Quantifying the benefits of vehicle pooling with shareability networks

Paolo Santi[a,b], Giovanni Resta[b], Michael Szell[a,1], Stanislav Sobolevsky[a], Steven H. Strogatz[c], and Carlo Ratti[a]

[a]Senseable City Laboratory, Massachusetts Institute of Technology, Cambridge, MA 02139; [b]Istituto di Informatica e Telematica del Consiglio Nazionale delle Ricerche, 56124 Pisa, Italy; and [c]Department of Mathematics, Cornell University, Ithaca, NY 14853

Taxi services are a vital part of urban transportation, and a considerable contributor to traffic congestion and air pollution causing substantial adverse effects on human health. Sharing taxi trips is a possible way of reducing the negative impact of taxi services on cities, but this comes at the expense of passenger discomfort quantifiable in terms of a longer travel time. Due to computational challenges, taxi sharing has traditionally been approached on small scales, such as within airport perimeters, or with dynamical ad hoc heuristics. However, a mathematical framework for the systematic understanding of the tradeoff between collective benefits of sharing and individual passenger discomfort is lacking. Here we introduce the notion of shareability network, which allows us to model the collective benefits of sharing as a function of passenger inconvenience, and to efficiently compute optimal sharing strategies on massive datasets. We apply this framework to a dataset of millions of taxi trips taken in New York City, showing that with increasing but still relatively low passenger discomfort, cumulative trip length can be cut by 40% or more. This benefit comes with reductions in service cost, emissions, and with split fares, hinting toward a wide passenger acceptance of such a shared service. Simulation of a realistic online system demonstrates the feasibility of a shareable taxi service in New York City. Shareability as a function of trip density saturates fast, suggesting effectiveness of the taxi sharing system also in cities with much sparser taxi fleets or when willingness to share is low.

carpooling | human mobility | urban computing | maximum matching

Vehicular traffic congestion—and the air pollution that results from it—is one of the greatest challenges facing cities all over the world. It comes at great monetary and human cost: in the 83 largest urban areas of the United States alone, the amount of wasted time and fuel caused by congestion has been placed at US$ 60 billion (1). At the same time, the World Health Organization has estimated that over one million deaths per year worldwide can be attributed to outdoor air pollution (2), which is to a large part caused by vehicular traffic (3). Further adverse effects include fatalities through road accidents and economic losses from missed business activities. For these reasons, great hope is placed today in the rapid deployment of digital information and communication technologies that could help make cities "smarter" (4), and, in particular, that could help manage vehicular traffic more efficiently. The use of real-time information allows the monitoring of the urban mobility infrastructure to an unprecedented extent, and opens up new potential for the exploitation of unused capacity. One major example is the public mobility infrastructure: taking advantage of the widespread use of smart phones and their capabilities for running real-time applications, it is possible to design new, smarter transportation systems based on the sharing of cars or minivans, effectively providing services that could replace public transportation with the on-demand qualities of individual mobility or taxis (5). However, although this option has been proposed in the past, municipal authorities, city residents, and other stakeholders may be reluctant to invest in it until its benefits have been quantified (6). This is the goal of the present paper.

At the basis of a shared taxi service is the concept of ride sharing or carpooling, a long-standing proposition for decreasing road traffic, which originated during the oil crisis in the 1970s (6). During that time, economic incentives outbalanced the psychological barriers on which successful carpooling programs depend: giving up personalized transportation and accepting strangers in the same vehicle. Surveys indicate that the two most important deterrents to potential carpoolers are the extra time requirements and the loss of privacy (7, 8). However, the lack of correlations between socio-demographic variables and carpooling propensity (8), the design of appropriate economic incentives (9), and recent practical implementations of taxi-sharing systems in New York City (http://bandwagon.io) give ample hope that many social obstacles might be overcome in newly emerging "sharing economies" (10, 11).

Besides psychological considerations, it is fundamental to understand the logistic limitations of realistic taxi-sharing systems, which is our focus here. From a theoretical perspective, trip sharing is traditionally seen as an instance of "dynamic pickup and delivery" problems (12, 13), in which a number of goods or customers must be picked up and delivered efficiently at specific locations within well-defined time windows. Such problems are typically solved by means of linear programming, in which a function of the system variables is optimized subject to a set of equations that describe the constraints. Whereas linear programming tasks can be solved with standard approaches of Operations Research or with constraint programming (14), their computational feasibility heavily depends on the number of variables and equations, e.g., the pickup and delivery time windows

---

Significance

Recent advances in information technologies have increased our participation in "sharing economies," where applications that allow networked, real-time data exchange facilitate the sharing of living spaces, equipment, or vehicles with others. However, the impact of large-scale sharing on sustainability is not clear, and a framework to assess its benefits quantitatively is missing. For this purpose, we propose the method of shareability networks, which translates spatio-temporal sharing problems into a graph-theoretic framework that provides efficient solutions. Applying this method to a dataset of 150 million taxi trips in New York City, our simulations reveal the vast potential of a new taxi system in which trips are routinely shareable while keeping passenger discomfort low in terms of prolonged travel time.

of each customer, used to describe the problem at hand. Most previous taxi studies have therefore focused on small-scale routing problems, such as within airport perimeters (15, 16). Large urban taxi systems, in contrast, involve thousands of vehicles performing hundreds of thousands of trips per day. A first step toward practical taxi ride-sharing systems is ref. 17, where the authors present the design of a dynamic ride-sharing system inclusive of a taxi dispatching strategy and fare management. Due to computational reasons trip sharing in ref. 17 is decided based on a heuristic approach tailored to the specific taxi dispatching strategy at hand. Our approach, by contrast, is the development of a framework which enables investigation in general terms the fundamental tradeoff between the benefit and the passenger discomfort induced by taxi-sharing systems at the city level, as an example from a wide class of spatial sharing problems.

Here we introduce the notion of shareability network to model trip sharing in a simple static way, and apply classical methods from graph theory to solve the taxi trip-sharing problem in a provably efficient way. The differences between static trip sharing as considered herein, and dynamic sharing as considered, e.g., in ref. 17, are discussed in detail in *SI Appendix*. The starting point of our analysis is a dataset composed of the records of over 150 million taxi trips originating and ending in Manhattan in the year 2011 by all 13,586 registered taxis. For each trip, the record reports the vehicle ID, the Global Positioning System (GPS) coordinates of the pickup and drop-off locations, and corresponding times. Pickup and drop-off locations have been associated with the closest street intersection in the road map of Manhattan (*Materials and Methods*). We impose a natural network structure on an otherwise unstructured, gigantic search space of the type explored in traditional linear programming. To this end we define two parameters: the shareability parameter k, standing for the maximum number of trips that can be shared, and the quality of service parameter Δ, which stands for the maximum delay a customer tolerates in a shared taxi service trip, mathematically equivalent to the notion of "time window" used in other approaches (13, 17). To ease the analysis, we use the Δ formalism; however, when presented in a real implementation to passengers, it might be psychologically more effective to use the neutral wording "time window" rather than explicitly mentioning the maybe more negatively connoted word "delay." The choice of defining the quality of service parameter as an absolute time, instead of as a percentage increase of the travel time, is in line with similar realizations in the literature (17), and is motivated by the fact that absolute delay information is likely more valuable than percent estimation of travel time increase

Fig. 1. Shareability networks translate spatiotemporal sharing problems into a graph-theoretic framework that provides efficient solutions. (*A*) Example of seven trips, $T_1, \ldots, T_7$, requested and to be shared in Manhattan, New York City. (*B*) Construction of shareability network for $k = 2$. Trips that could potentially be shared are connected, given the necessary time constraints to hold which we assume here to be the case. Trips 1 and 4 cannot be shared because the total length of the best shared route would be longer than the sum of the single routes. Likewise, trip 7 is an isolated node because it cannot possibly be shared with other trips. (*C*) Maximum matching of the shareability network gives the maximum number of trip pairs, i.e., the maximum number of shared trips. (*D*) Implementation (routing) of the maximum matching solution. (*E*) Alternatively, maximum weighted matching of the shareability network gives the solution with the minimal total travel time, which in this case leads to a different solution than unweighted maximum matching. Here only two pairs of trips are shared, but the amount of travel time saved, given by the sum of link weights of the matching, 30 + 16, is optimal. (*F*) Implementation (routing) of the weighted maximum matching solution. (*G*) k sharing and taxi capacity. Each of the three cases involves a number of trips T_i to be shared, but ordered differently in time t. (*Top*) This case corresponds to a feasible sharing according to our model with $k = 2$, and the trips can be accommodated in a taxi with capacity ≥ 2. (*Middle*) This case corresponds to a model with $k = 3$ because three trips are combined, but the three trips can be combined in a taxi with capacity = 2 because two of the trips are nonoverlapping. (*Bottom*) This case corresponds to $k = 3$, but here a taxi capacity ≥ 3 is needed to accommodate the combined trips. Here we are assuming one passenger per trip, in line with the data reported in ref. 18, according to which the average number of passengers per trip is 1.3.

for potential customers of a shared taxi service. Further, let $T_i = (o_i, d_i, t_i^o, t_i^d)$, $i = 1 \ldots k$ 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 the o_i and d_i in any order where each o_i precedes the corresponding d_i, except for configurations where single trips are concatenated and not overlapped like $o_1 \rightarrow d_1 \rightarrow o_2 \rightarrow d_2$, such that each customer is picked up and dropped off at the respective origin and destination locations with delay at most Δ, with the delay computed as the time difference to the respective single, individual trip. Imposing a bound of k on shareability implies that the k trips can be combined using a taxi of corresponding capacity (Fig. 1G). Deciding whether two or more trips can be shared necessitates knowledge of the travel time between arbitrary intersections in Manhattan, which we estimated using an ad hoc heuristic (*SI Appendix*, Fig. S2 and Table S1).

For the case $k = 2$, the shareability network associated with a set $\mathcal{T}$ of trips is obtained by assigning a node T for each trip in $\mathcal{T}$, and by placing a link between two nodes T_i and T_j if the two trips can be shared for the given value of Δ (Fig. 1 A and B). The value of Δ has a profound impact on topological properties of the resulting shareability network. Increasing Δ capitalizes on well-known effects of time-aggregated networks such as densification (19, 20), capturing the intuitive notion that the more patient the customers, the more opportunities for trip sharing arise (Fig. 2 A and B). For values of $k > 2$, the shareability network has a hypergraph structure in which up to k nodes can be connected by a link simultaneously. Because of computational reasons, the shareability parameter k has a substantial impact on the feasibility of solving the problem. A solution is tractable for $k = 2$, heuristically feasible for $k = 3$, whereas it becomes computationally intractable for $k \geq 4$ (*SI Appendix*). This constraint implies that taxi-sharing services, and social-sharing applications in general, will likely be able to combine only a limited number of trips. However, as we show below, even the minimum possible number of trip combinations ($k = 2$) can provide immense benefits to a dense enough community like the city of New York.

With the shareability network, classical algorithms for solving maximum matching on graphs (21, 22) can be used to determine the best trip-sharing strategy according to two optimization criteria: (*i*) maximizing the number of shared trips, or (*ii*) minimizing the cumulative time needed to accommodate all trips. To find the best solution according to (*i*) or (*ii*), it is sufficient to compute a maximum matching or a weighted maximum matching on the shareability

network, respectively (Fig. 1 C and E, *Materials and Methods*). Because a shared trip can be served by a single taxi instead of two, the number of shared trips can be used as a proxy for the reduction in number of circulating taxis. For instance, an 80% rate of shared trips translates into a 40% reduction of the taxi fleet. Other important objectives such as total system cost and emissions are reasonably approximated by criterion (*ii*).
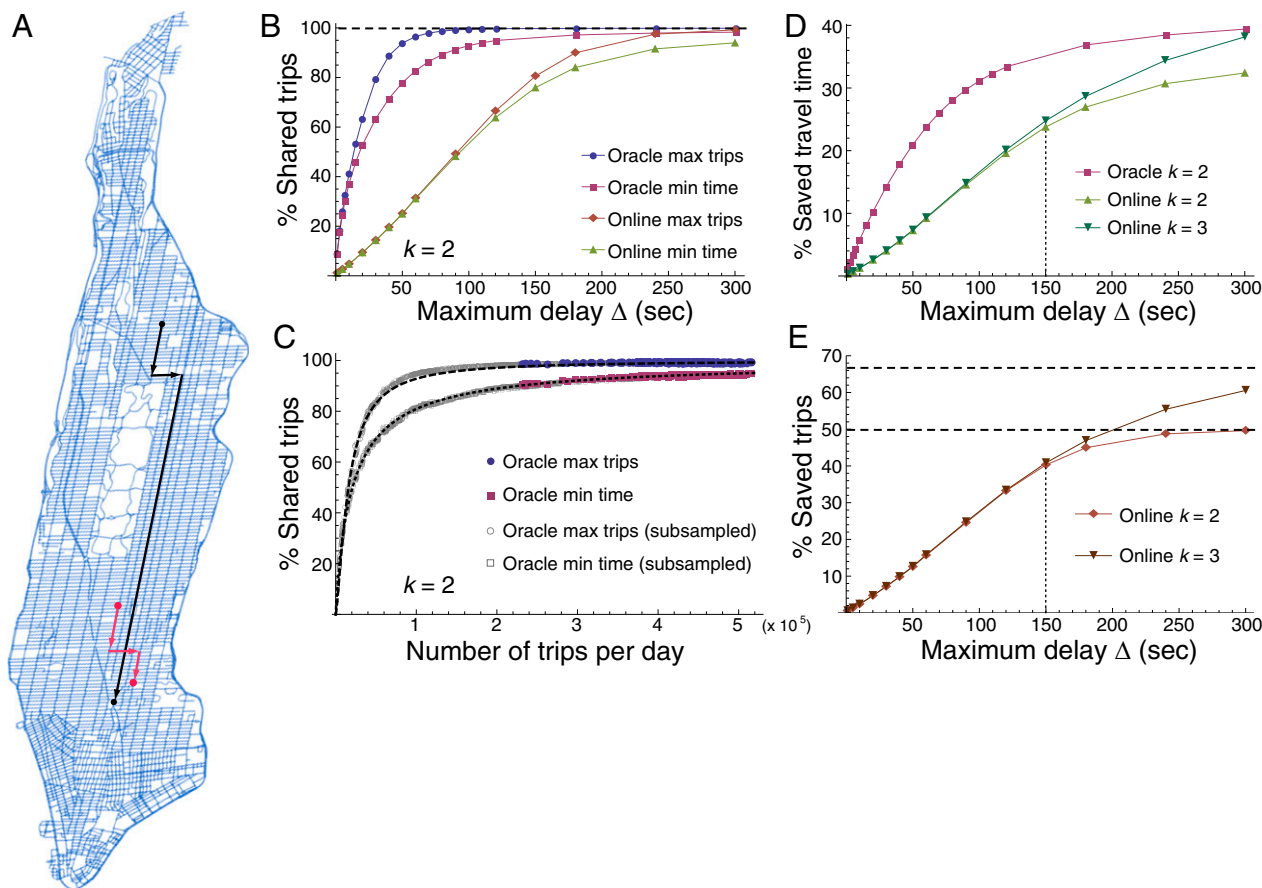
Results

Using a maximum value of $\Delta = 10$ min and all trips performed in New York City in the year 2011, the resulting shareability network has more than 150 million nodes and over 100 billion links. We first consider trip-sharing opportunities under a model in which the entire shareability network is known beforehand, and maximum matchings are computed on the entire network. This omniscient Oracle approach models an artificial scenario in which trip-sharing decisions can be taken considering not only the current taxi requests, but also all future ones, serving as a theoretical upper bound for sharing opportunities. In practice, the Oracle model is useful to assess the benefits of social-sharing systems where bookings are placed well ahead of time (Fig. 3A). Because of this foreknowledge, even with the low and reasonable value of $\Delta = 2$ min, the average percentage of shareable trips is close to 100% (Fig. 3B).

In practical systems however, the Oracle approach is of limited use, as only trip requests issued in a relatively short time window are known at decision time, corresponding to a small time-slice of the shareability network. In the following, we therefore focus on trip-sharing opportunities in a realistic model in which the trip-sharing decision for a trip T_1 considers only trips that start within a short interval around its starting time t_1^o. More formally, we retain in the shareability network only links connecting trips T_i and T_j such that $|t_i^o - t_j^o| \leq \delta$, where δ is a time window parameter. This Online model is representative of a scenario in which a customer, using an "e-hailing" application, issues a taxi request reporting pickup and drop-off locations, and after the small time window δ receives feedback from the taxi management system on whether a shared ride is available. This parameter is fundamental in the Online model: the larger δ, the more trip-sharing opportunities can be exploited, for the same reasons of network time aggregation as with Δ (*SI Appendix*, Fig. S3). However, δ should be kept reasonably small to be acceptable by a potential customer, and to allow real-time computation of the shared trip matching (*SI Appendix*). Therefore, in what follows, we set $\delta = 1$ min.

As expected, reducing the time horizon δ from practically infinite in the Oracle model to 1 min in the Online model considerably reduces trip-sharing opportunities for low values of Δ. For instance, when $\Delta = 1$ min, the Oracle model allows sharing of 94.5% of the trips, but the Online model only less than 30%. However, the situation is much less penalizing for the Online model when the delay parameter is increased within reasonable range. When $\Delta = 5$ min, the Online model can exploit virtually all available trip-sharing opportunities (Fig. 3B). Concerning saved travel time, results are similarly promising (Fig. 3D). When $\Delta = 5$ min, we can save 32% of total travel time with the Online model, compared with 40% savings in the optimal Oracle model. Note that our method only concerns the sharing of nonvacant trips, but these make up the majority of taxi traffic (18, 23). In fact, the fraction of time during which taxis are serving customers corresponds to the high value of about 75% of the on-service time of a taxi (*SI Appendix*, Fig. S1). Accounting for the effect of empty trips would approximately reduce the total travel time savings from 40% and 32% to the still substantial values of 30% and 24% in the Oracle and Online model, respectively.

Is it possible to even further improve efficiency by increasing the number k of shareable trips? When $k = 3$, the shareability network becomes a shareability hypernetwork, for which

Fig. 2. Shareability networks densify with longer time aggregation, increasing sharing opportunities. This exemplary subset of the shareability network corresponds to 100 consecutive trips for values of (A) $\Delta = 30$ s and (B) $\Delta = 60$ s. Open links point to trips outside the considered set of trips. Isolated nodes are represented as self-loops. Node positions are not preserved across the networks. A similar, although visually not insightful, densification effect is observed in shareability networks obtained when $k = 3$.

Fig. 3. Benefits of trip sharing. (*A*) Street network of Manhattan, and examples of two trips that can be shared under the omniscient Oracle model, but not under the Online model. The starting time of the red trip is much later than that of the black trip, but in the Online model trip sharing decisions must be taken within a very short time window $\delta = 1$ min to notify customers of trip-sharing opportunities as soon as possible after their order. (*B*) Percentage of shared trips as a function of the trip time delay Δ in the Oracle and in the Online model for the two considered optimization criteria of maximizing shared trips (max trips) and minimizing total travel time (min time), when up to $k = 2$ trips can be shared. (*C*) Shareability as a function of trips per day in the Oracle model. Typical days in New York City feature around 400,000 trips with near-maximum shareability. Subsampling data by randomly removing vehicles reveals the underlying saturation curves, fit (dashed lines) by a simple function of type $f(x) = Kx^n/(1 + Kx^n)$ with the two parameters K and n well-known in adsorption processes and biochemical systems (*SI Appendix*). The fast, hyperbolic saturation implies that taxi sharing could be effective even in cities with vehicle densities much lower than New York, or when the willingness to share is low. (*D*) Percentage of saved travel time as a function of Δ for $k = 2$ and $k = 3$. Although δ is reduced from practically infinite in the Oracle model to $\delta = 1$ min in the Online model, saved travel time is well above 30% for $\Delta = 300$ s, for $k = 3$ almost reaching the maximum possible value from the Oracle model with $k = 2$. (*E*) Percentage of saved trips as a function of Δ with $k = 2$ and $k = 3$. The theoretically possible maximum (dashed lines) of 50% for $k = 2$ and 66.7% for $k = 3$ are closely approximated. For $\Delta < 150$ s (dotted line), the benefits of 3-sharing over 2-sharing are negligible.

maximum matching is solvable only in approximation using a heuristic algorithm which is computationally feasible for relatively small networks only (24, 25). Because of this methodological issue and the combinatorial explosion of sharing options, we calculated the number of shared trips and the fraction of saved travel time for $k = 3$ only in the Online model—which by definition features much smaller shareability networks. Simulations show that increasing the number of shareable trips k provides noticeable benefits only when the quality of service parameter Δ crosses a threshold around $\Delta_{\text{crit}} \sim 150$ s (Fig. 3 *D* and *E*). When $\Delta = 300$ s, the number of saved taxi trips is increased from about 50% with $k = 2$ to about 60% with $k = 3$, which is however well below the 66.7% maximum theoretical percentage of shared trips. This suboptimal result suggests that the effort for implementing a service for sharing $k > 2$ trips may not be well-justified. Further, to become widely accepted, a multishared taxi service might require vehicles of higher capacity and/or physically separated, private compartments, possibly inflating overhead for $k > 2$. Because the benefit of multisharing is not that high, it might not cover these additional expenses.

Discussion

Our analysis shows that New York City offers ample opportunities for trip sharing with minimal passenger discomfort, without having to resort to a computationally demanding sharing strategy in which already started trips would be rerouted on the fly, and that these opportunities are realistic to be implemented in a new taxi system. From a computational standpoint, the polynomial runtimes of our algorithms suggest that there should be no issues with designing systems in which taxi companies calculate sharing options within $\delta = 1$ min of the request and immediately dispatch their taxis. By implementing a system that is 40% more efficient and affordable, the ultimate goal is to make taxi systems a more attractive and sustainable mode of transportation, able to generate increased demand and to satisfy it with the current or an even higher number of vehicles.

To assess to what extent our results could be generalized to cities with lower taxi densities than New York, or to account for situations where willingness to share or where market penetration of an accompanying software application is low, we studied how the number of shareable trips in a given day changes as

a function of the total number of trips (Fig. 3C). The average number of daily trips in New York is highly concentrated around 400,000. Hence, we have generated additional low-density situations by subsampling the dataset, randomly removing increasing fractions of vehicles from the system (*Materials and Methods*). The resulting shareability values are excellently fit by saturation curves of the form $f(x) = Kx^n/(1 + Kx^n)$. These curves are well-known to describe binding processes in biochemical systems, providing an interesting link to general pairing problems (*SI Appendix*). At around 100,000 trips, or 25% of the daily average, we already reach saturation and near-maximum shareability. This fast saturation suggests that taxi-sharing systems could be effective even in cities with taxi fleet densities much lower than New York.

Future work should aim to assess in more detail the psychological limitations of taxi sharing, to understand the conditions and appropriate incentive systems under which individuals are willing to be seated in the same vehicle. This includes the design of suitable faring systems aimed at fairly distributing the economic benefits of sharing between drivers and customers, such as the one proposed in ref. 17. Moreover, the sharing analysis should be extended to other cities to better understand the generalizability of the results, and if possible, to measure and incorporate currently unknown data such as the actual search or waiting times of passengers who are trying to find an empty taxi, or the number of passengers that are being transported per vehicle. Finally, the framework of shareability networks could be used to study more generally other social sharing scenarios (26) such as ride sharing of cars, bikes, etc. or the communal use of equipment which is characterized by considerable unit cost and infrequent use, stimulating new forms of sharing and models of ownership (10).

Materials and Methods

Trip Data. The dataset contains origin-destination data of all 172 million trips with passengers of all 13,586 taxicabs in New York during the calendar year of 2011. There are 39,437 unique driver IDs in the dataset, which corresponds to 2.9 drivers per taxi on average. The dataset contains a number of fields from which we use the following: medallion ID, origin time, destination time, origin longitude, origin latitude, destination longitude, and destination latitude. Times are accurate to the second; positional information has been collected via GPS technology by the data provider. Out of our control are possible biases due to urban canyons which might have slightly distorted the GPS locations during the collection process (27). All IDs are given in anonymized form; origin and destination values refer to the origins and destinations of trips, respectively.

Map Data and Map Matching. To create the street network of Manhattan we used data from www.openstreetmap.org. We filtered the streets of Manhattan, selecting only the following road classes: primary, secondary, tertiary, residential, unclassified, road, and living street. Several other classes were deliberately left out, such as footpaths, trunks, links, or service roads, as they are unlikely to contain delivery or pickup locations. Next we extracted the street intersections to build a network in which nodes are intersections and directed links are roads connecting those intersections (we use directed links because a nonnegligible fraction of streets in Manhattan are one-way). The extracted network of street intersections was then manually cleaned for obvious inconsistencies or redundancies (such as duplicate intersection points at the same geographic positions), in the end containing 4,091 nodes and 9,452 directed links. This network was used to map match the GPS locations from the trip dataset. We only matched locations for which a closest node in the street intersection network exists with a distance less than 100 m. Finally, from the remaining 150 million trips we discarded about 2 million trips that had identical starting and end points, and trips that lasted less than 1 min.

Maximum Matching of Shareability Networks. Given a graph $G = (V, E)$, a matching M in G is a set of pairwise nonadjacent edges. A maximum matching is a matching that contains the largest possible number of edges. A weighted maximum matching is a matching in which the sum of edge weights is maximal. In the context of shareability networks, maximum matching solves optimizing the number of shared trips, whereas weighted maximum matching minimizes the cumulative time needed to accommodate all trips if the weights on the shareability network are taken as the travel time that is saved by sharing. Given that shareability networks are sparse, for the case $k = 2$ maximum matching and weighted maximum matching can be solved in polynomial times $O(n\sqrt{n})$ and $O(n^2 \log n)$ (22), respectively, where n is the number of nodes in the network. For higher dimensions, $k > 2$, fast approximations to the optimal solutions exist (24), which however become computationally unfeasible for $k > 3$. For details see *SI Appendix*.

Subsampling of Vehicles. To assess to which extent our results could be generalized to cities with lower taxi densities than New York, or to situations where willingness to share is low, we have generated additional low-density situations by subsampling our dataset, randomly removing various fractions of vehicles from the system in the following way: For each day in the dataset, we randomly selected a percentage c of the taxis in the trace, and deleted the corresponding trips from the dataset. We varied c from 95% down to 1%, generating a number of trips per day as low as 1,962. Note that by subsampling the vehicles we filter both taxis and the trips which represent the demand.

1. Arnott R, Small K (1994) The economics of traffic congestion. *Am Sci* 82(5):446–455.
2. World Health Organization (2011) *World Health Statistics 2011* (World Health Organization, Geneva).
3. Caiazzo F, Ashok A, Waitz IA, Yim SH, Barrett SR (2013) Air pollution and early deaths in the united states. part i: Quantifying the impact of major sectors in 2005. *Atmos Environ* 79:198–208.
4. Batty M (2013) *The New Science of Cities* (MIT Press, Cambridge, MA).
5. Dowling R, Simpson C (2013) Shift–the way you move: Reconstituting automobility. *Continuum (N Y)* 27(3):421–433.
6. Handke V, Jonuschat H (2013) *Flexible Ridesharing* (Springer, Berlin).
7. Dueker KJ, Bair BO, Levin IP (1977) Ride sharing: Psychological factors. *Transp Engrg J* 103:685–692.
8. Teal RF (1987) Carpooling: Who, how and why. *Transp Res A* 21(3):203–214.
9. Ben-Akiva M, Atherton TJ (1977) Methodology for short-range travel demand predictions: Analysis of carpooling incentives. *J Transp Econ Policy* 11(3):224–261.
10. Botsman R, Rogers R (2010) *What's Mine Is Yours: The Rise of Collaborative Consumption* (HarperCollins, New York).
11. John NA (2013) The social logics of sharing. *Commun Rev* 16(3):113–131.
12. Yang J, Jaillet P, Mahmassani H (2004) Real-time multivehicle truckload pickup and delivery problems. *Transport Sci* 38(2):135–148.
13. Berbeglia G, Cordeau JF, Laporte G (2010) Dynamic pickup and delivery problems. *Eur J Oper Res* 202(1):8–15.
14. Shaw P (1998) *Principles and Practice of Constraint Programming-CP98* (Springer, Berlin), pp 417–431.
15. Marin A (2006) Airport management: Taxi planning. *Ann Oper Res* 143(1):191–202.
16. Clare GL, Richards AG (2011) Optimization of taxiway routing and runway scheduling. *Proc IEEE Intelligent Transportation Systems* 12(4):1000–1013.
17. Ma S, Zheng Y, Wolfson O (2013) T-Share: A large-scale dynamic taxi ridesharing service. *Proc IEEE ICDE* (IEEE, Brisbane, QLD), pp 410–421.
18. Bloomberg M, Yassky D (2014) New York City 2014 Taxicab FactBook (New York City Taxi and Limousine Commission, New York).
19. Kossinets G, Watts DJ (2006) Empirical analysis of an evolving social network. *Science* 311(5757):88–90.
20. Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrinking diameters. *ACM Trans Knowl Discov Data* 1(1):1–41.
21. Cormen TH, Leiserson CE, Rivest RL (1990) *Introduction to Algorithms* (MIT Press, Cambridge, MA; McGraw-Hill, New York).
22. Galil Z (1986) Efficient algorithms for finding maximum matching in graphs. *ACM Comput Surv* 18(1):23–38.
23. Phithakkitnukoon S, Veloso M, Bento C, Biderman A, Ratti C (2010) Taxi-aware map: Identifying and predicting vacant taxis in the city. *Ambient Intelligence: Lecture Notes in Computer Science* (Springer, Berlin), Vol 6439, pp 86–95.
24. Chandra B, Halldorsson M (2001) Greedy local improvement and weighted set packing approximation. *J Alg* 39(2):223–240.
25. Johnson DS (1974) Approximation algorithms for combinatorial problems. *J Comput Syst Sci* 9(3):256–278.
26. Carvalho R, Buzna L, Just W, Helbing D, Arrowsmith DK (2012) Fair sharing of resources in a supply network with constraints. *Phys Rev E Stat Nonlin Soft Matter Phys* 85(4 Pt 2):046101.
27. Grush B (2008) The case against map-matching. *Eur J Nav* 6(3):2–5.

Santi et al.