



## Research Article

# Modeling and Prediction of Ride-Sharing Utilization Dynamics

Tal Altshuler,<sup>1</sup> Yaniv Altshuler ,<sup>2</sup> Rachel Katoshevski,<sup>1</sup> and Yoram Shiftan ,<sup>1</sup>

<sup>1</sup>Department of Civil and Environmental Engineering, Israel Institute of Technology, Israel

<sup>2</sup>MIT Media Lab, USA

Correspondence should be addressed to Yaniv Altshuler; yanival@mit.edu

Received 22 February 2019; Accepted 29 May 2019; Published 28 November 2019

Guest Editor: Olivier Orfila

Copyright © 2019 Tal Altshuler et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The potential of an efficient ride-sharing scheme to significantly reduce traffic congestion, lower emission level, and drivers' stress, as well as facilitating the introduction of *smart cities* has been widely demonstrated in recent years. Furthermore, ride sharing can be implemented within a sound economic regime through the involvement of commercial services that creates a win-win for all parties (e.g., *Uber*, *Lyft* or *Sidecar*). This positive thrust however is faced with several delaying factors, one of which is the volatility and unpredictability of the potential benefit (or utilization) of ride-sharing at different times, and in different places. Better understanding of ride-sharing dynamics can help policy makers and urban planners in increase the city's "ride-sharing friendliness" either by designing new ride-sharing oriented systems, as well as by providing ride-sharing service operators better tools to optimize their services. In this work the following research questions are posed: (a) Is ride-sharing utilization stable over time or does it undergo significant changes? (b) If ride-sharing utilization is dynamic can it be correlated with some traceable features of the traffic? and (c) If ride-sharing utilization is dynamic can it be predicted ahead of time? We analyze a dataset of over 14 million taxi trips taken in New York City. We propose a dynamic travel network approach for modeling and forecasting the potential ride-sharing utilization over time, showing it to be highly volatile. In order to model the utilization's dynamics, we propose a network-centric approach, projecting the aggregated traffic taken from continuous time periods into a feature space comprised of topological features of the network implied by this traffic. This feature space is then used to model the dynamics of ride-sharing utilization over time. The results of our analysis demonstrate the significant volatility of ride-sharing utilization over time, indicating that any policy, design, or plan that would disregard this aspect and chose a static paradigm would undoubtedly be either highly inefficient or provide insufficient resources. We show that using our suggested approach it is possible to model the potential utilization of ride sharing based on the topological properties of the rides network. We also show that using this method the potential utilization can be forecasting a few hours ahead of time. One anecdotal derivation of the latter is that perfectly guessing the destination of a New York taxi rider becomes nearly three times easier than rolling a "Snake Eyes" at a casino.

## 1. Introduction

The increasing availability of portable technologies gives new fuel to studies on metropolitan transportation optimization, pushing urban design one step closer towards the long sought concept of "smart cities" [1, 2]. Mobile devices and ubiquitous connectivity make it easier than ever to collect data on the way people live in cities and big-data analytic methods facilitate the extraction of actionable insights from it. City administrators and policy makers can in turn act upon such results to enhance city management, channeling current advancements in data analysis for the immediate improvement of urban quality of life.

Many of the fundamental problems in big cities nowadays relate to cars. The high number of vehicles congests the streets,

vehicles standing in traffic jams increase air pollution while also increasing traveling times, significantly increasing passengers' stress levels. Availability of large-scale datasets accompanied with recent advancements in the analysis of big data and the development of novel models of human mobility give rise to new possibilities to study urban mobility.

Such new models include for example the work of [3] in which large-scale mobile phone data were analyzed in order to characterize individual mobility, show that human travel patterns are far from random, and are efficiently describable by a single spatial probability distribution. Similarly, [4] show that mobile phone data can be used as a proxy to examine urban mobility and [5] analyzes social network data of different cities to find that mobility highly correlates with the

distribution of urban points of interest. Mobile technologies are also the enablers of many successful consumer applications, such as Waze [6], that provide traffic-aware city navigation by using data provided by the community. Alternative ways of moving in the city, such as autonomous mobility-on-demand and short-term car rental have been identified among the possible solutions to the ever-growing transport challenge [7].

Ride sharing has the potential of improving traffic conditions by reducing the number of vehicles on the roads, reducing the emission of CO<sub>2</sub> and the fuel consumption per person, and giving the riders the opportunity to socialize with people (that otherwise would have been fierce “road competitors”). A recent study [8] shows that traffic in the city of Madrid can be reduced by 59% if people are willing to share their home-work commute ride with neighbors. Even if they are not willing to ride with strangers, but only with friends of friends (for safety issues), the potential reduction is still up to 31%. Another recent study [9] had shown that on-demand route-free public transportation based on mobile phones outperforms standard fix-route assignment methods when comparing traveling times. These results encourage the deployment and policies supporting ride sharing in urban settings.

However, despite such evidence and others, ride-sharing adoption rate in cities worldwide is slower than what can be expected given the clear benefits of ride-sharing [10, 11]. One important reason, as suggested by [12–14] and others is the uneven, and often unstable, potential benefits associated with ride-sharing. When the value that can be extracted from using a service such as *Lyft* [15], *Uber* [16], or *Sidecar* [17] is high at one part of the city, but significantly lower at another neighborhood, or worse—suddenly decreases for a period of two days—potential users of the service are much likely to opt for a private car usage [18].

In this work we propose a data-driven framework to dynamically predict the impact, or *potential utilization*, of ride sharing in a city, at different times, and in different regions. Specifically, the technique we propose provides both policy makers as well as ride-sharing operators tools for assessing the future benefit of ride-sharing, encapsulated through the percent of rides saveable through merging of nearby departures and destinations. Simply put, a shared taxi service can use this proposed technique in order to know ahead of time what the ride-sharing demand is going to be (at various places in the city), whereas municipal services can dynamically change tolls and service fees in order to incentivize the use of ride-sharing in “low hours” that are predicted in advance.

Our method is based on analyzing the network features of the dynamic O-D matrix as represented by data collected by various sources, such as mobile phone call records, or sensors mounted on the taxis themselves. In our research, we show a clear correlation between such properties and the portion of “merge-able rides”. We have analyzed the efficacy of our proposed network-oriented method using a dataset of over 14 million taxi trips taken in New York City during January 2013 [19].

This work is structured as follows: Section 2 presents an overview of the relevant related research in the field. In Section 3, we discuss the data and analytic methodologies that

were used for this work: starting with the calculation of the average ride-sharing potential as a function of the maximum delay a taxi-user would be willing to sustain, we demonstrate that more than 70% of the rides can be shared when users are willing to undertake up to 5 minute delay. We then demonstrate that urban ridesharing potential is not only highly dynamic, but that it can also be predicted using the analysis of the rides that took place in the city a few hour beforehand. We present a method for comprising a dynamically changing network using the taxi-rides, and analyzing the topological properties of this network (Section 4). We analyze the dynamics of these properties over time, and demonstrate our ability to accurately predict changes in the utilization of ride-sharing several hours in advance. Concluding remarks and suggestions for future works are contained in Section 5.

## 2. Related Work

Network features can signal and are often used to predict events or properties that are external to the network, but influence it. A network can often be built on easily available data and serve as an important source for predictions regarding various (seemingly unrelated) events and large-scale decision-making processes [20–22]. Features of a phone call network can signal the occurrence of an emergency situation or predict trust among individuals [23], and specific behaviors in a Twitter account can identify a spammer [24]. Such discoveries had sparked the interest of researchers in different research fields, who could benefit from this new ability to model large-scale human dynamics. One of the fields most influenced by this evolving research thrust was the data-driven study of human mobility and its potential application for Intelligent Transportation Systems [11, 25–28].

It has been recently shown that in trying to detect semantic network events (such as an accident or a traffic jam) it is crucial to understand the underlying structure of the network these events are taking place at [29, 30], the role of the link weights [31], as well as the response of the network to node and link removal [32]. Past research [33] had pointed out the existence of powerful patterns in the placement of links, or that clusters of strongly tied together individuals tend to be connected by weak ties [31]. It was also shown that this finding provides insight into the robustness of the network to particular patterns of link and node removal, as well as into the spreading processes that take place in the network [34, 35]. In addition, recent work had demonstrated the trade-off between the number of individuals (the width of the data) and the amount of information available from each one (the depth of the data), with respect to the ability to accurately model crowds behavior [36–38]. An analytical approach to this problem discussing the (surprisingly large) amount of personal information that can be deduced by an “attacker” who has access to one’s personal interactions’ meta-data can be found in [39–41].

One of the first works that examined the statistical distribution of event appearance in mobility and communication networks have found that these follow a power law principle [42], and that such distribution is significantly affected by anomalous events that are external to the networks [43]. A

method for filtering mobile phones Call Data Records (CDRs) in space and time using an agglomerative clustering algorithm in order to reconstruct the origin-destination urban travel patterns was recently suggested in [44].

Recent works that have been analyzing data collected by the pervasive use of mobile phones have broadly supported the notion that most of human mobility patterns are affected by a relatively small number of factors, easily modeled, and very predictable [4, 45–47]. A comprehensive survey of ride-sharing literature can also be found in [48] and another recent relevant study that developed spatial, temporal, and hierarchical decomposition solution strategy for ride-sharing is presented in [49].

To-date, much of the research related to ride sharing has focused on understanding the characteristics of ride-sharing trips and users. In a recent survey of app-based, on-demand rideshare users in San Francisco, researchers found that 45% of ridesharers stated they would have used a taxi or driven their own car had ridesharing not been available, while 43% would have taken transit, walked, or cycled [50].

A recent work by Santi et al. [51] introduces a way of quantifying the benefits of sharing. The study applies to a GPS dataset of taxi rides in New York City and uses the notion of *shareability network* to quantify the impact and the feasibility of taxi-sharing. When passengers have a 5 minutes flexibility on the arrival time, and they are willing to wait up to 1 minutes after calling the cab, over 90% of the sharing opportunities can be exploited and 32% of travel time can be saved. The authors have also shown that the problem is computationally tractable when we look for sharing a taxi among two people with the option of in-route picking up. Furthermore, sharing solutions involving more people are not tractable, but do not provide a significant improvement with respect to solutions involving only two people. Similar results have been demonstrated using a theoretical model analyzing *Autonomous Mobility On Demand* system, demonstrating that a combined predictive positioning and ridesharing approach is capable of reducing customer service times by up to 29% [52].

An extensive simulation infrastructure for ride-sharing analysis is suggested in [53], allowing the initialization and tracking of a wide variety of realistic scenarios, monitoring the performance of the ride-sharing system from different angles, considering different stakeholders interests and constraints. The simulative infrastructure is claimed to use an optimization algorithm that is linear in the number of trips and makes use of an efficient and fully parallelized indexing scheme.

In another study by Cici et al. [8] mobile phone data and social network data were used to estimate the benefits of ride sharing on the daily home-work commute. Mobile phone data are easier to collect than GPS traces, and have a higher penetration, providing a good sample of a city mobility. Social network data are used to study the effect of friendship on the potential of ride sharing, showing that if people want to travel only with friends then expected ride-sharing benefits are negligible. On the other hand, when people are willing to ride with friends of friends the achieved efficiency resembles this of the variant that also allows riding with strangers (implying

that safety issues may have significant effect on the actual success of a ride-sharing solution).

A similar study has been presented by [54] calculating shareability curves using millions of taxi trips in New York City, San Francisco, Singapore, and Vienna, showing that a natural rescaling collapses them onto a single, universal curve. The authors presented a model that predicts the potential for ride sharing in any city, using a few basic urban quantities and no adjustable parameters.

The issue of pricing policies in ride-sharing services have gained significant attention recently, with the booming expansion of commercial ride-sharing services such as *Uber*, *Lyft* and others. The work of [55] studies dynamic pricing policies for ride-sharing platforms. As such platforms are two-sided this requires economic models that capture the incentives of both drivers and passengers. In addition, such platforms support high temporal-resolution for data collection and pricing. The combination of the latter requires stochastic models that capture the dynamics of drivers and passengers in the system.

In [56] the authors highlight the impact of the demand pattern of the underlying network on the platforms optimal profits and aggregate consumer surplus. In particular, the authors establish that both profits and consumer surplus are maximized when the demand pattern is balanced across the networks locations. In addition, the authors show that profits and consumer surplus are monotonic with the “balancedness” of the demand pattern (as formalized by the patterns structural properties).

The work of [57] proposes a recommendation framework to predict and recommend whether and where should ride-sharing users wait in order to maximize their chances of getting a ride. In the framework, a large-scale GPS data set generated by over 7,000 taxis in a period of one month in Nanjing, China was utilized to model the arrival patterns of occupied taxis from different sources.

The recent work of Alexander and Gonzalez [11] uses smart-phone data in order to model the behavior of an urban population in Boston, in an attempt to assess the impact of efficient ride-sharing service on the urban traffic, and specifically on the expected levels of congestion. This data-centric approach leads to a highly accurate modeling of the mobility patterns in the city. However, much like most of the recent work on this subject, the researchers have followed an aggregative modeling, that tries to find the static long-term definitive mobility patterns, purposely omitting any dynamic fluctuations.

In another study, researchers from the Microsoft Research Center [58] analyzed the ride data of 12,000 taxis during 110 days in order to model the mobility patterns of potential passengers. Using this probabilistic model, the researchers were able to build a recommendation system for taxi drivers that would maximize their profits (yielding an overall 10% improvement in the overall profits) and a second recommendation system for passengers, advising them where to turn in order to maximize their chances of finding a vacant taxi (with 67% accuracy). A similar research can be found in [59].

A recent review of dynamic ridesharing systems [60] focused on the optimization problem of finding efficient

matches between passengers and drivers. This ride-matching optimization problem determines vehicle routes and the assignment of passengers to vehicles considering the conflicting objectives of maximizing the number of serviced passengers, minimizing the operating cost, and minimizing passenger inconvenience. Another study [61] presented an algorithm that increases the potential destination choice for ride-sharing schemes set by considering alternative destinations that are within given space-time budgets.

On a similar note, a recent study [62] analyzed the benefits of meeting points in ride-sharing systems, investigating the potential benefits of introducing meeting points in a ride-sharing system. With meeting points, riders can be picked up and dropped off either at their origin and destination or at a meeting point that is within a certain distance from their origin or destination. The increased flexibility results in additional feasible matches between drivers and riders, and allows a driver to be matched with multiple riders without increasing the number of stops the driver needs to make. A similar approach for the optimization of such meetings points was discussed in [63].

The challenge of rides-matching was also discussed in works such as [64, 65] or [66], which have demonstrated that 2,000 vehicles (15% of the taxi fleet in New York) of capacity 10 passengers (or 3,000 vehicles of capacity of 4 passengers) can serve 98% of the New York taxi demand within a mean waiting time of 2.8 minutes and mean trip delay of 3.5 min.

A *path merging* approach, which instead of merging rides to and from the same locations calculate new paths which go through the same locations of the original trips, at the same order, and thus improves the ability to merge rides, was discussed in [67].

In a recent theoretical study [68] where the combinatorial optimization of ridesharing matching problem was tackled using the proof of the equivalence between classical centroid clustering problems and a special case of set partitioning called metric k-set partitioning, in which an efficient expectation maximization algorithm was used to achieve a 69% reduction in total vehicle distance, as compared with no ridesharing.

A fully decentralized reputation-based approach is discussed in [69], using a peer-to-peer architecture to provide self-assembling ride-sharing infrastructure capable of functioning with no central authority or regulator.

### 3. Dataset and Methodology

Our analysis was performed using a dataset of 14,776,615 taxi rides collected in New York City over a period of one month (January 2013) [19]. Each ride record consists of the following fields: pick-up time, pick-up longitude, pick-up latitude, drop-off longitude, drop-off latitude, number of passengers per ride, average velocity, and overall trip duration. Times granularity is second-based and positional information has been collected via GPS technology by the data provider. From this raw data sample, we omit records containing missing or erroneous GPS coordinates, as well as records that represent rides that started or ended outside Manhattan, yielded a cleaned dataset containing 12,784,243 rides.

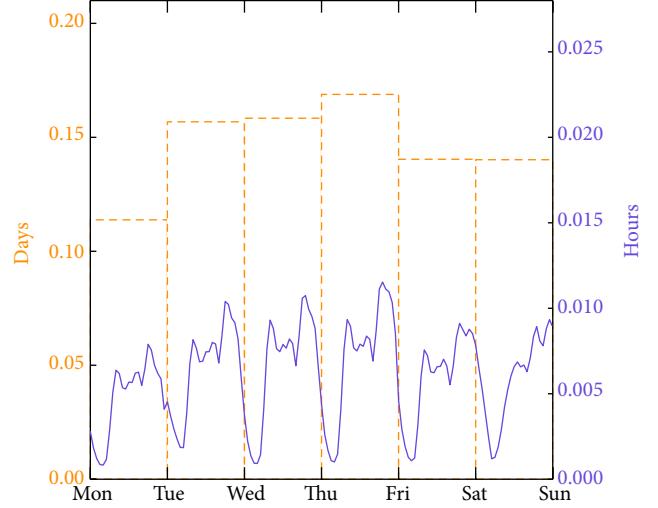


FIGURE 1: Probability Density Function (PDF) of the number of rides per day of week/hour of day. Afternoon peaks are centered on average around 7 pm.

As a first step in modeling the feasibility and efficiency of ride-sharing schemes using taxi rides in New York City, a comprehensive understanding of the data itself is required. How do the rides distribute over the various geographic locations? Are there patterns that emerge when observing the O'D matrix of the various rides? Can we use those in order to predict the destinations of passengers when they board a taxi at a certain location? The figures below attempt to answer some of the Power Law distribution) strongly implying on the potential of a network-centric approach as the method of choice with respect to the modeling of the dynamics of the data.

Some of the following illustrations analyzing the dataset's statistical properties were first presented in our previous publication [70]. These illustrations appear here to contribute to the reader's understanding of the nature of the data and the behavior dynamics it encapsulates.

Figure 1 reports the distribution of rides per day of the week and per hour of the day. As can be seen in the figure, the number of rides has a far-from-uniform time distribution. More specifically, the number of rides is higher in the middle of the week and is lower during the weekend. In addition, the daily rides distribution peaks, as expected, in the morning hours and around 6-7 pm.

We use the set of taxi ride records to construct a "rides network"  $G_{T_1, T_2}$ , comprising of  $|V|$  nodes representing equally sized squared regions of New York City, and a set of  $|E|$  edges, such that each edge  $(u, v) \in E$  corresponds to a connection between two regions  $u, v \in V$  if and only if there exists at least one ride from region  $u$  to region  $v$  in the time-frame referred to by the network. Such a connection exists if and only if a ride started at some time  $t$  departing at  $u$  and reaching  $v$ , or vice versa, such that  $T_1 \leq t \leq T_2$  is contained in the time period defined for the network  $G_{T_1, T_2}$ .

As we create edges only based on rides that were created during a certain period of time the network may change (and quite significantly so) for various values selected for  $T_1$  and  $T_2$ . As the time period defined by these values increases the

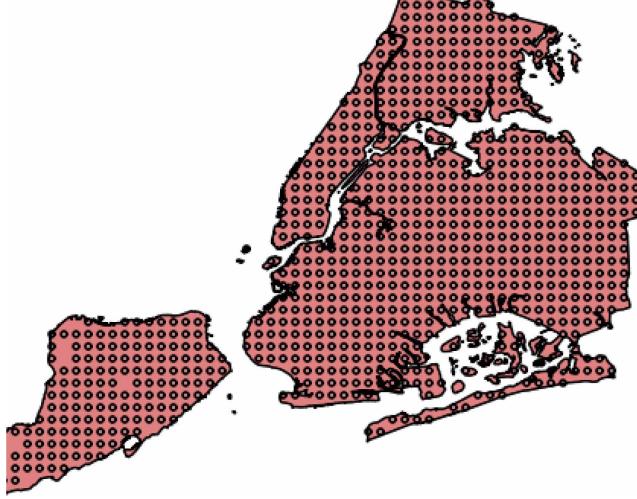


FIGURE 2: Illustration of the rides network  $G$ , portrayed on the map of NYC. It can be seen that the network has high density through the city, with a few empty spots in Staten Island.

network is expected to contain more edges, with the densest network received for  $G = G_{\infty,\infty}$  being the network that is based on the complete aggregation of all the rides. In order to encapsulate the traffic properties of a certain point in time  $T$  we would observe the time period circumventing  $T$ . Similarly, in order to analyze the network dynamics, that is – the way it changes over time, we would analyze the evolution of the network properties for networks created in nonidentical, yet partially-overlapping time periods. This methodology is extensively used in Section 4.

For different granularity of city partitioning (reflecting through the use of different sizes of the square regions) different ride networks would be produced. However, Network Theory implies that changing this parameter would not affect the existence of various mathematic invariants such as the network’s “Scale Free-ness” or its expected small diameter [71], but rather – mainly change the sparsity of the network and its number of nodes. During this work we have examined several sizes of squared-regions, ranging from rectangular regions of 0.0156 square miles in size, to 1 square mile, obtaining similar results. The analysis below is based on square tiles of 0.39 square mile (i.e. 1 square kilometer). In such a case, when taking  $G = G_{\infty,\infty}$  the network that aggregates all the rides, it comprises 813 nodes and 58,014 edges. Figure 2 illustrates the geographical distribution of the nodes  $V_G$  on the map of New York.

Figure 3 illustrates the distribution of the number of trips on the various O-D routes in the taxi network. By *weight* we refer to the number of trips that took place through this edge and by *Frequency* we refer to the number of edges that have a specific weight. Note the small number of edges who have more than 500 rides (approximately 5,000 edges out of 58,000 edges). Similarly, over 47,000 edges have less than 50 rides passing through them. This observation coincides well with the fact that human mobility is known to follow a power law distribution [3].

As we analyze the network properties of graph implied by the taxi rides, it is interesting to observe the characteristics of

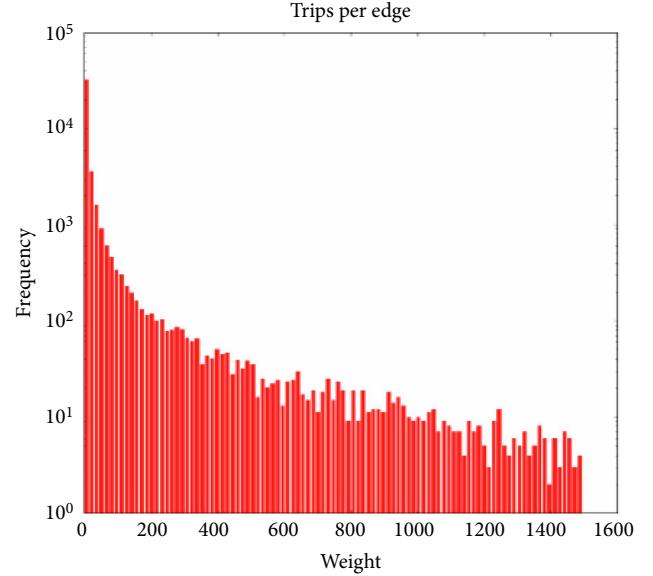


FIGURE 3: Edge weights of the taxi rides network  $G$ , denoting the number of trips per edge (namely, between every two nodes in the city). The X-axis denotes the number of trips per edge (representing a pair of origin-destination nodes), and the Y-axis (shown in a log-scale) represents the number of edges who have such number of trips.

the degrees of the nodes of the network  $G$ . A ‘degree’ of a node  $v \in V$  is the number of nodes  $v$  is connected to through edges in  $E$ , where such nodes represent the actual destinations passengers who boarded a taxi at location  $v$  chose to go to. Namely, a degree of a node  $v$  represents, therefore, the number of possible destinations a passenger boarding a taxi on location  $v$  may choose to go to. An important observation is that the popularity of a node  $v$  as reflected both by its in-degree (i.e. the number of origins passengers depart from in order to get to  $v$ ) as well as by its out-degree (i.e. the number of destinations passengers leaving  $v$  may go to) is independent of the geographic size or shape of node  $v$  – as all nodes refer to equally-sized square regions.

Interestingly, analyzing the distribution of this property reveals that whereas there are some nodes with a high degree (probably corresponding to main train stations or large administration facilities) the vast majority of the nodes have a very low degree. In other words – for the vast majority of the locations in New York, it is extremely easy to predict the destination of a passenger starting his ride there (as a low degree implies a low number of possible destinations, and a high chance of guessing the correct one). This observation is quite remarkable, as it implies that taxi users are *much more predictable* than may seem. Indeed, it seems that when one boards a taxi, one’s destination can quite accurately be predicted.

Specifically, in 24% of the possible origins of a taxi ride in New York City, the number of possible destination of a passenger leaving these origins is on average 5, and in 43% of the origins it is 10. A quick arithmetic yields that if at some point in time we would pick a random person just boarding a taxi anywhere in New York, we would have more than 7.5% of

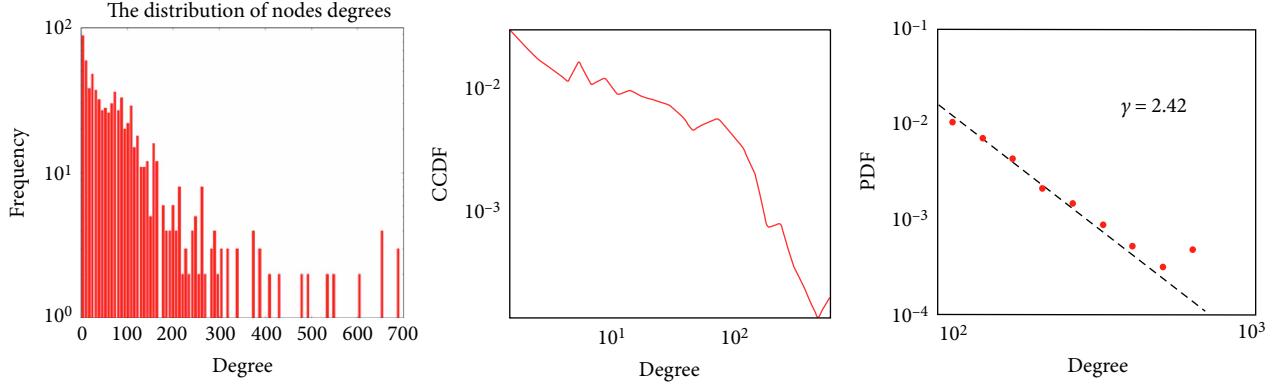


FIGURE 4: The distribution of nodes degrees in the taxi rides network, representing the number of possible destinations a passenger boarding a taxi at some location in the city may chose to go to. Note the surprisingly high number of origins with very low degrees – number of possible destinations. The nodes' degrees are unaffected by the size or shape of the actual geographic region they refer to as all nodes refer to equally-sized square patches of the city.

guessing precisely his or her destination. This probability is about three times *higher* than rolling a “Snake Eyes” (two 1’s in a 6-sides dice). See Figure 4 for more details.

In this context, it is also important to note that in this work we are less interested in the specific characterization of nodes having high (or low) degrees, but rather – in the dynamics those values represent over time, as discussed in detail in the following sections.

In order to analyze the “sharability”, or the ability to merge rides using the same vehicle at an overlapping times, we applied a simplified version of the methodology used by Santi et al. [51] to calculate the potential benefits of ride sharing: Let  $T_i = (o_i, d_i, t_i^o, t_i^d)$ ,  $i = 1 \dots k$  be  $k$  trips where  $o_i$  denotes the origin of the trip,  $d_i$  the destination, and  $t_i^o, t_i^d$  the starting and ending times, respectively. We say that multiple trips  $T_i$  are shareable if there exists a route connecting all of their origins and destinations in any order where each  $o_i$  precedes the corresponding  $d_i$ .

Sharability, or ‘ridesharing utilization’ is expressed in terms of the number of rides that can be ‘merged’, as a function of the guaranteed quality of service, expressed through the number of latency minutes agreeable by the passengers – the maximum time delay in catching a ride and arriving at destination, representing the maximum discomfort that a passenger can experience using the service. In other words, given a predefined level of discomfort passengers are willing to undertake (expressed in a prolonged wait-time), the ride-sharing utilization depicts the portion of rides that are redundant and can be saved by merging with other rides to and from the same locations.

Our analysis aims at finding pairs of rides, which are represented in the network by the same edge (i.e., have the same origin and destination), that can be shared. For each edge, we examine its corresponding set of originating rides, and count the number of ride pairs that can be merged, taking into consideration the maximum time delay parameter.

The main difference between our approach and the one discussed in [51] is that we only merge rides that leave the same origin ‘tile’ and go to the same destination ‘tile’. There are several advantages for this approach:

- (1) The routing-agnostic scheme is significantly less sensitive to the temporary changes in the infrastructure, such as detours, traffic jams, accidents, and so on.
- (2) Merging rides based only on their origin and destination makes our ride-sharing policy entirely agnostic to the routing decision of the driver. Alternatively, the approach that is based on allowing rides to be merged even if they do not leave from the same origin, but are rather partially overlapping, depends on the assumption that the route of the “containing ride” indeed passes through the origin of the second ride. This assumption in turn depends on either perfectly guessing the routing decisions of the driver, or – dictating those decisions to the driver by the ridesharing service.
- (3) As a result, our routing-agnostic approach is also expected to be easier to implement in real-life scenario, as it requires less cooperation from the drivers.
- (4) In addition, the increased simplicity of the routing-agnostic approach makes it easier to optimize from a computational point of view. The routing-aware approach discussed in [51] has a time complexity of  $O(n^2 \log(n))$  when merging pairs of rides [72], becomes much harder when triple rides merging is allowed [73], and eventually becomes computationally unfeasible for larger numbers of rides-to-be-merged [51].
- (5) When comparing the merging efficiencies of our proposed routing-agnostic approach with the routing-aware one, it is shown that whereas the latter is slightly more efficient when long wait-times are allowed (increasing our proposed 73% sharability to 93% for 5 minutes maximal delay), the improvement for shorter wait times becomes significantly smaller (this is illustrated by comparing Figure 5 to Figure 3 in [51]).

Figure 6 shows the probability density function (pdf) of the number of rides per edge. As can be seen from the figure, the distribution is heavy tailed and seems to follow a

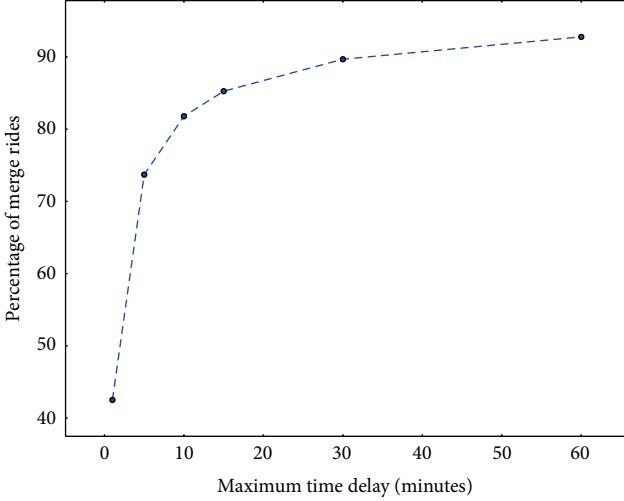


FIGURE 5: Percentage of merged rides (for the entire network).

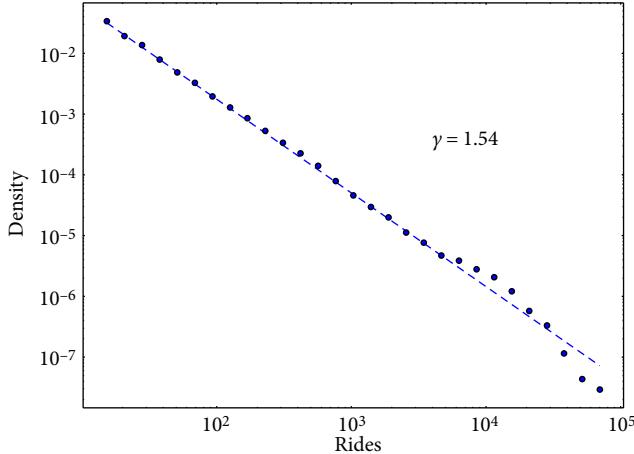


FIGURE 6: Probability Density Function of the number of rides per edge.

power-law. In other words, most of the edges (i.e., pairs of origin-destination) induce a small number of rides, while a small number of edges induce an extremely high number of rides.

Figure 5 presents the percentage of shareable rides as a function of the maximum time delay parameter. Results are encouraging: more than 70% of the rides can be shared when passengers can accept a delay of up to 5 minutes. As expected, the benefit of ride sharing increases when the passengers are willing to take a higher discomfort, and the percentage of shareable rides is more than 90% when passengers can wait 30 minutes or more.

It should be noted that the simplified analysis illustrated in Figure 5 assumes that two rides that took place at the same time can always be merged, regardless of the number of passengers in each ride. Since the average number of passengers per ride is 1.7 and most of the rides involve a single passenger, the number of saved rides could have been even higher by merging more than 2 rides at a time. On the other hand, in

some cases, even the merging of two rides at a time might have resulted in overcrowding of the vehicle.

In order to assess the effect of these two potential phenomena over our analysis, we can observe the distribution of the number of passengers per trip in the data. While doing so, we artificially segregate trip made using private taxi caps (that can board up to 4 passengers) and trips made with larger vehicles (capable of boarding from 5 to 48 passengers):

- (i) 49.22 percent of the trips have 1 passenger.
- (ii) 24.22 percent of the trips have 2 passengers.
- (iii) 15.72 percent of the trips have 3 passengers.
- (iv) 10.84 percent of the trips have 4 passengers.

We examine two approaches for the assessment of the actual theoretical ride-sharing utilization.

*Greedy merging, assuming an even distribution of number of passengers:* in this approach, we analyze the merging process in a two-phase greedy approach. In the first phase, we assume that all the original trips that can be merged are indeed merged, and are done so under the assumption that the number of passengers is distributed approximately uniformly, with respect to the various geographic locations. Then, the resulting merged trips are merged again, if possible. This analysis approach should result in a lower bound for the actual ride-sharing utilization, as in real life our ride-matching algorithm would aspire for maximizing the number of merged rides, where possible.

*Optimal merging:* in this approach we assume that whenever two rides are merged, the number of passengers they have receives the value that would result in the most efficient merging scheme possible (confined to the overall distribution of the numbers of passengers for rides). This analysis approach should result in an upper bound for the actual ride-sharing utilization, as in real life there will be times where the only way to merge rides would be in a suboptimal way.

Following is a detailed analysis of both approaches:

*Greedy merging:* the expected distribution of the merged trips for the first phase would be:

- (i) In 24.23 percent of the pairs, we would merge a trip that has 1 passenger with a trip that has 1 passenger. This results in a merged trip of 2 passengers.
- (ii) In 23.84 percent of the pairs, we would merge a trip that has 1 passenger with a trip that has 2 passengers. This results in a merged trip of 3 passengers. These trips cannot be merged, assuming the greedy 2-step approach.
- (iii) In 15.48 percent of the pairs, we would merge a trip that has 1 passenger with a trip that has 3 passengers. This results in a merged trip of 4 passengers, that cannot be further merged.
- (iv) In 5.87 percent of the pairs, we would merge a trip that has 2 passengers with a trip that has 2 passengers. This results in a merged trip of 4 passengers, that cannot be further merged.
- (v) In 30.58 percent of the pairs, we would not be able to merge the trips, has these would be pairs that either

- (a) have one of the trips with 4 passengers, or (b) having a trip with 2 passengers and a trip with 3 passengers, or (c) having two trips having 3 passengers each.

The second phase will, therefore, be able to merge another  $0.2423 \cdot 0.2423 \cdot 100 = 5.87$  percent of the original pairs, which reflects a  $5.87 \cdot 2 = 11.74$  percent increase. Overall, this would sum up to  $100 - 30.58 + 11.74 = 81.16$  percent of the naive potential utilization (namely, the utilization that is calculated under the assumption that all rides are merge-able, and that we do not merge more than two rides).

*Optimal merging:* assuming an optimal merging scheme we can calculate the merging of the relevant New York City data as follows:

- (i) The 10.84 percent of the rides that have 4 passengers cannot be merged at all.
- (ii) The 24.22 percent of the rides that have 2 passengers would be merged among themselves.
- (iii) The 15.72 percent of the rides that have 3 passengers would be merged with a matching 15.72 percent of the rides that have 1 passenger.
- (iv) This would leave another  $(49.22 - 15.72) = 33.5$  percent of the rides, that have 1 passenger. These rides would be merged in a 4-to-1 ratio, virtually implying a  $33.5 \cdot 1.5 = 50.25$  percent save.

Altogether, the actual optimal theoretical utilization would sum up to  $24.22 + 15.72 + 15.72 + 50.25 = 105.91$  percent (namely, under the assumption of optimal merging the benefit from merging 4 rides of a single passenger more than compensates the loss due to rides with 4 passengers).

Therefore, the actual theoretical utilization for the New York City taxi dataset, denoted as  $U$ , would be bounded by:

$$0.8116 \cdot \alpha \leq U \leq 1.0591 \cdot \alpha \quad (1)$$

such that  $\alpha$  is the potential utilization that is calculated throughout this work, using the method that was described above, ignoring the effect of multiple merges, as well as the effect of over-population of rides.

## 4. Analyzing the Dynamic Ride-Sharing Network

In the previous section we have described the taxi data that were used for this study, illustrated various mathematical properties of these, and discussed the way they can be analyzed for the purpose of assessing the potential ability of ride-sharing schemes to merge rides between similar locations (denoted as the *ride-sharing potential utilization*). In this section we demonstrate the inability of static analytic approaches to efficiently model this utilization and suggest an alternative approach, that is based on the construction of multiple network-snapshots, derived using a sliding-window based aggregation of the taxi rides. We show that this technique can serve as a valuable methodology for both (a) assessing the potential

ride-sharing utilization of the current supply and demand scheme (as appears in Section 4.2), as well as (b) serve as a prediction method for estimating *changes* in this utilization, in the near future, up to a few hours (as shown in Section 4.3).

**4.1. The Need for Dynamic Ridesharing Optimization and Prediction.** Mainstream transportation analysis models (such as [74–78] and many more) approach the problem of transportation forecasting and analysis through the use of long-term data aggregation. Simply put, the dominating approach today sees the accurate approximation of the “steady state”, or “average state”, of the transportation system as the most efficient way to understand the behavior of the system, and to use this understanding in order to reach better decisions [79]. Such decisions are often concerned with the locations, type, or size of new infrastructures that should be built, large-scale budgets investment alternatives or long-term policy revisions [80].

When examining the rapidly expanding field of ridesharing this approach suffers an inherent limitation, as it is not well adequate for the nature of decisions ridesharing operators and regulators are required to make. As ridesharing uses existing roads and metropolitan infrastructure, does not require setting fixed-place stations of fixed-paths, and often uses existing vehicles, it is mostly located “outside” the realm of these analysis methodologies. Furthermore, ridesharing introduces a new set of factors that traditional methods usually cannot easily cope with, such as dynamic changes in fares, which may significantly influence network properties such as global congestion [81].

Analyzing ridesharing using the existing models would be inefficient at best. Taking the static approach using a long-term aggregation of the supply and demand would inevitably result in a model that would be optimized for the *average states* of the rides network, ignoring its inherent volatility (that is caused due to daily and weekly patterns as well as irregular spikes created by events such as street-parties, sports events, etc.).

Interestingly, as shown in Section 4.2, the dynamic rides network spends only an extremely *small portion* of the time in those average network states. Furthermore, our analysis demonstrates that overlooking the dynamic nature of the traffic scheme disregards the vast majority of the network states, as manifested in the O-D matrix, as well as the possible ridesharing utilization of it. Specifically, this phenomenon is demonstrated in Figure 7 that reveals that the system spends approximately 33% of the time in states that have a potential utilization of either 50% above the monthly average, or 50% below it.

Ignoring this dynamic nature of the urban rides system through the use of a static analysis model (which is the mainstream approach of today) will be inherently limited in its efficiency. The key to unlocking the development of effective next generation ridesharing systems, therefore, lays in an analysis that is rooted in the understanding of its dynamic nature, and the way to use it in order to develop proactive strategies that dynamically adapt their forecast using an ad-hoc analysis of the network’s state.

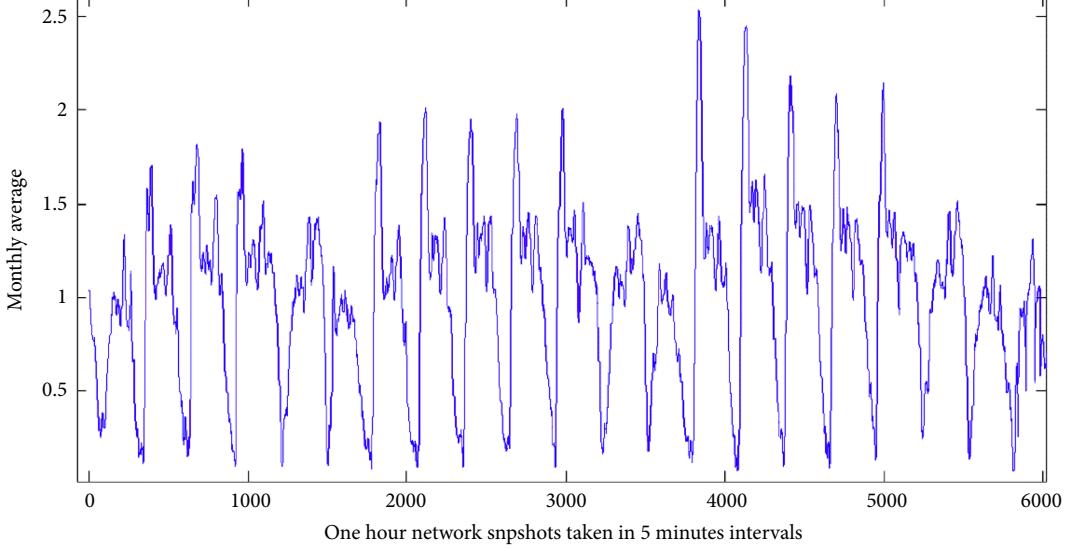


FIGURE 7: Dynamics of the potential ride-sharing utilization over time. X-axis denotes the time, given in 5-minute granularity. Y-axis denotes the change of the potential utilization compared to its monthly average.

A potential example for this approach can be found in [82], containing a computational study aimed at identifying environments in which the use of “dedicated drivers” are most useful. As urban supply and demand environments are constantly (and significantly) changing (as demonstrated in our analysis of the New York taxi data), it is therefore likely that a strategy that detects the *times* where the use of such drivers is most efficient and upon such detection – launches these drivers to supply the demand (this can be done using a dynamic change in the commission drivers are required to pay, giving such drivers a temporary priority in certain roads, or forbidding them from granting service on a regular basis except from when their service is required) – would achieve a superior performance compared to a static strategy that does not react to such changes.

Another example can be the work of [83] in which the size of a carsharing fleet is optimized in order to maximize the monetary operational savings. Again, such an approach reaches the global optimization assuming a static approach, whereas the incorporation of the dynamic nature of the system could yield a significant. This could be done for example by allowing the fleet operators to dynamically use the services of a public service (such as Uber or Lyft), rented cars, or private drivers. Using such service when needed will allow to reduce the ongoing basic cost.

**4.2. Dynamic Network Analysis.** As discussed in previous sections, one of the main hurdles that prevents the wide adoption of ride-sharing might be the high volatility of its potential utilization, and the extreme unpredictability of it. In this section, we propose to mitigate this problem by using a dynamic network that represents the evolving travel patterns in the city. That is, a multitude of rides-networks, representing data of fixed-length periods of time, each of which starting at different points in time of equal distances. Such “sliding window” approach is useful for tracking changes in various

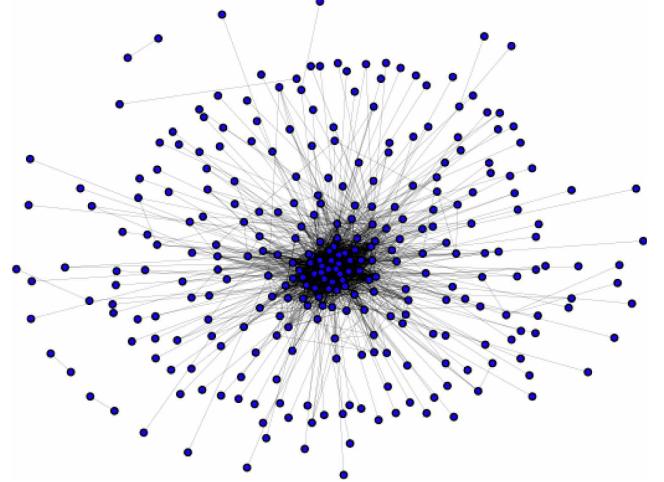


FIGURE 8: An illustration of the rides sub-network  $G_{T_{144}, T_{145}}$ , denoting the structure that is implied by the aggregation of the rides between the 144-th and the 145-th hour of the month.

properties of this dynamic network, which we show are not only highly correlated with the potential ride-sharing utilization at the corresponding points in time, but can also predict the utilization few hours ahead of time.

We divide the rides dataset into hourly aggregated snapshots, creating  $31 \times 24 = 744$  sub-networks, each is denoted by  $G_{T_n, T_{n+1}}$ , such that  $T_n$  represents the  $n$ -th hour in the month. An illustration of one such sub-network is shown in Figure 8. Intuitively, we see that most of the nodes are highly connected, but a considerable number of nodes are connected to only one other node in the network.

Similarly to Figure 5 in which the potential benefit of ride-sharing over the entire data was shown, we have performed the same calculation for every hourly network separately. Figure 9 presents the average potential ride-sharing

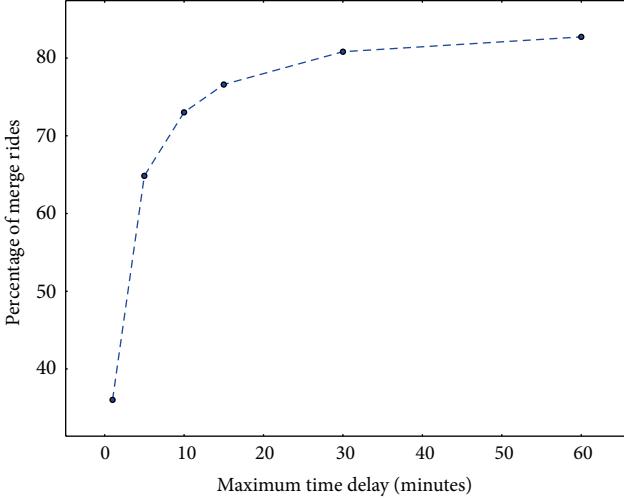


FIGURE 9: Potential of ride-sharing utilization, measured as the percentage of potentially merged rides (averaged over all sub-networks), as a function of maximal delay agreeable by the passengers. The figure is based on the result presented [70].

utilization taken on all hourly networks, as a function of the maximal delay allowed (notice that this is in fact a lower bound, since we artificially prevent passengers from being merged with rides “outside” their hourly network). It can be seen that this produces a lower utilization than the previous calculation using the overall aggregation (approximately 10% decrease), caused by the fact that each pair of nodes has a lower probability of being connected.

We now extract a set of six common network properties for each traffic-network  $G_{T_i, T_j}$ , to be used as the features values representing each network. These features encapsulate various topological aspects of the network and enable us to project each hourly-collection of traffic data (containing a large and apriorically unknown number of rides) into a single coordinate in a 6-dimensional feature-space.

- (1) *Number of Nodes*: the number of nodes in the network  $G$ , denoted as  $|V|$ , representing the number of unique pick-up and drop-off locations of rides made during this time window. Note that although all the networks refer to the same dataset, and the same geographic environment, different networks may have different values of  $|V|$ , since at different time-segments different locations may be “active”.
- (2) *Number of Edges*: the number of edges in the network  $G$ , denoted as  $|E|$ , representing the number of unique pick-up to drop-off pairs of rides made during this time window. This is also the number of nonzero elements of the temporal O-D matrix that is derived from this network.
- (3) *Network Density*: the average degree of the network’s nodes, defined as  $|E|/|V|$ . This property represents the average number of unique drop-off locations per pick-up location (and vice versa) and is associated with the predictability of rides made during this time window, and is also related to the system’s entropy.

- (4) *Average Betweenness Centrality*: each node  $v$  in the network  $G$  has a calculate-able betweenness centrality score [84], representing the portion of “shortest paths” between all the node-pairs in the network, that pass through  $v$ . Formally, for a network node  $v \in V$  this is defined as:

$$\sum_{s \neq v \neq t} \frac{\sigma_{s,t}(v)}{\sigma_{s,t}}, \quad (2)$$

where  $\sigma_{s,t}$  is the total number of shortest paths from node  $s$  to node  $t$  and  $\sigma_{s,t}(v)$  is the number of those paths that pass through  $v$ .

Averaging these values yields an estimation of the network’s efficiency, with respect to the number of nodes whose adequate availability is required in order to preserve the network’s ability to maintain efficient flow without increasing the length or durations of trips between arbitrary points [28, 85].

- (5) *Average Closeness Centrality*: the closeness centrality of a node [86] is a measure of centrality in a network, calculated as the sum of the length of the shortest paths between the node and all other nodes in the graph. Thus the more central a node is, the closer it is to all other nodes. For a node  $v \in V$ , the measure is defined as:

$$\frac{1}{\sum_x d(v, x)} \quad (3)$$

Averaging the closeness centrality over all the network’s nodes yields an estimation of the compactness of the network, that is – how short it is to travel between an arbitrary pair of network nodes.

- (6) *Average Eigenvalue Centrality*: eigenvalue centrality [87] (also called eigencentrality or eigenvector centrality) is a measure of the influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

For a given graph  $G$  with an adjacency matrix  $A$  the centrality score of a node  $v \in V$ , denoted as  $x(v)$ , is defined as

$$\frac{1}{\lambda} \sum_{u \in M(v)} x_u, \quad (4)$$

where  $M(v)$  is a set of the neighbors of  $v$  and  $\lambda$  is the graph’s largest positive real eigenvalue. This can be accurately estimated by taking the  $v^{\text{th}}$  component in the eigenvector that corresponds to the largest positive real eigenvalue.

The use of eigenvalues to analyze propagation phenomena over networks can be see for example in [88], where its usability for predicting the epidemic potential of viruses is demonstrated.

We use a linear regression to fit these features for the calculated potential utilization, as well as a multiple linear regression to fit the potential utilization for the entire set of network properties. As can be seen in Figure 10, these features show a

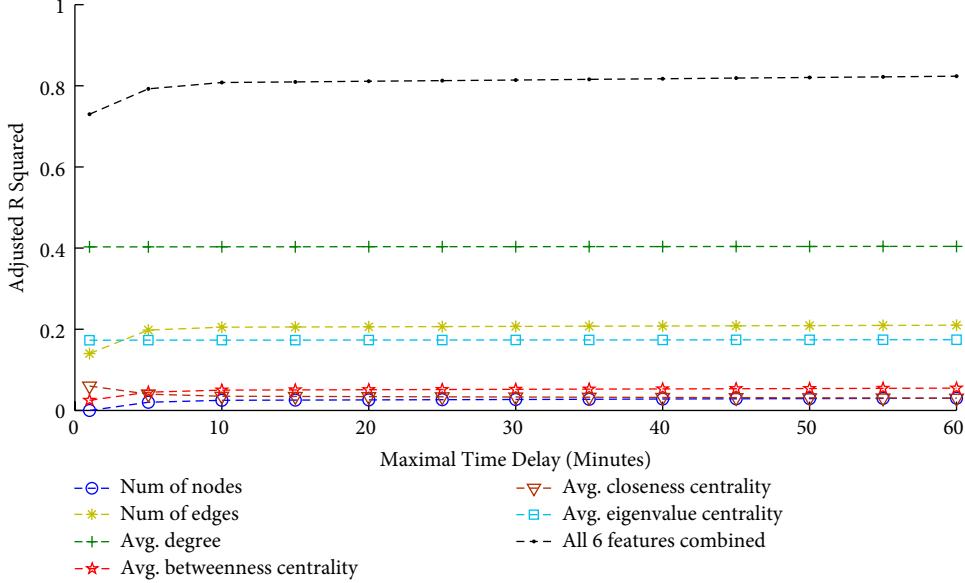


FIGURE 10: Adjusted  $R^2$  of the correlation between seven features of the hourly rides network and the potential ride-sharing utilization for this network. Most features have low quality of fit, but the combined mixture of all seven results in a remarkably high correlation ( $R^2 = 0.82$ ). Features are (1) the number of nodes in the network, (2) the number of edges, (3) the averaged degree, (4) the averaged betweenness centrality, (5) the averaged closeness centrality, and (6) the averaged eigenvector centrality.

high correlation with the potential utilization for this hourly network (the figure reports the adjusted  $R$  squared to account for the different number of predictors).

**4.3. Ride-Sharing Potential Prediction.** In the previous section, we have shown that the monthly rides can be partitioned into hourly aggregative snapshots, each of different characteristics (and specifically, network oriented ones), and different ride-sharing potentials. In addition, we have demonstrated the correlation between these network properties and the ride-sharing potentials of the rides the corresponding networks are implied from (as appears in Figure 10). In this section, we discuss whether this correlation can also be used for predictive purposes. Specifically, can we deduce from the current values of various network properties how the *change* in the ride-sharing potential compared to its current value.

In order to do so, we first analyze the evolution of various network properties of the hourly aggregative rides network  $G_{T_n, T_{n+1}}$  over time. Figure 11 illustrates the evolution of the mean nodes' degree of the rides network as a function of time (that is, the average over all of the network's nodes' degrees, for all the dynamic hourly networks). For the sake of clarity, we have increased the time granularity used in the analysis, so that the hourly networks are now generated with 5-minute intervals, thus significantly overlapping, and subsequently generating a smoother and easier to read graph. The change from the monthly average of the mean degree as a function of time is portrayed, clearly showing a dominant daily pattern. However, on top of this pattern we can see significant hourly fluctuations, tens of percent in magnitude. This reveals the existence of strong volatility in the rides dynamics alongside the predicted daily and weekly dynamics.

A similar dynamics is observed when analyzing the evolution of the largest eigenvalue of the rides-network' adjacency

matrix over time. The use of eigenvalues to analyze propagation phenomena over networks can be seen for example in [88, 89], where its usability for predicting the epidemic potential of viruses (both human and computer-based) is demonstrated. Additional mathematical analysis on the role of eigenvalues in the analysis of network structures can be found in [90]. This property, known to encapsulate various behavioral characteristics of the people whose mobility patterns the network is depicting, displays a clear (and easy to predict and understand) daily pattern, on top of which significant and erratic spikes are added, as can be seen in Figure 12. These spikes seem to appear sporadically, lacking any clear patterns or internal regularity, implying again the need for understanding the dynamic aspects of the network.

Now, let us perform a similar analysis over the potential ride-sharing utilization, looking at its evolution over time. The results of this analysis, presented in Figure 7, clearly demonstrate a similar dynamics to the couple of network properties mentioned earlier. Specifically, it can be seen that alongside the dominating daily pattern (and weaker, but still easy to see, weekly one), there are clear changes in the potential utilization. These changes take various shapes and forms, from sudden decrease in the daily peak (as can be seen around  $X = 1400$ ), to changes in the intra-weekly peaks (the first week analyzed showing a 'U-shaped' form among its days, the second week showing an equal-peaks dynamics, and the third week showing an extremely high Monday and Tuesday, and weaker Wednesday, Thursday and Friday), and others. Surprisingly, the magnitude of these changes may even exceed the dominating daily pattern. For example, the change between the first Tuesday (around  $X = 200$ ) and the third Tuesday ( $X = 4100$ ) is 90% compared to the monthly average, whereas the average change in potential utilization between workdays and weekends is only 70%.

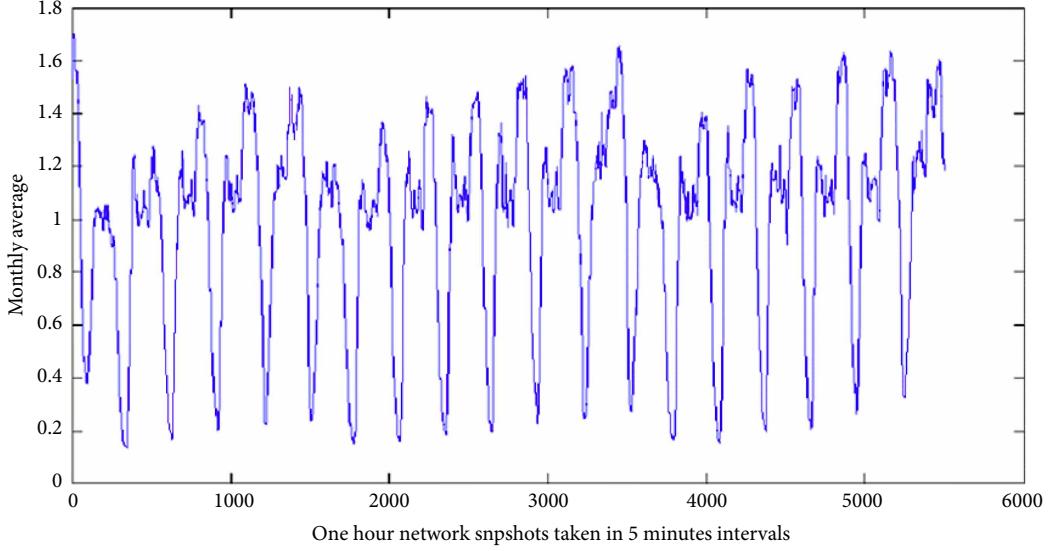


FIGURE 11: Dynamics of the mean degree of the rides network nodes. X-axis denotes the time, given in 5 minutes granularity. Y-axis denotes the change of the mean degree of the network compared to its monthly average.

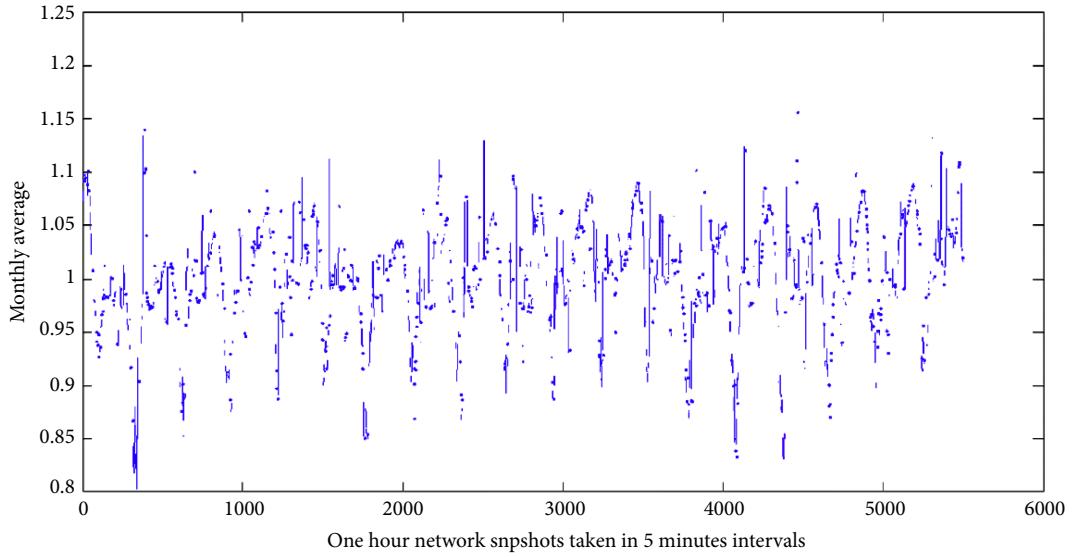


FIGURE 12: Dynamics of the largest eigenvalue of the rides-dynamic network over time. X-axis denotes the time, given in 5-minute granularity. Y-axis denotes the change of the largest eigenvalue of the network compared to its monthly average.

At this point, we ask the following question: “can we find a statistical correlation between *current* values of the rides network properties and *future* values of the potential ride-sharing utilization?”. This question is of interest, as such a correlation would allow us to predict future changes in the potential utilization, providing valuable tools for both ride-sharing users, operators, and regulators.

We first address this question by comparing network properties values at time  $T$  with potential utilization of at time  $T + 1$  (1 hour prediction). Figure 13 presents an example of such a comparison, in the form of a scatter plot showing for each point in time  $T$  a dot whose X-axis is the mean nodes’ degree of the network  $G_{T,T+1}$  and whose Y-axis is the change in the potential utilization of the rides between  $T + 1$  and  $T + 2$

compared to the rides between  $T$  and  $T + 1$ . That is, the change in the momentary ride-sharing utilization between “now” (time  $T$ ) and “in an hour” (time  $T + 1$ ). It is easy to see that this representation reveals a clear and strong negative correlation between the two.

Trying to increase our lookahead and predict the change in the dynamic ride-sharing utilization from a 2 hours time-frame, Figure 14 illustrates the correlation between the value of the largest eigenvalue of the rides network at time  $T$  and the change in the potential utilization between time  $T$  (aggregated to  $T + 1$ ) and  $T + 2$  (aggregated to  $T + 3$ ). Again, a clear strong negative correlation is easily visible. For example, in times where the value of the largest eigenvalue of the rides network is smaller than 0.012, the potential ride-sharing utilization was

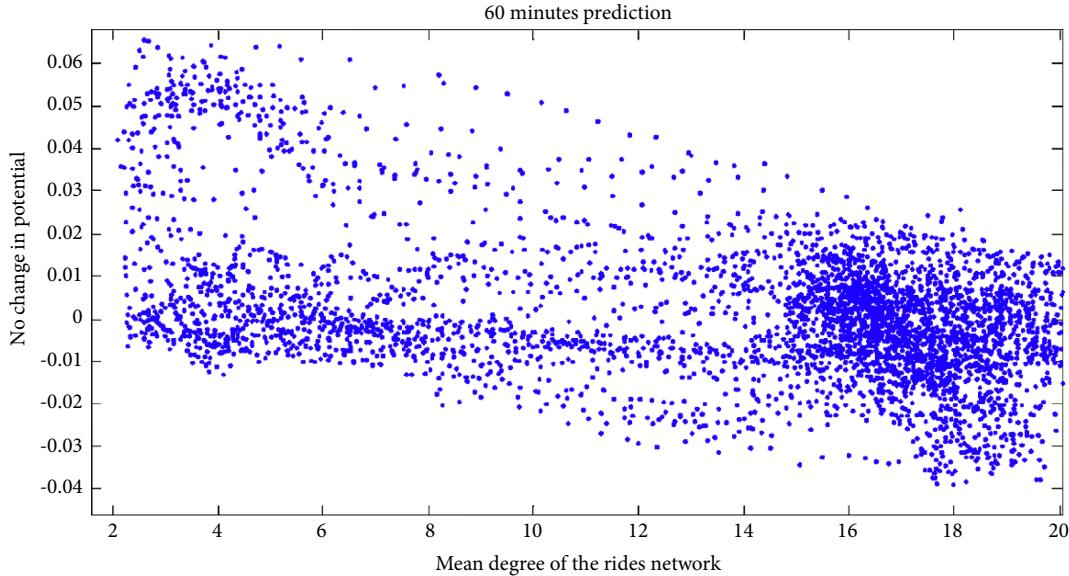


FIGURE 13: Change in potential ride-sharing utilization (Y-axis), 1 hour prediction, as a function of the mean degree of the rides network (X-axis).

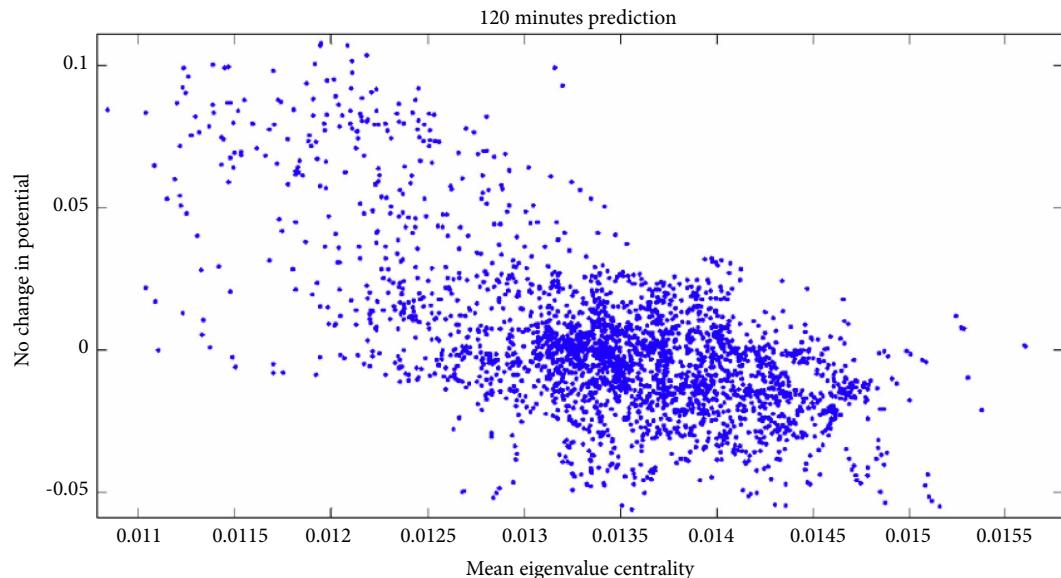


FIGURE 14: Change in potential ride-sharing utilization (Y-axis), 2 hours prediction, as a function of the largest eigenvalue of the rides network (X-axis).

statistically guaranteed (during the month of the observation) to significantly increase in the coming 2 hours. Similarly, largest eigenvalue of 0.014 would indicate a significant decrease in the ride-sharing potential within the next 2 hours.

Figures 13 and 14 are based on the analysis of the first 3 weeks of January 2013. These observations were then validated using the last week of January, as can be seen in Figures 15 and 16.

Once demonstrating the predictive power of the dynamic network's properties with respect to the network's future ride-sharing potential, we can now construct a multiple linear regression model that would fit all of these 6 properties. We have created 18 models, for 2 values of distance tolerance

(400 m and 800 m, denoting the pick-up and drop-off distances that still allow rides to be merged), 3 values of time tolerance (30 s, 2 minutes and 5 minutes, denoting the time passengers would be willing to wait in order to merge their rides) and 3 values of prediction horizon (no prediction, 1 hour prediction and 2 hours prediction). The results include a scatter plot of the data, effects of the various properties, ANOVA, and other statistical analyses as appearing in Supplementary Figures 17–34.

The effectiveness of the prediction as a function of the prediction horizon (i.e., the distance between the point in time where the prediction is calculated and the point in time this

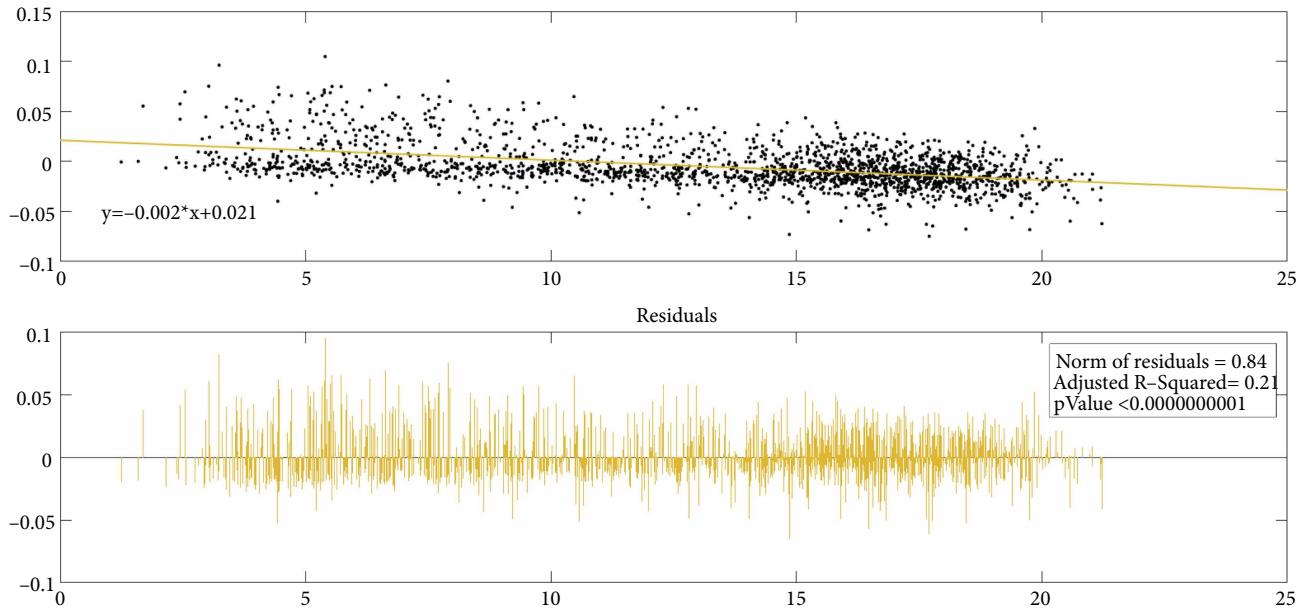


FIGURE 15: Change in potential ride-sharing utilization (Y-axis), 1 hour prediction, as a function of the mean degree of the rides network (X-axis), created for the last week of the data.

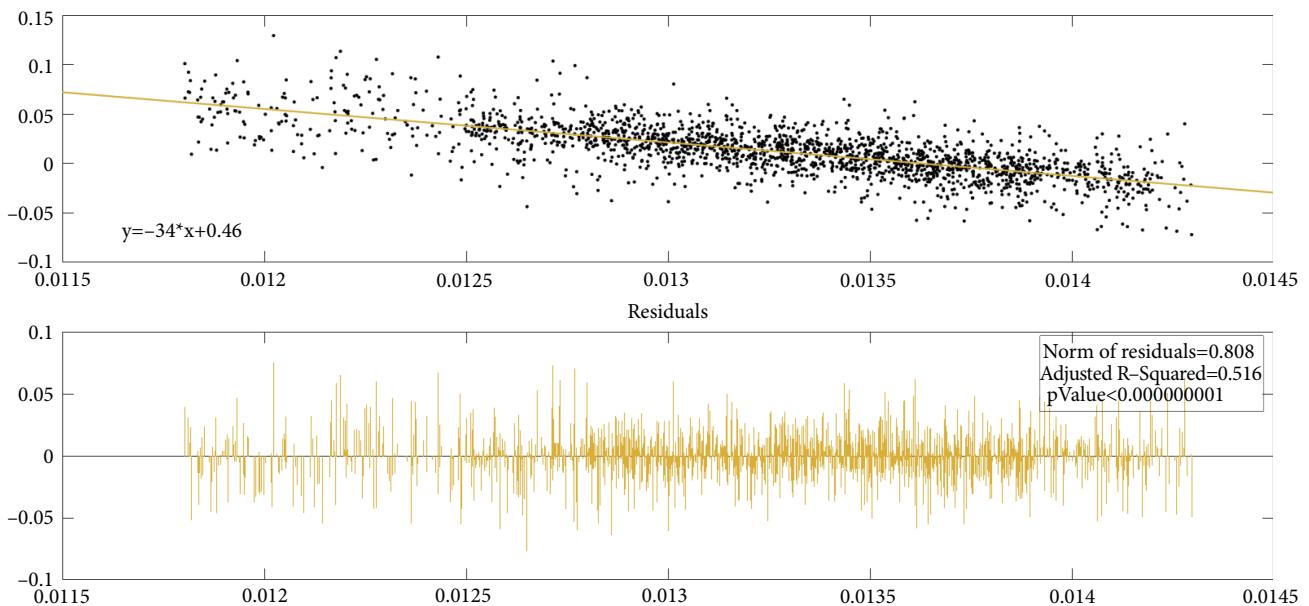


FIGURE 16: Change in potential ride-sharing utilization (Y-axis), 2 hours prediction, as a function of the largest eigenvalue of the rides network (X-axis), created for the last week of the data.

prediction refers to) is illustrated in Supplementary Figures 35–40, showing the  $R^2$  of the model (both ordinal and adjusted) as a function of the time horizon (between 0 and 12 hours), for several values of distance tolerance and time tolerance. It can clearly be seen that in general (and as expected) the accuracy of the model decreases with the increase in the prediction horizon used (that is, when the model tries to predict the behavior of the system further into the future).

The effect of each feature, depicted by the adjusted response plot for its various values, is presented in

Supplementary Figures 41–46, created for a scenario with distance tolerance of 800 meters, time tolerance of 5 minutes, and prediction horizon of 2 hours.

## 5. Summary and Future Work

As the popularity of ride-sharing systems grow, its users-base gradually transform from early adopters to mainstream consumers. Whereas the first are characterized by a keen affection

for innovative solution that are powered by cutting edge technologies and aim to disrupt the governing paradigm in the field, the latter are often interested mainly in the advantages these services can offer them with as smallest change in their habits as possible. With respect to ride sharing, these new users are willing to sustain far less wait-time and are extremely more susceptible to inconvenience than their preceding tech-savvy innovation-hungry early users. The key to a scalable mature ride-sharing infrastructure is found in the level of service such systems will provide, mainly measured by the availability of vehicles when they are needed. Alas, the availability maximization is immediately linked to a reduction in the financial savings that the service can offer. In other words, a further expansion of ride-sharing is being constrained among others by the ability to offer high utilization, defined as the ability to “merge” similar rides in a way that would not require the passengers to sustain more than a minimal delay in their trips.

This optimization problem was extensively discussed in previous literature (comprehensive literature review can be found in Section 2). However, the conventional approach to this problem assumed a static environment which needs to be optimized. By finding the optimal number of cars, or optimal pricing policy, the efficiency (or potential) of the system was assumed to be calculable in a robust way – a key component in the decision of operators where to deploy new systems, in the design of relevant urban legislations by municipal policy makers, and of course in the likelihood of passengers to use these services.

In this work we discussed the dynamic nature of ride-sharing systems. Specifically, we were interested whether ride-sharing utilization is stable over time (which coincides with the implicit assumption of most previous works in this field) or does it undergo significant and often rapid changes (which would imply the inherent inefficiency of schemes assuming a static nature). We modeled the ride-sharing utilization using the known New York Taxi dataset and clearly show that it is highly dynamic, and that any system that would be designed for the “average” utilization would be highly inefficient.

We then show that assuming a dynamic approach the taxi data can be modeled as a sequence of data-snapshots, resulting in a dynamic traffic-network model. Several recent works have shown that network features can effectively be used to predict a variety of events and properties, e.g., emergency situations, individuals’ personality and spending behaviors [91, 92]. We used a similar technique in order to project the taxi data as into a feature space comprised of topological features of the dynamic network implied by this traffic. This (dynamic) feature space is then used to model the dynamics of ride-sharing utilization over time.

Using this approach we were able to demonstrate a clear correlation between the utilization of the ride-sharing system over time and several topological features of the network it creates. In addition, we demonstrated that the potential benefit of ride sharing expressed as the percentage of rides that can be shared with a limited discomfort for riders can also be predicted a few hours in advance. Such prediction can be used as a tool for an accurate short-term forecasting of the ride-sharing potential in cities and metropolitan areas.

Researchers in [8, 51, 93] and others have focused on addressing the computational challenges of trip-matching

(an NP-hard optimization problem) in real-time and developed heuristics to quantify potential ride-sharing demand. These algorithms reroute trips in order to match them with similar, overlapping trips, explicitly capturing demand for ridesharing relative to passenger’s willingness to experience prolonged travel time. However, finding an optimal solution to this problem is not computationally plausible (even under extreme limitations of the problem’s space [94]), and even the calculation of approximation heuristics would be computationally intense when done ad-hoc. Therefore, the ability to use current traffic dynamics in order to predict properties of an efficient near-future ride-sharing scheme – such as the method we propose in this work – can be used to make this process significantly more efficient [95, 96].

Future work should focus on the analysis of the correlation we find in this paper, trying to detect traces of possible causalities. Are network properties merely correlated with ride-sharing utilization, or do they possess an active influence over it? Evidence of the latter would enable us to offer urban designers and policy makers an innovative tool for encouraging and facilitating the adoption of ride-sharing systems. Alternatively, incentives and fees could be better moderated, used as “remedies” in the case of a change in the travel patterns, in order to balance it and maintain a sustainable ride-sharing paradigm. Another approach could be the pipelining of the dynamic ride-sharing utilization forecast as the input of models intended to predict the benefits of ride-sharing on the overall traffic [97].

Recent works have demonstrated the benefit of tracking the network’s dynamics in order to improve collaborative decision making [98, 99]. A possible continuation of the current work can analyze ride-sharing optimization as a case of decentralized decision-making process, using the technique that is presented here.

As the prediction of future ride-sharing potential is ultimately needed for optimization purposes (of the overall travel time, congestion or any other utilization metric) of a dynamic coverage problem, comparing the performance of any proposed method to the theoretical results that are available for various types of such decentralized collaborative coverage challenges (see [100–105] and specifically [106]) can also be of value.

Finally, as our suggested approach is agnostic to the actual route taken by the drivers it would be interesting to see whether the introduction of ride-sharing affects additional factors such as detours (that for a merged ride may become cost-effective), usage of toll-routes, etc.

## Data Availability

The taxi data used to support the findings of this study, encompassing a dataset of over 14 million individual taxi trips taken in New York City, are accessible at the NYC Taxi repository [19].

## Conflicts of Interest

The authors declare that they have no conflicts of interest.

## Acknowledgments

This research was funded in part by the Israeli Institute of Technology graduate scholarship and Israel Science Foundation (ISF).

## Supplementary Materials

Prediction results. (*Supplementary Materials*)

## References

- [1] A. Caragliu, C. Del Bo, and P. Nijkamp, "Smart cities in Europe," *Journal of Urban Technology*, vol. 18, no. 2, pp. 65–82, 2011.
- [2] H. Chourabi, T. Nam, S. Walker et al., "Understanding smart cities: an integrative framework," in *Proceedings of the 45th Hawaii International Conference on System Science (HICSS)*, IEEE, USA, pp. 2289–2297, 2012.
- [3] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [4] F. Calabrese, M. Diao, G. Di Lorenzo, J. Ferreira Jr., and C. Ratti, "Understanding individual mobility patterns from urban sensing data: a mobile phone trace example," *Transportation Research Part C: Emerging Technologies*, vol. 26, pp. 301–313, 2013.
- [5] A. Noulas, S. Scellato, R. Lambiotte, M. Pontil, and C. Mascolo, "A tale of many cities: universal patterns in human urban mobility," *PLoS One*, vol. 7, no. 5, Article ID e37027, 2012.
- [6] <http://www.waze.com>, [Online; accessed 15-October-2014].
- [7] Solving transport headaches in the cities of 2050, BBC Future [Online; accessed 15-October-2014].
- [8] B. Cici, A. Markopoulou, E. Frias-Martinez, and N. Laoutaris, "Assessing the potential of ride-sharing using mobile and social data: a tale of four cities," in *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, ACM, USA, pp. 201–211, 2014.
- [9] T. Altshuler, Y. Shiftan, R. Katsovsky, N. Oliver, A. S. Pentland, and Y. Altshuler, "Mobile phones for on-demand public transportation," in *NetSci*, 2014.
- [10] D. Graziotin, "An analysis of issues against the adoption of dynamic carpooling," <http://arxiv.org/abs/1306.0361>.
- [11] L. Alexander and M. Gonzalez, "Assessing the impact of real-time ridesharing on urban traffic using mobile phone data," 2015.
- [12] S. Hobrink, "Explaining regional adoption differentials in Dutch car sharing markets," 2014.
- [13] L. Coenen, P. Benneworth, and B. Truffer, "Toward a spatial perspective on sustainability transitions," *Research Policy*, vol. 41, no. 6, pp. 968–979, 2012.
- [14] S. Shaheen, A. Cohen, and M. Chung, "North American carsharing," *Transportation Research Record: Journal of the Transportation Research Board*, vol. 2110, no. 1, pp. 35–44, 2009.
- [15] lyft.com, <http://www.lyft.com>.
- [16] uber.com, <http://www.uber.com>.
- [17] side.cr, "(currently closed, and acquired by General Motors)," <http://www.side.cr>.
- [18] D. Efthymiou, C. Antoniou, and P. Waddell, "Factors affecting the adoption of vehicle sharing systems by young drivers," *Transport Policy*, vol. 29, pp. 64–73, 2013.
- [19] NYC Open Data Website, <http://data.ny.gov/>. Accessed September, 2014.
- [20] Y. Altshuler, A. S. Pentland, and G. Gordon, "Social behavior bias and knowledge management optimization," *Social Computing, Behavioral-Cultural Modeling, and Prediction*, Springer, Cham, pp. 258–263, 2015.
- [21] P. M. Krafft, J. Zheng, W. Pan et al., "Human collective intelligence as distributed bayesian inference," <http://arxiv.org/abs/1608.01987>.
- [22] D. Lazer, A. S. Pentland, L. Adamic et al., "Computational social science," *Science*, vol. 323, no. 5915, pp. 721–723, 2009.
- [23] E. Shmueli, V. K. Singh, B. Lepri, and A. S. Pentland, "Sensing, understanding, and shaping social behavior," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 1, pp. 22–34, 2014.
- [24] A. Almaatouq, A. Alabdulkareem, M. Nouh et al., "Twitter: Who gets caught? observed trends in social micro-blogging spam," in *Proceedings of the 2014 ACM Conference on Web Science, WebSci '14*, ACM, pp. 33–41, 2014.
- [25] M. C. Gonzalez, C. A. Hidalgo, and A.-L. Barabasi, "Understanding individual human mobility patterns," *Nature*, vol. 453, no. 7196, pp. 779–782, 2008.
- [26] Y. Altshuler, R. Puzis, Y. Elovici, S. Bekhor, and A. S. Pentland, "On the rationality and optimality of transportation networks defense: a network centrality approach," *Securing Transportation Systems*, pp. 35–63.
- [27] J. Armstrong and A. Khan, "Modelling urban transportation emissions: role of GIS," *Computers, Environment and Urban Systems*, vol. 28, no. 4, pp. 421–433, 2004.
- [28] R. Puzis, Y. Altshuler, Y. Elovici, S. Bekhor, Y. Shiftan, and A. Pentland, "Augmented betweenness centrality for environmentally-aware traffic monitoring in transportation networks," *Journal of Intelligent Transportation Systems*, vol. 17, no. 1, pp. 91–105, 2013.
- [29] B. Waclaw, *Statistical mechanics of complex networks*, <http://arxiv.org/abs/0704.3702>.
- [30] S. Wassermann and K. Faust, *Social network analysis: Methods and applications*, Cambridge University Press, 1994.
- [31] M. Granovetter, "The strength of weak ties," *American Journal of Sociology*, vol. 78, no. 6, pp. 1360–1380, 1973.
- [32] R. Albert, H. Jeong, and A. Barabasi, "Error and attack tolerance of complex networks," *Nature*, vol. 406, no. 6794, pp. 378–382, 2000.
- [33] J.-P. Onnela, J. Saramäki, J. Hyvönen et al., "Structure and tie strengths in mobile communication networks," *Proceedings of the National Academy of Sciences*, vol. 104, no. 18, pp. 7332–7336, 2007.
- [34] R. Pastor-Satorras and A. Vespignani, "Epidemic spreading in scale-free networks," *Physical Review Letters*, vol. 86, no. 14, pp. 3200–3203, 2001.
- [35] R. Pastor-Satorras and A. Vespignani, *Evolution and Structure of the Internet: A Statistical Physics Approach*, Cambridge Univ Pr, 2007.
- [36] Y. Altshuler, M. Fire, N. Aharony, Y. Elovici, and A. Pentland, "How many makes a crowd? on the correlation between groups' size and the accuracy of modeling," in *International Conference on Social Computing, Behavioral-Cultural Modeling and Prediction*, Springer, pp. 43–52, 2012.

- [37] Y. Altshuler, N. Aharony, M. Fire, Y. Elovici, and A. Pentland, "Incremental learning with accuracy prediction of social and individual properties from mobile-phone data," in *2012 International Conference on Privacy, Security, Risk and Trust and 2012 International Conference on Social Computing*, IEEE
- [38] Y. Altshuler, E. Shmueli, G. Zyskind, O. Lederman, N. Oliver, and A. Pentland, "Campaign optimization through behavioral modeling and mobile network analysis," *IEEE Transactions on Computational Social Systems*, vol. 1, no. 2, pp. 121–134, 2014.
- [39] Y. Altshuler, N. Aharony, A. Pentland, Y. Elovici, and M. Cebrian, "Stealing reality: when criminals become data scientists (or vice versa)," *IEEE Intelligent Systems*, vol. 26, no. 6, pp. 22–30, 2011.
- [40] C. Jernigan and B. F. Mistree, "Gaydar: Facebook friendships expose sexual orientation," *First Monday*, vol. 14, no. 10, 2009.
- [41] W. Pan, Y. Altshuler, and A. Pentland, "Decoding social influence and the wisdom of the crowd in financial trading network," in *Privacy, Security, Risk and Trust (PASSAT), 2012 International Conference on and 2012 International Conference on Social Computing (SocialCom)*, IEEE, pp. 203–209, 2012.
- [42] A. Barabasi, "The origin of bursts and heavy tails in human dynamics," *Nature*, vol. 435, no. 7039, pp. 207–211, 2005.
- [43] J. Candia, M. Gonzalez, P. Wang, T. Schoenharl, G. Madey, and A. Barabasi, "Uncovering individual and collective human dynamics from mobile phone records," *Journal of Physics A: Mathematical and Theoretical*, vol. 41, Article ID 224015, 2008.
- [44] L. Alexander, S. Jiang, M. Murga, and M. C. Gonzalez, "Origin-destination trips by purpose and time of day inferred from mobile phone data," *Transportation Research Part C: Emerging Technologies*, vol. 58, no. Part B, pp. 240–250
- [45] D. Brockmann, L. Hufnagel, and T. Geisel, "The scaling laws of human travel," *Nature*, vol. 439, no. 7075, pp. 462–465, 2006.
- [46] Y.-A. de Montjoye, C. A. Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the crowd: the privacy bounds of human mobility," *Scientific Reports*, vol. 3, no. 1, 2013.
- [47] C. Song, Z. Qu, N. Blumm, and A. Barabasi, "Limits of predictability in human mobility," *Science*, vol. 327, no. 5968, pp. 1018–1021, 2010.
- [48] M. Furuhata, M. Dessouky, F. Ordóñez, M.-E. Brunet, X. Wang, and S. Koenig, "Ridesharing: the state-of-the-art and future directions," *Transportation Research Part B: Methodological*, vol. 57, pp. 28–46, 2013.
- [49] K. Ghoseiri, *Dynamic rideshare optimized matching problem [thesis]*, University of Maryland, College Park, MD, 2012.
- [50] L. Rayle, S. Shaheen, N. Chan, D. Dai, and R. Cervero, *App-based, on-demand ride services: comparing taxi and ridesourcing trips and user characteristics in San Francisco* University of California Transportation Center (UCTC), Tech. rep., UCTC-FR-2014-08, University of California Transportation Center, Berkeley, Calif., 2014.
- [51] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, "Quantifying the benefits of vehicle pooling with shareability networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 37, pp. 13290–13294, 2014.
- [52] J. Miller and J. P. How, *Predictive positioning and quality of service ridesharing for campus mobility on demand systems*, <http://arxiv.org/abs/1609.08116>.
- [53] M. Ota, H. Vo, C. Silva, and J. Freire, "Stars: simulating taxi ride sharing at scale," *IEEE Transactions on Big Data*, vol. 3, no. 3, 2017.
- [54] R. Tachet, O. Sagarra, P. Santi et al., "Scaling law of urban ride sharing," <http://arxiv.org/abs/1610.09921>.
- [55] S. Banerjee, R. Johari, and C. Riquelme, "Dynamic pricing in ridesharing platforms," *ACM SIGecom Exchanges*, vol. 15, no. 1, pp. 65–70, 2016.
- [56] K. Bimpikis, O. Candogan, and S. Daniela, "Spatial pricing in ride-sharing networks," *Operations Research*, vol. 67, no. 3, 2019.
- [57] C. Dai, "Ridesharing recommendation: Whether and where should i wait?" in *International Conference on Web-Age Information Management*, Springer, pp. 151–163, 2016.
- [58] J. Yuan, Y. Zheng, L. Zhang, and X. G. Xie Sun, "Where to nd my next passenger?" in *Ubicomp'11*, 2011.
- [59] X. Li, G. Pan, Z. Wu et al., "Prediction of urban human mobility using large-scale taxi traces and its applications," *Frontiers of Computer Science*, vol. 6, no. 1, pp. 111–121, 2012.
- [60] N. Agatz, A. Erera, M. Savelsbergh, and X. Wang, "Optimization for dynamic ride-sharing: a review," *European Journal of Operational Research*, vol. 223, no. 2, pp. 295–303, 2012.
- [61] Y. Wang, R. Kutadinata, and S. Winter, "Activity-based ridesharing: increasing exibility by time geography," in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, p. 12016.
- [62] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "The benefits of meeting points in ride-sharing systems," *Transportation Research Part B: Methodological*, vol. 82, pp. 36–53, 2015.
- [63] P. Goel, L. Kulik, and K. Ramamohanarao, "Optimal pick up point selection for effective ride sharing," *IEEE Transactions on Big Data*, vol. 32, pp. 154–168, 2017.
- [64] L. Alarabi, B. Cao, L. Zhao, M. F. Mokbel, and A. Basalamah, *A demonstration of sharek: an efficient matching framework for ride sharing systems*, in *Proceedings of the 24th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, ACM, p. 95, 2016.
- [65] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "Making dynamic ride-sharing work: the impact of driver and rider exibility," *Transportation Research Part E: Logistics and Transportation Review*, vol. 91, pp. 190–207, 2016.
- [66] J. Alonso-Mora, S. Samaranayake, A. Wallar, E. Frazzoli, and D. Rus, "On-demand high-capacity ride-sharing via dynamic trip-vehicle assignment," *Proceedings of the National Academy of Sciences*, vol. 114, no. 3, pp. 462–467, 2017.
- [67] E. D'Andrea, D. Di Lorenzo, B. Lazzerini, F. Marcelloni, and F. Schoen, "Path clustering based on a novel dissimilarity function for ride-sharing recommenders," in *IEEE International Conference on Smart Computing (SMARTCOMP)*, IEEE, pp. 1–8, 2016.
- [68] C. Wu, E. Kamar, and E. Horvitz, "Clustering for set partitioning with a case study in ridesharing," in *2016 IEEE 19th International Conference on Intelligent Transportation Systems (ITSC)*, IEEE, pp. 1384–1388, 2016.
- [69] D. Sánchez, S. Martínez, and J. Domingo-Ferrer, "Co-util p2p ridesharing via decentralization and reputation management," *Transportation Research Part C: Emerging Technologies*, vol. 73, pp. 147–166, 2016.
- [70] T. Altshuler, Y. Shiftan, R. Katoshevski et al., "The network dimension of ride sharing," in *International Conference on Computational Social Science (ICCSS)*, 2015.

- [71] A.-L. Barabasi and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999.
- [72] Z. Galil, "Efficient algorithms for finding maximum matching in graphs," *ACM Computing Surveys (CSUR)*, vol. 18, no. 1, pp. 23–38, 1986.
- [73] B. Chandra and M. M. Halldorsson, "Greedy local improvement and weighted set packing approximation," *Journal of Algorithms*, vol. 39, no. 2, pp. 223–240, 2001.
- [74] M. E. Ben-Akiva and S. R. Lerman, *Discrete Choice Analysis: Theory and Application to Travel Demand*, vol. 9, MIT Press, 1985.
- [75] S. Erlander and N. F. Stewart, *The Gravity Model in Transportation Analysis: Theory and Extensions*, vol. 3, VSP, 1990.
- [76] C.-H. Wen and F. S. Koppelman, "The generalized nested logit model," *Transportation Research Part B: Methodological*, vol. 35, no. 7, pp. 627–641, 2001.
- [77] J. G. Wardrop, "Some theoretical aspects of road traffic research," *Proceedings of the Institution of Civil Engineers*, vol. 1, pp. 325–378, 1952.
- [78] M. G. McNally, "The four step model," *Handbook of Transport Modelling*, vol. 1, pp. 35–41, 2007.
- [79] T. L. Friesz, "Transportation network equilibrium, design and aggregation: key developments and research opportunities," *Transportation Research Part A: General*, vol. 19, no. 5, pp. 413–427, 1985.
- [80] M. D. Meyer and E. J. Miller, *Urban Transportation Planning: A Decision-Oriented Approach*, Osgoode Course Casebooks, 2nd edition, 1984.
- [81] H. Xu, F. Ordóñez, and M. Dessouky, "A traffic assignment model for a ridesharing transportation market," *Journal of Advanced Transportation*, vol. 49, no. 7, pp. 793–816, 2015.
- [82] A. Lee and M. Savelsbergh, "Dynamic ridesharing: Is there a role for dedicated drivers?" *Transportation Research Part B: Methodological*, vol. 81, pp. 483–497, 2015.
- [83] D. J. Fagnant and K. M. Kockelman, *Dynamic ride-sharing and optimal fleet sizing for a system of shared autonomous vehicles*, in *Transportation Research Board 94th Annual Meeting*, no. 15–1962, 2015.
- [84] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, 1977.
- [85] Y. Altshuler, R. Puzis, Y. Elovici, S. Bekhor, and A. Pentland, "Augmented betweenness centrality for mobility prediction in transportation networks," in *International Workshop on Finding Patterns of Human Behaviors in NEtworks and MObility Data, NEMO11*, 2011.
- [86] K. Stephenson and M. Zelen, "Rethinking centrality: methods and examples," *Social Networks*, vol. 11, no. 1, pp. 1–37, 1989.
- [87] P. Bonacich, "Factoring and weighting approaches to status scores and clique identification," *Journal of Mathematical Sociology*, vol. 2, no. 1, pp. 113–120, 1972.
- [88] B. A. Prakash, D. Chakrabarti, M. Faloutsos, N. Valler, and C. Faloutsos, *Got the flu (or mumps)? check the eigenvalue!*, <http://arxiv.org/abs/1004.0060>.
- [89] D. Chakrabarti, Y. Wang, C. Wang, J. Leskovec, and C. Faloutsos, "Epidemic thresholds in real networks," *ACM Transactions on Information and System Security*, vol. 10, no. 4, pp. 1–26, 2008.
- [90] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D.-U. Hwang, "Complex networks: structure and dynamics," *Physics Reports*, vol. 424, no. 4, pp. 175–308, 2006.
- [91] Y.-A. de Montjoye, J. Quoidbach, F. Robic, and A. S. Pentland, "Predicting personality using novel mobile phone-based metrics," *Social Computing Behavioral-Cultural Modeling and Prediction*, Springer, pp. 48–55, 2013.
- [92] V. K. Singh, L. Freeman, B. Lepri, and A. S. Pentland, *Predicting spending behavior using socio-mobile features*, in *Proceedings of the International Conference on Social Computing (SocialCom)*, IEEE, pp. 174–179, 2013.
- [93] S. Ma, Y. Zheng, and O. Wolfson, "T-share: a large-scale dynamic taxi ridesharing service," in *Proceedings of the 29th International Conference on Data Engineering (ICDE)*, IEEE, pp. 410–421, 2013.
- [94] M. R. Garey, D. S. Johnson, and L. Stockmeyer, "Some simplified NP-complete graph problems," *Theoretical Computer Science*, vol. 1, no. 3, pp. 237–267, 1976.
- [95] F. A. Kuipers and P. F. Van Mieghem, "Conditions that impact the complexity of QoS routing," *IEEE/ACM Transactions on Networking (TON)*, vol. 13, no. 4, pp. 717–730, 2005.
- [96] Y. Naveh, Y. Richter, Y. Altshuler, D. L. Gresh, and D. P. Connors, "Workforce optimization: Identification and assignment of professional workers using constraint programming," *IBM Journal of Research and Development*, vol. 51, no. 3.4, pp. 263–279, 2007.
- [97] O. Bahat and S. Bekhor, "Incorporating ridesharing in the static traffic assignment model," *Networks and Spatial Economics*, pp. 1–25, 2015.
- [98] Y.-A. de Montjoye, A. Stopczynski, E. Shmueli, A. Pentland, and S. Lehmann, "The strength of the strongest ties in collaborative problem solving," *Scientific Reports*, vol. 4, no. 1, 2015.
- [99] D. Lazer and A. Friedman, "The network structure of exploration and exploitation," *Administrative Science Quarterly*, vol. 52, no. 4, pp. 667–694, 2007.
- [100] Y. Altshuler, V. Yanovsky, A. Bruckstein, and I. Wagner, "Efficient cooperative search of smart targets using uav swarms," *Robotica*, vol. 26, pp. 551–557, 2008.
- [101] Y. Altshuler and A. M. Bruckstein, "Static and expanding grid coverage with ant robots: complexity results," *Theoretical Computer Science*, vol. 412, no. 35, pp. 4661–4674, 2011.
- [102] J. Svennebring and S. Koenig, "Building terrain-covering ant robots: a feasibility study," *Autonomous Robots*, vol. 16, no. 3, pp. 313–332, 2004.
- [103] S. Koenig, B. Szymanski, and Y. Liu, "Efficient and inefficient ant coverage methods," *Annals of Mathematics and Artificial Intelligence*, vol. 31, pp. 41–76, 2001.
- [104] Y. Altshuler, I. Wagner, V. Yanovski, and A. Bruckstein, "Multi-agent cooperative cleaning of expanding domains," *International Journal of Robotics Research*, vol. 30, no. 8, pp. 1037–1071, 2010.
- [105] E. Regev, Y. Altshuler, and A. M. Bruckstein, *The cooperative cleaners problem in stochastic dynamic environments*, <http://arxiv.org/abs/1608.01987>.
- [106] Y. Altshuler, A. Pentland, S. Bekhor, Y. Shiftan, and A. Bruckstein, *Optimal dynamic coverage infrastructure for large-scale fleets of reconnaissance UAVs*, <http://arxiv.org/abs/1611.05735>.

