

An Open-Data Approach for Quantifying the Potential of Taxi Ridesharing

1. Introduction

While the basic idea of sharing is as old as humankind [1,2], recent developments in information and communication technology (ICT) have led to the emergence and proliferation of new forms of business models based on sharing [1,3]. For example, the companies Uber and Lyft have developed mobile applications, which allow consumers with smartphones to make trip requests, which are then matched to private drivers who use their own cars to fulfill the service requests. These online transportation networks are disrupting the taxi industry, which had remained largely unchanged since its inception [1,4].

The emergence of new sharing-based transportation networks is also in line with the rising need for more sustainable consumption [1,5]. In particular, there are political ambitions to implement more efficient transportation systems and sustainable mobility [6] to counter urban congestion, fuel-wasting, and air-pollution [7,8]. Especially in mega cities, like New York City (NYC), the taxi industry is a considerable contributor to these challenges [9].

Taxi ridesharing (TRS), which matches at least two separate ride requests with similar spatio-temporal characteristics to a joint taxi trip, has been proposed as one means towards more sustainable transportation [10,11]. The idea of TRS is not entirely new; in fact, the idea to share rides has been used in developing countries for several decades [11,12]. However, in recent years several technology-driven approaches have been proposed in order to improve the feasibility of TRS in the developed world. The recent literature has

particularly focused on dynamic ‘many-to-many’ TRS approaches, which allow trips with different start and/or end points to be shared [6,9,11–16]. While these approaches promise a large sharing potential, they also make rigid assumptions on the TRS system, including an a priori availability of data on all taxi trips and a consent of customers to allow dynamic ridesharing. Until now, only a few field tests of dynamic many-to-many ridesharing have been conducted and documented in practice (e.g., *Uber Pool*, *Maaxi*).

Against this backdrop, we argue that a less rigid approach towards TRS may actually be more promising. Our TRS approach, which requires common, or closely co-located, start and end points as the basis for shared rides (one-to-one), was not yet considered in the literature. Although a one-to-one TRS approach has a lower ride sharing potential—compared to many-to-many TRS—we argue that it is easier to implement for transportation providers and likely to enjoy higher acceptance by consumers. It could, as will be seen later, ideally be implemented at airports to share trips to hotels or other points of interest lying in the same area, or, vice versa, to go from points of interest in the city to airports.

In this paper, we develop an innovative concept for one-to-one taxi ridesharing and demonstrate its utility by applying it to an open dataset of more than 5 million historical taxi trips in NYC. Our empirical analysis reveals that the proposed approach can match up to 48.34% of all taxi rides in NYC, saving 2,892,036 kilometers of travel distance, 231,363 liters of gas, and 532,135 kilograms of CO₂ emissions per week. From a decision-support perspective, our proposed approach and our empirical results contribute to enhancing the decision-making of actors at various levels. First, the proposed approach enables taxi operators to estimate the feasibility of one-to-one taxi ridesharing for a given city and various scenarios. The conceptual framework presented here can easily be instantiated to form the backbone of a decision support system that allows for calculating the overall

economic potential of taxi ridesharing as well as for conducting detailed what-if analyses. Second, our approach and empirical results can inform policy making related to urban transportation. Policy makers with access to taxi trip data, for example, national or local transportation authorities, can use the approach presented here and replicate our analyses to support decisions regarding sustainable urban transportation (e.g., subsidizing taxi ridesharing). Third, our findings may convince individual customers of the potential of taxi ridesharing and thereby influence their decision making related to whether to use existing taxi ridesharing systems or not.

The paper unfolds as follows. In Section 2, we discuss the related literature, position TRS as a form of collaborative consumption, and compare and contrast alternative TRS approaches. In Section 3, we describe our one-to-one TRS approach and our data collection and analysis procedures. In Section 4, we present our empirical results regarding the sharing potential of a one-to-one TRS system in NYC. In Section 5, we discuss our findings and implications for research and practice. Section 6 concludes the paper.

2. Ridesharing Approaches

Ridesharing refers to the joint and simultaneous trip of at least one driver and one rider by assigning individuals with similar schedules and itineraries to a shared vehicle [1,4,17–22]. Ridesharing not only reduces travel costs, but by increasing seat occupancy also reduces the total distance traveled and the number of cars on the street. Hence, traffic congestion and environmental pollution can be reduced through ridesharing [8,11,17,19,21,23].

While unorganized ridesharing has existed for decades, organized ridesharing became increasingly popular in recent years. Approaches range from simple online bulletin boards

to complex decision support systems that offer automated matching [11,17,18,20,21,24]. Organized ridesharing can be performed through centralized or decentralized asset provisioning, meaning that matches can either be provided by matching agencies or by ridesharing operators [17].

When matching is performed by matching agencies (such as *Uber*), a two-sided market emerges that matches riders with drivers of self-owned cars [17,20]. This market can involve static carpooling with pre-arranged regular trips, flexible but pre-arranged long-distance ridesharing, or dynamic ridesharing. Through advances in ICT, the latter allows an automatic matching of single, non-recurring, short-notice, on-demand requests. It is also known as ad-hoc ridesharing, real-time ridesharing, instant ridesharing, or dynamic carpooling [1,17,19,21,23,25].

When matching is performed by ridesharing operators, operators use dedicated vehicles and drivers to pick-up and drop-off passengers. Several vehicle sharing services, such as shuttle services, vanpools, or fixed-route transportation exist [17]. Furthermore, they build on other shared vehicle services like traditional taxis, which already have characteristics similar to dynamic ridesharing [6,11,17].

TRS, also known as shared taxi or collective taxi, is an advanced form of public transportation with flexible routing and scheduling that matches at least two separate ride requests with similar spatio-temporal characteristics in real-time to a jointly used taxi, driven by an employed driver without own destination [6,10,16,18,23,26]. TRS, therefore, differs from private ridesharing, which refers to sharing rides among private people [27]. TRS is a more restricted dynamic dial-a-ride approach, which considers the requirements of both multiple passengers and the service provider [6,11]. Because of the pooled simultaneous utilization of a taxi [7], TRS is a form of collaborative consumption.

Modern TRS is enabled by ICT, including the Global Positioning System (GPS), smartphones, mobile apps, and internet platforms [9,19,25,28]. Using these technologies, taxi seekers can submit their desired pick-up and drop-off times and locations. An information system then matches the requests considering the transportation network's state and additional predefined, or customer-defined, matching criteria [11,14,16,23].

There are several advantages connected to TRS. Customers can profit from the same mobility, accessibility, and efficiency as provided by traditional taxis, and from some of the advantages of public transportation [6,10,29]. For example, one advantage is the cost reduction through split fares [6,9,13,14,16,23,30 p. 409]. Furthermore, the overall waiting time in high demand situations can be reduced [6,13,14,30 p. 409]. From a provider's perspective, TRS increases seat occupancy, reduces the number of taxis required, and enables cutting the cumulative trip length and travel time. In turn, this results in the reduction of operational costs, such as fuel consumption and car depreciation [6,8,9,11,14,23]. Shuo Ma et al. [14] and Sun et al. [30 p. 409] argue that even increased revenues may be possible by implementing favorable pricing mechanisms and by better utilizing delivery capacities [14]. Societal benefits include diminishing the negative impacts of taxis on cities by reducing noise and traffic congestion, a reduction of greenhouse emission, and less energy consumption [6,9,11,13,14,16,23,30, p. 409].

On the downside, TRS can also lead to disadvantages. First, there may be an increase in the overall service and waiting time [6,9], leaving passengers concerned about the reliability of TRS [26]. Second, there may be problems related to the privacy and security of traveling with strangers in the same vehicle [9,30 p. 415]. Third, the potential positive environmental effects need to be treated with caution, as the reduced cost of taxi rides could also result in increased demand, or other rebound effects [31,32].

While there are already several commercial operators of TRS services, such as *Bandwagon* in NYC, *Share the Fare* in Australia, or *Taxi for Two* in the UK [9,26], no dominant approach has been established in the field yet [16]. However, researchers have studied various TRS methods over the last ten years. Approaches towards TRS [6–14,16,23,28,29] differ regarding the proposed routing patterns, matching constraints, matching dynamics, and matching objectives. However, most approaches are based on the idea that passengers can embark and disembark anytime during a trip. While implementing such a flexible and dynamic approach has the potential to solve the TRS assignment problem optimally, we argue that implementing such a system is impractical, since customers maybe reluctant to accept picking-up and dropping-off multiple passengers, nearly all of them strangers, during a trip. In addition, a many-to-many approach also comes with operational challenges and complex decision problems for taxi operators (e.g., for optimally combining and rerouting trips on-the-fly).

Therefore, we propose an approach that is based on less rigid assumptions. Our one-to-one approach is focused on sharing rides that start and end at approximately the same location and time, such that people can decide whether to collaboratively consume a taxi ride when embarking or shortly before (e.g., when requesting a taxi via a mobile app). This approach only requires static taxi routing, as the start and endpoints become a bundle of unchangeable requests when a matching is established [8,9,13,33]. Even if the returns to be generated from this approach are a subset of the total—theoretical—TRS benefits, we argue that it would make TRS more feasible to implement. Furthermore, the TRS service would be easier to sell to customers because it reduces their perceived inconvenience [13].

3. Research Method

3.1. Data Collection

To assess the potential of a one-to-one TRS system for large urban areas, we analyzed an open dataset collected by the NYC City Taxi & Limousine Commission (TLC), which records data about their taxi fleet operations on a per-trip basis. Taxi trip data from NYC was first released and described by Donovan and Work [34]. Later the TLC [35] published a more comprehensive dataset, spanning from 2009 to 2015 and covering several hundred million trips completed by the yellow and green cab companies (the latter starting from August 2013). The volume of this dataset makes our approach a spatial big data analysis.

Several authors have analyzed the yellow cab dataset before. Donovan and Work [36] used this dataset to analyze the resilience of the taxi systems to Hurricane Sandy, Ferreira et al. [37] developed a model to visually explore the taxi trips, Zhan et al. [38] develop a model to estimate the link travel times of the taxi trips, and Yazici et al. [39] used it to improve pick-up decisions at John F. Kennedy (JFK) Airport. Other authors analyzed the potentials of more complex TRS methods using this dataset [e.g.: 9,11]. The frequent use of the dataset suggests that it is a reliable and well-accepted source of taxi trip data.

The dataset includes, amongst others, data on taxi license, driver ID, taximeter rate types, start and end time of each trip, number of passengers in each trip, trip duration in seconds, trip distance in miles, GPS coordinates of the pick-up and drop-off locations, meter fares, extras and surcharges, taxes, tips, tolls, and the total amount of money paid [34,40].

We subsampled the overall dataset (consisting of more than 160 million records per year) to reduce processing time and to increase analysis flexibility [41 p. 10]. We selected the first full week of April 2014 (7th to 13th), which was an "ordinary week" without any

holidays or special events. In NYC, 3,364,351 trips took place during this week, which is close to the average of 3.39 million trips per week, as reported by Bloomberg and Yassky [42]. To compare and contrast this week with other time periods, and assess the reliability of our results, we also extracted data for an "extraordinary week" [43 p. 24]. We chose the Christmas week in 2013, which had significantly fewer trips (2,469,270).

3.2. Data Cleansing

Next, the dataset was checked for data quality and cleansed. Following Donovan and Work [34,36], the following data cleansing steps were performed. Some trips were conducted outside of the city borders. Therefore, all trips outside a given coordinate range were dropped. The south-western border was set to [40.477399, -74.25909] and the north-eastern border was specified as [40.917577, -73.700009]. Further cleansing operations were performed to remove unrealistic entries. All trip distances were checked according to their great circle distance, which is the direct distance considering the spherical shape of the earth. There were a few trips where the actual trip distance was smaller than this calculated distance, which is geometrically impossible [36]. Additionally, trips shorter than 200 meters were discarded. Upon manual investigation, those trips often had unrealistic times and/or trip distances. Moreover, all trips with identical start and end points were dropped [9]. Trips with unrealistic durations (i.e., trips lasting less than a minute, longer than 2 hours, and trips with velocities above 144 km/h, which is 40 km/h above the speed limit of 65mp/h) were dropped. Finally, all trips which had 0 passengers and those which had a `rate_code` equal to 5 (individual negotiation price) or 6 (group fare on fixed route) were dropped. These data preparation activities removed around 7.3% of all trips in the ordinary week (7.14% in the extraordinary week).

Figure 1 shows a hexagonal binning histogram of the remaining pick-up and drop-off spots [44 p. 1]. Trips had an average length of 4.8 km (4.9 km in the extraordinary week), while 85% of all trips were smaller or equal to 8 km. Yet, we also observed that taxis were often (about 30% of the times) used for short trips below or equal to 2 km. The average trip duration was 13:27 (12:17 in the extraordinary week) minutes. 46.04% (51.42% in the extraordinary week) of all trips were shorter than 10 minutes. The average price paid for a trip was \$13.46 (\$13.47 in the extraordinary week).

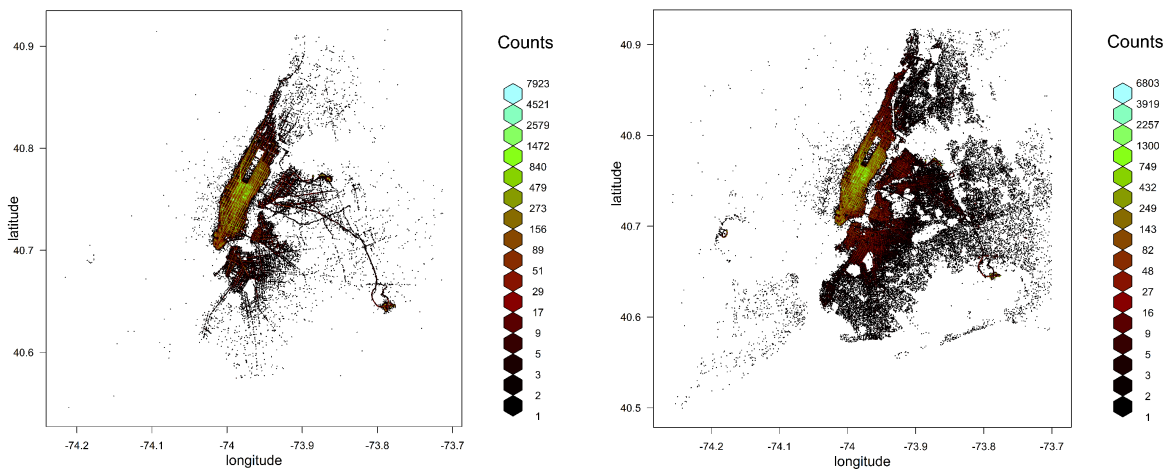


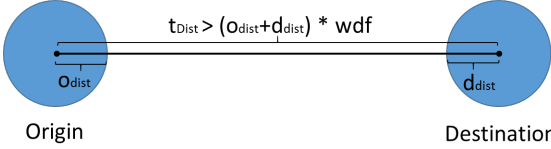
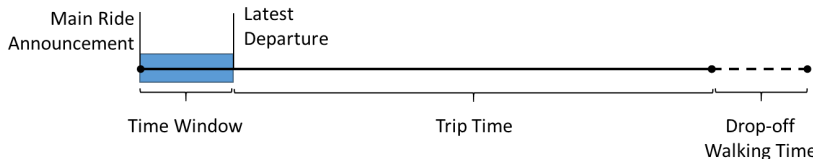
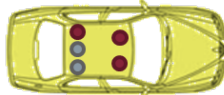

Figure 1. Pick-up (left) and drop-off (right) heat maps (ordinary week)

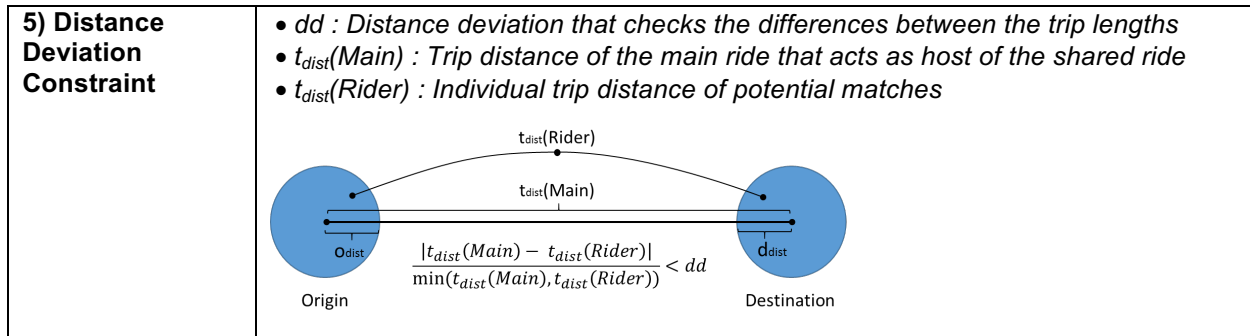
3.3. Matching Constraints

Recall that our goal is to assess the sharing potential of a one-to-one TRS system, where trips are merged if they start at a common origin and end at a common destination. Just as in more complex many-to-many TRS systems, our approach is bound by several constraints. Most importantly, spatio-temporal overlaps of trips are required to be able to merge trips. While in a real-world scenario one might set spatio-temporal constraints on an individual basis, in our analysis we treated them as configurable but static values (Table 1).

First, the distances between the origins (o) and destinations (d) of the individual trips need to be checked [18]. Two trips should only be matched if the walking distance between their origins is smaller than o_{dist} and the walking distance between their destinations is smaller than d_{dist} . Also, we ensured that the total trip distance (t_{dist}) is longer than a portion of the overall walking distance ($wdf * (o_{dist} + d_{dist})$).

Table 1. Constraints of the proposed one-to-one TRS approach

Constraint	Parameter / Formula
1) Spatial Constraints: Pick-up Distance Drop-off Distance Walking Distance Factor	<ul style="list-style-type: none"> • o_{dist} : Maximal walking distance between pick-up locations • d_{dist} : Maximal walking distance between drop-off locations • t_{dist} : Individual trip riding distance • wdf : Walking distance factor, $wdf < 1$ 
2) Temporal Constraint: Time-Window	<ul style="list-style-type: none"> • $timeWindow$: Difference between the trip announcement and the latest possible departure time • $arrivalTime$: Time of the main ride announcement plus the walking time the sharing participants require to reach the taxi • $arrivalTime \leq latestDepartureTime$ 
3a) Car Capacity Constraint	<ul style="list-style-type: none"> • $carCapacity$: Maximum capacity of a taxi, $carCapacity$ is set to 5 for NYC • $sum(passengerAmount) \leq carCapacity$ 
3b) Multi-Ridesharing Constraint	<ul style="list-style-type: none"> • $maxTripShare$: Max amount of trips merged to a shared ride • $count(trips) \leq maxTripShare \in 2-4$ 
4) Feasibility Constraint	<ul style="list-style-type: none"> • $Main-rider$: $individualCosts / 2 + sharingSurcharge \leq individualCosts$ • $Participants$: $mainRideCosts / tripAmount + sharingSurcharge \leq individualCosts$



Second, matched trips need to start at approximately the same time [21]. We added a time-window constraint that specifies the time between the trip announcement and the latest possible departure time (*timeWindow*). To ensure that customers would arrive on time ($arrivalTime \leq latestDepartureTime$), the approximate walking time between the pickup points was calculated and added to the difference. While other TRS approaches define a constraint on the maximum delay in trip duration or the actual drop-off time [6,9,11,13,14], a one-to-one approach as presented here does not require this extra constraint, since by definition shared trips end at approximately the same destination and time. Since the walking time to the drop-off location is already bounded by the distance constraint and a constant walking speed is assumed, an additional constraint is not required.

Third, our approach also considers the capacity of taxis (*carCapacity*). While the car capacity may vary in reality, a fixed car size was assumed. As stated by D'Orey et al. [6] and the TLC [45], the maximum taxi capacity in NYC is five, which we used to restrict the number of combined passengers in a shared ride ($sum(passengerAmount)$). Similar to Chen et al. [8] and Santi et al. [9], another constraint defines the maximum number of individual trips (*maxTripShare*) that can be shared in a single ride (e.g. pairwise, triple, or quadruple matches). We did not consider personal constraints such as smoking attitudes or

gender, as suggested by Lalos et al. [23], Santos and Xavier [16] and Tao and Wu [28], since our dataset did not provide these attributes.

Fourth, we designed a simple and easy-to-understand pricing mechanism that evenly splits the cost of a shared ride between the participants. To prevent that the reduction of individual rides had a negative impact on the operators' and drivers' incomes, we added a constant sharing surcharge that needs to be paid by each passenger. Similar to the approach of Shuo Ma et al. [14], this simple pricing scheme provides taxi companies and drivers with more profit per ride, while reducing the expenses for individual passengers. Other TRS approaches either ensure that trips are profitable, that the price paid is smaller than the price of the original trip, or that the price is high enough to pay the driver [7,11,13,16].

Similarly, our one-to-one TRS approach ensures that sharing is worthwhile for the passengers. Especially on short rides, the additional sharing surcharge might result in shared costs higher than the costs of individual trips. For this reason, we assume that a rider is only willing to host a shared trip if a single match (*tripAmount* = 2) is enough to reduce his/her individual costs. Comparable to this, a rider would only be willing to join a shared ride if the current amount of matched rides (*tripAmount*) is high enough to reduce his/her expenses (*individualCosts*).

Fifth, only trips that ended at the same destination without long detours to intermediate destinations were matched. Hence, a distance deviation constraint (*dd*) was added that checks the differences between the individual trip lengths. Without this check, unrealistically profitable trips would be merged, biasing the analysis results.

3.4. Matching Process

The matching constraints described in the last chapter served as a basis for our matching process. Instead of preferring candidates that satisfy a given optimization criterion (such as the saved distance or time) with mixed integer programs or maximum matching approaches [e.g.: 9,10,14], similar to Shuo Ma et al. [14], all trips were analyzed in chronological order, using a greedy first-come-first-served (FCFS) heuristic. Therefore, we ordered and numbered the trips according to their pick-up times. This method simulates a real-time TRS system and thereby mimics the properties of a real-world system, in which trips are merged as quickly as possible and in which future trips are unknown. For each incoming ride, our matching algorithm searched for matching candidates within the trips that have been announced in the specified time-window. If a candidate was found that satisfied all constraints, the rides were shared. Else, the new ride itself was considered as a host for a shared ride with trips that were announced later. Overall, our trip matching process can be divided into four phases (Figure 2):

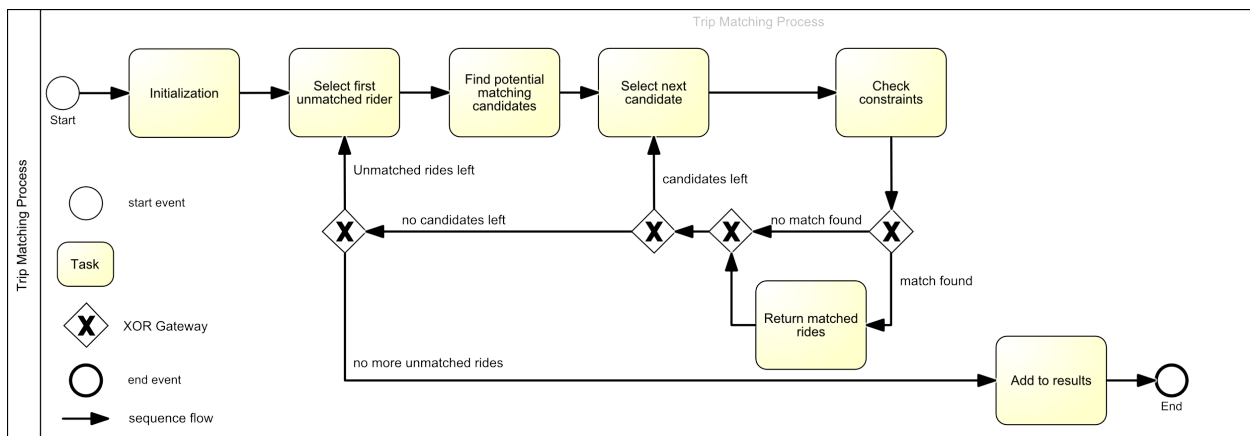


Figure 2. The trip matching process modelled with Business Process Model and Notation (BPMN)

1. **Initialization:** This step ensured that the prerequisites of the algorithm, such as clearing old cached results, preparing the input and output data structures, and initializing algorithm components—such as a spatial index for the pick-up locations—were met.

2. **Loop over unmatched riders:** We considered each unmatched rider as a potential host for a shared trip. Therefore, we processed the trips in ascending order of their pick-up times. In this step, we used a spatial index to approximate the pick-up distance constraint and thus, optimized the search strategy. This method was inspired by the idea of national grid reference systems [46] and can be compared to a simplified version of the “lazy shortest path calculation” presented by Shuo Ma et al. [14]. With this approach, we restricted the distance calculation and the check of the other constraints to a quickly identified narrow set of matching candidates.
3. **Loop over potential matching candidates:** For each matching candidate we checked the constraints explained before (including the pick-up distance constraint). If all constraints were satisfied, we added the matching candidate to the trip of the host.
4. **Return the overall matching results:** In the last phase, we evaluated and stored the result data structure containing the shared rides for further analyses.

The properties of our one-to-one TRS approach made the calculation less complex in comparison to other approaches that assume different start or destination spots [e.g., 6,13,36]. Since a one-to-one approach does not have to calculate the actual routes, the necessity to calculate the shortest path became redundant. Likewise, the static routing characteristic removed the need to track the current status and position of the taxis that were already matched [e.g.: 11,14]. This property not only significantly simplified our matching algorithm, but it also reduces the runtime complexity of a potential real-world system. While the many-to-many approach with dynamic routing suggested by Shuo Ma et al. [14] required a spatio-temporal index to find matches quickly, our one-to-one approach with static routing just needs a spatial index. In a real-world system, each incoming ride

request would be assigned to a fixed pick-up grid cell and already started or fully matched trips would not need to be re-considered by the algorithm.

Our approach is based on a few assumptions. First, we assume that all taxis that can be shared are actually shared and that every participant arrives on time. Thus, our results calculate an upper bound of the total sharing potential that can be realized with the proposed approach. Second, since the distance constraints focus on short walking distances (between 200 and 3,000 meters), we assumed that only the straight-line distance between the coordinate points needs to be considered. However, as the Earth is a three-dimensional object with a spherical surface, we could not calculate the distance using the Euclidean distance. Instead, we used the “great circle distance” to represent the direct path on the earth’s surface called “geodesic” [46, p. 21-23].

3.5. Scenarios and What-if Analysis

We specified a reference scenario for the proposed one-to-one TRS service and then varied the parameters (Table 1) of this scenario to quantify their influence on the results. For this reference scenario, it seemed reasonable to assume a maximum walking distance of 500 meters for both the pick-up and the drop-off area to introduce some flexibility, while limiting inconvenience for passengers. We assumed that passengers are not willing to share a ride if he/she would need to walk more than half of the trip length just to participate in TRS. We set the time-window to 10 minutes since it seemed likely that passengers would have to wait anyways, either standing in queue at a taxi stand or waiting for a taxi to arrive. To reduce the waiting times, passengers were matched pair-wise, as soon as the first fitting trip was found. Besides that, the maximum capacity of a taxi was set to 5 passengers. In

line with taxi fares in NYC, \$2.50 was assumed to be an acceptable sharing surcharge. The distance derivation constraint was set to 0.2.

Starting from this reference scenario, we introduced variations of all input parameters to assess their influence on the results. Usually, a sensitivity analysis would be used to measure the behavior of a system to small changes of parameters. However, since this analysis highly depends on the simulation and modeling of an artificial TRS service, the approach utilized for the analysis here can more closely be compared to a what-if analysis. In general, a what-if analysis allows the data-intensive simulation of a complex system and inspects the results under different scenarios [47].

4. Data Analysis and Results

4.1. Reference Scenario

First, the reference scenario was evaluated both for the ordinary and the extraordinary week (Table 2). The data indicate that a considerable number of trips can be shared (48% for the ordinary week and 43% for the extraordinary week), resulting in a substantial ride reduction (753,860 rides for the ordinary week and 497,734 rides for the extraordinary week). Passenger occupancy is increased, saving almost 3 million kilometers of trip distance, 231,363 liters of gas, and 532,135 kg of CO₂ emissions in an ordinary week. Taxi companies would save 22.42% of travel time at the expense of 12.19% of total fares, while the fares for shared trips would increase by 53.24%. The results for both the ordinary and the extraordinary week were similar. Even though the April week had 26.49% more individual trips, the relative matching results of the extraordinary week were quite similar for both the taxi companies and the customers, suggesting that our approach scales well with an increasing number of trips.

Our TRS approach also benefits customers (Table 3). The average walking distance for passengers to the pick-up location and from the drop-off location is 550 meters. In total, this results in an average travel time increase of 5:31 minutes. Assuming a sharing surcharge of \$2.50, those inconveniences are compensated by an average cost saving of 23.77%.

Table 2. Reference scenario, taxi company metrics

Metrics	Ordinary week	Extraordinary week
Total individual trips	3,119,254	2,292,934
Total shared rides	1,507,720	995,468
% shared rides	48.34	43.41
Total ride reduction	753,860	497,734
% ride reduction	24.17	21.71
Ø passenger occupancy (individual)	1.71	1.8
Ø passenger occupancy (shared)	2.25	2.3
Total distance saved (km)	2,892,036.11	1,864,658.11
Ø saved kilometers per shared ride	3.84	3.75
% distance saved	18.98	16,47
Saved gas (liters)	231,362.89	149,172.65
Saved CO2 (kg)	532,134.64	343,097.09
Saved ride time (h)	156,965	94,628.98
% saved ride time	22.42	20.14
% revenue reduction	12.19	10.68
Ø increase trip fare for shared rides (%)	53.24	53.53

Table 3. Analysis results for the basic case, customer metrics

Metrics	Ordinary week	Extraordinary week
Total increase travel time (h)	138,741.02	90,689.97
Ø increase travel time (min)	5.52	5.47
% increase travel time (min)	2.69	2.8
Total waiting time (h)	72,007.09	47,751.64
Ø waiting time (min)	5.73	5.76
Total walking time (h)	69,445.03	44,700.34
Ø walking time (min)	5.53	5.39
Total walking distance (km)	416,670.20	268,202.02
Ø walking distance (km)	0.55	0.54
Ø participant cost savings (%)	23.77	23.56

4.2. What-if Analysis

4.2.1. Influence of the Distance Constraint

Reducing the distance constraint to 200 meters results in a huge reduction of shared trips in both the ordinary and the extraordinary week. Instead of a trip reduction of 24.17% (reference scenario for the ordinary week), a reduction of only 6.25% can be reached. In contrast, when increasing the distance constraint to 2,000 or 3,000 meters, saturation is reached quickly. The highest matching rate, with a 34.74% trip reduction and an average passenger occupancy of 2.62, is achieved with the distance constraint set to 3,000 meters. With a distance constraint of 1,000 meters the amount of saved rides is 32.06%, with an average passenger amount of 2.51 per trip.

The waiting time increased from 5:14 minutes at 200 meters to 6:05 for 1,000 meters and then stabilizes at around 6 minutes for greater distances. Also, the walking distance including the pick-up and drop-off distance increase with greater distance constraints. Overall, the participants' financial savings increase from 22.21% at 200 meters to 26.31% at 3,000 meters. The average distance saved increases from 3.54 km to 4.85 km.

4.2.2. Influence of the Length of the Time Window

To identify a suitable time window for the TRS service, the pick-up time constraint was varied. The reference scenario used a time constraint of 10 minutes and reached a 24.17% reduction in rides. In contrast, a 20-minutes time window leads to a 30.80% reduction and a 30-minutes window results in a reduction of 33.48%. For smaller time windows, the matching rate decreased quickly. A window of two minutes leads to a reduction of only 2.83%, and a five-minute window to a reduction of 13.85%.

While the average waiting time is 1:20 for a 2-minutes time window, longer windows lead to longer waiting times (10:58 minutes for a 30-minutes window). On average, it takes 3:03 minutes to find a match. The walking distance is between 340 and 570 meters.

In the extraordinary week, a lower trip density causes a longer waiting time for a small time window of 2 minutes. Yet, a time window of 30 minutes results in a longer waiting time of 11:41 minutes. Also, the average distance saved per ride increases with the time window, increasing the participants' cost savings.

4.2.3. Influence of Car Capacity

We assumed that car capacity would have a considerable influence on shared trips in a densely populated mega-city like NYC. To test this assumption, the reference scenario was tested under decreased and increased car capacities. We found that even with increased car capacity most shared rides would still have carried only two passengers. But the seat capacity had an impact on the matching rate. While the difference between a capacity of 4 and 5 was quite low (0.8% difference in ride reduction), increasing it to 6 improved the result by 1.9%. With a maximum seat capacity of 7, an improvement of 3.3% could be reached (with an average passenger occupancy of 2.36 instead of 1.71 without the TRS service). Increasing car capacity had no negative impact on waiting time and walking distance.

4.2.4. Influence of Multi-Ridesharing Approaches

When allowing triple TRS, 39.26% of all shared trips consist of three trips (Figure 3). Allowing quadruple matching results in 12.32% of all shared rides being quadruple rides, while reducing the volume of pairwise and triple matches. In sum, allowing multi-ridesharing improves the matching rate in the ordinary and extraordinary week. While the reference scenario achieves a 24.17% reduction in rides and an average seat occupancy of 2.25,

triple matching (28.6% reduction, average seat occupancy 2.39) and quadruple matching (29.53% reduction, average seat occupancy 2.42) slightly improve these figures.

However, multi-ridesharing has a negative impact on waiting times. By allowing quadruple matching, the average waiting time increases from 5:44 minutes to 6:54 minutes. In contrast, the average walking distance remains stable at around 550 meters. However, the inconveniences for passengers are reduced by increasing the savings per passenger due to a higher total distance saved per trip. Also, the operator profits from multi-ridesharing by increasing the fare of shared trips by 63.74% in the triple case (66.63% in the quadruple case), instead of 53.24% in the pairwise case.

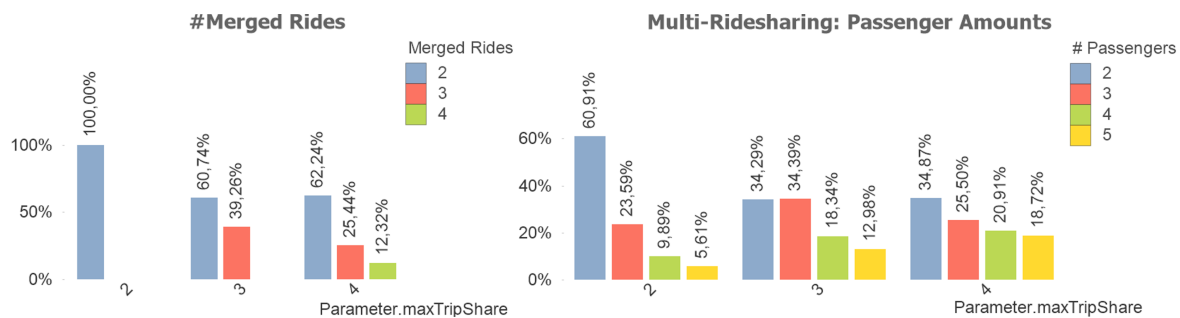


Figure 3. Amount of shared rides (left) and capacity usage (right) in multi-ridesharing

4.2.5. Influence of Sharing Surcharges

By analysing of various surcharge amounts, we found a negative correlation between the sharing surcharge and the amount of matched trips, as can be expected.

By reducing the surcharge to \$1.5, a small increase in the amount of reduced rides can be reached (0.95% in the ordinary week). This reduction, however, has a positive impact on the average cost savings for passengers, increasing them from 23.77% to 33.51%. But, a surcharge reduction has an even bigger impact on the average fare of the combined ride.

Whereas in the reference scenario the income of a driver from shared rides increased by 53.24%, a surcharge of \$1.5 leads to an increase of only 32.73%.

The opposite behavior is observed when increasing the surcharge. A surcharge of \$3.5 results in a drop of the ride reduction from 24.17% to 19.42%. It has a negative impact on the passengers' savings (18.13%), but a positive impact on the average increase of the shared trip fares (65.59%). While the slope of the average passenger's savings slightly decreased, a clear saturation could however not be discovered.

4.2.6. Influence of Week Days and Time of Day

In the dataset, the number of trips increases from Mondays to Saturdays and reduces again on Sundays. Even though weekdays have a slightly better chance for sharing trips, no single day has a clearly better matching rate.

Interestingly, when analyzing the data on an hourly basis, we discovered a correlation between trip density and the matching rate (Figure 4), such that there is a huge reduction of trips and matches between 4 and 5 a.m. However, the influence of the trip volume on the matching rate decreases for higher volumes. For instance, the increase in the amount of trips from 5 p.m. to 11 p.m. did not result in a large increase of the matching rate in comparison to the time between 7 a.m. and 5 p.m. This could suggest a saturation for larger trip volumes.

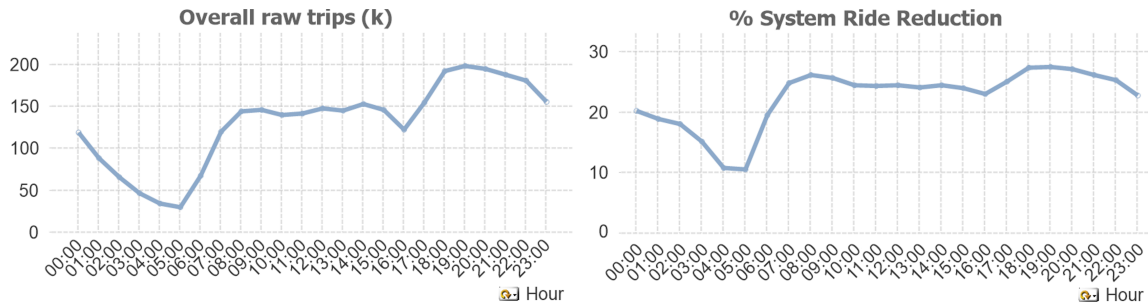


Figure 4. Raw trips vs. relative ride reduction per hour

Even if the matching rate is smaller during the night, it has to be noted that during this time the highest average distance savings can be realized (6.85 km at 4 a.m., 7.24 km at 5 a.m.). This is probably due to the fact that the average distance is higher during that period.

4.2.7. Influence of Pick-Up and Drop-Off Locations

Most taxi trips are located in central Manhattan, but the best ridesharing opportunities exist in the southern part of Manhattan, and, to a lesser extent, in Brooklyn and Queens. Almost no potential for sharing yellow cabs was found in New Jersey and Newark, although it hosts the Newark airport. Outside of the city center, most matches were found at the LaGuardia and JFK airports.

Considering these findings, a TRS service in NYC should mainly focus on the central areas and the two main airports. By restricting the TRS service only to the southern part of Manhattan, a global ride reduction of 23.48% would have been reached. In fact, 24.93% of the trips that start in Manhattan could have been saved, increasing the average passenger occupancy from 1.71 to 2.28. This would have meant 2,513,373 saved kilometers and a reduction of trip times of 23.42%. The average distance saved by a shared ride would have been 3.43 km. Therefore, the average participant savings would have ended up at 23.27% with a 54.27% increase of the ride fare for the taxi driver. The participants would have needed to wait on average 5:44 minutes or walk 560 meters.

Operating the TRS service only at the LaGuardia airport would result in a 19.48% reduction of the trips that start at the airport. This would have increased the average passenger occupancy from 1.7 to 2.11 and contributed additional 0.39% to the system-wide reduction. Shared trips would have saved on average 16.03 km, summing up to a total of 192,618 km in the ordinary week and reducing the overall distance from the airport by 19.9% and the travel time by 20.41%. The average waiting time for the passengers would have been 5:02 minutes and the average walking distance 450 meters. Participants would have saved 42.70% of their money, increasing the total fare of the shared trips by 14.54%.

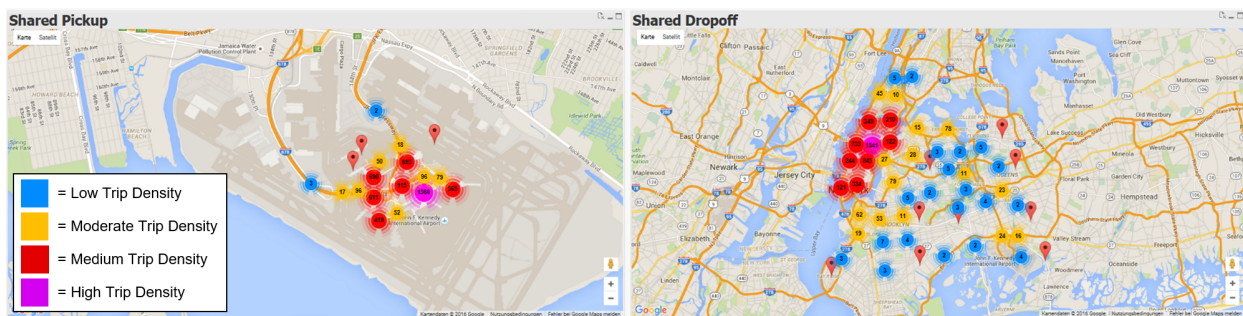


Figure 5. Cluster map of potentially shared trips from the JFK Airport (hot spots highlighted)

JFK Airport would have achieved lower results. By operating the TRS service only from this spot, 12.55% of all trips that started at the airport could have been saved, increasing the average occupancy from 1.73 to 1.98. Figure 5 shows the start and endpoints of these trips on a cluster map as generated in our data discovery tool. Most of the trips started at the JFK main entrance (1366 trips) and most ended at the central of Manhattan (1841 trips). This trip reduction would have contributed an additional 0.18% to the overall result. Thereby, 13.53% of the trip distance and 13.83% of the trip time could have been saved. The average waiting time for the main riders would have been 5:40 minutes and the walking distance 520 meters. Overall, the passengers could have saved 45.22% of the individual trip fare, and the total fare of the shared ride would have been increased by 9.49%.

A taxi company could also consider operating a TRS service for trips that go to certain destinations and start at arbitrary locations. In fact, 23.8% of all trips that went to *LaGuardia* Airport could have been saved by such a service, increasing the passenger occupancy to the airport from 1.7 to 2.23. This reduction would have meant a 24.11% reduction of the trip distance (111184 km in total) and 24.86% less ride time. The hosts would have needed to wait for 5:55 minutes or walk 440 meters. The participant savings of this service would have been 34.80%, and the total trip fare of the shared ride would have been increased by 13.13%. In the case of the *John F. Kennedy* airport, 12.75% of all trips could have been saved. Thereby, the average passenger occupancy could have been increased from 1.78 to 2.04. Also here, 13.34% of the ride time and 12.66% of the trip distance could have been saved with an average distance saving of 27.87 km per rideshare. The participants on average would have waited for 6:14 minutes, walked 530 meters and saved 45.43% of their expenses. The income from a single shared ride would have increased by 9.09%.

5. Discussion

5.1. Basic Properties of a One-to-One TRS Service in NYC

As outlined by Chen et al. [8], our analysis revealed that the number of vehicles in the pick-up area is an important factor for the successful matching of rides in our proposed one-to-one approach. From all constraints, the distance constraint, which directly influences the number of trips available for matching, had the strongest impact on the overall matching rate. While a maximum walking distance of 200 meters leads to a ride reduction of only 6%, increasing the distance constraint to 1000 meters, the maximum distance a customer can reasonably walk within 10 minutes, leads to a ride reduction of about 32%.

Furthermore, our results showed that a TRS provider needs to tweak both the distance and time constraint simultaneously. Naturally, the more patient a passenger is, the more matching opportunities arise; an effect that is called densification [9]. Our results show that shortening the time window constraint to 5 minutes has a strong negative impact on the matching rate. Hence, it could be argued that a 10-minute time window is a good choice for a TRS service in NYC. Yet, the time constraint does not only limit the impact of the distance constraint, but its configuration also decides upon a fair distribution of the inconveniences of the hosts and guests of a shared ride. While the distance constraint increases the walking distance for passengers who want to join a ride, the time constraint mainly impacts on the waiting time of the host passenger. In the reference scenario, the waiting and walking times were almost equal (5:44 and 5:32 minutes). However, increasing the distance constraint to 1,000 meters created a disadvantage for passengers joining a hosted ride. Hence, according to our analysis, taxi service providers should not increase the distance constraint any further than 750 meters without simultaneously increasing the time window constraint.

We also found that multi-ridesharing has a positive impact on TRS. While it only slightly increased the amount of matched rides, it also increased the average seat occupancy, the saved distance, the average participant savings, and the fare of combined rides. Yet, it also caused longer waiting times for customers. For this reason, a one-to-one TRS provider could consider using multi-ridesharing in situations where demand exceeds supply, or if the customer is willing to wait longer to increase savings. Our results also showed that multi-ridesharing has higher business potential in dense settings.

A TRS provider would also have to decide on the right sharing surcharge. In dynamic many-to-many ridesharing, increasing the cost-saving threshold impacts more negatively on the matching rate than on the saved distances [25]. In our one-to-one TRS approach,

increasing the sharing surcharge had a smaller impact on the overall distance saved than on the matching rate. In fact, it reduced the amount of matched short distance trips and thus improved the average increase of the combined trips' total fares. Choosing an appropriate sharing surcharge is, therefore, an essential part of implementing TRS [26].

5.2. Implementing TRS in Scenarios with Low Spatio-Temporal Density

Santi et al. [9] showed that 25% of the daily taxi trips in NYC (around 100,000 trips) are enough to reach a near-maximum matching with a flexible many-to-many TRS approach. Hence, they concluded that TRS could also be efficient in cities with low densities. In contrast, Stiglic et al. [18] showed that the amount of matched rides in classic ridesharing greatly depends on the spatio-temporal distribution density of the announced trips.

Our results showed that trip density has a considerable impact on one-to-one TRS. A lower trip density leads to a reduced matching rate and, thus, decreased most of the metrics we tracked. For example, while we reached a matching rate of 52.52% at 8 a.m., at 5 a.m. we were only able to match 21.57% of the trips. Yet with an increasing amount of trips, our results also showed a decreasing impact of the trip density on the matching rate.

Related research suggests that ridesharing approaches may fail because of a low probability to find other passengers to share a ride with [18,26]. Since passengers have to wait for potential matches, this observation especially applies to our proposed one-to-one TRS approach. Thus, the approach is based on the assumption that enough route overlaps exist [28,30 p. 409,48]. Our analysis showed that the proportion of matched rides in NYC can reach up to 48.34%, which provides a strong indication that the proposed approach can be implemented successfully. However, as our temporal and spatial analysis showed, taxi

companies have to perform thorough robustness checks in order to identify the best configuration of a one-to-one TRS system.

Taxi companies should also consolidate shared rides by considering hotspots like transportation hubs or special landmarks as start and end points of their rides [17,21]. In theory, it has been shown that carefully selecting meeting points can significantly increase the amount of matched rides in classic ridesharing scenarios [18]. Especially at intermodal transportation hubs, approaches of resource pooling like ridesharing are good strategies to deal with irregular, uncertain, and distributed demand [7]. Also in practice, (manually) coordinated taxi sharing at transportation hubs or events is already commonplace [30 p. 409]. Our results show that intermodal transportation hubs like train stations and airports, tourist attractions, and hospitals are good starting points to establish a one-to-one TRS service. Also, considering airports as destinations of shared rides could be a good strategy to improve the matching success in both dense and less dense situations. Such locations must allow people to meet and taxis to wait for their final departure. Another possibility to deal with low-density situations could be to use fixed route TRS, as is already done at some group ride taxi stands in NYC or by an operator in Iran [49,50].

5.3. Generating Business Success with TRS

The main motivation for a TRS service may be economic benefits [3,26]. These benefits not only apply to customers, but also to car owners and taxi drivers. In fact, every participant involved in taxi operations can be assumed to strive for profit maximization [6].

Our results show that in the reference scenario fares from combined trips would increase by 53.43%, due to multiple paying customers in one car and the sharing surcharge. At the same time, we have to acknowledge that taxi drivers may resist

participating in ridesharing due to an overall reduction of demand for trips (24.17% in the reference scenario) [26]. Yet, the decrease in demand for trips may be alleviated by several factors. First, a fare reduction may lead to a long-term increase in the demand for trips (rebound effect). Flores-Guri [51] estimated the price elasticity of the taxi demand in NYC to be between -0.22 and -1.05 [31]. Consequently, while the reference scenario would have removed 24.17% of the original rides, it can be assumed that the passengers' savings of 23.77% could have increased the long-term demand by 5.23% to 24.95%. Furthermore, the demand for taxis in NYC clearly exceeds supply [52]. Hence, it can be assumed that the reduction of individual trips is smaller than the number of saved trips.

In addition, the disruptive nature of the sharing economy can be a strong motivator to establish TRS [1,53]. In NYC, the number of taxi licenses is capped, significantly increasing their value. In fact, a license in NYC is traded for up to 1 million US dollars. However, the emergence of sharing economy business models like Uber pose a risk to the value of those licenses [12,52]. Besides that, Uber claims that their drivers earn more [54]. A TRS approach could help the traditional taxi companies to stay competitive against such rivals.

Finally, there maybe further incentives for both customers and taxi companies to participate in TRS. For example, the government could subsidize the lost revenue as a reward for the reduction in CO₂ emissions and traffic congestion. The sharing economy and climate change promote practices like sustainable consumption and so-called sharing cities [1,3,53]. This is why cities could be interested in subsidizing TRS business models.

5.4. Comparison of One-to-One and Many-to-Many TRS approaches

Our proposed one-to-one TRS service can be compared to the more complex approach proposed by Santi et al. [9]. Since the approach used different settings and time spans, they

are only comparable to a limited extent. First, our results show clearly that a one-to-one TRS approach is feasible in NYC. A time window of 10 minutes and a walking distance constraint of 500 meters is sufficient to match 48.34% of the trips and thereby save 2,892,036 km trip distance in just one week. In contrast, the approach of Santi et al. [9] is able to match 30% of rides with a one-minute system response time and a maximal travel time increase of 2 minutes. By changing the maximal increase in the travel time to 5 minutes, a matching rate of about 90% was possible, resulting in a saving of 32% of the total travel time. In contrast, with considerably less rigid constraints, our one-to-one approach was able to save 22.42% of overall travel time and resulted in an average increase of individual travel time by 5.52 minutes. Comparing these findings, it is questionable whether the efforts of implementing a many-to-many TRS approach, which is based on much more rigid assumptions regarding data availability, customer acceptance, and computational complexity, are worthwhile in densely populated mega-cities like NYC.

6. Conclusion

In this paper, a data-driven approach for quantifying the potential of one-to-one taxi-ridesharing (TRS) was presented. Our results show that a one-to-one TRS approach can generate considerable benefits for passengers, drivers, taxi companies, and society. Amongst others, our approach reduced the number of rides by 20-25%.

To test the applicability of our approach in other cities, we used a publically available taxi data from Porto [55]. While the dataset contains data from only one taxi company, the results suggest that, due to lower trip density, our one-to-one approach would have achieved a ride reduction of 3.39% in an ordinary week. Therefore, we assume that the approach could rather be introduced for a limited, yet frequently used set of spots (e.g.,

intermodal transportation hubs like airports and train stations). These spots can be identified with our decision support system, too (see Section 4.2.7). While our approach allows for selectively implementing one-to-one TRS, further research is required to identify the minimum trip density that is required for fully implementing the proposed approach in a city.

Our approach is based on a few assumptions. First, while the data-driven approach mimics a dynamic matching approach, our approach assumes that all trips that could have been shared were matched. This is unlikely to happen in reality since participants may be unaware of the possibility to share a taxi, or unwilling to share a ride with strangers. Also, the additional inconveniences induced by sharing a ride may impede acceptance.

Ultimately, the proposed TRS approach needs to be subjected to field tests. A field test would allow measuring the actual success of the proposed service under realistic circumstances. First, the impact of the approach on the number of satisfied ridesharing requests could be measured, to test the assumption that the price elasticity and the excess demand do indeed encounter the reduction of individual trips [14,18]. Also, the field test could be used to figure out incentives that drive the early acceptance of the approach [6]. As an alternative, the presented data-driven implementation could be replaced by simulations that mimic the real-world behavior of the different participants more closely.

7. References

- [1] B. Cohen, P. Muñoz, Sharing cities and sustainable consumption and production: towards an integrated framework, *J. Clean. Prod.* (2015) 1–11. doi:10.1016/j.jclepro.2015.07.133.
- [2] J.A. Price, Sharing: The Integration of Intimate Economies, *Anthropologica.* 17 (1975) 3. doi:10.2307/25604933.
- [3] J. Hamari, M. Sjöklint, A. Ukkonen, The sharing economy: Why people participate in collaborative consumption, *J. Assoc. Inf. Sci. Technol.* (2015) 1–13. doi:10.1002/asi.23552.

- [4] H.A. Posen, Ridesharing in the Sharing Economy: Should Regulators Impose Über Regulations on Uber?, *Iowa Law Rev.* 101 (2015) 405–433. http://ilr.law.uiowa.edu/files/ilr.law.uiowa.edu/files/ILR_101-1_Posen.pdf.
- [5] R. Botsman, *The Sharing Economy Lacks A Shared Definition*, Co.EXIST. (2013). <http://www.fastcoexist.com/3022028/the-sharing-economy-lacks-a-shared-definition#20> (accessed March 18, 2016).
- [6] P.M. D'Orey, R. Fernandes, M. Ferreira, Empirical evaluation of a dynamic and distributed taxi-sharing system, in: 2012 15th Int. IEEE Conf. Intell. Transp. Syst., IEEE, 2012: pp. 140–146. doi:10.1109/ITSC.2012.6338703.
- [7] S.-F. Cheng, D.T. Nguyen, H.C. Lau, A Mechanism for Organizing Last-Mile Service Using Non-dedicated Fleet, in: 2012 IEEE/WIC/ACM Int. Conf. Web Intell. Intell. Agent Technol., IEEE, 2012: pp. 85–89. doi:10.1109/WI-IAT.2012.254.
- [8] P.-Y. Chen, J.-W. Liu, W.-T. Chen, A Fuel-Saving and Pollution-Reducing Dynamic Taxi-Sharing Protocol in VANETs, in: 2010 IEEE 72nd Veh. Technol. Conf. - Fall, IEEE, 2010: pp. 1–5. doi:10.1109/VETEFC.2010.5594422.
- [9] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S.H. Strogatz, C. Ratti, Quantifying the benefits of vehicle pooling with shareability networks, *Proc. Natl. Acad. Sci.* 111 (2014) 13290–13294. doi:10.1073/pnas.1403657111.
- [10] C.-C. Tao, C.-Y. Chen, Heuristic Algorithms for the Dynamic Taxipooling Problem Based on Intelligent Transportation System Technologies, in: Fourth Int. Conf. Fuzzy Syst. Knowl. Discov. (FSKD 2007), IEEE, 2007: pp. 590–595. doi:10.1109/FSKD.2007.346.
- [11] H. Hosni, J. Naoum-Sawaya, H. Artail, The shared-taxi problem: Formulation and solution methods, *Transp. Res. Part B Methodol.* 70 (2014) 303–318. doi:10.1016/j.trb.2014.09.011.
- [12] L.M. Martinez, G.H.A. Correia, J.M. Viegas, An agent-based simulation model to assess the impacts of introducing a shared-taxi system: an application to Lisbon (Portugal), *J. Adv. Transp.* 49 (2015) 475–495. doi:10.1002/atr.1283.
- [13] D.O. Santos, E.C. Xavier, Taxi and Ride Sharing: A Dynamic Dial-a-Ride Problem with Money as an Incentive, *Expert Syst. Appl.* 42 (2015) 6728–6737. doi:10.1016/j.eswa.2015.04.060.
- [14] Shuo Ma, Yu Zheng, O. Wolfson, T-share: A large-scale dynamic taxi ridesharing service, in: 2013 IEEE 29th Int. Conf. Data Eng., IEEE, 2013: pp. 410–421. doi:10.1109/ICDE.2013.6544843.
- [15] Y. Lin, W. Li, F. Qiu, H. Xu, Research on Optimization of Vehicle Routing Problem for Ride-sharing Taxi, *Procedia - Soc. Behav. Sci.* 43 (2012) 494–502. doi:10.1016/j.sbspro.2012.04.122.
- [16] D.O. Santos, E.C. Xavier, Dynamic Taxi and Ridesharing: A Framework and Heuristics for the Optimization Problem, in: F. Rossi (Ed.), *IJCAI 2013, Proc. 23rd Int. Jt. Conf. Artif. Intell. August 3-9, 2013, IJCAI/AAAI, Beijing, China, 2013*: pp. 2885–2891. <http://www.aaai.org/ocs/index.php/IJCAI/IJCAI13/paper/view/6779>.

- [17] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, S. Koenig, Ridesharing: The state-of-the-art and future directions, *Transp. Res. Part B Methodol.* 57 (2013) 28–46. doi:10.1016/j.trb.2013.08.012.
- [18] M. Stiglic, N. Agatz, M. Savelsbergh, M. Gradisar, The benefits of meeting points in ride-sharing systems, *Transp. Res. Part B Methodol.* 82 (2015) 36–53. doi:10.1016/j.trb.2015.07.025.
- [19] K. Ghoseiri, A. Haghani, M. Hamedi, Real-Time Rideshare Matching Problem, Maryland, 2011. <http://www.mautc.psu.edu/docs/UMD-2009-05.pdf>.
- [20] T. Teubner, C.M. Flath, The Economics of Multi-Hop Ride Sharing, *Bus. Inf. Syst. Eng.* 57 (2015) 311–324. doi:10.1007/s12599-015-0396-y.
- [21] N. Agatz, A. Erera, M. Savelsbergh, X. Wang, Optimization for dynamic ride-sharing: A review, *Eur. J. Oper. Res.* 223 (2012) 295–303. doi:10.1016/j.ejor.2012.05.028.
- [22] C. Morency, The ambivalence of ridesharing, *Transportation (Amst.)* 34 (2007) 239–253. doi:10.1007/s11116-006-9101-9.
- [23] P. Lalos, A. Korres, C.K. Datsikas, G.S. Tombras, K. Peppas, A Framework for Dynamic Car and Taxi Pools with the Use of Positioning Systems, in: 2009 Comput. World Futur. Comput. Serv. Comput. Cogn. Adapt. Content, Patterns, IEEE, 2009: pp. 385–391. doi:10.1109/ComputationWorld.2009.55.
- [24] D. Woskow, Unlocking the sharing economy: An independent review, Department for Business, Innovation and Skills, London, 2014. https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/378291/bis-14-1227-unlocking-the-sharing-economy-an-independent-review.pdf.
- [25] N.A.H. Agatz, A.L. Erera, M.W.P. Savelsbergh, X. Wang, Dynamic ride-sharing: A simulation study in metro Atlanta, *Transp. Res. Part B Methodol.* 45 (2011) 1450–1464. doi:10.1016/j.trb.2011.05.017.
- [26] S. Wright, J.D. Nelson, An investigation into the feasibility and potential benefits of shared taxi services to commuter stations, *Urban, Plan. Transp. Res.* 2 (2014) 147–161. doi:10.1080/21650020.2014.908736.
- [27] R. Belk, You are what you can access: Sharing and collaborative consumption online, *J. Bus. Res.* 67 (2014) 1595–1600. doi:10.1016/j.jbusres.2013.10.001.
- [28] C. Tao, C. Wu, Behavioral responses to dynamic ridesharing services - The case of taxi-sharing project in Taipei, in: 2008 IEEE Int. Conf. Serv. Oper. Logist. Informatics, IEEE, 2008: pp. 1576–1581. doi:10.1109/SOLI.2008.4682777.
- [29] C.-C. Tao, Dynamic Taxi-Sharing Service Using Intelligent Transportation System Technologies, in: 2007 Int. Conf. Wirel. Commun. Netw. Mob. Comput., IEEE, 2007: pp. 3204–3207. doi:10.1109/WICOM.2007.795.
- [30] X. Sun, D. Golightly, S. Sharples, B. Bedwell, User Requirements and Constraints for On-Demand Taxi Sharing Technology, in: *Contemp. Ergon. Hum. Factors 2012 Proc. Int. Conf. Ergon. Hum. Factors*, Blackpool, UK, 2012: pp. 409–416. <https://books.google.com/books?hl=de&lr=&id=IUXLBQAAQBAJ&pgis=1>

(accessed December 7, 2015).

- [31] L.A. López, T. Domingos, M.Á. Cadarso, J.E. Zafrilla, The smarter, the cleaner? Collaborative footprint: A further look at taxi sharing, *Proc. Natl. Acad. Sci.* 111 (2014) E5488–E5488. doi:10.1073/pnas.1420242112.
- [32] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S.H. Strogatz, C. Ratti, Reply to Lopez et al.: Sustainable implementation of taxi sharing requires understanding systemic effects, *Proc. Natl. Acad. Sci.* 111 (2014) E5489–E5489. doi:10.1073/pnas.1421300112.
- [33] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S.H. Strogatz, C. Ratti, Supporting Information for: Quantifying the benefits of vehicle pooling with shareability networks, *Proc. Natl. Acad. Sci.* 111 (2014) 13290–13294. <http://www.pnas.org/content/suppl/2014/08/28/1403657111.DCSupplemental/pnas.1403657111.sapp.pdf>.
- [34] B. Donovan, D.B. Work, New York City Taxi Trip Data (2010-2013), (2014). doi:10.13012/J8PN93H8.
- [35] NYC Taxi & Limousine Commission, TLC Trip Record Data, (2016). http://www.nyc.gov/html/tlc/html/about/trip_record_data.shtml (accessed March 18, 2016).
- [36] B. Donovan, D.B. Work, Using coarse GPS data to quantify city-scale transportation system resilience to extreme events, in: *Transp. Res. Board 94th Annu. Meet.*, 2015. <http://arxiv.org/abs/1507.06011>.
- [37] N. Ferreira, J. Poco, H.T. Vo, J. Freire, C.T. Silva, Visual Exploration of Big Spatio-Temporal Urban Data: A Study of New York City Taxi Trips, *IEEE Trans. Vis. Comput. Graph.* 19 (2013) 2149–2158. doi:10.1109/TVCG.2013.226.
- [38] X. Zhan, S. Hasan, S. V. Ukkusuri, C. Kamga, Urban link travel time estimation using large-scale taxi data with partial information, *Transp. Res. Part C Emerg. Technol.* 33 (2013) 37–49. doi:10.1016/j.trc.2013.04.001.
- [39] M.A. Yazici, C. Kamga, A. Singhal, A big data driven model for taxi drivers' airport pick-up decisions in New York City, in: *2013 IEEE Int. Conf. Big Data*, IEEE, 2013: pp. 37–44. doi:10.1109/BigData.2013.6691775.
- [40] NYC Taxi & Limousine Commission, Data Dictionary – Yellow Taxi Trip Records, (2015). http://www.nyc.gov/html/tlc/downloads/pdf/data_dictionary_trip_records_yellow.pdf (accessed March 9, 2016).
- [41] R. Bivand, E. Pebesma, V. Gomez-Rubio, *Applied Spatial Data Analysis with R*, Springer New York, New York, NY, 2008. doi:10.1007/978-0-387-78171-6.
- [42] M.R. Bloomberg, D. Yassky, *2014 Taxicab Fact Book*, New York, New York, USA, 2014. http://www.nyc.gov/html/tlc/downloads/pdf/2014_taxicab_fact_book.pdf.
- [43] C. Pete, C. Julian, K. Randy, K. Thomas, R. Thomas, S. Colin, R. Wirth, *Crisp-Dm 1.0 - Step by step data mining guide*, 2000. <https://www.the-modeling-agency.com/crisp-dm.pdf>.

- [44] N. Lewin-Koh, Hexagon Binning: an Overview, (2011). https://cran.r-project.org/web/packages/hexbin/vignettes/hexagon_binning.pdf (accessed February 2, 2016).
- [45] NYC Taxi & Limousine Commission, FAQs - Passenger, (2016). http://www.nyc.gov/html/tlc/html/faq/faq_pass.shtml (accessed March 18, 2016).
- [46] J.J. Mwemezi, Y. Huang, Optimal facility Location on Spherical Surfaces: Algorithm and Application, *New York Sci. J.* 4 (2011) 21–28. <http://isomase.org/JOMase/Vol.1 Nov 2013/1-3.pdf>.
- [47] M. Golfarelli, S. Rizzi, A. Proli, Designing what-if analysis, in: *Proc. 9th ACM Int. Work. Data Warehous. Ol. - Dol. '06*, ACM Press, New York, New York, USA, 2006: pp. 51–58. doi:10.1145/1183512.1183523.
- [48] B. Cici, A. Markopoulou, E. Frías-Martínez, N. Laoutaris, Quantifying the potential of ride-sharing using call description records, in: *Proc. 14th Work. Mob. Comput. Syst. Appl. - HotMobile '13*, ACM Press, New York, New York, USA, 2013: p. 1. doi:10.1145/2444776.2444799.
- [49] City of New York, Yellow Taxicab Rate of Fare, (2016). http://www.nyc.gov/html/tlc/html/passenger/taxicab_rate.shtml (accessed March 18, 2016).
- [50] A. Gholami, A.S. Mohaymany, Analogy of fixed route shared taxi (taxi khattee) and bus services under various demand density and economical conditions, *J. Adv. Transp.* 46 (2012) 177–187. doi:10.1002/atr.157.
- [51] D. Flores-Guri, An Economic Analysis of Regulated Taxicab Markets, *Rev. Ind. Organ.* 23 (2003) 255–266. doi:10.1023/B:REIO.0000031368.93775.0a.
- [52] S. Bryant, How NYC's Yellow Cab Works and Makes Money, (2015). <http://www.investopedia.com/articles/professionals/092515/how-nycs-yellow-cab-works-and-makes-money.asp> (accessed March 18, 2016).
- [53] C.J. Martin, The sharing economy: A pathway to sustainability or a nightmarish form of neoliberal capitalism?, *Ecol. Econ.* 121 (2015) 149–159. doi:10.1016/j.ecolecon.2015.11.027.
- [54] R. Lawler, Uber Study Shows Its Drivers Make More Per Hour And Work Fewer Hours Than Taxi Drivers, (2015). <http://techcrunch.com/2015/01/22/uber-study/> (accessed March 18, 2016).
- [55] Taxi Service Trajectory Prediction Challenge, Porto Taxi Trajectory Dataset Specifications, (2015). <http://www.geolink.pt/ecmlpkdd2015-challenge/dataset.html> (accessed March 18, 2016).