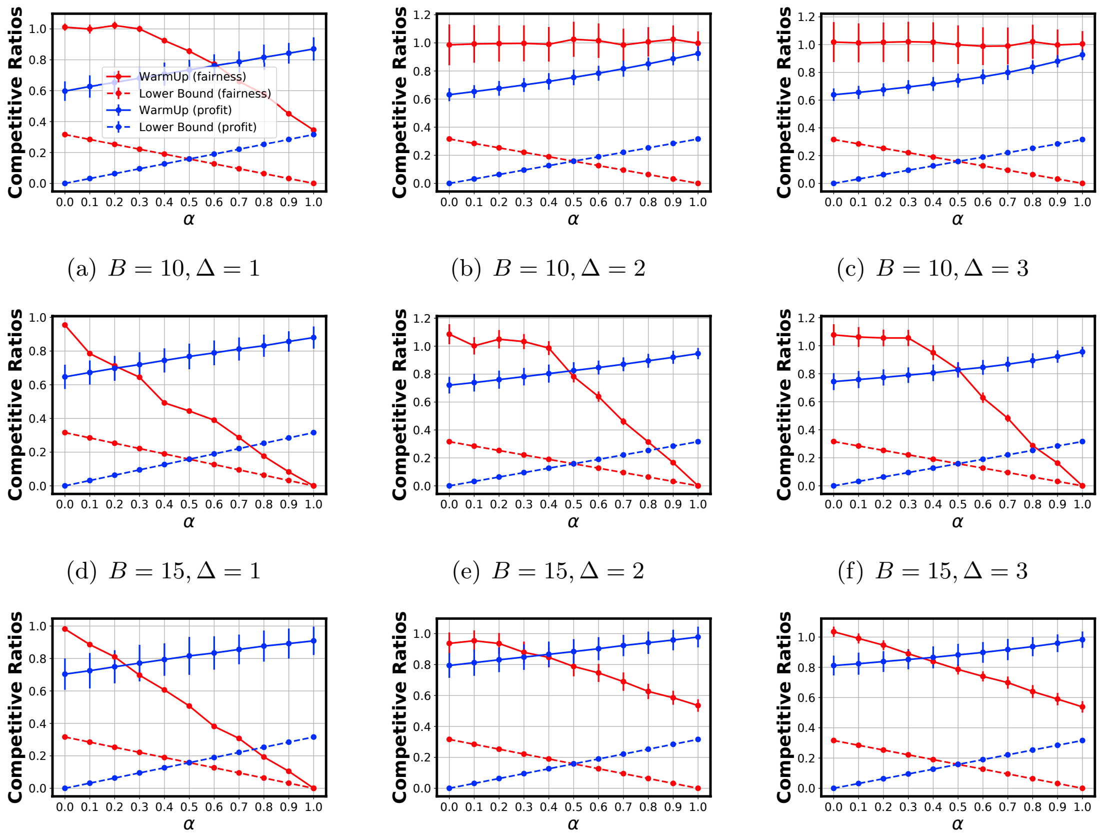


(c) $B = 20$



(d) $B = 25$

Figure 5: Real dataset: performance comparisons with Greedy_P and Greedy_F. (Blue: performance of WarmUp; Red: performance of Greedy_P; Green: performance of Greedy_F)



(a) $B = 10, \Delta = 1$ (b) $B = 10, \Delta = 2$ (c) $B = 10, \Delta = 3$

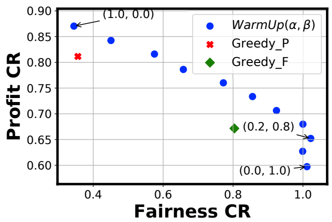
(d) $B = 15, \Delta = 1$ (e) $B = 15, \Delta = 2$ (f) $B = 15, \Delta = 3$

(g) $B = 20, \Delta = 1$ (h) $B = 20, \Delta = 2$ (i) $B = 20, \Delta = 3$

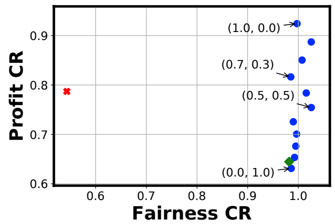
Figure 6: Synthetic dataset: competitive ratios for profit and fairness with different values of α and β with $\alpha + \beta = 1$. (Red solid lines: fairness competitive ratios for WarmUp; Red dotted lines: lower bound of fairness competitive ratios for WarmUp; Blue solid lines: profit competitive ratios for WarmUp; Blue dotted lines: lower bound of profit competitive ratios for WarmUp)

Figure 4 illustrates the competitive ratios of the proposed algorithm with different values of α while keeping β fixed at $1 - \alpha$. Notably, the profit and fairness competitive ratios of WarmUp consistently surpass the corresponding theoretical lower bounds (represented by dotted lines), validating the predictions of Theorem 1. The observed gaps between the performance and the lower bounds indicate that worst-case scenarios, as described by the theory, are rare in real-world settings. It is worth noting that Figure 4(d) demonstrates a tight lower bound match with the fairness performance when $B = 25$ and $\alpha = 1$.

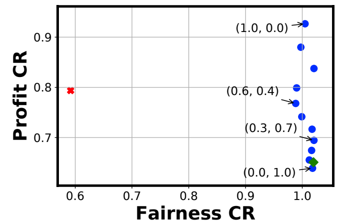
In Figure 5, the profit and fairness performances of WarmUp are compared to those of Greedy_P and Greedy_F. Several interesting observations emerge from this comparison. Firstly, regarding profit, Greedy_P consistently outperforms Greedy_F, while the advantage of Greedy_P over WarmUp becomes more pronounced with larger values of B and less significant with smaller values of B. Since the expected total number of rider arrivals remains constant in our experiments, the parameter B directly controls the driver-rider imbalance. Consequently, when B is larger, indicating more available drivers compared to riders, Greedy_P



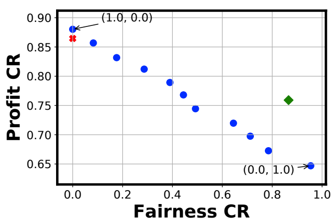
(a) $B = 10, \Delta = 1$



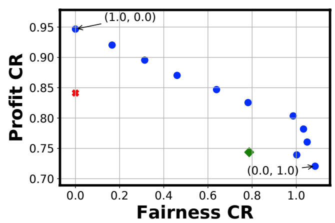
(b) $B = 10, \Delta = 2$



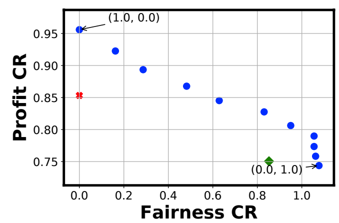
(c) $B = 10, \Delta = 3$



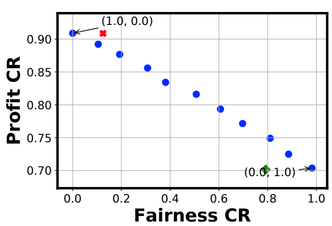
(d) $B = 15, \Delta = 1$



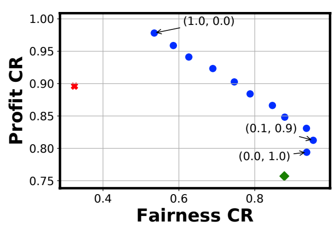
(e) $B = 15, \Delta = 2$



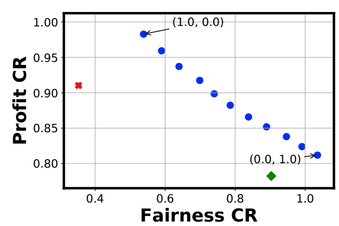
(f) $B = 15, \Delta = 3$



(g) $B = 20, \Delta = 1$



(h) $B = 20, \Delta = 2$



(i) $B = 20, \Delta = 3$

Figure 7: Synthetic dataset: performance comparisons with Greedy_P and Greedy_F. (Blue: performance of WarmUp; Red: performance of Greedy_P; Green: performance of Greedy_F)

emerges as the top performer in terms of profit. Conversely, when B is small, careful policy design is necessary to optimize profit, making WarmUp the dominant choice. Secondly, with respect to fairness, Greedy_F consistently outperforms the other algorithms, although WarmUp exhibits greater flexibility in achieving fairness objectives. Notably, WarmUp demonstrates relatively low sensitivity to the first parameter, α, for profit, but high sensitivity to the second parameter, β, for fairness. This sensitivity is particularly evident when B is large.

Figure 6 showcases the competitive ratios of WarmUp on the synthetic dataset. Since all values are randomly generated, the observed trends align closely with our theoretical results. Generally, as α increases, the profit competitive ratios of WarmUp rise, while the fairness competitive ratios decrease, as depicted in Figure 6. This aligns well with the control power exerted by α and β. Specifically, as α increases, WarmUp becomes more profit-oriented, and vice versa. Another notable observation is that the profit competitive ratios are less sensitive to the value of B, but the gaps between the profit competitive ratios and their theoretical lower bounds tend to narrow as B increases. This phenomenon arises because, with a fixed expected total number of request arrivals, injustices become more prevalent as the average capacities of driver types increase, which occurs when B is larger. It is worth mentioning that, even when α is large, WarmUp can achieve a good fairness competitive ratio by leveraging a uniform patience parameter Δ, as demonstrated in Figure 6(a),6(b), and 6(c).

Finally, Figure 7 showcases the smooth tradeoff between the two objectives achieved by the proposed algorithm when compared to the other benchmarks on the synthetic dataset. Notably, in some special cases, WarmUp outperforms both Greedy_F and Greedy_P in terms of profit and fairness simultaneously, such as when $B = 10$ and $\Delta = 2$ or $\Delta = 3$.

8. Conclusion

This paper presents a comprehensive method that offers flexibility in matching ride requests to drivers, aiming to reconcile the conflicting objectives of maximizing income equality among rideshare drivers while maximizing overall system revenue. Our proposed approaches allow the policy creator to customize the system's fairness and profitability by leveraging two distinct parameters. Through rigorous competitive ratio analyses, we demonstrate that our algorithm, denoted as AttenAlg, achieves a nearly optimal ratio for each individual objective. In other words, there is minimal room for improvement in terms of the competitive ratio for both fairness and profit. Furthermore, we provide extensive experimental results based on both synthetic and real-world datasets. These results not only surpass the theoretical lower bounds but also showcase the ability of our approaches to effectively balance the two objectives by employing natural heuristics. This capability allows for a smooth tradeoff between fairness and profit. Our work suggests several intriguing directions for future research. One immediate avenue involves narrowing the gap between the sum of ratios of profit and fairness achieved by AttenAlg, which currently stands at 0.46. It would be of great interest to develop a more refined online analysis or establish a sharper hardness result indicating that the sum of the two ratios should be significantly lower than 1. Additionally, we invite further exploration into deriving a competitive ratio bound for Boosting. This would likely involve introducing an auxiliary balls-and-bins model for the purpose of analysis.

Acknowledgment

Pan Xu's research was partially supported by NSF CRII Award IIS-1948157. Evan Yifan Xu's research was partially supported by the National Key Research and Development Program of China (No.2022YFB3105002) and the Fundamental Research Funds for the Central Universities. A preliminary version appeared in the 29th International Joint Conference on Artificial Intelligence (IJCAI), 2020 (Xu & Xu, 2020). Compared to the conference version, we added the following into the current work: i) We provide a more formal competitive analysis of the attenuation-based algorithm; ii) We introduce an illustrative example and formally state and prove hardness results; iii) To validate their wide applicability, we rigorously test our algorithms on randomly generated synthetic datasets; iv) We have thoroughly revised significant theoretical analyses and clarifications to enhance the precision of our logical theory demonstration; v) We have incorporated fresh results for both the attenuation-based method and the newly introduced boosting-based algorithm.

References

Aggarwal, G., Cai, Y., Mehta, A., & Pierrakos, G. (2014). Biobjective online bipartite matching. In *International Conference on Web and Internet Economics*, pp. 218–231. Springer.

Ashlagi, I., Burq, M., Dutta, C., Jaillet, P., Sholley, C., & Saberi, A. (2019). Edge weighted online windowed matching. In *Proceedings of the Nineteenth ACM Conference on Economics and Computation*.

Bai, J., So, K. C., Tang, C. S., Chen, X., & Wang, H. (2019). Coordinating supply and demand on an on-demand service platform with impatient customers. *Manufacturing & Service Operations Management*, *21*(3), 556–570.

Bansal, N., Gupta, A., Li, J., Mestre, J., Nagarajan, V., & Rudra, A. (2012). When lp is the cure for your matching woes: Improved bounds for stochastic matchings. *Algorithmica*, *63*(4), 733–762.

Bei, X., & Zhang, S. (2018). Algorithms for trip-vehicle assignment in ride-sharing. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 3–9.

Beirigo, B. A., Schulte, F., & Negenborn, R. R. (2022). A learning-based optimization approach for autonomous ridesharing platforms with service-level contracts and on-demand hiring of idle vehicles. *Transportation Science*, *56*(3), 677–703.

Bernstein, F., DeCroix, G. A., & Keskin, N. B. (2021). Competition between two-sided platforms under demand and supply congestion effects. *Manufacturing & Service Operations Management*, *23*(5), 1043–1061.

Bertsimas, D., Farias, V., & Trichakis, N. (2011). The price of fairness. *Operation Research*, *59*, 17–31.

Bertsimas, D., Farias, V. F., & Trichakis, N. (2012). On the efficiency-fairness trade-off. *Management Science*, *58*, 2234–2250.

Brubach, B., Sankararaman, K. A., Srinivasan, A., & Xu, P. (2020). Attenuate locally, win globally: Attenuation-based frameworks for online stochastic matching with timeouts. *Algorithmica*, *82*(1), 64–87.

Chen, X., Wang, T., Thomas, B. W., & Ulmer, M. W. (2023). Same-day delivery with fair customer service. *European journal of operational research*, *308*(2), 738–751.

Chen, X., Li, Z., Ming, L., & Zhu, W. (2022). The incentive game under target effects in ridesharing: A structural econometric analysis. *Manufacturing & Service Operations Management*, *24*(2), 972–992.

Chen, Y., & Wang, H. (2018). Why are fairness concerns so important? lessons from a shared last-mile transportation system. *Available at SSRN: https://ssrn.com/abstract=3168324*.

Cohen, M. C., Elmachtoub, A. N., & Lei, X. (2021). Price discrimination with fairness constraints. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, p. 2.

Cook, C., Diamond, R., Hall, J., List, J. A., & Oyer, P. (2018). The gender earnings gap in the gig economy: Evidence from over a million rideshare drivers. Tech. rep., National Bureau of Economic Research.

Curry, M., Dickerson, J. P., Sankararaman, K. A., Srinivasan, A., Wan, Y., & Xu, P. (2019). Mix and match: Markov chains and mixing times for matching in rideshare. In *International Conference on Web and Internet Economics*, pp. 129–141. Springer.

Danassis, P., Sakota, M., Filos-Ratsikas, A., & Faltings, B. (2022). Putting ridesharing to the test: Efficient and scalable solutions and the power of dynamic vehicle relocation. *Artificial Intelligence Review*, 1–64.

Dickerson, J. P., Sankararaman, K. A., Srinivasan, A., & Xu, P. (2021). Allocation problems in ride-sharing platforms: Online matching with offline reusable resources. *ACM Transactions on Economics and Computation (TEAC)*, *9*(3), 1–17.

Dickerson, J. P., Sankararaman, K. A., Srinivasan, A., & Xu, P. (2018). Assigning tasks to workers based on historical data: Online task assignment with two-sided arrivals. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*, pp. 318–326.

Didi (2023). More than a journey. https://www.didiglobal.com. [Online]. Accessed: 2023-10-17.

Esfandiari, H., Korula, N., & Mirrokni, V. (2016). Bi-objective online matching and submodular allocations. In *Advances in Neural Information Processing Systems*, pp. 2739–2747.

Fata, E., Ma, W., & Simchi-Levi, D. (2019). Multi-stage and multi-customer assortment optimization with inventory constraints. *CoRR*, *abs/1908.09808*.

Feldman, J., Mehta, A., Mirrokni, V. S., & Muthukrishnan, S. (2009). Online stochastic matching: Beating 1-1/e. In *50th Annual IEEE Symposium on Foundations of Computer Science*, FOCS '09, pp. 117–126.

Feng, G., Kong, G., & Wang, Z. (2021). We are on the way: Analysis of on-demand ride-hailing systems. *Manuf. Serv. Oper. Manag.*, *23*(5), 1237–1256.

FinancesOnline (2023). 47 lyft statistics in 2023: Data on revenue, riders & drivers.. https://financesonline.com/lyft-statistics/. Accessed: 2023-10-17.

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

Grandoni, F., Ravi, R., & Singh, M. (2009). Iterative rounding for multi-objective optimization problems. In *European Symposium on Algorithms*, pp. 95–106. Springer.

Guo, Y., Zhang, Y., Boulaksil, Y., Qian, Y., & Allaoui, H. (2023). Modelling and analysis of online ride-sharing platforms–a sustainability perspective. *European Journal of Operational Research*, *304*(2), 577–595.

H, D. (2023). Charged a fee on uber? how to get a refund on the cancel fee, cleaning fee, and more. https://www.ridesharingdriver.com/uber-fees-cancellation-booking-cleaning-fees/. [Online]. Accessed: 2023-10-17.

Haeupler, B., Mirrokni, V. S., & Zadimoghaddam, M. (2011). Online stochastic weighted matching: Improved approximation algorithms. In *Internet and Network Economics - 7th International Workshop*, WINE '11, pp. 170–181.

Hinchliffe, E. (2017). 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/. [Online]. Accessed: 2023-10-17.

Hosseini, H., Huang, Z., Igarashi, A., & Shah, N. (2023). Class fairness in online matching. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37, pp. 5673–5680.

Jaillet, P., & Lu, X. (2013). Online stochastic matching: New algorithms with better bounds. *Mathematics of Operations Research*, *39*(3).

Korula, N., Mirrokni, V. S., & Zadimoghaddam, M. (2013). Bicriteria online matching: Maximizing weight and cardinality. In *International conference on web and internet economics*, pp. 305–318. Springer.

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

Li, Y., Fu, K., Wang, Z., Shahabi, C., Ye, J., & Liu, Y. (2018). Multi-task representation learning for travel time estimation. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 1695–1704.

Lowalekar, M., Varakantham, P., & Jaillet, P. (2018). Online spatio-temporal matching in stochastic and dynamic domains. *Artificial Intelligence*, *261*, 71 – 112.

Lyft (2023) https://www.lyft.com. [Online]. Accessed: 2023-10-17.

Lyu, G., Cheung, W. C., Teo, C.-P., & Wang, H. (2019). Multi-objective online ride-matching. *SSRN Electronic Journal*.

Ma, J., Li, X., Zhou, F., & Hao, W. (2017). Designing optimal autonomous vehicle sharing and reservation systems: A linear programming approach. *Transportation Research Part C: Emerging Technologies*, *84*, 124–141.

Ma, W., Xu, P., & Xu, Y. (2020). Group-level fairness maximization in online bipartite matching. *ArXiv*, *abs/2011.13908*.

Ma, W., Xu, P., & Xu, Y. (2023). Fairness maximization among offline agents in online-matching markets. *ACM Transactions on Economics and Computation*, *10*(4), 1–27.

Manshadi, V. H., Gharan, S. O., & Saberi, A. (2012). Online stochastic matching: Online actions based on offline statistics. *Mathematics of Operations Research*, *37*(4).

Nanda, V., Xu, P., Sankararaman, K. A., Dickerson, J., & 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*, Vol. 34, pp. 2210–2217.

Ravi, R., Marathe, M. V., Ravi, S., Rosenkrantz, D. J., & Hunt III, H. B. (1993). Many birds with one stone: Multi-objective approximation algorithms. In *Proceedings of the 25nd Annual ACM Symposium on Theory of Computing*, pp. 438–447. Citeseer.

Review, W. P. (2023). New york city population. http://worldpopulationreview.com/us-cities/new-york-city-population/. Accessed: 2023-10-17.

Rosenblat, A., Levy, K., Barocas, S., & Hwang, T. (2016). Discriminating tastes: Customer ratings as vehicles for bias. *Data & Society*, 1–21.

Shao, S., Mittal, A., Twumasi-Boakye, R., & Gupta, A. (2021). Fair pricing of ridehailing services with asymmetric demand and travel time. *IEEE Transactions on Control of Network Systems*, 1–1.

Singer, Y., & Mittal, M. (2013). Pricing mechanisms for crowdsourcing markets. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 1157–1166.

Singla, A., & Krause, A. (2013). Truthful incentives in crowdsourcing tasks using regret minimization mechanisms. In *Proceedings of the 22nd international conference on World Wide Web*, pp. 1167–1178.

Sühr, T., Biega, A. J., Zehlike, M., Gummadi, K. P., & Chakraborty, A. (2019). Two-sided fairness for repeated matchings in two-sided markets: A case study of a ride-hailing platform. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pp. 3082–3092.

Sumita, H., Ito, S., Takemura, K., Hatano, D., Fukunaga, T., Kakimura, N., & Kawarabayashi, K.-i. (2022). Online task assignment problems with reusable resources. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, pp. 5199–5207.

Taylor, T. A. (2018). On-demand service platforms. *Manufacturing & Service Operations Management*, *20*(4), 704–720.

Tong, Y., Shi, D., Xu, Y., Lv, W., Qin, Z., & Tang, X. (2021). Combinatorial optimization meets reinforcement learning: Effective taxi order dispatching at large-scale. *IEEE Transactions on Knowledge and Data Engineering*.

Uber (2023) https://www.uber.com. [Online]. Accessed: 2023-10-17.

Xu, P., & Xu, Y. (2022). Equity promotion in online resource allocation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36, pp. 9962–9970.

Xu, Y., & Xu, P. (2020). Trade the system efficiency for the income equality of drivers in rideshare. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence*, pp. 4199–4205.

Yao, H., Wu, F., Ke, J., Tang, X., Jia, Y., Lu, S., Gong, P., Ye, J., & Li, Z. (2018). Deep multi-view spatial-temporal network for taxi demand prediction. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 2588–2595.

Zhao, B., Xu, P., Shi, Y., Tong, Y., Zhou, Z., & Zeng, Y. (2019). Preference-aware task assignment in on-demand taxi dispatching: An online stable matching approach. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 2245–2252.