

## MIT Open Access Articles

### *Home-work carpooling for social mixing*

The MIT Faculty has made this article openly available. **Please share** how this access benefits you. Your story matters.

**Citation:** Librino, Federico et al. "Home-work carpooling for social mixing." *Transportation* (July 2019): doi.org/10.1007/s11116-019-10038-2 © 2019 Springer Science+Business Media, LLC

**As Published:** <http://dx.doi.org/10.1007/s11116-019-10038-2>

**Publisher:** Springer Science and Business Media LLC

**Persistent URL:** <https://hdl.handle.net/1721.1/126887>

**Version:** Author's final manuscript: final author's manuscript post peer review, without publisher's formatting or copy editing

**Terms of use:** Creative Commons Attribution-Noncommercial-Share Alike



## Home-work Carpooling for Social Mixing

Federico Librino · M. Elena Renda · Paolo Santi · Francesca  
Martelli · Giovanni Resta · Fabio Duarte · Carlo Ratti ·  
Jinhua Zhao

Received: date / Accepted: date

**Abstract** Shared mobility is widely recognized for its contribution in reducing carbon footprint, traffic congestion, parking needs and transportation-related costs in urban and suburban areas. In this context, the use of carpooling in home-work commute is particularly appealing for its potential of lessening the number of cars and kilometers traveled, consequently reducing major causes of traffic in cities. Accordingly, most of the carpooling algorithms are optimized for reducing total travel time, cost, and other transportation-related metrics. In this paper, we analyze carpooling from a new perspective, investigating the question of whether it can be used also as a tool to favor social integration, and to what extent social benefits should be traded off with transportation efficiency. By incorporating traveler's social characteristics into a recently introduced network-based approach to model ride-sharing opportunities, we define two social-related carpooling problems: how to maximize the number of rides shared between people belonging to different social groups, and how to maximize the amount of time people spend together along the ride. For each of the problems, we provide corresponding optimal and computationally efficient solutions. We then demonstrate our approach on two datasets collected in the city of Pisa, Italy, and Cambridge, US, and quantify the potential social benefits of carpooling, and how they can be traded off with traditional transportation-related metrics. When collectively considered, the models, algorithms, and results presented in this paper broaden the perspective from which carpooling problems are typically analyzed to encompass multiple disciplines including urban planning, public policy, and social sciences.

**Keywords** Shared Mobility · Social Mixing · Social Carpooling · Mobility Planning.

---

A very short preliminary version of the researches reported in this paper has been presented to the 97<sup>th</sup> Transportation Research Board Annual Meeting

Corresponding Author: M. Elena Renda. Present address: erenda@mit.edu; permanent address: elena.renda@iit.cnr.it

F. Librino, M. E. Renda, P. Santi, F. Martelli, and G. Resta  
Istituto di Informatica e Telematica – CNR, Pisa, Italy  
E-mail: {federico.librino, elena.renda, paolo.santi, francesca.martelli, giovanni.resta}@iit.cnr.it

M. E. Renda and J. Zhao  
JTL Mobility Lab – MIT, Cambridge, USA  
E-mail: {erenda, jinhua}@mit.edu

P. Santi, F. Duarte, and C. Ratti  
Senseable City Lab – MIT, Cambridge, USA  
E-mail: {psanti, fduarte, ratti}@mit.edu

F. Duarte  
Pontificia Universidade Catolica do Parana, Curitiba, Brazil

## 1 Introduction

Car- and ride-sharing programs have been gaining momentum in transportation planning, for their contribution to reduce cost, carbon footprint, traffic congestion and parking needs. They have also proved to be commercially successful, with companies such as Uber and Didi Chuxing expanding their operations in developed and developing countries, and some cities even incorporating them as part of their transit options. The success of car-sharing programs came somehow as a surprise in transportation studies, which for decades have been trying to cope with the negative environmental externalities of the automobile while acknowledging that a door-to-door transportation system, such as the private car, brings powerful benefits to its users. The gap between the awareness of negative effects of private cars at the societal level (traffic, pollution) and the personal attachment to the benefits it brings at the individual level is well known [13, 20, 24, 35].

Shared mobility programs have been promoting a half-way solution: by sharing trips, users of the system reduce the negative externalities tied to the use of private cars, while still enjoying the comfort and some level of privacy way beyond what is offered in public transportation. Still, a higher acceptance of car-sharing faces the challenge of convincing more users to share a car with a total stranger who is not a professional driver. This human aspect of shared mobility is still vastly understudied in transportation scholarship, which mostly understands mobility primarily by its instrumental value, i.e. an activity undertaken to move from point A to point B. Only recently, the intrinsic values of the time spent traveling and the social dimensions of mobility have been gaining traction [8]. Some scholars have been pointing to the “social capital” of sharing daily trips with others. For instance, [7] compared users of public and private transportation in the Netherlands, considering the first as more exposed to social diversity and the latter as less exposed, concluding that while more exposure to diversity does not necessarily create more connections among people, non-exposure certainly blocks opportunities of person-to-person interactions. Using travel smart card data in Singapore, [33] demonstrates that encounters between strangers in public transportation show regular temporal patterns, creating a sort of a small-world network of “familiar strangers”.

Acknowledging the growing interest into the human and social aspects related to shared mobility, this paper studies for the first time the potential of a specific car sharing program – carpooling – to become a tool for promoting social mixing between heterogeneous social categories. Introducing social mixing into car-sharing schemes builds upon a recognized new feature of contemporary sharing economy: the “stranger sharing” [15]. While traditional forms of sharing were usually confined to a closed network of friends, family and trusted individuals, sharing apps have been successful in creating ways to make strangers share resources and services. The success of sharing platforms indicates that mixing people with different characteristics into a shared ride is better accepted than previously imagined.

The focus on carpooling is motivated by the fact that, different from app-based car-sharing schemes, in this shared mobility model the driver is not only transporting other people, but she is also purposefully travel from an origin to a destination, often recurrently. Thus, travel endpoints can be considered as “social anchor points” in social network analysis [9]. While psychological and social aspects still influence the willingness to participate in carpooling schemes [14], a common trace of carpooling groups, or clubs, is that they “provide a common minimum level of trust between its members” [11]. Thus, carpooling, especially if recurrently used for home-work commuting, is the ideal setting for attempting to bridge the gap between societal and individual utility of car transportation.

This paper investigates the question of whether carpooling can be used as a tool to *actively* promote social integration. We thus revert the traditional perspective that considers social ties as possible constraints or impediment to ride sharing [11, 26], and analyze the potential of carpooling to act as “social mixer”, in which social characteristics are taken into consideration [5]. A thorough discussion of this novel viewpoint through which the performance of carpooling is analyzed is reported in Section 2.

**Table 1** Reference Table.

Name	Description
$T_i$	Trip $i$
$S(i)$	Starting location of trip $i$ in $(lat, long)$ coordinates
$D(i)$	Destination location of trip $i$ in $(lat, long)$ coordinates
$d_{AB}$	Distance between two locations $A$ and $B$
$\tau_{AB}$	Time required to travel from $A$ to $B$
$\Delta$	Maximum allowed sharing delay
$M_c$	Maximum Cardinality Matching Algorithm
$M_d$	Maximum Saved Distance Matching Algorithm
$M_t$	Maximum Mingling Time Matching Algorithm
$M_x$	Maximum Social Mixing Time Matching Algorithm
CDF	Cumulative distribution function
$\rho$	Assortativity coefficient
$\sigma_{ij}$	Kilometers saved while sharing trips $i$ and $j$
$\sigma$	Normalized saved mileage achieved with trip sharing
$\tau_{ij}$	Mingling time spent together by the travelers associated to the shared trips $i$ and $j$
$\tau$	Average mingling time achieved with trip sharing
$\psi$	Fraction of matched trips
$\phi$	Fraction of dissortative links achieved with trip sharing

In order to answer the question above, we devise a novel methodology and a set of algorithmic tools to *actively* promote social integration through carpooling. The devised methodology is then applied to data sets collected in the city of Pisa, Italy, and Cambridge, US, to provide accurate quantification of the potential social benefits of carpooling, and discuss how social-based metrics can be traded off with traditional transport-based metrics in identifying the most desirable carpooling strategy.

The paper is organized as follows: Section 2 explains the main idea behind the paper, namely to explore carpooling as a possible means for promoting social integration. Section 3 discusses related work and highlight the contribution of this paper. Section 4 describes the used methods and material, in particular: Section 4.1 describes how carpooling could be modeled through shareability networks, a computationally efficient network-based approach; Section 4.2 describes four possible metrics defined for optimizing the matching of rides, two of which geared towards maximization of environmental benefits, while the other two oriented to maximize potential social benefits; Section 4.3 describes the data we used for the analyses. Section 5 shows the results of our analyses, while Section 6 concludes the paper summarizing the main findings and discussing possible future research directions. Table 1 reports all the symbols and specific terms used throughout the paper for immediate reference of the reader.

## 2 Carpooling for Social Integration

This paper investigates whether and to what extent carpooling can be used as a tool for not only reducing the carbon footprint of home-work commuting, but also to favor social integration. Although ride-sharing systems must be designed to minimize the inconvenience for both drivers and passengers in terms of waiting and travel time to provide an acceptable service, researchers have tested certain degrees of inconvenience (such as including meeting points along the journey [32]) to optimize the efficiency of the system without decaying the participants' satisfaction with the ride-sharing experience. Here, we introduce another dimension for analysis: by coupling network-based shared mobility models with a social profile of the travelers performing the trips, we demonstrate that it is possible to define social-based metrics linked to sharing opportunities, and ultimately to prioritize the ride matching choices accordingly.

It is important to observe that this paper does not attempt to claim that travelers of different profiles interact more. For the social interaction to happen, what matters is not similarity but compatibility

between two individuals. More broadly speaking, for social mixing two steps need to happen: 1) to create opportunities for social interaction; and 2) given the opportunity, assess whether and to what extent that opportunity fostered a fruitful social interaction. This paper is focused with 1), and in particular to investigate to what extent carpooling can be used to create opportunities for social interaction.

Social mixing is highly valued in urban studies and transit studies and policies. Although social mixing is an umbrella term that include different characteristics in different contexts, overall its goal is to bring together social groups with different social, economic, ethnic characteristics within the same urban unit, what is achieved with the implementation of specific programs [18]. Attempts to use transit to foster social mixing have a long history. Early 20th century, transportation policies in New York actively fostered social mixing, including open seating, single price, first-served basis, lateral seats giving more room to physical proximity inside the cars [19]. Research on carpooling and ride-sharing and social mixing is understandably restrict, since these are relatively new forms of mobility and not as common as transit. Still, surveying 2,041 Uber and Lyft users, in [23] the authors showed that rider-to-rider discriminatory attitudes do not present a significant barrier to getting individuals in shared vehicles, and are actually negatively predictive of the passengers level of service among dynamic ridesharing users (when the riders do not know who their ride companion will be).

Our approach is aimed at evaluating the *potential* for social integration and its fundamental trade off with traditional transport-related metrics. In current ride-sharing systems, the equilibrium between the number of available cars and potential users is the resultant of a day-to-day adjustment process that aims to reach an optimal operational point between offer and demand. The equilibrium point is tied to the daily choices of travelers and operators, with implications on several aspects of flexible transport services, such as fare price, operator's profit, and wait time for passengers. Such analyses could inform operating policies - in particular in cases where there is a clear cut between who provides the service and who uses the service, such as with taxis [1]. By adding social characteristics of the users as an element in the matching algorithms, as we propose here, the process of optimally matching rides is taken out of the control of customers and assumed to be performed by a central server that collects trip requests, computes the optimal matching, and returns the results to the carpooling users. In this regard, stimulating social integration becomes part of a transportation policy and strategy, not simply a personal choice.

We also would like to emphasize that this paper deliberately proposes a centralized approach to ride matching optimization since recent studies in social aspects of transportation have clearly shown that, when left free to choose, user preferences tend to induce social segregation, if not explicit race and gender biases [17,29]. Similar evidences come from housing policy experiences, where initiatives aimed favoring social inclusion are customarily centrally enacted and often locally enforced [6]. Thus, we believe that, in order to fully exploit the potential of carpooling as a tool for social integration, ride matching decision should be mostly taken out of user control and centrally optimized.

More generally speaking, how to translate the potential benefits investigated in this paper into actual societal and ecological returns becomes a new research problem by itself, bearing some resemblance to the field of mechanism design in game theory where, known the socially optimal outcome of a game, the goal is to design its rules in such a way that selfish players converge to the social optimum [27]. Analogously, this paper then opens the problem of how to design policy, incentives, and intervention actions so that the socially and economically desirable outcome of carpooling defined herein becomes acceptable for the individuals participating in the carpooling program.

We stress that the specific social integration metric and objectives reported herein – e.g., the choice to favor dissortative over assortative shared trips (see Section 4.2) – shall be intended for exemplifying reasons only. The methodology and ideas presented in this paper easily extend to other social metrics and optimization objectives, as defined for the specific urban scenario at hand. For instance, the approach can be used to favor travel time between members of the same social groups, thus favoring social cohesion within a group instead of mixing across groups.

Returning to the problem of evaluating the potential social benefits of carpooling, the main challenge is to identify datasets that comprise not only travel data, but also some information regarding social characteristics of the travelers. While GPS traces would provide extensive travel data and accuracy in mobility tracking, none of the GPS trace for academic study we are aware of contains also information about travelers. Thus, we resume to survey data to perform the study reported herein. While smaller, yet significant, in size, the datasets used in this study have the unique feature of providing both accurate, street-level information about the origin and destination of trips, and rich profile information about the travelers.

### 3 Related work and Contribution

The social and human behavioral aspects of shared mobility have consistently attracted attention of the research community only recently. Typically, the traveler behavior dimension in transportation studies (including ride-sharing models) is restricted to waiting times -both in traveler's strategies to reduce waiting times and as a parameter in the optimization algorithms [21]. Interestingly, [36] propose an incentive-based active demand management (ADM) method that, based on requests from automobile commuters, provides multiple departure times and route choices as a way to create incentives for potential car-sharing, with users receiving more benefits if the chosen departure time and route are more beneficial to the entire system. [38] propose a ride-matching method that incorporates personal preferences of carpool commuters. In their case, the travel behaviors taken into account are route preferences, departure time, walking distance to the carpool location, and the size of the carpool group. In spite of its contribution and its argument on the importance of considering commuter behaviors, the social characteristics of commuters are not included in their model and are not considered in the ride matching process, which is precisely the contribution of our paper. In a related paper [28], De Palma et al. investigate how preferences related to spending more time with a partner tend to modify the starting time of trips, with an ultimate negative effect on congestion.

In [12], the authors investigate the dynamics of reaching optimal equilibrium points in shared mobility markets. While their shared mobility model, based on a clear distinction between mobility providers (drivers) and users (passengers), is different from the carpooling scenario considered here where a participant can take either role, integrating the ideas and approach of [12] to include social-related metrics in the shared ride formation process is an interesting avenue for further research that is outside the scope of this paper.

From a methodological viewpoint, the modeling and analytic approach presented herein has the merit of bridging the gap between socially-oriented [2, 24] and mathematically-oriented [4, 31, 37] approaches to carpooling, presenting a mathematically clean optimization framework. Specifically, we extend the computationally efficient model of shareability networks introduced in [31] to incorporate social characteristics of the travelers. In doing so, we expand the scope of matching algorithms from traditional transport-related metrics [16] to include the potential social capital of the trips. This is the main conceptual contribution of this paper.

## 4 Material and Methods

### 4.1 Modeling Carpooling

We model carpooling opportunities based on spatial and temporal constraints. Notice that the definition of carpooling opportunity reported herein is independent of the social profile of travelers, which will be introduced in the next section.

To model carpooling we use a network representation of the sharing opportunities we call *carpooling shareability network*, defined as  $N = (T, E)$ , where  $T$  is the set of nodes corresponding to the trips in the dataset, and  $E$  is the set of directed links between nodes corresponding to shareable trips.

Given a set of commuting trips, each trip  $T_i$  in the dataset  $T$  is defined by the following quantities:

- $S(i)$ , the starting location, expressed as a pair of  $(lat, long)$  coordinates;
- $D(i)$ , the destination location, also expressed as a pair of  $(lat, long)$  coordinates;
- $s_t(i)$ , the typical starting time of the trip, as reported by respondents in the survey;
- $t_t(i)$ , the tolerance on the starting time. This value, together with  $s_t(i)$ , identifies a time interval  $[s_t(i) - t_t(i), s_t(i) + t_t(i)]$  which represents the time window within which the traveler is willing to start the trip. Consequently, we set  $s_t^{min}(i) = s_t(i) - t_t(i)$  and  $s_t^{MAX}(i) = s_t(i) + t_t(i)$ .

Furthermore, we define  $d_{AB}$  to be the distance between any two locations  $A$  and  $B$ , as determined by the underlying road network; and  $\tau_{AB}$ , the time required to travel from  $A$  to  $B$ . Travel time estimations have been obtained through a public, OpenStreetMap-based API. Using the above notation, for a given trip  $T_i$ , its length is denoted  $d_{S(i)D(i)}$ , and its duration is  $\tau_{S(i)D(i)}$ .

Similarly to what have been done by [31] and [34], in the remainder of this paper we focus on pairwise, static matching of rides: i.e., at most two trips can be combined into a single ride and, when the shared ride is formed, no further modification to the route is possible. Although simplified with respect to some ride-sharing models and existing services, this approach has proven to be computationally effective and to provide impressive opportunities for ride sharing and carbon footprint reduction.

#### 4.1.1 When to share a trip

Serving two trips with a single vehicle has the potential advantage of halving the number of vehicles required for the two travelers to reach their destinations, which can in turn lead to a substantial reduction in the travel expenses. Different from Uber-like car-sharing programs, in which drivers roam the city picking up and dropping off passengers, in the context of carpooling travelers have a different role of *driver* and *passenger*, and the driver might incur a higher travel cost due to the extra time typically required to detour from the optimal route and pick up (and, possibly, drop off) the passenger. As in [31], we call this the *sharing delay*  $\Delta$ , and incorporate this notion in the definition of the carpooling shareability network.

Specifically,  $\Delta$  is defined as the maximum amount of extra time allowed by the driver for the detour needed to pick a passenger up. Correspondingly, given a driver trip  $T_1$  and a passenger trip  $T_2$ , we say that  $T_2$  is shareable with  $T_1$  if the two following conditions hold:

$$\tau_{S(1)S(2)} + \tau_{S(2)D(2)} + \tau_{D(2)D(1)} \leq \tau_{S(1)D(1)} + \Delta \quad (1)$$

$$s_t^{min}(2) - s_t^{MAX}(1) \leq \tau_{S(1)S(2)} \leq s_t^{MAX}(2) - s_t^{min}(1) \quad (2)$$

The first condition states that the required detour time for the driver to pick the passenger up and drop her off does not exceed the threshold value  $\Delta$ . Note that the value of  $\Delta$  is only taking into account the extra time necessary to detour with respect to the standard driver's route, but not the time the passenger would need to physically get into the car (quantifiable in 30 seconds, 1 minute at most).

The second condition ensures that the two starting time windows are properly overlapped. More precisely, the two inequalities in (2) ensure that there exists a time  $t_1$  in the interval  $[s_t^{min}(1), s_t^{MAX}(1)]$  such that the driver, leaving at  $t_1$ , can reach the passenger start location at a time  $t_2 \in [s_t^{min}(2), s_t^{MAX}(2)]$ .

While the second condition establishes the temporal compatibility of the two trips, this is not enough to ensure that carpooling actually contributes to reduce traffic and carbon footprint. In fact, it is possible that the driver performs a very long detour to pickup the passenger, thus making the length of the shared trip longer than the sum of the lengths of the two individual trips. This would be a detrimental condition

for traffic and carbon emissions, that is likely to occur especially for relatively large values of the sharing delay  $\Delta$ . Therefore, while focusing our attention on the social aspects of carpooling, we do not want this to come at the expense of negative environmental externalities; so, we add a third condition that ensures that the above described possible negative externalities of shared mobility are avoided:

$$d_{S(1)D(1)} > d_{S(1)S(2)} + d_{D(2)D(1)} \quad (3)$$

To sum up, in the carpooling shareability network  $N = (T, E)$  a directed link  $e_{1,2} \in E$  between  $T_1$  and  $T_2$  exists if and only if the three conditions (1), (2) and (3) stated above are satisfied. The carpooling shareability network as defined herein is substantially different from the one defined in [31] to analyze taxi ride sharing opportunities in New York City, the main distinctions being the dissimilar criteria used to determine the existence of a link in the network, and the directed nature of carpooling links in contrast to the bi-directional links used for taxi rides. However, for the sake of conciseness, and when clear from the context, in the following we will drop the word ‘‘carpooling’’ and simply use the term ‘‘shareability network’’ when referring to  $N$ . In the Supplementary Material (Appendix A) we report the results of the analyses of the shareability networks generated by applying the above described shared conditions to the data described in Section 4.3. Due to space reason, the Appendix reports only the results for the Pisa data set, being the others quite similar.

Notice that the role of the traveler corresponding to a trip  $T_i$  is not predefined in the shareability network  $N$ . In other words, the traveler corresponding to  $T_i$  can be either a driver or a passenger. The role of the traveler in a shared trip is determined by the direction of the link incident into  $T_i$ . If the link is outgoing, the traveler corresponding to  $T_i$  is designated to be the driver; if the link is incoming, she is designated to be the passenger. The choice of not preassigning roles to travelers has been done with the goal of maximizing trip sharing opportunities, and it is consistent with the choice of selecting, among those available in the datasets, only the trips performed by people that declared the car as either the primary or the secondary mode of commuting. On the other hand, the role of each traveler becomes well-defined *after* the execution of the trip matching algorithms, and it is determined by the direction of the incident link of the matching as described above. The fact that shared trips are selected according to a (maximum) matching algorithm ensures that no traveler is assigned a double role of passenger and driver at the same time, and, hence, the correctness of our model.

## 4.2 Matching metrics and algorithms

We now define a number of metrics that can be used to optimize the trip matching process. Some of them are borrowed from [31] and based on transport-related metrics, while others are introduced here for the first time to account for the social dimension of carpooling. The social metrics here defined must be intended as examples of the type of information that is possible to assign to each traveler to enrich the shareability network so to be able to perform social-aware trip matching. Note that the algorithms reported in the following matches at most 2 rides in a carpool. However, it is possible to extend the method to matching of  $k$  rides, with  $k$  a small number such as 3,4, following the methodology of [31]. Technically speaking, all the algorithms reported in the following are instances of maximum weighted matching on graphs, which can be solved in time  $O(\sqrt{T} \cdot E)$  using Micali and Vazirani’s algorithm [22]. The difference between the algorithms lies in the way link weights are defined, and the resulting matching properties.

### 4.2.1 Transport-related metrics

The first and most commonly used metric for optimizing the trip matching process is the number of paired trips. The rationale is that, by increasing as much as possible the number of shared trips, it is possible to



reduce to the maximum extent the number of vehicles needed to serve all the trips. It is easy to see that the number of paired trips can be maximized by computing the maximum cardinality matching on the shareability network. The corresponding matching algorithm is called *Maximum Cardinality* and denoted  $M_c$ .

While optimally reducing the number of vehicles needed to serve the commuting trips, the matching resulting from Maximum Cardinality is not necessarily optimal from a carbon footprint viewpoint. In fact, it is possible that the two travelers share only a small portion of the trip, while the total number of traveled kilometers is not optimized. To account for this, we assign a weight to each link in the shareability network, where the link weight is defined as follows:

- the amount of kilometers  $\sigma_{ij}$  saved if a shared trip is performed by the travelers associated to trips  $i$  and  $j$ , with respect to performing the two separate trips.

Notice that  $\sigma_{ij}$  is equal to the sum of the length of the two trips  $i$  and  $j$  minus the length of the shared trip, which is always positive due to the fact that condition (3) is fulfilled when building the shareability network.

It is easy to see that the total traveled distance can be minimized (or, similarly, the total saved distance can be maximized) by computing the maximum weighted matching on the shareability network, where the network links are weighted with quantities  $\sigma_{ij}$  defined as above. The resulting matching algorithm is called *Maximum Saved Distance* and denoted  $M_d$ .

#### 4.2.2 Social-related metrics

We now define two social-related metrics aimed at promoting social interactions between carpooling participants.

The first metric is rooted on the observation that carpooling, by locating two travelers into a physically constrained and private environment like the car, can become an active means for starting new social relationships or strengthening the existing ones, if any: a form of social mingling. To this purpose, we define a metric aimed at maximizing the time travelers spend together on their way to the destination, called the *mingling time*. More specifically, we assign a weight to the links in the shareability network defined as:

- the mingling time  $\tau_{ij}$  that the two travelers associated to trips  $i$  and  $j$  spend together when they share their trip.

Notice that the herein defined quantity  $\tau_{ij}$  is *different* from the previously defined metric  $\sigma_{ij}$ . In fact, the latter quantity aims at minimizing the total traveled distance, which includes also the detour necessary to pick the passenger up and drop her off, while the former quantity specifically maximizes the time the two travelers spend together on-board, while still under the constraints imposed by condition 3 defined in Section 4.1. Thus, while related, these two quantities are inherently different and serve distinct optimization goals.

Next in this section, we formally prove that the total mingling time can be maximized by applying maximum weighted matching algorithms on the shareability network, where links are weighted with the  $\tau_{ij}$  values. The corresponding matching algorithm is called *Maximum Mingling Time* and denoted  $M_t$ .

With the above defined metrics we are still not taking into account any social trait of travelers, which could be used to define *social categories*. In accordance with [3], by social category we mean here a “collection of people that have certain characteristics or traits in common, but they do not necessarily interact with each other on a regular basis”. Following this definition, social categories could be defined to classify travelers with respect to the age, the education level, the gender, the race, and so on. For this reason, the second social metric defined accounts for the fact that, while maximizing the mingling time, the metric  $\tau_{ij}$  defined above is oblivious to the social category of the travelers: i.e., travelers of

same/different categories are neither favored nor hampered when forming shared trips. On the other hand, if carpooling is to be proposed as a possible tool for favoring social mixing, traveler's social categories cannot be disregarded in the shared trip formation process. Motivated by this observation, we now define a social metric aimed at favoring shared trips across different social categories and, thus, *social mixing*.

Let us define a link  $e_{i,j}$  in the shareability network  $N$  to be *assortative* if both its endpoints (trips  $i$  and  $j$ ) are performed by travelers belonging to the same social category, and *dissortative* otherwise. The *Maximum Social Mixing* matching, denoted as  $M_x$ , is the one which contains the largest number of dissortative links, and hence maximizes the social mixing between trips whose travelers belong to different social categories. In order to find this matching, we assign a weight to each link in the shareability network. The weight depends on whether the link is assortative or dissortative, and is equal to:

- $z$  for assortative links;
- $W$  for dissortative links,

where  $z$  and  $W$  can be any positive integers, with  $W > \ell z$ , being  $\ell$  the total number of assortative links. In our work, we set  $z = 1$ , but any other choice is admissible as well.

The lemmas reported in the following prove that, by computing the maximum weighted matching on the shareability network with link weights as defined above, it is possible to obtain the *Maximum Social Mixing* matching. Note that, if the goal would instead be to favor assortative links to strengthen ties within same social categories, the above weights should be assigned exactly in the opposite way. In fact, these weights do not represent real quantities or metrics. They are instead tuned in order to privilege the dissortative (or assortative) links and to ensure that the maximum weighted matching, applied on the shareability network, gives the the desired result.

It is also possible to combine some of the above defined matchings to further increase a particular kind of benefit. For instance, if we wish to use carpooling to foster social inclusion, we may wish to increase at the same time the number of dissortative links in the matching and the mingling time. We can obtain this by combining algorithms  $M_x$  and  $M_t$ . In the Supplementary Material (Appendix B) we also present this new matching, called  $M_{xt}$ .

The Maximum Social Mixing matching  $M_x$  over a shareability network, as well as its variant  $M_{xt}$ , are meant to maximize the number of included dissortative links. Therefore, they belong to the class of the *Maximum Dissortative (MD)* matchings, which we define here as the set of all the matchings over a multi-class network  $N$  which contain the highest possible number of dissortative links.

If the network  $N$  is weighted, we can define, for  $N$  and for any matching over it, the *assortative weight* as the sum of the weights of all the assortative links. An analogous definition can be given to the *dissortative weight*, while the total weight is the sum of the two quantities. Among the MD matchings, we can thus identify three matchings of interest, namely: the *MD-Maximum Assortative Weight (MD-MAW)*, which has the highest assortative weight, the *MD-Maximum Dissortative Weight (MD-MDW)*, which has the highest dissortative weight, and the *MD-Maximum Total Weight (MD-MTW)*, which has the highest total weight.

The Maximum Social Mixing matching  $M_x$  corresponds to the MD-MAW matching over the shareability network, with all the weights set to 1; its variant  $M_{xt}$  is instead the MD-MTW over the shareability network where the weight of each link is given by the mingling time. Now, we show how the MD-MAW and the MD-MTW over a network  $N$  are equivalent to the maximum weighted matching over the same network, with properly modified weights.

Consider a weighted undirected network  $N = (\mathcal{V}, \mathcal{E})$  whose nodes can be divided into  $K$  disjoint subsets  $\mathcal{V}_1, \mathcal{V}_2, \dots, \mathcal{V}_K$ . Correspondingly, the links can be partitioned into two subsets, namely:

$$\mathcal{E}_a = \{(a, b) \in \mathcal{E} : a, b \in \mathcal{V}_i\} \text{ for some } i = 1, \dots, K \quad (4)$$

$$\mathcal{E}_d = \{(a, b) \in \mathcal{E} : a \in \mathcal{V}_i, b \in \mathcal{V}_j\} \text{ for some } i \neq j \quad (5)$$

We call *dissortative links* the links belonging to  $\mathcal{E}_d$ , and *assortative links* those belonging to  $\mathcal{E}_a$ . Consider now a generic matching  $\mathcal{M} \subset \mathcal{E}$  over the whole network  $N$ , whose weight is defined as  $M = \sum_{e \in \mathcal{M}} w(e)$ , being  $w(e)$  the weight of link  $e \in \mathcal{E}$ . As any matching over  $N$ ,  $\mathcal{M}$  can be partitioned into two subsets  $\mathcal{S}_a$  and  $\mathcal{S}_d$ :

$$\mathcal{S}_a = \mathcal{M} \cap \mathcal{E}_a \quad \mathcal{S}_d = \mathcal{M} \cap \mathcal{E}_d \quad (6)$$

whose weights are  $S_a$  and  $S_d$ , respectively. Therefore,  $M = S_a + S_d$ , where  $S_a$  is the assortative weight, while  $S_d$  is the dissortative weight.

The maximum weighted matching  $\mathcal{M}_N$  is the one whose weight  $M_N$  is the highest possible value for  $M$ . Now, let us call  $\Gamma$  the set of all the MD matchings, that is, the set of all the matchings which contain the highest possible number  $C$  of dissortative links. Clearly,  $C \leq |\mathcal{E}_d|$ , where the equality holds if no node in  $\mathcal{V}$  is touched by more than one dissortative link. The exact value of  $C$  can be computed by finding the maximum matching over the non weighted sub-network of  $N$  which contains only the dissortative links (all with the same unitary weight).

The MD-MAW matching  $\mathcal{M}_e \in \Gamma$  is the MD matching with the highest assortative weight. Notice that in general:

- $\mathcal{M}_e$  is different from the maximum weighted matching  $\mathcal{M}_N$ ;
- $\mathcal{M}_e$  does not necessarily maximize the dissortative weight  $S_d$  (unless all the dissortative links have the same weight): it is only guaranteed to maximize the cardinality of  $\mathcal{S}_d$ , irrespective of the weight.

If we define as  $N'(W)$  a network obtained from  $N$  by setting the weight of all the dissortative links to a given value  $W > 0$ , we can state the following Lemma:

**Lemma 1.** *For any  $W > S_a$ , being  $S_a$  the total assortative weight of network  $N$ , the MD-MAW matching over the network  $N$  is the maximum weighted matching  $\mathcal{M}'_N$  over the network  $N'(W)$ .*

*Proof.* See the Supplementary Material (Appendix C.1). □

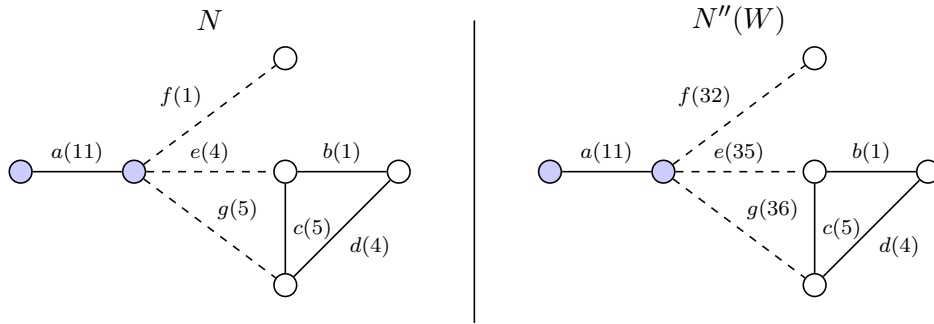
The MD-MTW matching is the one with the highest total weight. Notice that the MD-MTW matching does not necessarily maximize neither the assortative weight nor the dissortative weight, but simply their sum. If we define as  $N''(W)$  a network obtained from  $N$  by adding  $W$  to the weight of any dissortative link, we can state the following Lemma:

**Lemma 2.** *For any  $W \geq M$ , being  $M$  the total weight of the links of network  $N$ , the MD-MTW matching over the network  $N$  is the maximum weighted matching  $\mathcal{M}''_N$  over the network  $N''(W)$ .*

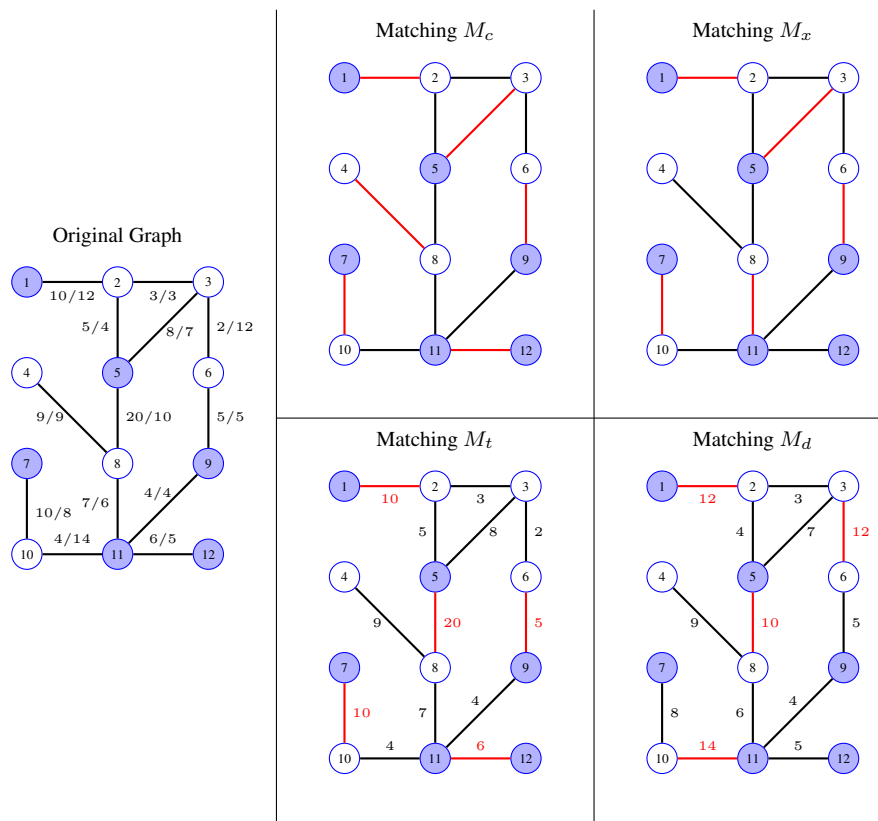
*Proof.* See the Supplementary Material (Appendix C.2). □

We report in Figure 1 an example of a network  $N$  partitioned into two groups. The total weight of the links is  $W = 31$ . Notice that the maximum weighted matching of  $N$  is  $\{a, c\}$ , whose weight is 16. Conversely, if we apply Lemma 1, we discover that the MD-MAW matching is  $\{f, c\}$ , whose weight is 6. The MD-MDW matching which maximizes the dissortative weight is  $\{g, b\}$ , with weight 6. Finally, the MD-MTW matching, obtained using Lemma 2, is  $\{e, d\}$ , whose weight is 8. Notice that MD-MTW does not maximize the assortative nor the dissortative weight, but only their sum.

The four proposed matching algorithms, which aim at maximizing different metrics, give different results in most cases. This is easily seen in Figure 2. Here, as an illustrative example, we build a small shareability network consisting of 12 trips belonging to two social categories, where each edge is labeled with two weights, representing the mingling time and the saved mileage. The four matchings are then computed as explained in this section, each leading to a different sharing configuration.



**Fig. 1** An example of weighted network  $N$ , partitioned into  $k = 2$  sub-networks, with the corresponding modified network  $N''(W)$ . The weight of each link is reported after its label, in brackets. Here  $C = 1$ , and in order for Lemma 2 to hold,  $W$  is required to be at least equal to  $M = 31$ .



**Fig. 2** An example of a small shareability network consisting of 12 trips, belonging to two different categories. Each edge is labeled with two weights, the former utilized by matching  $M_t$ , the latter by matching  $M_d$ . The four matchings lead to four different sharing configurations.

### 4.3 Datasets

The results presented in the following sections are based on two mobility surveys performed in the city of Pisa, Italy, and at MIT campus, Cambridge, MA, US. The nature and quality of survey data are described in the following subsections. From the datasets, we extracted a set of  $N_P = 1966$  morning commuting trips in Pisa, and a set of  $N_M = 1968$  morning commuters towards MIT in Cambridge. These trips were performed by commuters that declared the car as a primary or secondary transport mode.

#### 4.3.1 Data from Pisa

In Pisa, the *MobilitandoPisa* joint initiative of the National Research Council with the city of Pisa and other academic partners [10] led to an anonymous mobility survey whose objective was to understand the daily commuting habits and verify potential attitude towards shared mobility and public transportation within the metropolitan area of the city of Pisa. The ultimate goal of the initiative was to design a new, ad-hoc mobility plan for the whole city.

The survey was extensive and structured into different parts, namely:

- *Demographic Data*, where information such as gender, age, marital status, educational level, profession, and owned transportation means, were asked;
- *Workers Commuting Data*, where respondents were asked which transportation mode(s) they use to commute, their commuting hours and typical travel times, and to provide information about their home and work addresses, stopovers (if there were any on the way to the working place), working times during the day and working days throughout the week, and their potential flexibility in departing earlier/later from/to home;
- *Students Specific Data*: since Pisa has several private and public Universities, students were asked specific information related to the way they commute to/from university, the preferred transportation mode, times, etc.;
- *Moving Around Pisa Metropolitan Area*: in this section people were asked when, how and how many days per week they travel, both within and to the city of Pisa, for activities like entertainment, study, family, sport;
- *Attitude towards*: car sharing/pooling, public transportation, biking;
- *Transportation Related Problems*, and in particular on using the bike, using transit, pooling, sharing, parking;
- *Suggestions and Criticisms*, where there were specific questions on suggested improvements, such as "would you like more improved personal safety within parking lots?", or more connections, transit availability, and on potential incentives they would like to shift their current transportation habits, such as subsidies, lower costs, prizes, as well as open-answer questions.

Among the 6,200+ respondents, for the study we considered the subset of  $N_P = 1966$  commuters who daily commute by car, and provided detailed information on their home and work addresses.

#### 4.3.2 Data from MIT campus in Cambridge

With converging motivations of the MIT Climate Action Plan, local sustainability goals and a pragmatic consideration of land use pressures in Cambridge, the Massachusetts Institute of Technology recognized an opportunity to jointly address these pressures through transportation reforms for its own community. To this end, every two years the MIT surveys its staff and student members to monitor commuting behavior and attitude towards a potential mode shift.

The 2016 survey was structured into different parts, namely:

**Table 2** Age and Education level classes.

Age Class ID	Age	Education Class ID	Education Level
1	18-25	1	Lower Secondary School
2	26-35	2	High School
3	36-45	3	Bachelor's Degree
4	46-55	4	Master's Degree
5	56-65	5	Ph.D.
6	Over 65	6	Professional Specialization

- *About You*, where information about flexibility and time were asked, such as time spent on campus, daily arriving and departure time;
- *Your Commute*, where respondents were asked which transportation mode(s) they use to commute, if and (possibly) why there was a mode shift compared to the previous academic year, and if they are considering/willing to shift and the reason why;
- *Use of Transportation Services*, where employees and students also answered to what extent current MIT transportation benefits could influence their commuting choices;
- *Knowledge of & Feedback on Transportation Services and MIT Current Benefits*, such as Subsidized MBTA Pass, Electric Vehicle Charging Stations, Emergency Ride Home Program, Carpools/Vanpool Parking Programs, Hubway (bike sharing), etc. ;
- *Driving, Car Ownership, and Bicycling*, where information such as car and bike ownership/use were asked;
- *Feedback and comments*.

The survey did not require any demographic section since it was not anonymous and each respondent is directly connected to her MIT account. Note that before being processed, the data in our analyses has been anonymized.

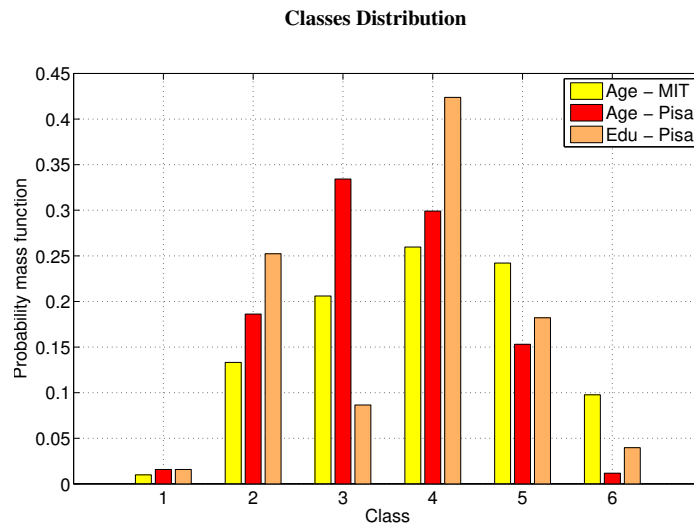
Among the 10,000+ respondents, for the study we considered the subset of  $N_M = 1968$  commuters belonging to the MIT staff who daily commute by car. Notice that in this scenario the work address is the same for each respondent. Home address information is also known but only at the census block level and, even if not as accurate as in the case of survey data in Pisa, it represents a good compromise between spatial accuracy and the need of preserving privacy.

#### 4.3.3 Node classification

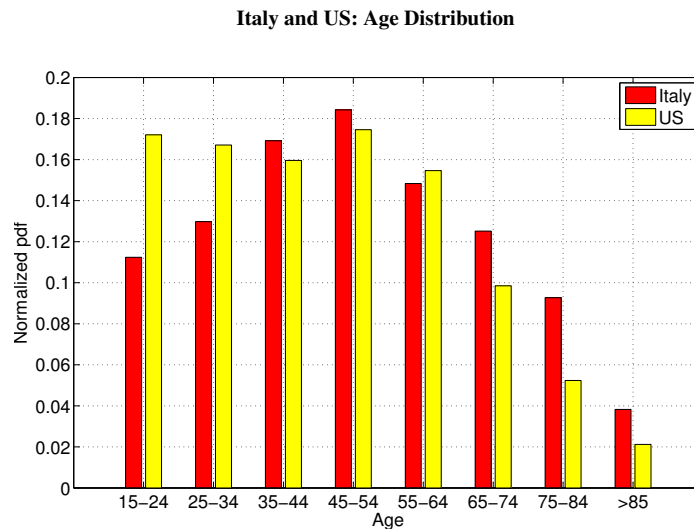
The datasets chosen for this study have the unique feature of containing not only trip information, but also additional information about travelers that can be used to classify them into different *social categories*.

In the following, we consider two specific traveler features to form distinct social categories: *age*, and *educational level*. Both features are routinely used in social sciences to define social categories [3]. For the former feature, we use 6 different age intervals to divide travelers into 6 corresponding social categories. For the latter feature, which is only available for the Pisa survey, we used the education degree as stated in the survey, which also comes into 6 increasing levels of education. The resulting social categories used in this study are reported in Table 2. We stress that the study is focused on a single social feature at a time, and does not consider their combination into more refined social categories (e.g., people with age in the 36-45 interval and a Bachelor degree) which would have required more extensive datasets. We also stress that these categories are used as examples, and should not be read as a constraint to the proposed method, nor any sort of particular policy recommendation.

We reiterate that the goal of this paper is to show the potential of a social category-based approach to carpooling, and not to test all the possible categories and/or metrics, nor to identify the best: it would



**Fig. 3** Travelers distribution according to age (Pisa and MIT) and education level (Pisa).



**Fig. 4** Age distribution in Italy and US derived from census data.

be up to policy makers to select, based on specific situations and the results provided by our framework, which one is better for the purpose at hand.

Figure 3 reports the distribution of travelers across the social categories based on age in Pisa (purple) and MIT (yellow), and education level in Pisa (orange). MIT education level info is not reported because the data was too sparse to be used here, since providing education level info is not mandatory for MIT staff. It is interesting to note that in Pisa while the age distribution hints to a Gaussian curve, centered around age interval 36 – 45, the education level distribution is bi-modal with two peaks for High School and Master Degree level. We stress that the above are not necessarily representative, e.g., of the whole population in Pisa, but are likely influenced by the way the survey was administrated. E.g., a large number of respondents in Pisa were University undergraduate students, which likely explains the peak in the High School education level. While existent, the bias in social categories distribution observed in the dataset

**Table 3** Average Node Degree and Assortativity Coefficient

Delay $\Delta$	Pisa			MIT	
	Average Degree	$\rho$ (Age)	$\rho$ (Edu)	Average Degree	$\rho$
0 min	0.396	-0.0112	0.0406	1.284	-0.0175
1 min	3.941	0.0309	0.0629	12.387	-0.001
2 min	11.788	0.0374	0.0482	33.941	-0.0001
3 min	23.618	0.0292	0.0371	55.296	0.0004
4 min	38.468	0.0272	0.0330	74.61	0.0003
5 min	53.713	0.0242	0.0327	92.601	0.0016
10 min	109.076	0.0196	0.0266	146.329	0.0045
15 min	134.911	0.0181	0.0249	171.055	0.0059

does not influence the validity of the proposed methodology, which is oblivious to social category definition and distribution. On the other hand, the bias might have an impact on the results of the considered case study, which is discussed later on in the paper.

Nevertheless, we also include Figure 4 reporting the age distribution of population in Italy and US, derived from census data. Even if different, by comparing the two distributions we can highlight similar shapes, but with relatively lower representation of both young and older ages. This is consistent with the fact that the surveys were focused on working population and university students, who have a narrower age range than general population. An interesting topic for future work is repeating the analysis with an unbiased population sample selected for an ad hoc study.

Based on the definition of social category according to one of the social features described above, each node in the shareability network is labeled with the social category corresponding to the traveler associated with the trip. Thus, in the following we consider the shareability network to be a node-labeled network.

## 5 Results

In the following two subsections we report the results of the performed analyses, both in terms of social mixing and algorithms performance.

### 5.1 Social analysis of shareability networks

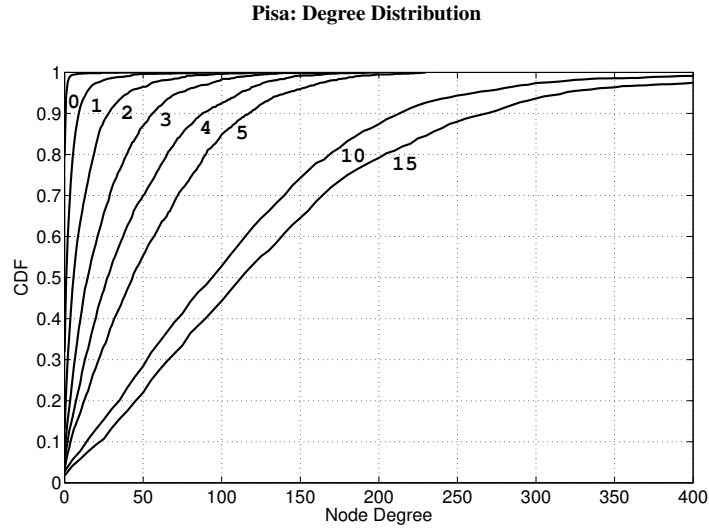
We start analyzing the structure of the resulting shareability networks, and their interplay with the above described social categories. We generate a number of different shareability networks by varying the shareability delay parameter  $\Delta$  within the interval  $[0, 1, 2, 3, 4, 5, \dots, 15]$ , measured in minutes. Correspondingly, we get 16 different shareability networks, one for each value of  $\Delta$ .

Figure 5 reports the cumulative distribution function of the node degree of the shareability networks obtained for Pisa data. For clarity, we report only some representative curves, using for  $\Delta$  the values  $\{0, 1, 2, 3, 4, 5, 10, 15\}$ . Each value reported in the plot corresponds to the sum of the out- and of the in-degree of a node. The amount of carpooling opportunities for even moderate values of  $\Delta$  is noteworthy: with  $\Delta = 10$  min, we already observe more than 50 carpooling opportunities for over 70% of the trips.

Table 3 reports the average node degree in the shareability network for different values of  $\Delta$ , i.e. the *average sharing opportunities* per trip, while varying the delay needed to detour for picking a carpooler up and dropping her off. Note that with only 5 minutes of delay  $\Delta$  each trip in Pisa has on average the opportunity to be shared with about 54 other trips, 92 if we consider MIT.

Next, we take into account the node labels based on social categories, and compute the *assortativity coefficient* of the shareability networks. Assortativity is a well-known metric in social network analysis





**Fig. 5** Pisa data: degree distribution for some representative shareability networks. Each value reported corresponds to the sum of the out- and of the in-degree of a node.

[25] which measures to what extent links between nodes belonging to the same category are more often established with respect to a random graph. More specifically, the assortativity coefficient  $\rho$  is a real number varying from -1 to 1: if  $\rho = -1$ , only links between nodes belonging to different categories exist, and we say that the corresponding network is *fully dissortative*; on the other hand, if  $\rho = 1$  only links between nodes of the same category exist, and the corresponding network is said to be *fully assortative*. Values of  $\rho$  close to 0 indicate that there is neither an assortative nor a dissortative trend in how links connect nodes, resembling entirely random connection.

For the 16 values of  $\Delta$ , we computed  $\rho$  as:

$$\rho = \frac{\sum_{i,j} (A_{ij} - k_i k_j / 2m) \delta(c_i, c_j)}{2m - \sum_{i,j} (k_i k_j / 2m) \delta(c_i, c_j)} \quad (7)$$

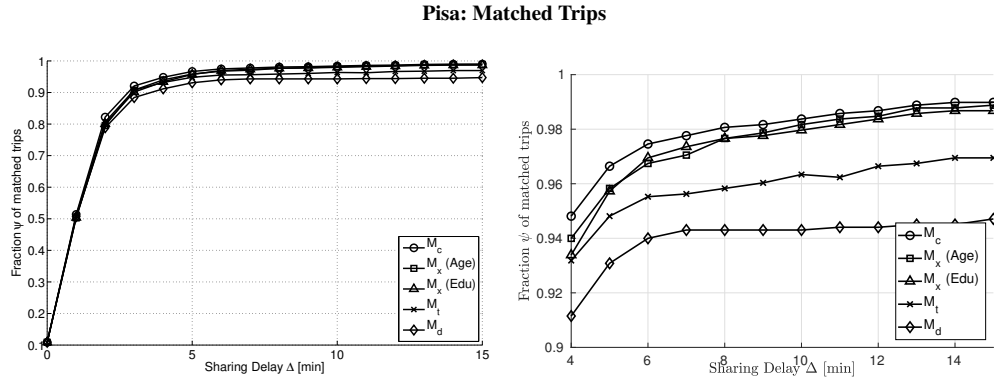
where  $A_{ij}$  is the element at row  $i$  and column  $j$  of the adjacency matrix  $\mathbf{A}$  representing the shareability network,  $k_i$  is the degree of node  $i$ ,  $m$  is the total number of edges,  $c_i$  is the class of node  $i$ , and  $\delta(a, b)$  is the Kronecker delta function, and is equal to 1 if  $a = b$  and 0 otherwise.

We have computed the assortativity coefficient for the shareability networks obtained with different values of  $\Delta$ , obtaining values very close to 0, as illustrated in Table 3.

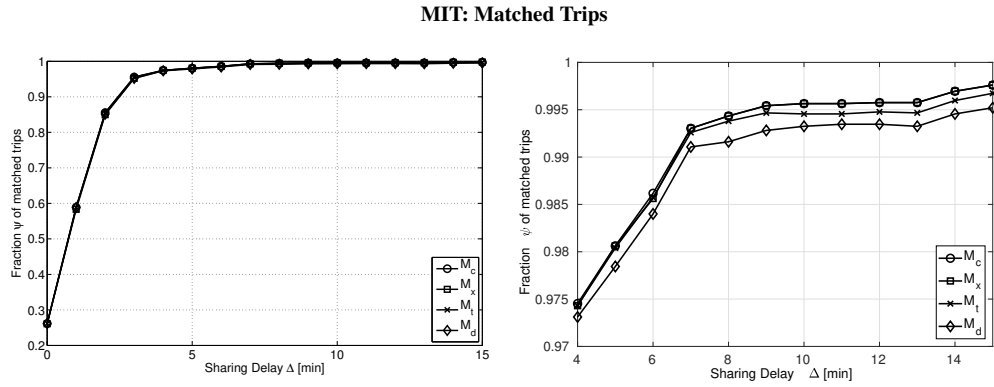
This indicates that the structure of the shareability network is neither assortative nor dissortative, and trip opportunities can equally be found between and across different social categories. This is a first interesting finding, indicating that carpooling “per se” neither favor nor impede social mixing between categories. So, unless social factors are explicitly taken into account when matching rides, we can expect a “socially neutral” effect of centrally-optimized carpooling when the two chosen dimensions are used (age and educational level). This intuition is confirmed by the results presented in the following.

## 5.2 Performance Analysis

In this Section we evaluate the performance of the four matching algorithms defined in Section 4.2, namely Maximum Cardinality ( $M_c$ ), Maximum Saved Distance ( $M_d$ ), Maximum Mingling Time ( $M_t$ ),



**Fig. 6** Pisa data: the fraction  $\psi$  of matched trips for increasing values of the sharing delay  $\Delta$ . The plot on the right reports a close-up view for values of the sharing delay above 4 min.



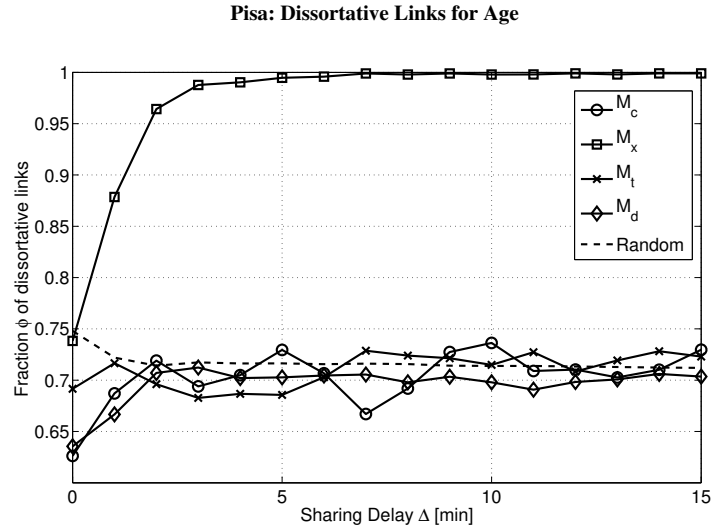
**Fig. 7** MIT data: the fraction  $\psi$  of matched trips for increasing values of the sharing delay  $\Delta$ . The plot on the right reports a close-up view for values of the sharing delay above 4 min.

and Maximum Social Mixing ( $M_x$ ). We use the following evaluation metrics, accounting for both transport and social aspects of carpooling:

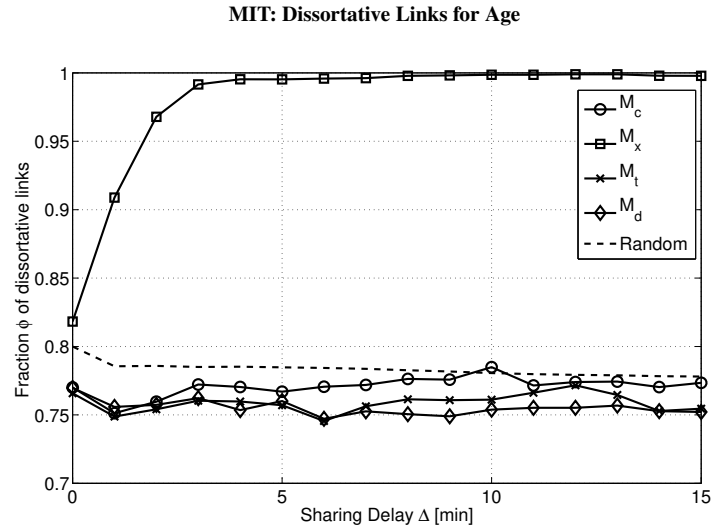
1. the fraction  $\psi$  of matched trips;
2. the fraction  $\phi$  of dissortative links included in the matching, considering social categories defined on both age and education level;
3. the average mingling time  $\tau$ ;
4. the overall saved mileage  $\sigma$  achieved with trip sharing.

The four metrics are computed as a function of the sharing delay  $\Delta$ .

Figures 6 and 7 reports the fraction  $\psi$  of matched trips for Pisa and MIT, respectively. As expected,  $\psi$  is increasing with  $\Delta$ , since a higher delay tolerance translates into a higher number of links in the shareability network, and consequently to a lower number of unmatched trips. It is worth noticing that, for values of  $\Delta = 5$  min or larger, the fraction of matched nodes is consistently above 0.9, with values for MIT slightly larger than those observed in Pisa. While algorithm  $M_c$ , which specifically maximizes the fraction  $\psi$  of matched trips, achieves a slightly higher value of  $\psi$  as expected (above 0.99 for  $\Delta \geq 10$  min in case of MIT), the other matching algorithms perform very similarly in terms of the fraction of matched trips. This is an important result, as it indicates that the space of solutions with maximal or nearly maximal number of matched trips is very large. In turn, this opens the way to the approach proposed herein of



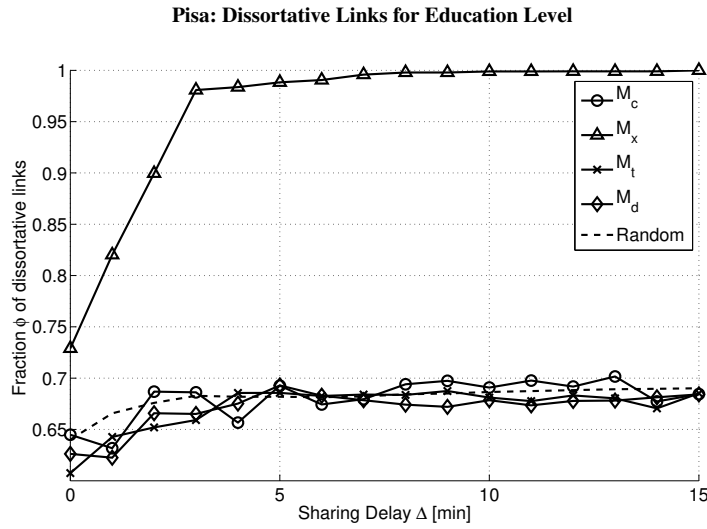
**Fig. 8** Pisa data: fraction  $\phi$  of dissortative links for increasing values of the sharing delay  $\Delta$ . Age social categories are considered.



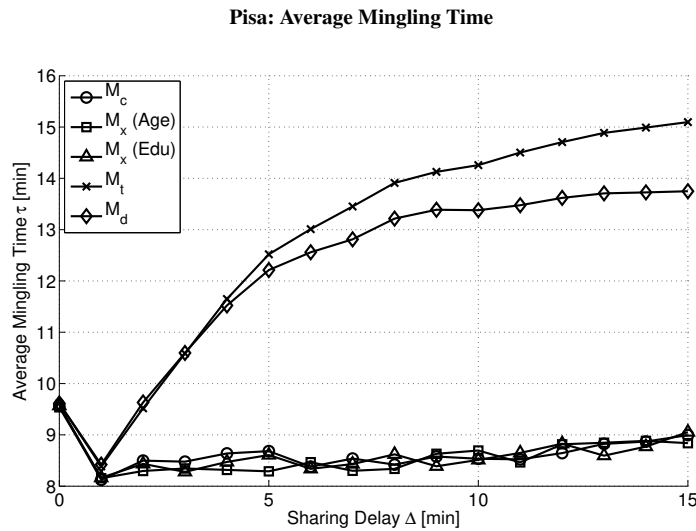
**Fig. 9** MIT data: fraction  $\phi$  of dissortative links for increasing values of the sharing delay  $\Delta$ . Age social categories are considered.

considering other metrics in the matching process that, while slightly sub-optimal from the viewpoint of the fraction  $\psi$  of matched trips, might incorporate environmental or social aspects.

Figures 8 and 9 report the fraction  $\phi$  of dissortative links, referred to age classes in Pisa and MIT, respectively, while Figure 10 reports those referred to education level in Pisa, as per Table 2. As a reference, the plots also report the probability of choosing a dissortative link when a link is randomly selected from the corresponding shareability network. It is interesting to observe that, while matching algorithms  $M_c$ ,  $M_t$ , and  $M_d$  do not significantly differ from the inherent dissortativity of the shareability network, represented by the “random” (dotted) curve, algorithm  $M_x$  achieves a nearly optimal dissortativity value of 1 for values of  $\Delta$  of at least 5 min. This means that potentially all the trips can be shared by people belonging to different age or education level categories. Comparing these results with those in Figures 6 and 7, we can also argue that a social mixing strategy can be quite effective, achieving near optimal social



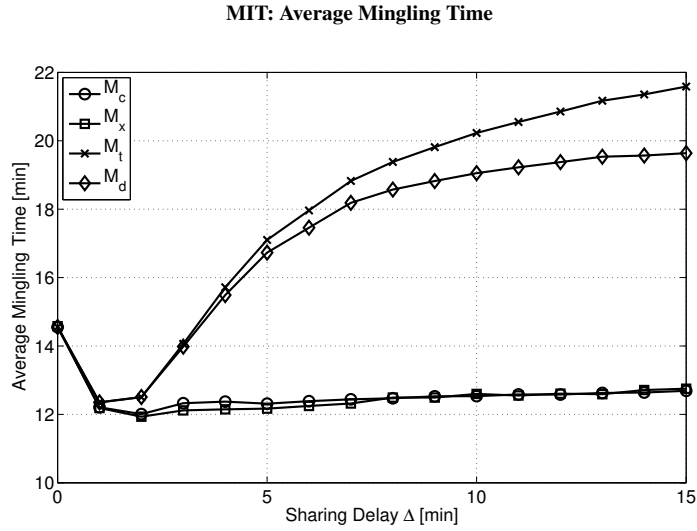
**Fig. 10** Pisa data: fraction  $\phi$  of dissortative links for increasing values of the sharing delay  $\Delta$ . Education level social categories are considered.



**Fig. 11** Pisa data: average mingling time for increasing values of the sharing delay  $\Delta$ .

mixing while, at the same time, matching almost all possible trips. Interestingly, the observations made above consistently apply in two different locations (Pisa and MIT), as well as when different definitions of social categories are used (age and educational level in Pisa). This hints to generality of the observed trends in both urban scenarios and definition of social category.

Figures 11 and 12 report average mingling time  $\tau$  for Pisa and MIT, respectively. The value reported in the plots is computed by averaging the duration of mingling time for each shared trip selected by the matching. Two trends can be clearly identified. Matchings  $M_c$  and  $M_x$  lead to an average value for  $\tau$  between 8 and 9 minutes for Pisa and between 12 and 13 minutes for MIT, with no relevant variations as  $\Delta$  varies. On the other hand,  $M_d$  and  $M_t$  achieve a much longer mingling time  $\tau$ , which is increasing with the sharing delay. The maximum value is attained for  $\Delta = 15$  min, where we get a maximum mingling



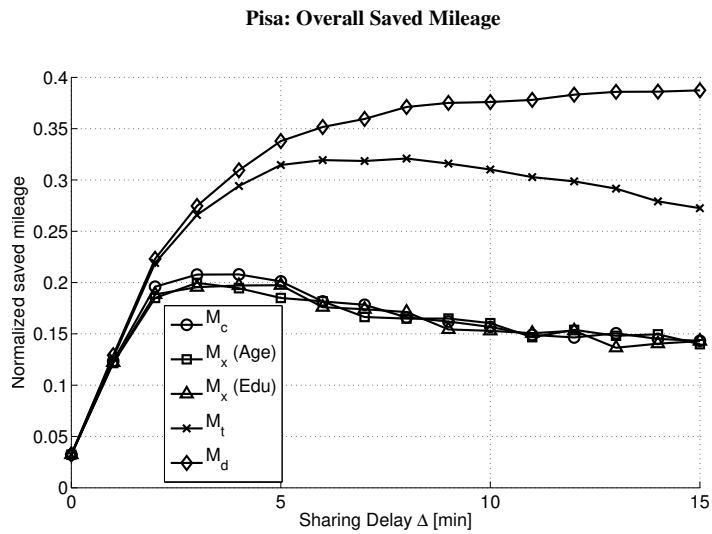
**Fig. 12** MIT data: average mingling time for increasing values of the sharing delay  $\Delta$ .

time of about 15 min in Pisa, and above 20 min for MIT. Thus, without exceeding the time constraints posed by the travelers and the mileage constraints imposed by the model, it is possible to match trips in a way that travelers can spend an average of at least 15 minutes together, which is a very significant time for mingling. The relatively higher values of the mingling time observed in the MIT vs. Pisa case study are consistent with the larger geographical footprint of the Boston metropolitan area as compared to the region surrounding Pisa. The similar behavior of algorithms  $M_d$  and  $M_t$  confirms that the two metrics used to weight links in the respective algorithms, namely,  $\sigma_{ij}$  and  $\tau_{ij}$ , are indeed related but distinct.

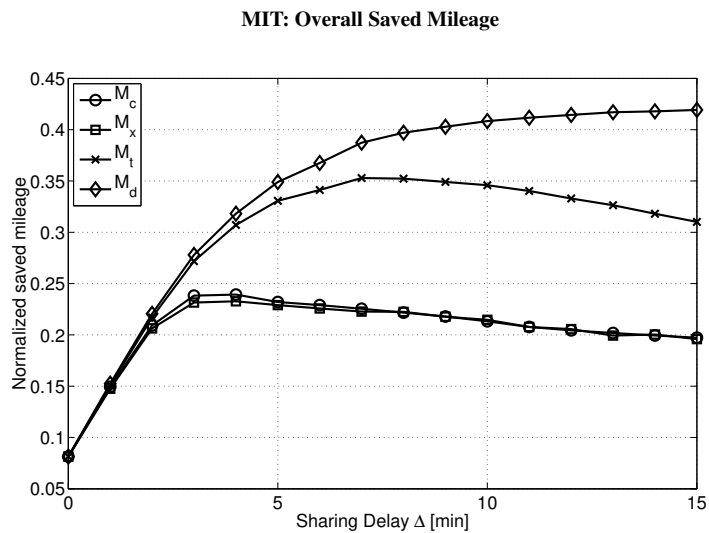
By comparing the results reported in Figures 8, 10, and 11 for Pisa, and Figures 9 and 12 for MIT, we can see that a trade-off is likely to emerge when we consider carpooling as a tool for social integration: we can either aim at a better mixing, with shorter mingling time, or at a longer time for interactions at the expense of fewer mixing opportunities.

Finally, Figures 13 and 14 show the amount of saved kilometers obtained through car pooling, normalized over the sum of the lengths of all trips. All the matchings for  $\Delta = 0$  min offer only minimal normalized savings, amounting to about 4% of total traveled distance in Pisa and about 8% for MIT. These savings, however, grow very substantially when the sharing delay is increased, up to about 40% reduction in Pisa and even higher reductions for MIT. Again, the relatively higher traveled distance savings observed in the MIT dataset as compared to Pisa are likely implied by the larger geographical footprint of the Boston metropolitan area, and seems to suggest that carpooling mechanism could be especially effective in large cities to reduce the carbon footprint of home-work commuting. Finally, we observe that as expected the highest performance is offered by  $M_d$ , and that ‘social mixing’ algorithms, while still providing substantial savings in the order of 15-20%, are well below the optimal travel distance savings provided by suitably optimized algorithms. Note that from a certain delay  $\delta$  on  $M_t$  starts to worsen in terms of saved kilometers. This is due to the fact that it is designed to maximize the time the poolers spend together on the car, thus not always choosing the shortest path for the shared trip, even though it is subject to the general mileage constraints of Equation 3.

Thus, a trade-off between environmental and social benefits of carpooling is likely to arise and shall be addressed at policy and regulatory level.



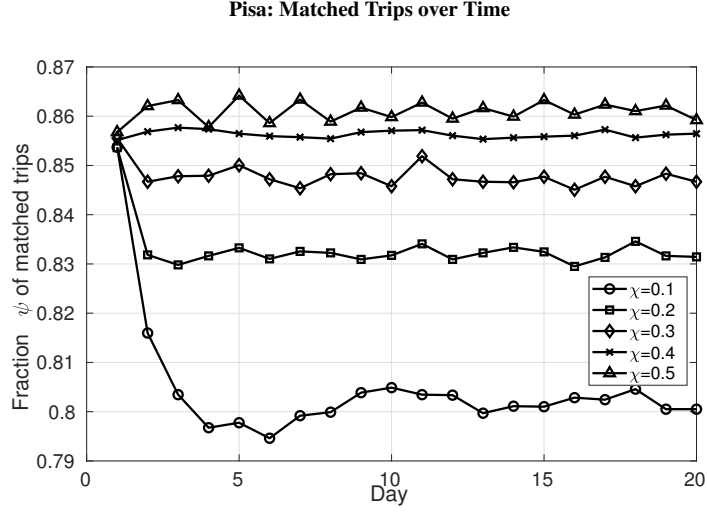
**Fig. 13** Pisa data: distance  $\sigma$  saved thanks to carpooling for increasing values of the sharing delay  $\Delta$ . To ease comparison, the metric is normalized over the total amount of traveled kilometers.



**Fig. 14** MIT data: distance  $\sigma$  saved thanks to carpooling for increasing values of the sharing delay  $\Delta$ . To easy comparison, the metric is normalized over the total amount of traveled kilometers.

### 5.3 Introducing temporal diversity

In the above sections, the trip sharing opportunities and the resulting matchings are considered as a static problem. Indeed, the commuting trips are likely to be performed daily and should be able to keep their performance characteristics across time. Nevertheless, offering the same sharing solution over several subsequent days, despite optimal from the traffic point of view, may appear limiting in terms of social mixing. In this section, we hence add a temporal dimension to the considered problem. We assume that the set of trips to be pooled is kept unchanged during a 20-day period, but we slightly modify our approach in order to favor sharing with different co-travelers across different days.



**Fig. 15** Fraction  $\psi$  of matched trips as a function of time, when time coefficients are used to favor different choices across different days. Here, the sharing delay is  $\Delta = 3 \text{ min}$ .

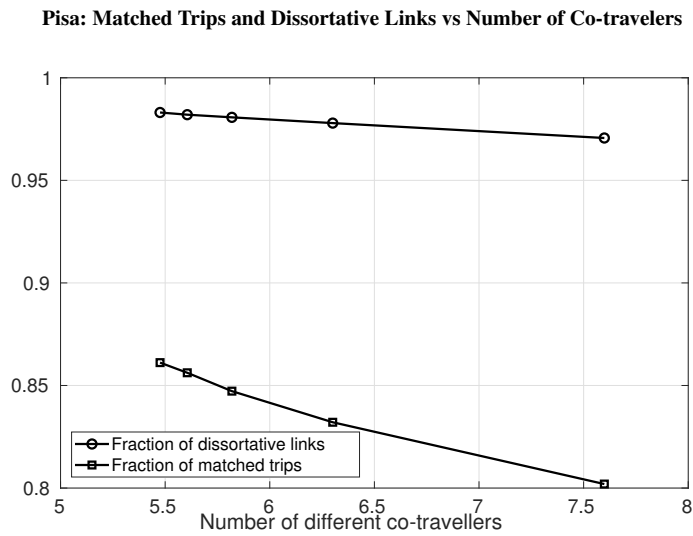
The key point is to avoid that a commuter is matched with the same co-traveler in subsequent days. To this aim, for each link  $e_{i,j}$  of the shareability network and for a given day  $d$ , we consider a *temporal coefficient*  $\omega_{i,j}(d)$ . The value of  $\omega_{i,j}(1)$  for the first day is randomly selected in the interval  $[0, 1]$ . For the subsequent days, it is instead updated as

$$\omega_{i,j}(d) = \begin{cases} 0 & \text{if link } e_{i,j} \text{ was selected on day } d-1 \\ \omega_{i,j}(d-1) + X & \text{otherwise,} \end{cases} \quad (8)$$

where  $X$  is a uniform random variable in the interval  $[0, 2\chi]$ , with expected value  $\chi$ . Each day, the weight assigned to each link by the selected matching algorithm ( $M_c$ ,  $M_x$ ,  $M_t$  or  $M_d$ ) is multiplied by the temporal coefficient of the same link. The matching is then computed on the resulting weighted graph, which now keeps into account also the past choices. By adopting this scheme, it is unlikely that a link  $e_{i,j}$  is chosen in subsequent days: after being selected in day  $d$ , its total weight drops to 0, thus making other links preferable when a maximum weighted matching is applied. In the subsequent days,  $\omega_{i,j}(d)$  gradually increases, until  $e_{i,j}$  becomes again a valid option.

The average coefficient increment  $\chi$  plays a key role in determining a trade-off between the time diversity of the scheme and the maximization of the other transport or social metrics. When  $\chi$  is high, the selected links become again advantageous after a short time. This means that most of the time coefficients are always close to 1, and it is possible to choose a nearly optimal solution in terms of cardinality, dissortativity, mingling time or saved distance. However, this also implies that only a small subset of links is repeatedly chosen, thus reducing the number of co-travelers met by a commuter across time. Conversely, a low value of  $\chi$  strongly increases the recovery time of a selected link. As a consequence, different sub-optimal solutions are to be selected in the following days. This limits the performance in terms of the above mentioned transport or social metrics, but it allows a larger number of different links to be chosen during subsequent days, thus making a commuter change her co-traveler more often from day to day.

In Figure 15, we show the fraction  $\psi$  of matched trips over a period of 20 days, for increasing values of  $\chi$ , using the Pisa dataset with  $\Delta = 3 \text{ min}$ . Matching  $M_c$  is employed for each day, and the results are averaged over 10 runs. The figure highlights the predicted reduction of  $\psi$  observed when the optimal links can no longer be always selected due to the dynamics of the time coefficients. A transient phase of around 5 days can be noticed, after which all the curves oscillate around steady state values. As expected,



**Fig. 16** Average fraction of matched trips and of dissortative links in the steady state phase as a function of the average number of different co-travelers per commuter.

low values of  $\chi$  make the best links unavailable for longer times, thus lowering the fraction of matched trips. However, even for  $\chi = 0.1$ , approximately 80% of the trips can be matched every day.

We now move to Figure 16 in order to illustrate the effectiveness of our strategy in improving the number of co-travelers per commuter over time. For each of the five curves of Figure 15, we consider the average value after the transient phase (from day 7 to day 20), and we plot it against the measured average number of different co-travelers selected for a commuter during the same two-week period. As the fraction  $\psi$  decreases (corresponding to lower values of  $\chi$ ) the number of co-travelers becomes almost 50% higher, growing from 5.5 to more than 7.5, implying that a commuter does not travel with the same person more than once a week.

We repeated the same experiment by applying matching  $M_x$  (using Age classes), and analyzing the trade-off between the fraction  $\phi$  of dissortative links among the selected ones and the average number of different co-travelers. The resulting curve, still shown in Figure 16, reveals that by lowering  $\chi$  (moving from left to right in the figure) it is possible to keep almost the same fraction of dissortative links while substantially increasing the number of co-travelers met by a commuter over subsequent days. Analogous results (not reported) are obtained for the normalized saved mileage. This suggests that it is possible to further boost the social impact of home-work carpooling by also increasing the number of co-traveling pairs over time, with a limited cost in terms of the other transport-related or social metrics.

#### 5.4 Exploring Commuters' preferences

As described above, our proposed sharing approach is centralized. This allows to leverage full information and hence to achieve optimal global performance in terms of traffic-related or social metrics. From the user perspective, however, this might seem undesirable: the user in fact is not aware of the various sharing options, and she can only accept the proposed co-traveler, regardless of her preferences. While this might be fine in certain scenarios, where users are only interested in reaching their destination within a predefined time interval, the possibility of being involved in the co-traveler selection is often a relevant point which can potentially improve the commuter's attitude towards carpooling. In the Supplementary



Material (Appendix D) we introduce an extension of our scheme to a semi-distributed approach, which accounts for both user preferences and the need for a valid solution at a global level.

## 6 Discussion and Conclusions

The aim of this paper is to investigate whether and to what extent carpooling can be used as a tool for not only reducing the carbon footprint of home-work commuting, but also to favor social integration. By coupling network-based shared mobility models with a social profile of the travelers performing the trips, we were able to define social-based metrics linked to sharing opportunities, and ultimately to prioritize the ride matching choices accordingly. One of the main lessons we have learnt from the proposed approach is that it brings a different perspective into current ride-sharing and carpooling methods, which consider drivers and passengers as abstract entities within optimization algorithms. By incorporating social characteristics of drivers and passengers, and demonstrating the potential of matching users with different backgrounds, we show that carpooling has the potential to promote the social aspects of urban mobility, and could become a social mixing tool to be embraced by transportation and planning agencies. Within the sharing economy, more than regulators, policy makers would play the role of facilitators, helping to design services and incentives to promote social mixing and carpooling [30].

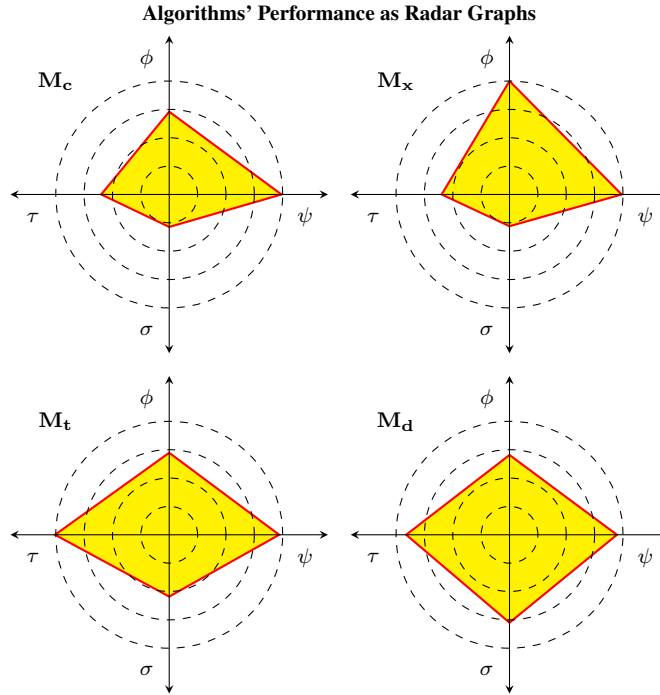
We applied the defined algorithms to two different commuting scenarios, the city of Pisa, in Italy, and the MIT campus, in Cambridge, MA, US, obtaining similar results with respect to all considered metrics. Although clearly non-conclusive, the observed similarities between the case studies in Pisa and Cambridge hint to a more general validity of carpooling as a social mixing tool.

When collectively considered, the results presented in this paper clearly indicate existence of non-trivial trade-offs between environmental and social benefits of carpooling. While not advocating any specific matching strategy amongst the ones proposed in this paper, and the many others possible, the methodology and tools presented in this paper allow for the first time an accurate quantification of these trade-offs.

To give a collective view of the performance of each matching algorithm in terms of the four considered performance metrics, Figure 17 reports the radar graphs corresponding to each algorithm  $M_c$ ,  $M_x$ ,  $M_t$  and  $M_d$ , with  $\Delta = 15 \text{ min}$ , while varying the metric. Generally speaking, a larger yellow area indicate higher performance.

While finding a definite optimal matching strategy is beyond the scope of this paper for the reasons described above, we want to sketch here a possible way of identifying the most suitable matching algorithm once priorities between the four different matching metrics considered herein have been defined. Assume for instance that policy makers decide that carpooling services operating in the city should prioritize social mixing, and then achieve the best travel distance savings possible. We could then assign weights to the matching performance metrics so that they sum up to 1, and reflect the priorities defined by policy makers, e.g.: weight .5 for  $\phi$ , weight .25 for  $\sigma$ , weight .15 for  $\tau$ , and weight .1 for  $\psi$ . Then, we can compute for each value of  $\Delta$  the aggregate performance of each matching algorithm using the defined weights, and select the best performing algorithm accordingly. Another approach to simultaneously account for different performance metrics in the optimization process could be to introduce a multi-objective optimization function directly in the ride matching algorithms, and to find a Pareto optimal matching of the rides. While more rigorous from a theoretical viewpoint, this approach is however computationally cumbersome. Assessing whether a similar approach can be applied at a large scale is an interesting direction for future work.

The one reported above is only an exemplification of how the wealth of results presented in this paper can help to better inform the growing debate about shared mobility and its implied social and environmental benefits. Moreover, it adds social aspects to be used in car-sharing and car-pooling optimization algorithms.



**Fig. 17** Radar graphs comparing the performance of the 4 selected matchings, for  $\Delta = 15 \text{ min}$ . The four metrics are the matched node fraction  $\psi$ , the fraction of age-based dissortative links  $\phi$ , the mingling time  $\tau$  (normalized to 15 min) and the overall normalized saved mileage  $\sigma$  (normalized to the maximum, 0.5).

## 7 Acknowledgements

P.S and C.R. would like to gratefully thank the members of the Senseable City Lab Consortium: UBER, Allianz, the Amsterdam Institute for Advanced Metropolitan Solutions, Ericsson, the Fraunhofer Institute, Liberty Mutual Institute, Philips, the Kuwait-MIT Center for Natural Resources and the Environment, Singapore-MIT Alliance for Research and Technology (SMART), Volkswagen Electronics Research Laboratory, and all the members of the MIT Senseable City Lab Consortium for supporting this research.

F.L., M.E.R., P.S., G.R., and F.M. would like to thank the City of Pisa and all members of the *Mobilizando Pisa* initiative.

M.E.R. and J.Z. would like to thank Adam Rosenfield for the help provided in getting the MIT data.

All the authors wish to thank the whole MIT JTL Mobility Lab members and Umberto Fugiglando for useful comments and suggestions.

## 8 Authors Contribution

The authors confirm contribution to the paper as follows: study conception and design: Authors F.L, M.E.R., P.S., G.R., J.Z.; data collection: Authors F.L., M.E.R., F. M., J.Z.; analysis and interpretation of results: Authors F.L., M.E.R., P.S., F. M., G.R., F.D., J.Z.; draft manuscript preparation: Authors F.L., M.E.R., P.S., F.M., G.R., F.D., J.Z.; research supervision: Authors C.R., J.Z. All authors reviewed the results and approved the final version of the manuscript.

## References

1. Donald N. Anderson. not just a taxi? for-profit ridesharing, driver strategies, and vmt. *Transportation*, 41(5):1099-1117, 2014.
2. Theo Arentze and Harry Timmermans. Social networks, social interactions, and activity-travel behavior: A framework for microsimulation. *Environment and Planning B: Planning and Design*, 35(6):1012–1027, 2008.
3. Michael Banton. A theory of social categories. *Sociology*, 45(2):187–201, 2011.
4. Gerardo Berbeglia, Jean-Franois Cordeau, and Gilbert Laporte. Dynamic pickup and delivery problems. *European Journal of Operational Research*, 202(1):8–15, 2010.
5. Evelyn Blumenberg and Michael Smart. Getting by with a little help from my friends and family: immigrants and carpooling. *Transportation*, 37(3):429-446, 2010.
6. Gideon Bolt, Deborah Phillips, and Ronald Van Kempen. Housing policy, (de)segregation and social mixing: An international perspective. *Housing Studies*, 25(2):129–135, 2010.
7. Willem R. Boterman and Sako Musterd. Cocooning urban life: Exposure to diversity in neighbourhoods, workplaces and transport. *Cities*, 59:139–147, 2016.
8. Marco Te Brömmelstroet, Anna Nikolaeva, Meredith Glaser, Morten Skou Nicolaisen, and Carmen Chan. Travelling together alone and alone together: mobility and potential exposure to diversity. *Applied Mobilities*, 2(1):1–15, 2017.
9. Juan Antonio Carrasco, Bernie Hogan, Barry Wellman, and Eric J Miller. Collecting social network data to study social activity-travel behavior: an egocentric approach. *Environment and Planning B: Planning and Design*, 35(6):961–980, November 2008.
10. City of Pisa. Mobilizando Pisa Initiative. 2016. Last accessed: August 24th 2017.
11. Gonalo Correia and Jos Manuel Viegas. Carpooling and carpool clubs: Clarifying concepts and assessing value enhancement possibilities through a stated preference web survey in Lisbon, Portugal. *Transportation Research Part A: Policy and Practice*, 45(2):81–90, 2011.
12. Shadi Djavadian and Joseph Y.J. Chow. An agent-based day-to-day adjustment process for modeling mobility as a service with a two-sided flexible transport market. *Transportation Research Part B: Methodological*, 104(Online First):36–57, 2017.
13. Fábio Duarte, Tatiana Gadda, Carlos Augusto Moreno Luna, and Fábio Teodoro Souza. What to expect from the future leaders of Bogotá and Curitiba in terms of public transport: Opinions and practices among university students. *Transportation Research Part F: Traffic Psychology and Behaviour*, 38:7–21, 2016.
14. Kenneth J Dueker. *Carpooling: Attitudes and Participation*. Technical Report, Institute of Urban and Regional Research, 1976.
15. Koen Frenken and Juliet Schor. Putting the sharing economy into perspective. *Environmental Innovation and Societal Transitions*, 23(2):3–10, 2017.
16. Masabumi Furuhata, Maged Dessouky, Fernando Ordóñez, Marc-Etienne Brunet, Xiaoping Wang, and Sven Koenig. Ridesharing: The state-of-the-art and future directions. *Transportation Research Part B: Methodological*, 57(C):28–46, 2013.
17. Yanbo Ge, Christopher Knittel, Don MacKenzie, and Stephen Zoepf. Racial and gender discrimination in transportation network companies. *NBER Working paper series*, Working paper 22776, 2016.
18. T.J. Hanlon, B. and Vicino, editor. *The Routledge Companion of the Suburbs*. Routledge, 2018.
19. Clifton Hood. Changing perceptions of public space on the new york rapid transit system. *Journal of Urban History*, 22(3):308–331, 1996.
20. Fanchao Liao, Eric Molin, Harry Timmermans, and Bert van Wee. Carsharing: the impact of system characteristics on its potential to replace private car trips and reduce car ownership. *Transportation*, First Online:1–36, 2018.
21. Louis A. Merlin. Comparing automated shared taxis and conventional bus transit for a small city. *Journal of Public Transportation*, 20(2):19–39, 2017.
22. Silvio Micali and Vijai Vazirani. An  $O(\sqrt{N} \cdot e)$  algorithm for finding maximum matching in general graphs. *Proc. Symposium on Foundations of Computer Science*, pages 17–27, 1980.
23. Joanna Moody, Scott Middleton, and Jinhua Zhao. Rider-to-rider discriminatory attitudes and ridesharing behavior. *Transportation Research Part F: Traffic Psychology and Behaviour*, 62:258–273, 2019.
24. Jun Guan Neoh, Maxwell Chipulu, and Alasdair Marshall. What encourages people to carpool? an evaluation of factors with meta-analysis. *Transportation*, 44(2):423–447, 2017.
25. M E J Newman. *Networks: An Introduction*. Oxford University Press, 2010.
26. Jesper Riber Nielsen, Harald Hovmøller, Pascale-L. Blyth, and Benjamin K. Sovacool. Of “white crows” and “cash savers”: A qualitative study of travel behavior and perceptions of ridesharing in denmark. *Transportation Research Part A: Policy and Practice*, 78:113–123, 2015.
27. Noam Nisan and Amir Ronen. Algorithmic mechanism design (extended abstract). In *Proceedings of the Thirty-first Annual ACM Symposium on Theory of Computing*, STOC '99, pages 129–140, New York, NY, USA, 1999. ACM.
28. Andre’ De Palma, Robin Lindsey, and Nathalie Picard. Trip-timing decisions and congestion with household scheduling preferences. *Economics of Transportation*, 4(1-2):118–131, 2015.
29. Brishen Rogers. The social cost of uber. *SSRN Electronic Journal*, 2608017, 2015.
30. Silvia Mazzucotelli Salice and Ivana Pais. *Sharing Economy as an Urban Phenomenon: Examining Policies for Sharing Cities: Examining Policies for Sharing Cities*, pages 199–228. Palgrave Macmillan, Cham, 2017.
31. Paolo Santi, Giovanni Resta, Michael Szell, Stanislav Sobolevsky, Steven H. Strogatz, and Carlo Ratti. Quantifying the benefits of vehicle pooling with shareability networks. *Proceedings of the National Academy of Sciences*, 111(37):13290–13294, 2014.
32. Mitja Stiglic, Niels Agatz, Martin Savelsbergh, and Mirko Gradisar. The benefits of meeting points in ride-sharing systems. *Transportation Research Part B: Methodological*, 82(03):36–53, 2015.

33. Lijun Sun, Kay W. Axhausen, Der-Horng Lee, and Xianfeng Huang. Understanding metropolitan patterns of daily encounters. *Proceedings of the National Academy of Sciences*, 110(34):13774–13779, 2013.
34. Remi Tachet des Combes, Oleguer J. Sagarra Pascual, Paolo Santi, Giovanni Resta, Michael Szell, Steven H. Strogatz, and Carlo Ratti. Scaling law of urban ride sharing. *Nature Scientific Reports*, 7(42868):web, 2017.
35. Philip J. Vergragt and Halina Szejnwald Brown. Sustainable mobility: from technological innovation to societal learning. *Journal of Cleaner Production*, 15(11):1104 – 1115, 2007. The Automobile Industry & Sustainability.
36. Yi-Chang Chiu Xianbiao Hu and Lei Zhu. Behavior insights for an incentive-based active demand management platform. *International Journal of Transportation Science and Technology*, 4(2):119 – 133, 2015.
37. Jian Yang, Patrick Jaillet, and Hani Mahmassani. Real-time multivehicle truckload pickup and delivery problems. *Transportation Science*, 38(2):135–148, 2004.
38. Hani S.Mahmassani Zihan Hong, Ying Chen and Shuang Xu. Commuter ride-sharing using topology-based vehicle trajectory clustering: Methodology, application and impact evaluation. *Transportation Research Part C: Emerging Technologies*, 85:573 – 590, 2017.