Comparison of Trip Matching Algorithms for Mobility Sharing Applications

Francesca Martelli Institute of Informatics and Telematics (IIT) Italian National Research Council (CNR) Pisa, Italy Email: francesca.martelli@iit.cnr.it Maria Elena Renda IIT-CNR, Pisa, Italy and DUSP - MIT Cambridge, MA Email: elena.renda@iit.cnr.it, erenda@mit.edu

Abstract—Trip matching algorithms used in ride-sharing and carpooling systems share the common goal of optimizing the number of used vehicles to satisfy a set of trips according to their temporal and spatial constraints, in order to better allocate resources and reduce traffic and congestion. However, each matching algorithm could be designed to pursue a different objective like, for instance, reducing users' waiting time for quality of service, reducing the total amount of traveled distance within the system to reduce carbon footprint, or maximizing the time two trips are shared to favor user interaction. Changing the final system objective could significantly change the performance of the system itself. In this paper, we compare the performance of matching algorithms with different objectives and show whether potential tradeoffs exist in pursuing these objectives, taking into consideration all the actors in play in a mobility sharing application. In particular, by applying the matching algorithms to two sets of daily trips performed in the cities of Pisa, Italy, and Cambridge, USA, our analysis shows that there is a matching algorithm among the ones we tested able to provide a good compromise between different optimization objectives.

I. INTRODUCTION

Mobility sharing systems such as ride-sharing (e.g., Uber-POOL and Lyft Shared) and carpooling (e.g., WazeCarpool and Scoop) have expanded transportation options in cities around the world. By encouraging the sharing of vehicles and/or trips, these systems hold promise for improving the efficiency of transportation and optimizing resources utilization. These systems, provided through web or smartphone apps, have proven to be disruptive technologies with the potential to redefine transportation. If properly deployed, they can reduce traffic and congestion, leading to positive environmental and public health outcomes [1]. However, in mobility sharing applications, the algorithms used to match the trips (and consequently the users) could significantly change based on the goal of the system, which could be determined by company interests, users specific needs, or requirements imposed by the city/country the system is operating in. Changing the overall goal of the mobility sharing application could significantly impact the performance the system achieves.

Usually the goal of mobility sharing applications is to reduce the number of utilized vehicles by matching as many users and sharing as many trips as possible, reducing as well the environmental impact of shared trips. Another way to reduce vehicle miles traveled could be to properly design the trip matching algorithm so to minimize the distance traveled by the shared trips. Besides these two classical transportation approaches for the design of a trip sharing system, users' needs could be taken into consideration, like, for instance, the discomfort they could experience in sharing the trips.

Similarly, traditional transport-based metrics to evaluate trip-sharing methodologies include the evaluation of matched trips or mileage reduction. Other potential metrics worth investigating could be represented by the physical proximity of shared trips, somehow allowing users to potentially know the person they are sharing the trip with [2], and the amount of time users share on the same vehicle. In this paper, we report a thorough comparison of trip matching algorithms utilizing different performance metrics, by taking into consideration all the actors in play in a ride-sharing system, namely the providers, the authorities/cities, the users, both in the algorithms' design and in their evaluation. To the best of our knowledge, this is the first paper providing such a comparison and on two real datasets, one derived from an Italian city-wide mobility survey, the other from a campus-wide commuting survey in the US.

II. RELATED WORK

Even though ride-sharing and car-pooling problems are widely studied [3], [4], there are not many works related to the comparative evaluation of different metrics of ride-sharing matching algorithms. In [5] the authors provide the comparison among trip-matching algorithms, focusing on their potential to favor social integration and the evaluation of transportation efficiency when social benefits are pursued. A similar study has been presented in [6].

Similarly to our work, different objective functions have been taken under consideration in [7], but the algorithms are only based on trips' spatial attributes. Furthermore, differently from this paper, the authors only consider carpooling services where the users have preassigned roles (as driver or passengers). The concept of trips proximity has been exploited in several ways: one common way to compute proximity is by checking the similarity of the paths from origins to destinations, as in [8], [9], [10], typically by comparing trajectories coming from GPS. In here, we consider two trips to be close only if their origins and destinations are within a certain radius so to guarantee users a more accessible service.

In [11] the authors present some analytical results on the performance of ride-sharing algorithms with respect to some measures expressed as functions of the detour time, which is taken as an indicator of quality of service: the matching probabilities, the expected vehicle-kilometers traveled savings of ride matching, and the expected passenger-kilometers traveled.

To design and implement trip matching algorithms, we use the Shareability Network model [3], proven to be so efficient to be run online, in an on-demand, real-time fashion as required by ride-sharing applications, even when the constraints to be taken into consideration considerably increase. Differently from other approaches (see [4], [12]), increasing the number of parameters and/or constraints does not impact the complexity of the problem nor the efficiency in finding the optimal matching solution.

III. MODEL AND ALGORITHMS

Santi et Al. firstly introduced the notion of *shareability* network to model the taxi-sharing problem in Manhattan [3]. In their model, a trip is represented by a node in a graph, and the link between two nodes represents the sharing opportunity between the corresponding trips. The set of nodes (trips) and links (sharing opportunities) is called a shareability network. As in [3], we only consider sharing trips in pairs, which leads to nearly reduce by half the circulating vehicles. Given a trip *i*, the following notations are defined:

- S(i) is the starting location;
- D(i) is the destination location;
- st(i) is the starting time;
- d_{AB} is the distance between location A and location B;
- τ_{AB} is the time required to travel from A to B;
- tt(i) is the flexibility the user has on the trip starting time. This value, together with st(i), identifies a time interval [st(i) - tt(i), st(i) + tt(i)] representing the time window within which the traveler is willing to start the trip. Consequently, we set $st_{min}(i) = st(i) - tt(i)$ and $st_{MAX}(i) = st(i) + tt(i).$

Sharing trips comes with some discomfort for the users, namely the detour time required to pick up (and drop off) the passenger; this *sharing delay* time is called Δ . Given the above trip modeling and a specific Δ , two trips T_1 and T_2 are shareable, with the user corresponding to trip T_2 being the passenger, if the following three conditions hold:

- 1) $\tau_{S(1)S(2)} + \tau_{S(2)D(2)} + \tau_{D(2)D(1)} \le \tau_{S(1)D(1)} + \Delta$ 2) $st_{min}(2) st_{MAX}(1) \le \tau_{S(1)S(2)} \le st_{MAX}(2) \Delta$ $st_{min}(1)$
- 3) $d_{S(1)D(1)} > d_{S(1)S(2)} + d_{D(2)D(1)}$

The first condition states that the required extra time for the driver to pick up and drop off the passenger is below the detour time Δ , providing the users with a certain level of Quality of Service. The second condition states the temporal compatibility of the two trips, guaranteeing that the two starting time windows are properly overlapped. The third condition is related to the trips spatial compatibility, to guarantee that the distance traveled by sharing the two trips is shorter than

the sum of the two individual trips distances. Note that, in the case of carpooling, one of the two users is identified as the driver while the other is the passenger. This could happen only if the trip of the passenger is entirely included in the trip of the driver, but the allowed detour.

The reason why we adopt the shareability network model for comparing trip matching algorithms is that it is a very efficient model, enough to allow us to run the algorithms several times with different parameter settings. The amount of constraints and parameters setting does not affect the algorithms' complexity. The shareability network-based trip matching algorithms are scalable, being able to provide realtime solution even in presence of thousands of trips [3].

Once the shareability network on a given set of trips is built. a matching algorithm could be applied to find the optimal set of trips pairs to be shared in order to maximize the goal of the system. Thanks to the shareability networks model, it is possible to define weights on the graph's links and use them to maximize a desired metric. The weight could be, for instance, the travel time, or the travel distance, or the trips distance. Changing the weight on the links will generate a new matching algorithm and a completely different optimal set of matched trips. In the following, we describe some of the matching algorithms we implemented by applying different types of weights to the links of the shareability network. This is only a limited example of the different possible options that could be implemented using the shareability network model.

From the ride-sharing system point of view a good optimization goal could be to maximize the number of matched trips, so to reduce the number of utilized vehicles. We implemented this algorithm by simply assigning a weight equal to 1 to all the links in the shareability network, and called it the Cardinality Matching algorithm (CM). Even if this algorithm intuitively seems to be more efficient to reduce the overall system carbon footprint, it has been proved it does not [5]. From the environment point of view it would be better to maximize the overall saved distance by maximizing users shared kilometers. For this reason, we also implemented the Saved Distance Matching algorithm (DM), where each link is assigned with a weight equal to the distance that could be saved if the trips associated to the connected nodes are matched.

Another objective could be to favor users' interaction while sharing the trips. To pursue this goal, we implemented the *Time* Matching algorithm (TM), where the weight of each link is set to the time the users would spend together by sharing the corresponding trips.

There is one more player that we have not taken into consideration so far: the user and their satisfaction. Users could be keener to use a ride-sharing system if the trips to be shared would be more convenient, not only in terms of required extra time (expressed by the detour time Δ), but also in terms of required extra distance, that could be expressed in terms of the proximity of the starting location and/or the destination of the two trips. In fact, proximity intended as convenience is another way to see ride sharing as "more accessible". Furthermore, since users hardly trust to travel with strangers [2], physical proximity of the trips could lead to know the potential person with whom they are going to share the trip. To account for this, we also implemented the *Proximity Matching* algorithm (PM), where, given a radius r > 0, the weight on the links is set to the distance between the two trips starting locations, if this is smaller than r, to a very large value otherwise (we used the sum of all the distances over the shareability network).

IV. PERFORMANCE EVALUATION

A. Datasets

For our simulations we used two different mobility datasets: one collected in Pisa, Italy, the other one in Cambridge, US. The former is related to an anonymous mobility survey issued in 2016 to citizens who live or work in Pisa; the latter one derives from a commuting survey issued to MIT employees in 2018. From both surveys, we extracted the answers of respondents who declared to use the car as a primary transport mode to reach their workplace. We obtained 1,966 complete and valid answers for Pisa, and 1,368 for Cambridge.

Both surveys were fed to the correspondent community of commuters to understand current mobility patterns and habits, and identify potential future development to lighten daily commuting traffic and congestion. Nevertheless, data has been collected in a slightly different way; in particular, in Pisa dataset the provided addresses are exact, while for MIT home addresses are expressed as census block group numbers. Thus, for each MIT answer, we randomly generated a position within the census block group and use it as the trip starting location; furthermore, while trip starting times and user flexibility in starting their trips have been exactly reported in the Pisa survey, MIT employees were allowed to select a time interval for the departure time (30' for peak hours, two or more hours otherwise), while there was no explicit mention of how much flexibility users had. For this reason, for the MIT dataset, for each data point, we first ran 20 instances by randomly generating a departure time within the time interval, then we averaged the results. Moreover, we set four different flexibility times: 0, 5, 15, and 30 minutes. The main difference between the two datasets is, however, that in the MIT survey all answers share the workplace (the MIT campus), while in the survey issued in Pisa the users do not share the destination location, even though there are three main major workplaces.

B. Simulations

Given the above four defined matching algorithms, CM, DM, TM, and PM, to evaluate efficiency differences and identify potential tradeoffs, we first build the shareability network related to the two datasets. Since a specific maximum allowed detour time is a required input to guarantee a certain level of quality of service to the users, we use different values of detour time Δ , namely $\Delta \in [0, 1, 2, ..., 15]$ minutes. Each value of Δ changes the set of potentially shareable trips, thus generates a different shareability network. For the PM algorithm, we also define different values of potential matches for a trip and then select the best one. In our simulations we set $r \in [1, 2, 3, 4, 5]$ km.

To compare the performance of the proposed matching algorithms, we evaluate four different metrics, computed as a function of the detour time Δ :

- 1) the percentage of matched trips;
- 2) the percentage of saved distance;
- 3) the average percentage of users' shared time per trip;
- 4) the percentage of close matched users.

We compute the normalized saved distance by summing up the distance saved in all shared trips and by dividing the result by the sum of the traveled distance required to do all the trips without sharing. For each shared trip s, obtained by matching trips a and b, its saved distance sd is given by: $sd(s) = d_{S(a)D(a)} - d_{S(a)S(b)} - d_{D(b)D(a)}$. The normalized saved distance of a matching set M is:

$$nsd(M) = \frac{\sum_{s \in M} sd(s)}{\sum_{a \in V(G)} d_{S(a)D(a)}}$$

where V(G) is the set of nodes in the shareability network.

The normalized shared time is computed by summing up the time the users spend together in the shared trips and by dividing the results by the number of shared trips. More precisely, given a shared trip s, the shared time is the time the passenger is on the trip, namely $\tau_{S(b)D(b)}$. Then we could compute the normalized shared time of a matching set M as:

$$nst(M) = \frac{\sum_{s \in M} \tau_s}{\#(M)}$$

Since for the MIT survey the departure times were expressed as time intervals, and people were only asked if they have or not flexibility in leaving to go to work, without providing a precise value or time interval, in the *MIT* data set we are missing some level of details with respect to the Pisa one. For this reason, we ran 20 instances of different settings and we averaged the results. This is the reason why the results for MIT reported in Figures 2 and 1 are expressed as an average. In particular, for each trip *i*, we randomly pick a starting time st(i) within the user selected time interval and we use four different values for flexibility tt(i), namely 0, 5, 15, and 30 minutes. Here, for space sake, we only show results for tt(i) = 15 for all *i*, representing a good average value for time flexibility. Furthermore, the average flexibility value provided in the Pisa survey is 18 minutes.

When running the proximity matching (PM) algorithm over the trips in Pisa, we apply the proximity radius on both trips origins and destinations, while for the trips in Cambridge we apply proximity radius only on the origins, since all the trips have the same destination, i.e. the MIT Campus.

C. Results

In the following, we show the outcomes of the simulations we executed on the two datasets. For each algorithm and



Fig. 1: Normalized saved distance metric (nsd) computed over the resulting matchings of the algorithms, with PM applied by varying the radius in 1-5 km.



Fig. 2: Matched trips metric computed over the resulting matchings of the four algorithms, with PM applied by varying the radius from 1 to 5 km.

for each detour time Δ , we find the best matching over the correspondent shareability network and compute the four metrics listed above.

a) Matched Trips: Maximizing the number of matched trips so to reduce the number of utilized vehicles is usually the main objective of a trip-sharing system. Figure 2 reports the percentage of matched trips for the two datasets by applying the designed algorithms. As expected and shown in the graphs, CM, built to maximize the number of matched trips, outperforms the other algorithms. Nevertheless, when $\Delta \geq 5$, almost all the algorithms (except for PM when applied with lower r values) achieve at least 80% of matched trips. This means that by slightly decreasing the quality of service provided to the users, here represented by the detour time Δ , it is possible to maximize the number of matched trips.

b) Saved Distance: When pursuing a more sustainable ride-sharing system goal, the matching algorithm should be able to reduce the overall traveled distance. In this scenario, DM clearly outperforms the other algorithms (Figure 1), especially for higher Δ values. The TM algorithm achieves similar results for low values of detour time, but as the detour time increases, in TM prevails the goal of maximizing the time the users spend together along the shared trip, thus generating more traveled mileage (even if within the distance constraints provided by the third condition for two trips to be shareable). CM and PM (especially PM with radius set to 1 km) are those achieving worse results. The overall behavior of all the algorithms is quite similar, with an initial increasing trend of nsd as Δ increases, followed by a stabilization and, for most algorithms, decline. These graphs clearly show that if the ride-sharing system objective is to achieve higher environmental benefits while providing shareable trips, other goals could be detrimental towards this objective. In particular, trying to match as much users as possible to reduce vehicles or close-by users, is not the best strategy to adopt. In the MIT dataset, we also observe relatively higher distance savings when compared to Pisa results; this is likely due to the larger geographical footprint of the considered area. A potential implication is that ride-sharing mechanisms could be especially effective in large cities to reduce their carbon footprint.

c) Users' Shared Time: In Figure 3 we report the results achieved by all the algorithms in terms of the users' shared time metric. TM, the algorithm designed to maximize users' shared travel time per trip, is the best, closely followed by DM. For PM, the higher is the proximity radius, the higher is the chance that users could share more time along a shared trip. This implies that in a ride-sharing system designed to favor users interaction, users should give up on convenience in terms of service accessibility. For Pisa dataset, a Δ value between 2 and 5 minutes does not result in a substantial differences between the algorithms (including the five PM variations).

d) Users' Position Proximity: In our model, Δ represents a certain level of quality of service provided to the user by bounding the detour time needed to share a trip. Another



Fig. 3: Average normalized users' shared time computed over the resulting matchings of the algorithms, with PM applied by varying the radius in 1-5 km.



Fig. 4: Close, far, and unmatched users for the four tested algorithms, with an increasing proximity radius, using Delta = 5.



Fig. 5: Performance radar graphs, given a proximity radius of 1 km, and $\Delta = 5$. Each corner of the radar graphs represent one of the metrics: the percentage of matched users (left), the percentage of close users (top), the normalized saved distance (right), and the normalized shared time (bottom).

level of quality of service we want to analyze is based on the other dimension of shareable trips, i.e., their distance. In Figure 4 we report, for each of the four algorithms and for each of the five radii considered, the resulting percentage of matched and unmatched users. Within the matched users, we also report the percentage of close (within the radius) and far (farther than the radius) matched users. PM is obviously the algorithm achieving better results in terms of close matched users, being the one built to optimize this metric, but it is also the one with higher percentage of unmatched users, meaning that the corresponding trips could not be shared. Among the algorithms not properly designed to maximize matched trips physical proximity, DM is the best and CM is the worse, while the opposite holds if we look at the percentage of unmatched trips (much clearer in the Pisa dataset graph).

e) Overall: In Figure 5 we report the performance achieved by the four algorithms applied to the two datasets, Pisa (a) and MIT (b), in terms of the four metrics we are evaluating (percentage of matched users, percentage of close users, normalized saved distance, normalized shared time). Here, for space sake, we only show the results obtained with a proximity radius of 1 km, and the shareability networks built considering a 5 minutes detour time Δ , being the other results quite similar. In these graphs a larger area indicate higher performance. By analyzing this Figure, it seems DM could provide a good tradeoff among the tested algorithms, being its performance a good compromise for the considered metrics.

V. CONCLUSION

In this paper we provide a comprehensive comparison of different ride matching algorithms, each pursuing a specific goal, applied to two datasets related to mobility surveys performed in Italy and US. In particular: CM algorithm has been designed to minimize the number of utilized vehicles by maximizing the number of matched trips; DM maximizes the amount of saved traveled distance within the overall system; PM provides the users with a more convenient service, by matching trips with close origins and destinations; and TM favors users' integration by maximizing the time users share in the vehicle. All the algorithms have been developed so to contain the traveled mileage and limit the sharing detour time within the threshold.

The performance results provide useful and meaningful insights. With no surprise, each algorithm outperforms the others when the corresponding metric is evaluated. But there is one algorithm, DM, showing promising average results in all the metrics (see Figure 5), suggesting that a potential tradeoff between different objectives exists. This insight could and shall be addressed not only in developing the system, but also at policy and regulatory level. In fact, if the authorities aim at a more sustainable transportation, they could require ridesharing systems to reduce carbon footprint. This way it could also be possible to partially meet the other goals, potentially achieving good results not only for the environment (and the society, in the end), but also for the operators (trough vehicles reduction) and for users (in terms of convenience).

For the future, we would like to develop new algorithms tackling different scenarios. For instance, to cope with COVID-19 and similar future threats, transportation policies have to be changed (as highlighted in [13], [14], [15] and references therein), since the main issue could be users' unwillingness to choose public transport or share a trip with strangers. From the point of view of shared mobility, such as car-pooling or ride-sharing, a new healthy aspect should be considered, favoring, for instance, trip matchings among users belonging to a pod or bubble, to limit contacts and virus's spread.

The set of algorithms and metrics presented here is only a small sample of what is achievable by adopting the shareability networks model for sharing trips. We also plan to evaluate a combination of these algorithms, in order to increase the benefits provided by each one while still maintaining a sustainable overall objective.

REFERENCES

[1] S. Shaheen and A. Cohen, "Shared ride services in north america: definitions, impacts, and the future of pooling," *Transport Reviews*, vol. 39, no. 4, pp. 427–442, 2019. [Online]. Available: https://doi.org/10.1080/01441647.2018.1497728

- [2] A. Rahimi, G. Azimi, and X. Jin, "Examining human attitudes toward shared mobility options and autonomous vehicles," *Transportation Research Part F: Traffic Psychology* and Behaviour, vol. 72, pp. 133–154, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1369847820304150
- [3] P. Santi, G. Resta, M. Szell, S. Sobolevsky, S. H. Strogatz, and C. Ratti, "Quantifying the benefits of vehicle pooling with shareability networks," *Proceedings of the National Academy of Sciences*, vol. 111, no. 37, pp. 13 290–13 294, 2014.
- [4] M. Stiglic, N. Agatz, M. Savelsbergh, and M. Gradisar, "Making dynamic ride-sharing work: The impact of driver and rider flexibility," *Transportation Research Part E: Logistics and Transportation Review*, vol. 91, pp. 190–207, 2016. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554515303033
- [5] F. Librino, M. E. Renda, P. Santi, F. Martelli, G. Resta, F. Duarte, C. Ratti, and J. Zhao, "Home-work carpooling for social mixing," *Transportation*, vol. 47, no. 5, pp. 2671–2701, 2020.
- [6] M. Berlingerio, B. Ghaddar, R. Guidotti, A. Pascale, and A. Sassi, "The graal of carpooling: Green and social optimization from crowd-sourced data," *Transportation Research Part C: Emerging Technologies*, vol. 80, pp. 20–36, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0968090X17300657
- [7] A. Najmi, D. Rey, and T. H. Rashidi, "Novel dynamic formulations for real-time ride-sharing systems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 108, pp. 122–140, 2017. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1366554517300017
- [8] N. Ta, G. Li, T. Zhao, J. Feng, H. Ma, and Z. Gong, "An efficient ridesharing framework for maximizing shared route," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 2, pp. 219–233, Feb 2018.
- [9] R. Trasarti, F. Pinelli, M. Nanni, and F. Giannotti, "Mining mobility user profiles for car pooling," in *Proceedings of the 17th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, ser. KDD '11. New York, NY, USA: Association for Computing Machinery, 2011, p. 1190–1198. [Online]. Available: https://doi.org/10.1145/2020408.2020591
- [10] W. He, D. Li, T. Zhang, L. An, M. Guo, and G. Chen, "Mining regular routes from gps data for ridesharing recommendations," in *Proceedings of the ACM SIGKDD International Workshop on Urban Computing*, ser. UrbComp '12. New York, NY, USA: Association for Computing Machinery, 2012, p. 79–86. [Online]. Available: https://doi.org/10.1145/2346496.2346510
- [11] C. F. Daganzo, Y. Ouyang, and H. Yang, "Analysis of ride-sharing with service time and detour guarantees," *Transportation Research Part B: Methodological*, vol. 140, pp. 130–150, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0191261520303660
- [12] N. Agatz, A. L. Erera, M. W. Savelsbergh, and X. Wang, "Dynamic ride-sharing: a simulation study in metro atlanta," *Procedia - Social and Behavioral Sciences*, vol. 17, pp. 532–550, 2011, papers selected for the 19th International Symposium on Transportation and Traffic Theory. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1877042811010895
- [13] M. V. Corazza and A. Musso, "Urban transport policies in the time of pandemic, and after: An arduous research agenda," *Transport Policy*, vol. 103, pp. 31–44, 2021. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0967070X21000160
- [14] A. Sharifi and A. R. Khavarian-Garmsir, "The covid-19 pandemic: Impacts on cities and major lessons for urban planning, design, and management," *Science of The Total Environment*, vol. 749, p. 142391, 2020. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0048969720359209
- [15] D. G. Costa and J. P. J. Peixoto, "Covid-19 pandemic: a review of smart cities initiatives to face new outbreaks," *IET Smart Cities*, vol. 2, no. 2, pp. 64–73, 2020. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/ietsmc.2020.0044