Ride Sharing: A Network Perspective

Erez Shmueli[1,2], Itzik Mazeh[1], Laura Radaelli[1], Alex "Sandy" Pentland[2], and Yaniv Altshuler[3]

[1]Department of Industrial Engineering, Tel-Aviv University
[2]MIT Media Lab
[3]Athena Wisdom

Abstract. Ride sharing's potential to improve traffic congestion as well as assist in reducing CO2 emission and fuel consumption was recently demonstrated by works such as [1]. Furthermore, it was shown that ride sharing can be implemented within a sound economic regime, providing values for all participants (e.g., Uber). Better understanding the utilization of ride sharing can help policy makers and urban planners in modifying existing urban transportation systems to increase their "ride sharing friendliness" as well as in designing new ride sharing oriented ones. In this paper, we study systematically the relationship between properties of the dynamic transportation network (implied by the aggregated rides) and the potential benefit of ride sharing. By analyzing a dataset of over 14 Million taxi trips taken in New York City during January 2013, we predict the potential benefit of ride sharing using topological properties of the rides network only. Such prediction can ease the analysis of urban areas, with respect to the potential efficiency of ride sharing for their inhabitants, without the need to carry out expensive and time consuming surveys, data collection and analysis operations.

1 Introduction

The increasing availability of portable technologies and ubiquitous connectivity makes the dream of having smart cities closer [2]. Easier collection of data on the way people live in a city and data analysis methods empower city administrators and policy makers that can better manage a city and improve the life in it.
Availability of large-scale datasets gives rise to new possibilities to study urban mobility. Calabrese et al. [3] show that mobile phone data can be used as a proxy to examine urban mobility and Noulas et al. [4] analyze social network data of different cities to find that mobility highly correlates with the distribution of points of interest. Mobile technologies support also successful applications, such as Waze [5], that provide traffic-aware city navigation by using data provided by the community.
Many of the most urgent problems of big cities relate to cars. Alternative ways of moving in the city, such as autonomous mobility-on-demand and short-term car rental have been identified among the possible solutions to the transport headaches [6].

Altshuler et al. [7] show that on-demand route-free public transportation based on mobile phones provides better traveling times than standard fix-route methods. Cici et al. [8] use mobile phone and social network data to show that traffic in the city of Madrid can be reduced by 59% if people are willing to share their home-work commute ride with neighbors. The authors study the effect of friendship on the potential of ride sharing, showing that if people are willing to ride with friends of friends the savings are close to the case of riding with strangers. Recent work by Santi et al. [1] introduces a new way to quantify the benefits of sharing, and studies a GPS dataset of taxi rides in New York City. The findings show that when passengers have a 5 minutes flexibility on the arrival time, and they are willing to wait up to 1 minute after calling the cab, over 90% of the sharing opportunities can be exploited and 32% of travel time can be saved. These results encourage the deployment and policies supporting ride sharing in urban settings.

Data analysis applied to large-scale datasets can reveal patterns of individual and group behaviors [9], and methods from social network analysis and graph theory show the structure and dynamics of social and communication networks [10]. A network can often be built on easily available data and becomes useful to predict apparently unrelated events or facts. A phone call network can signal an emergency situation [11], and social network of a Twitter account can identify a spammer [12]. Network analysis has been used to predict installation rates of mobile applications [13], spending behaviors of couples [14], and personality of individuals [15].

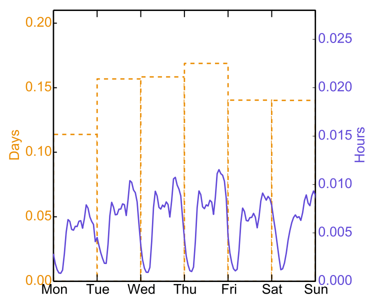
In this work, we propose a network-based approach to the prediction of benefit of ride sharing based on features of the transportation network of a city. We study its efficacy on a dataset of over 14 Million taxi trips taken in New York City during January 2013. First, we compute the benefits as a function of the maximum delay experienced by a rider, and we find encouraging results, e.g., more than 70% of the rides can be shared when riders are willing to accept a delay up to 5 minutes. Second, we apply the network-based approach to predict the benefit for a given maximum delay. We use topological features of the dynamic rides network to predict the benefit, and find that the combination of selected features has a high predictive power. Such prediction can enable analysis of urban areas, with respect to the potential efficiency of ride sharing for their inhabitants, without the need to carry out expensive and time consuming surveys or data collection.

The remainder of this paper is structured as follows: Section 2 describes the dataset; Section 3 presents the results of our study; and Section 4 summarizes our contributions and indicates future directions.

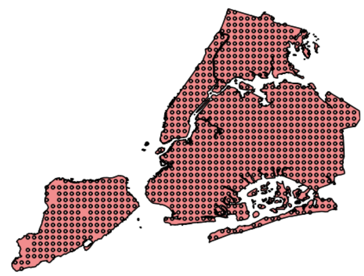
2 Datasets

We analyze a dataset of 14,776,615 taxi rides collected in New York City during January 2013 [16]. Each ride record consists of five fields: origin time, origin longitude, origin latitude, destination longitude, destination latitude. Removing rides with missing or erroneous GPS coordinates, we ended up with 12,784,243 rides.

Figure 1a reports the distribution of rides per day of the week and per hour of the day. The time distribution is far from uniform: the number of rides is higher in the middle of the week and it is lower during the weekend, and the daily distribution has peaks in the morning and around 6-7pm corresponding to start/end of working hours.



(a) Probability Density Function (PDF) of the number of rides per day of week/hour of day.



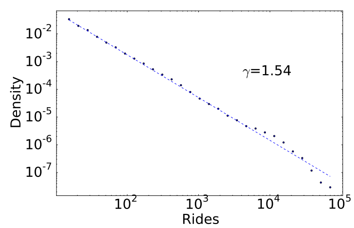
(b) Illustration of the network nodes on the map of NYC.

Fig. 1: Temporal and geographical distribution of data.

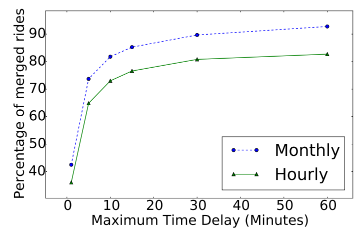
3 Results

We construct the rides network from the dataset, where the set of nodes represents equally divided regions of New York City, and an edge (u, v) connects the two regions u and v if there exists at least one ride from region u to region v. The resulting network comprises 813 nodes and 58,014 edges. Figure 1b shows the geographical distribution of the nodes on a map, where New York is densely covered by the nodes with only a few empty spots in Staten Island.
We applied a simplified version of the methodology used by Santi et al. [1] to calculate the potential benefits of ride sharing. Benefits are expressed in terms of how many rides can be shared (therefore, how many rides can be saved), and are computed as a function of the guaranteed quality of service. The quality of the service is measured by the maximum time delay in catching a ride and arriving at destination, and it represents the maximum discomfort that a passenger can experience using the service. Our analysis aims at finding pairs of rides, which are represented in the network by the same edge (i.e., have the same origin and destination), that can be shared. For each edge, we examine its corresponding set of originating rides, and count the number of ride pairs that can be merged, taking into consideration the maximum time delay parameter. Figure 2a shows the probability density function (pdf) of the number of rides per edge. As can be seen from the Figure, the distribution is heavy tailed and seems to follow a power-law. In other words, most of the edges (i.e.,

pairs of origin-destination) induce a small number of rides while a small number of edges induce an extremely high number of rides.



(a) Probability Density Function of the number of rides per edge.



(b) Percentage of merged rides for the entire network (blue dashed line) and averaged over all sub-networks (green solid line).

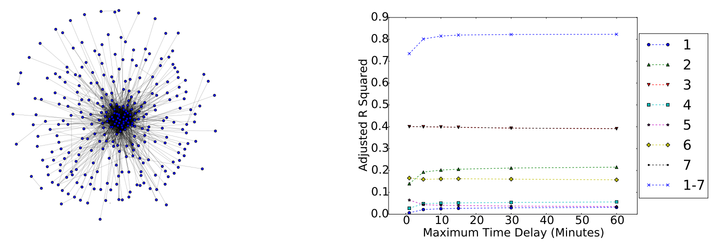
Fig. 2: Analysis of rides

Figure 2b (blue dashed line) reports the percentage of shareable rides as a function of the maximum time delay parameter. Results are encouraging, more than 70% of the rides can be shared when passengers can accept a delay of up to 5 minutes. As expected, the benefit of ride sharing increases when the passengers are willing to take a higher discomfort, and the percentage of shareable rides is more than 90% when passengers can wait 30 minutes or more.

In this simplified analysis we considered the merging of two rides at a time, without taking into account the number of passengers in each ride. Since the average number of passengers per ride is 1.7, the number of saved rides could have been even higher by merging more than 2 rides at a time. On the other hand, in some cases, even the merging of two rides at a time might have resulted in overcrowding of the vehicle.

We model the high dynamic nature of the network by dividing the dataset into hourly snapshots, and creating $31 \times 24 = 744$ sub-networks. An illustration of one such sub-network is shown in Figure 3a. Intuitively we see that most of the nodes are highly connected, but a considerable number of nodes are connected to only one other node in the network.

We continue our analysis by considering each sub-network separately and averaging the results. Figure 2b (green solid line) shows the potential benefit of ride sharing which is first calculated for each sub-network separately and then averaged over all sub-networks. When averaging the sub-networks the benefits are constantly lower (around 10% less) than for the entire network. This reduction is expected as rides that start at the end of the hour have lower chances of being merged (i.e. they cannot be merged with rides that start at the beginning of the consecutive hour). Nevertheless, the curves have a similar trend in the two scenarios. This suggests that the effect of the quality of service is similar even on different networks.



(a) One of the rides sub-networks.

(b) Adjusted R^2.

Fig. 3: Network analysis.

Finally, for each one of the sub-networks, we extract a set of seven topological features — (1) the number of nodes, (2) the number of edges, (3) the averaged degree centrality score, (4) the averaged betweenness centrality score, (5) the averaged closeness centrality score, (6) the averaged eigenvector centrality score, and (7) the density score. Next, we use a linear regression model to fit the percentage of merged rides (dependent variable) to the various extracted features (independent variables) for the 744 instances. This process is then repeated multiple times, each time using a different maximum time delay. Figure 3b shows the obtained quality of fit as a function of the maximum time delay. As can be seen in the figure, the results are quite encouraging. When using all seven features as independent variables the quality of fit is remarkably high (R squared value close to 1). Clearly, when using each feature separately as the independent variable, the quality of fit drops.

4 Summary and Future Work

We aim at predicting the benefit of ride sharing expressed as the percentage of rides that can be shared with a limited discomfort for riders. We perform empirical analysis on a large-scale dataset, showing that we can predict the benefit of ride sharing based on the topological properties of the rides network only. In particular, we identify seven network topological features that combined can effectively predict the benefit.
Assessing the benefit of ride sharing is only a first step in devising a ride sharing solution. A direction for future work is to use the data-driven approach to study and incentivize participation in ride sharing. A complete solution should aim at guaranteeing benefits both for the riders (quality of service and low prices) and for the drivers (reducing the inconvenience of sharing the personal car) and at achieving the true potential of ride sharing by reducing the number of cars on the roads.

References

1. Santi, P., Resta, G., Szell, M., Sobolevsky, S., Strogatz, S.H., Ratti, C.: Quantifying the benefits of vehicle pooling with shareability

networks. Proceedings of the National Academy of Sciences **111**(37) (2014) 13290–13294
2. Caragliu, A., Del Bo, C., Nijkamp, P.: Smart cities in europe. Journal of urban technology **18**(2) (2011) 65–82
3. Calabrese, F., Diao, M., Di Lorenzo, G., Ferreira Jr, J., Ratti, C.: Understanding individual mobility patterns from urban sensing data: A mobile phone trace example. Transportation research part C: emerging technologies **26** (2013) 301–313
4. Noulas, A., Scellato, S., Lambiotte, R., Pontil, M., Mascolo, C.: A tale of many cities: universal patterns in human urban mobility. PloS one **7**(5) (2012) e37027
5. waze.com. `http://www.waze.com` [Online; accessed 15-Oct-2014].
6. Chin, R.: Solving transport headaches in the cities of 2050. BBC Future (June 2013) [Online; accessed 15-Oct-2014].
7. Altshuler, T., Shiftan, Y., Katoshevski, R., Oliver, N., Pentland, A.S., Altshuler, Y.: Mobile phones for on-demand public transportation. In: NetSci. (2014)
8. Cici, B., Markopoulou, A., Frias-Martinez, E., Laoutaris, N.: Assessing the potential of ride-sharing using mobile and social data: a tale of four cities. In: Proceedings of the 2014 International Conference on Pervasive and Ubiquitous Computing, ACM (2014) 201–211
9. Lazer, D., Pentland, A.S., Adamic, L., Aral, S., Barabási, A.L., Brewer, D., Christakis, N., Contractor, N., Fowler, J., Gutmann, M., Jebara, T., King, G., Macy, M., Roy, D., Van Alstyne, M.: Computational social science. Science **323**(5915) (2009) 721–723
10. Onnela, J.P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., Kertész, J., Barabási, A.L.: Structure and tie strengths in mobile communication networks. Proceedings of the National Academy of Sciences **104**(18) (2007) 7332–7336
11. Altshuler, Y., Fire, M., Shmueli, E., Elovici, Y., Bruckstein, A., Pentland, A.S., Lazer, D.: The social amplifier—reaction of human communities to emergencies. Journal of Statistical Physics **152**(3) (2013) 399–418
12. Almaatouq, A., Alabdulkareem, A., Nouh, M., Shmueli, E., Alsaleh, M., Singh, V.K., Alarifi, A., Alfaris, A., Pentland, A.S.: Twitter: Who gets caught? observed trends in social micro-blogging spam. In: Proceedings of the 2014 ACM Conference on Web Science. WebSci '14, ACM (2014) 33–41
13. Pan, W., Aharony, N., Pentland, A.S.: Composite social network for predicting mobile apps installation. In: AAAI. (2011)
14. Singh, V.K., Freeman, L., Lepri, B., Pentland, A.S.: Predicting spending behavior using socio-mobile features. In: Proceedings of the International Conference on Social Computing (SocialCom), IEEE (2013) 174–179
15. de Montjoye, Y.A., Quoidbach, J., Robic, F., Pentland, A.S.: Predicting personality using novel mobile phone-based metrics. In: Social Computing, Behavioral-Cultural Modeling and Prediction. Springer (2013) 48–55
16. NYC Open Data. `http://data.ny.gov/` [Online; accessed 01-Sept-2014].