Unravelling System Optimum Structure by trajectory data analysis

Ruiwei Chen, Ludovic Leclercq

Unravelling System Optimum Structure by trajectory data analysis

Ruiwei Chen and Ludovic Leclercq

University of Lyon, IFSTTAR, ENTPE, Lyon, France
{ruiwei.chen, ludovic.leclercq}@ifsttar.fr

Extended abstract submitted for presentation at the hEART 2019 8th Symposium
Sept. 4–6, 2019, Budapest, Hungary

Word count: 2016 words (excluding the references)
April 23, 2019

Abstract

This work investigates network-related trajectory features to unravel trips that the most contribute to the system under-performance. When such trips are identified, features analysis also permits to identify the best alternatives in terms of routes to make the system to its optimum. First, data mining is carried out on trajectories obtained from reference dynamic traffic assignment (DTA) simulations in a real-world network, based on User-Equilibrium (UE) and System-Optimum (SO). This helps us (i) to target the trajectories to be changed, and (ii) to identify their main features (trip lengths, experienced travel time, path marginal costs, and network-related features such as betweenness centrality and traffic light parameters, etc.). Similarity analysis based on Longest Common Subsequence, Principle Component Analysis are the main methods that are performed to carry out descriptive analysis of trajectories. Supported Vector Machine is then used to determinate the features with regards to their contribution to better network performance.

1 Introduction

In urban areas, the conflict between the increasing mobility demand and limited infrastructures degrades the level of service of road networks. The resulting consequences include (i) economic loss resulting from wasted time and fuel in traffic jams and (ii) environmental pollutions. A key element to determinate the network level of service is the traffic assignment (TA) process as it describes how users spread over the network. Different levels of equilibrium may result from different TA: User equilibrium (UE) and system optimum (SO) (Wardrop, 1952, Beckmann et al., 1956, Smith, 1979, Mahmassani and Peeta, 1993). In UE, network users choose their route by minimizing their own travel cost when traveling from Origin to Destination (O-D). Under SO equilibrium, users choose their travel paths in such a way that the total travel costs of the whole network are minimized.

Over the past few decades, increasing sources of traffic data are becoming available: GPS-based floating car data, Bluetooth data, GPS data from cellphones, etc. (Treiber and Kesting, 2013) . A variety of vehicle trajectory data gives new insights for better understanding the network, user mobility patterns, and the congestion mechanism. This rich data helps engineers, decision makers, and researchers to propose corresponding strategies for improving urban mobility (Gonzalez et al., 2008, Saeedmanesh and Geroliminis, 2016, Lopez et al., 2017). For example, with detailed GPS data from mobile phones, Wang et al. (2012) show that the congestion of a given network is mostly due to very few network users who are on the most congested road segments. However, this conclusion is obtained by decreasing the traffic demand from a certain number of O-D pairs, without giving alternative routing solution. Çolak et al. (2016) use mobile phone GPS-data to compute path travel time and calibrate TA models. They show that if 10 % of drivers adjust their routing behavior under SO condition instead of selfish routing, the average travel cost of the whole network drops 40 %. Nevertheless, their static TA model ignores the dynamic interactions of the traffic, especially the spillback of queues in congested situations.

2 Objective and main contribution

The objective of this work is to investigate network-related trajectory features, in order to unravel trips that the most contribute to the system under-performance. When such trips are identified, features analysis also permits to identify the best alternatives in terms of routes to make the system to its optimum. Re-routing strategies are given to target trips in order to improve network performance by only considering network-related features. This avoids computational burden of DTA simulations. The contribution of this work is threefold:

- By analyzing trajectories in UE and SO equilibrium from DTA simulations, define the network-related trajectory features, that determinate the users who contribute the most to the network congestion.
- With the defined network-related trajectory features, propose re-routing strategies for target users in order to improve the total network performance (e.g., the total travel times of all vehicles).
- Assess through simulations the performance of the solution and re-routing process in a real-world test case.

3 Methodology

Figure 1 presents the framework of our methodology. First, descriptive analysis of UE and SO trajectories is carried out. We define trajectory features from two reference DTA simulations, under UE and SO condition, with the same traffic volume and departure time. Trajectories from the SO-based simulation are considered as the optimal travel pattern. We identify the most influential features that differ the SO trajectories from UE trajectories. Principal component analysis (PCA) is carried out to reveal similar trajectory features under both equilibrium. Longest common subsequence (LCS) is also used to measure the similarity of trajectories (Kim and Mahmassani, 2015). The users whose trajectories are of the largest dissimilarity are then targeted to give re-routing strategies.

Once the users are identified, new DTA simulations are carried out, with pre-defined optimal patterns for the target users. The others are assigned under UE condition. We then quantify the network total travel time (TTT) reduction with respect to the reference UE simulation. This defines the final target trajectories that contribute the most to the network performance improvement. Then, for these trajectories, the features related to traffic characteristics from SO simulation are considered as training samples$\boldsymbol{y}$. The network-related features are considered as training points $\boldsymbol{p}$. Supervised learning with Supported Vector Machine (SVM) (Ben-Hur et al., 2001) is carried to this training dataset, so that a relation $f : \boldsymbol{p} \rightarrow \boldsymbol{y}$, mapping network-related trajectory features, to trajectory features that define target users.

At last, a new set of target trajectories can be defined by only using network-related features. We give them pre-defined optimal patterns and carry out DTA in UE condition to evaluate the TTT reduction. Furthermore, instead of re-routing by optimal patterns, f can also help us to define the best alternative paths to make the system to its optimum, based on the identified network-related trajectory features. The proposed re-routing strategies are validated by UE simulation.

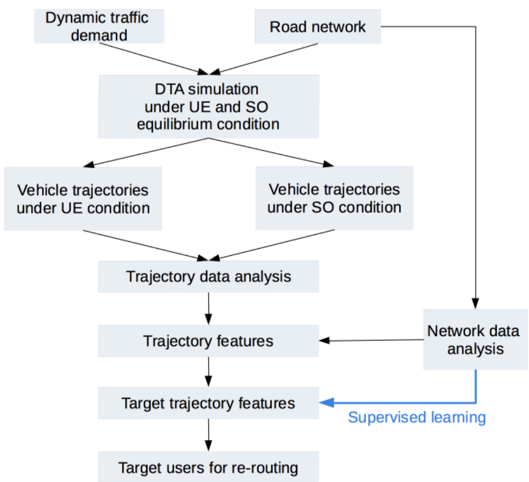

Figure 1: Flow chart of methodology

4 Case Study on a real-world network

4.1 Network and demand

The first test case is carried out for the road network of the 6^{th} district of Lyon (*Lyon6*), France. Figure 2 shows the area of Lyon6. Figure 3 (left) shows the link level representation of the main road network. There are in total 786 links, 205 intersections and 710 OD pairs in the network. Figure 3 (right) presents the time-dependent traffic demand in the network. The total simulation period is 4 hours.

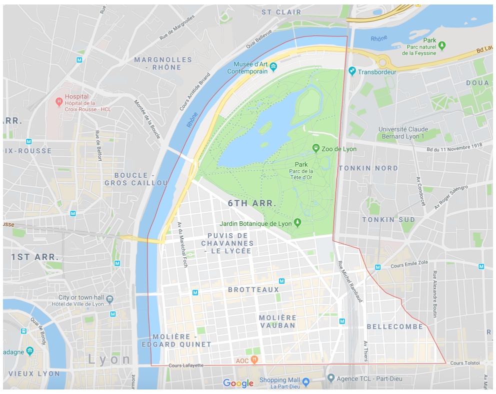

Figure 2: Area of the 6^{th} district of Lyon, France, © Google Maps 2019

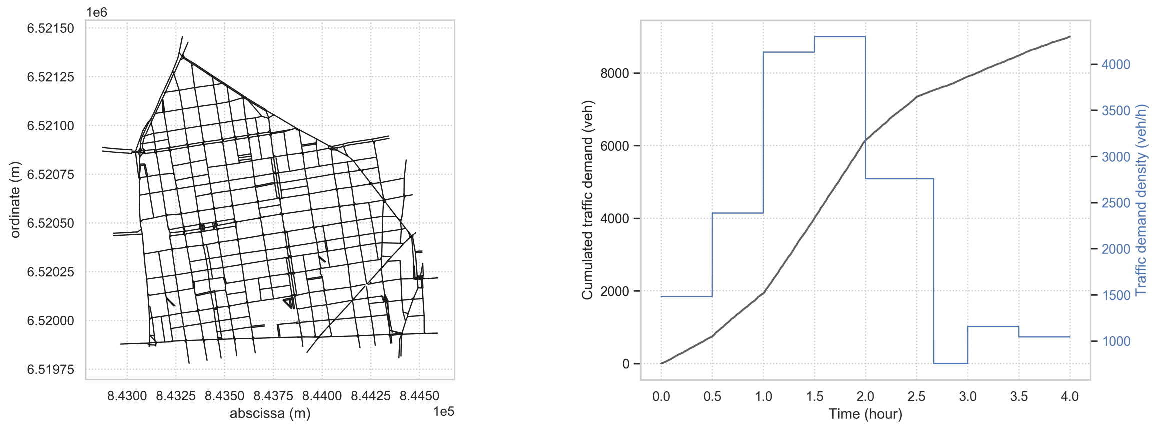

Figure 3: Flow chart of methodology

4.2 Discriptive analysis

A trajectory i $\mathcal{L}_i$ is composed by a set of links (l_j) and intersections (n_k): $\mathcal{L}_i = \{\{l_{i,1}, \dots, l_{i,j}\}, \{n_{i,1}, \dots, n_{i,k}\}\}$. We focus on path marginal costs (PMC), betweenness centrality (BC) of intersections. PMC is computed from SO simulation, while the BC can directly obtained from network topological features.

We use the solution algorithm proposed by (Peeta and Mahmassani, 1995) to solve the SO problem. Instead of minimizing path travel time, we minimize the path marginal costs in SO problem. The path marginal costs are computed based on time-dependent link marginal costs (LMC). The latter can be obtained from microscopic simulator SYMUVIA, developped by LICIT laboratory. The time step in the numerical simulation is $\Delta t = 60\,\mathrm{s}$. The total number of time steps is T. LMC of l_j at time t is denoted as $c_{j,t}$. The PMC of $\mathcal{L}_i$ is denoted as $\mathcal{C}_i$. It is obtained by summing up all the time-dependent LMC on the trajectory, i.e., $\mathcal{C}_i = \sum_t \sum_j c_{j,t}\delta_{j,t}$, where $\delta_{j,t}$ is the incidence indicator. $\delta_{j,t}$ equals to 1 if user i enters link j at time t, and equals to 0 otherwise. Figure 4 shows the distribution of PMC of trajectories from UE and SO reference simulations. It can be observed that the PMCs decrease from UE simulation to SO simulation.

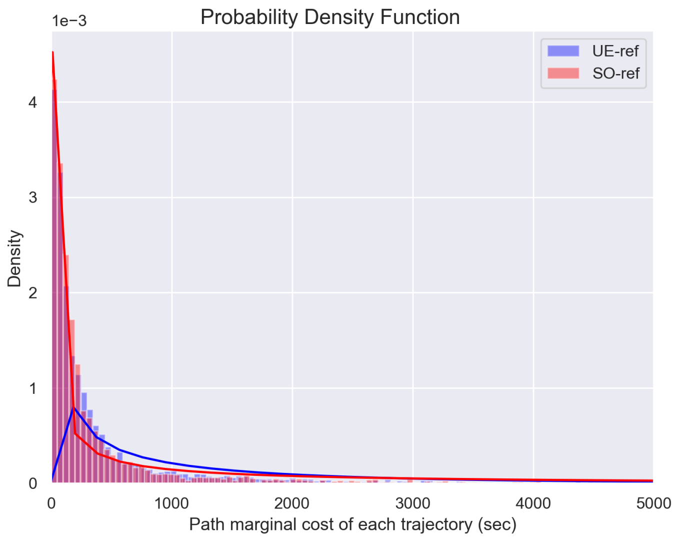

Figure 4: Distribution of path marginal costs (sec) of trajectories from UE-ref and SO-ref simulation.

The BC of a node n corresponds to the ratio of shortest paths crossing n over all possible shortest paths for all origin-destination pairs of the network (Freeman, 1977, Girvan and Newman, 2002). A graph $\mathcal{G} = (\mathcal{N}, \mathcal{A})$ has K nodes and J links. $\mathcal{N} = \{n_1, n_2, \ldots, n_i, \ldots, n_N\}$ is the set of nodes and $\mathcal{A} = \{a_1, \ldots, a_k, \ldots, a_K\}$ is the set of links with $a_{ij} \neq a_{ji}$. The BC of node n is calculated by

$$BC(n) = \sum_{i \neq j} \frac{\sigma_{ij}(n)}{\sigma_{ij}}, \tag{1}$$

where $\sigma_{ij}(n)$ is the number of shortest paths from node i to node j crossing node n, and σ_{ij} is the total number of shortest paths from i to j. In our case study, the *shortest paths* for calculating BC are measured by distance defined directly based on th(e topological parameters of the network. Therefore, for the trajectory $\mathcal{L}_i$ with K_i nodes, we have a vector of node BC denoted as $BC_{\mathcal{L}_i} = \{BC_{(n_{i,1})}, \ldots, BC_{(n_{i,K_i})}\}$ and obtain several statistical values such as it mean, median and standard deviation, etc. Figure 5 shows the distribution of mean node BC of trajectories from UE and SO reference simulations. It can be observed that the mean values of node BC in SO simulation are smaller than that in the UE simulation.

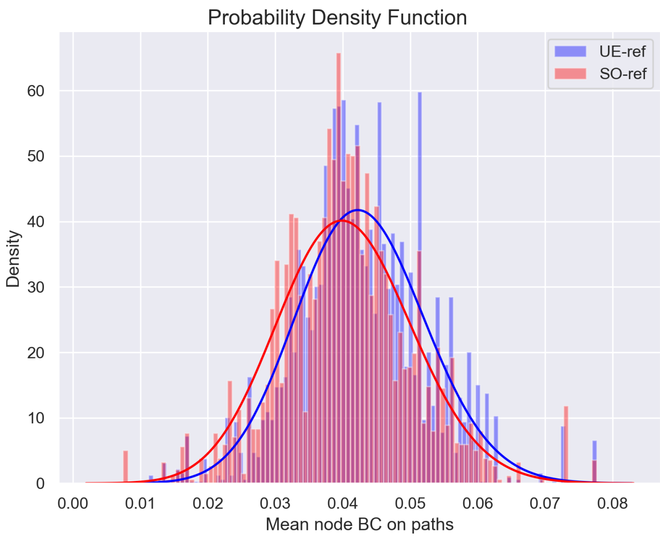

Figure 5: Distribution of mean node BCs of trajectories from UE-ref and SO-ref simulation.

Laval and Castrillón (2015) present that the mean capacity of a corridor can be determined by three dimensionless values: (i) mean red time to mean green time ratio : $\rho = \frac{\mu_{red}}{\mu_{green}}$, (ii) mean block length to mean green time ratio: $\lambda = \frac{\mu_l}{\mu_g reen}$ and (iii) the coefficient of variance of green light time, red light time, and block length. For the dimensionless, the traffic flow is in unit of saturation capacity (Q), the density in units of jam density (κ). The mean capacity of a corridor can be approximated by

$$Cap_{avg} = min\{\frac{1}{1 + \rho(0.58\delta\lambda + 1.64\lambda^2 - 5.3\lambda + 4.99)}; \frac{\mu_{greed}}{\mu_{cycle}} \times Q\}. \quad (2)$$

Figure 6 shows the distribution of mean node BC of trajectories from UE and SO reference simulations. It can be observed that the mean MFD capacities in SO simulation are smaller than that in the UE simulation.

4.3 Results

Descriptive analysis shows that trajectories with high path marginal costs (PMC) are among the first trajectories to be targeted. In addition, by analyzing different trajectories from UE and SO simulations, there are many trajectories with large difference of mean BC and mean MFD capacity from the two reference simulations. In order to test which features are the most influential ones in terms of improving network performance, several scenarios are carried out with following groups of target users. The O-D matrix and network are the same as those in reference UE simulation (UE-ref) and SO simulations (SO-ref). The cumulated traffic demand is 9005 vehicles during 4 hours. The targeting strategies are:

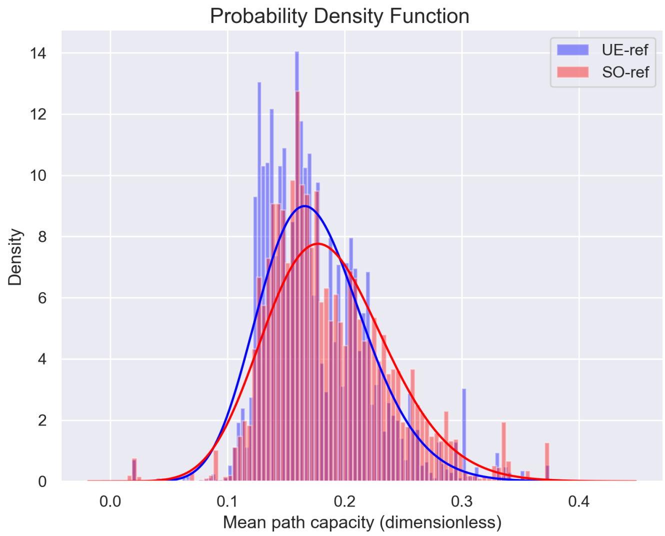

Figure 6: Distribution of mean MFD capacity of trajectories from UE-ref and SO-ref simulation.

(i) PMC-based targeting: 10 % trajectories with biggest PMC reduction from UE-ref to SO-ref. The total number of target trajectories is 900.

(ii) BC-based targeting: 10 % trajectories with the largest mean node BCs in the reference UE simulation.

(iii) DiffBC-based targeting: 10 % trajectories with the largest *reduction* of mean node BCs in from UE-ref to SO-ref simulation.

(iv) Cap-based targeting: 10 % trajectories with the largest mean corridor MFD capacity in the reference UE simulation.

(v) DiffCap-based targeting: 10 % trajectories with the largest *reduction* mean corridor MFD capacity from SO-ref to UE-ref simulation.

The results of the above simulation scenarios are presented in Table 1. In the reference DTA simulations, the TTT reduces 5.0510^5 seconds in the reference SO simulation, compared to the TTT in UE simulation. Results of the above scenarios show that if we change trajectories of 10 % users by targeting trajectories with big difference of node BC (scenario (iii)) or with big node BC (scenario (ii)), the TTT reduction reaches 71 % and 28 % of the TTT reduction in the reference cases, respectively. The relative TTT reduction is computed by Equation 3. We have also carried out 6 scenarios with random targeting strategy: to randomly choose 10 % of the users and give them predefined *optimal* routes. Among these, there are two scenarios that result in 49 % and 43 % of relative TTT reduction. There are two others bring about 10 % of relative TTT reduction. There is one scenario where we get approximatly the same TTT as in UE-ref simulation, and there is another one where we get larger TTT when compared to the UE-ref simulation.

$$\Delta TTT_{relative} = 100\,\% \times \frac{(TTT_{UE-ref} - TTT_{UE-predefine})}{TTT_{UE-ref} - TTT_{SO-ref}}. \tag{3}$$

Table 1: Simulation results of UE-reference and SO- reference senarios of Lyon6 network. (*TTT: total travel time* (s)*. The relative TTT reduction is computed by Equation 3.)*

Statistics	UE-ref	SO-ref	PMC-based	BC-based	DiffBC-based	Cap-based	DiffCap-based
Unfinished trips	91	91	93	88	90	94	87
TTT (s)	3.62×10^6	3.12×10^6	3.56×10^6	3.48×10^6	3.27×10^6	3.58×10^6	3.50×10^6
ΔTTT	–	5.05×10^5	0.68×10^5	1.41×10^5	3.59×10^5	0.46×10^6	1.23×10^6
$\Delta TTT_{\text{relative}}$	–	–	13 %	28 %	71 %	9 %	24 %
TT per user (min)	7.25	6.24	6.58	6.45	6.04	6.62	6.48

These results show that the node BC is one of the network-related trajectory features that the most contribute to the network under-performance. Although the descriptive analysis shows that the mean corridor MFD capacity has significantly changed from UE simulation to SO simulation, the case study results indicate that by targeting the largest improvement of mean MFD might not be the *best* strategy. Ongoing works are being carried out to identify other network-related trajectory features, for example, traffic signal characteristics. The objective is to define the best combination of identified network-related trajectory features, in order to define the best re-routing alternatives to make the system to its optimum.

Keywords: trajectory data analysis; network-related trajectory features; system optimum (SO); supervised learning

Acknowledgment

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 646592 – MAGnUM project).

References

Martin Beckmann, Charles B McGuire, and Christopher B Winsten. Studies in the economics of transportation. Technical report, 1956.

Asa Ben-Hur, David Horn, Hava T Siegelmann, and Vladimir Vapnik. Support vector clustering. *Journal of machine learning research*, 2(Dec):125–137, 2001.

Serdar Çolak, Antonio Lima, and Marta C González. Understanding congested travel in urban areas. *Nature communications*, 7:10793, 2016.

Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the national academy of sciences*, 99(12):7821–7826, 2002.

Marta C Gonzalez, Cesar A Hidalgo, and Albert-Laszlo Barabasi. Understanding individual human mobility patterns. *nature*, 453(7196):779, 2008.

Jiwon Kim and Hani S Mahmassani. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Procedia*, 9:164–184, 2015.

Jorge A Laval and Felipe Castrillón. Stochastic approximations for the macroscopic fundamental diagram of urban networks. *Transportation Research Procedia*, 7:615–630, 2015.

Clélia Lopez, Ludovic Leclercq, Panchamy Krishnakumari, Nicolas Chiabaut, and Hans Lint. Revealing the day-to-day regularity of urban congestion patterns with 3d speed maps. *Scientific Reports*, 7(1):14029, 2017.

Hani S Mahmassani and Srinivas Peeta. *Network performance under system optimal and user equilibrium dynamic assignments: implications for ATIS*. Transportation Research Board, 1993.

Srinivas Peeta and Hani S Mahmassani. System optimal and user equilibrium time-dependent traffic assignment in congested networks. *Annals of Operations Research*, 60(1):81–113, 1995.

Mohammadreza Saeedmanesh and Nikolas Geroliminis. Clustering of heterogeneous networks with directional flows based on "snake" similarities. *Transportation Research Part B: Methodological*, 91:250–269, 2016.

Mt J Smith. The existence, uniqueness and stability of traffic equilibria. *Transportation Research Part B: Methodological*, 13(4):295–304, 1979.

Martin Treiber and Arne Kesting. Trajectory and floating-car data. In *Traffic Flow Dynamics*, pages 7–12. Springer, 2013.

Pu Wang, Timothy Hunter, Alexandre M Bayen, Katja Schechtner, and Marta C González. Understanding road usage patterns in urban areas. *Scientific reports*, 2:1001, 2012.

John Glen Wardrop. Road paper. some theoretical aspects of road traffic research. *Proceedings of the institution of civil engineers*, 1(3):325–362, 1952.