Use of localization technologies and traffic prediction for the transportation assignment problem

Fabio Pompei, *GL Group, Rome, Italy*
Romeo Giuliano, *Università degli Studi Guglielmo Marconi, Rome, Italy*

ABSTRACT. This study develops an optimization model for bus transit network based on road network and zonal ODM (Origin Destination Matrix). The source information has been obtained via recorded phone data of transit users collected by an Italian telecommunication operator. The model aims at achieving minimum bus transfers and maximum passenger flow per unit length with line length and non-linear rate as constraints. An operations research algorithm is used to solve the problem (ACS). To further optimize results, the model uses Google APIs for traffic optimization. The model is tested with survey data of Rome city. The results show that an optimized bus network with less transfers and travel time can be obtained, and the application of operations research algorithm effectively increases the calculation speed and quality.

KEYWORDS: *Bus network optimization, Direct-through passenger flow density, Operations research*

Introduction

Passenger transport in large- and medium-sized cities mainly relies on transit system. The rationale of the transit network planning, therefore, directly influences the travel time and transfer rate of the passengers, and the overall transport system costs.
An ideal transit network, which is featured by large service area, high direct-through trips, small non-linear rate, short travel time, and high accessibility, should be able to meet the needs of the majority of people. However, as urban layout and population distribution change, the service level of the transit network may be gradually reduced, which has adverse impact on the development of public transportation and

the benefits of transit enterprises. To solve this problem, the existing transit network must be adjusted. One of the most adoptable techniques is to artificially change partial routes regardless of the transit network as a whole. In addition, such manual adjustment is largely dependent on the practical experience of the designer(s). Basically, there are two types of models for urban transit network design. One is by combining the transit route design and the departure frequency. The other is solely concentrating on the transit network optimization, based on which the departure frequency of each transit line is studied. Both network design and frequency planning are vital to the transit operating cost and passenger travel convenience. However, in comparison with the departure frequency, the route network is much more stable and less liable to external influences, and does not easily get changed once it is established. It is in this sense that the transport network design calls for the utmost circumspection. The departure frequency, on the other hand, is highly sensitive to factors such as passenger flow, weather and road conditions, and therefore needs to be adjusted in accordance with the different situations. Therefore, the quality of the network design may be adversely influenced if transit network and departure frequency are simultaneously optimized, when the network design determines the efficiency of the entire transit system.
The apparent ignorance of the departure frequency here does not equal to neglect the benefits of the transit enterprises. The utmost goal of transit network design is to facilitate passenger trips and to reduce the operating costs by some constraints; in addition, the unstableness of the departure frequency can lead to some uncertain factors in the optimization process. Therefore, transit network design is prioritized in our research.
To sum up, the solutions to the transit network optimization can be divided into five categories: analytic method, heuristic arithmetic, hybrid algorithm, experience-based arithmetic (Dashora et al, 1998; Fernandez, 1993), and simulation model (Senevirante, 1990).
Despite the prolific optimization models and solutions in previous researches, one problem lies in that most of them are theory-oriented and not practically implementable. Bearing this context in mind, we focus on the integrated transit network and on passenger demands. Then, the objective is to facilitate passengers' trips as well as to foster the transit enterprises' profits. In this paper, an optimization model is developed, which aims to maximize the direct-through passenger density. A vacant network is first established, followed by adding routes to the network according to the principle of maximizing the direct-through passenger density, until all passengers are distributed to the network or some given constraints are overrun. This method differs from previous researches (Simonis, 198; Michael et al. 1997; Wang, 2001) in that most of the previous researches. In fact, literature models are mainly based on the identification of the shortest path between the origin station and the terminal, and then the search of the route bearing the largest through passenger density among these shortest routes. The proposed model is not limited to the shortest paths between the origin station and the terminal, but seeks the through passenger density maximization path in all possible routes. This can be explained by two reasons. First, the passenger flow is not always the largest on the shortest path, which means it is not reasonable to lay all routes along the shortest path. Although this can simplify the models and reduce the calculations, it unfortunately damages the quality of the design. On the other hand, when the objective is set to maximize the through passenger flow, due to the increase of the passenger flow in accordance with the increase of the length of the transit line, a certain route may be abandoned because it is comparatively shorter, even if this short route abounds in passenger flow. This can result in a deviation of the route from the needs of the passenger flow. In addition, the overall length of the entire network increases where longer line is laid, thus consequently increasing the operating costs without full utilization of the network or fleet. Therefore, this paper employs direct-through passenger flow density maximization as the optimization objective, so as to enhance the network utilization rate as high as possible. The densest routes are laid first to facilitate passenger trips and to benefit the transport enterprises (Yu Bin, et al. 2005), The use of an optimization algorithm (maximum flow minimum cost) to a real case implies going beyond the logical steps of the computer language translation. The data flow (ODM matrices) must be elaborated in order to provide appropriate information to the software. The steps of the algorithm may not accurately correspond the general theoretical formulation but must be reviewed and changed, in order to obtain the optimization of the parameters concerning the study.
The paper is organized as follows. In Section II similar studies presented in literature are described. In Section III Materials and methods, Data sources and Sub-division of the analysis area are described. In Section IV the application of the optimization algorithm, the use of Google APIs, the simulation process and conclusions.

Section II

Similar studies

In this Section we provide an overview of key implementation steps, starting from data arrangement and following with the saving of information with specific focus on the core of the algorithm (Pompei, 2017). The study carried out and the operating contribution, also provided through the realization of the LPT (Local Public Transport) geo-referenced Analysis platform, as well as an appropriate data base, offer a useful support for the development of further investigations and analysis and for the planning of the LPT circuits optimization strategies, as part of the local public transport reorganization plans.
In 2014, an Italian company (ACI Informatica), conducted a similar experiment for the Inmoto (INformation and MObility for TOurism) R&D Project, funded by the Ministry of Education (MIUR) as part of the Research and Competitiveness 2007-2013 National Operative Plan (PON), dedicated to Smart Cities & Communities. The need was to figure out tourist flows and movements in the district of Calabria, in order to improve the tourist service and optimize motor traffic. In cooperation with Vodafone, ACI Informatica tried to develop, as a trial, a system prototype based on innovative and specific big data algorithms, in order to collect and process anonymized, encrypted and aggregated location data, coming from mobile network. The findings, which have been shown through an information dashboard, supported institutions in better planning of tourism infrastructure, based on the real number of users in certain areas. ACI Informatica and the University of Calabria based on those information tried to discriminate tourists from residents and detect aggregated properties of their movements. The amount of data was enormous and required specific Big Data algorithms. Processing the collected data turned out information about the origin of tourist flows, the typical movements on the territory, the daily average presence by area and the most common routes (Pompei, 2017).

Rome transport condition

Before proceeding with the description of the technological solution, it is worthwhile illustrating the transport situation in Rome. Termini Station is the largest railway hub of the metropolitan city of Rome, from which both regional and national main lines branch of. Both high-speed train to Turin, Milan, Florence and Naples, and the railway to Leonardo Da Vinci Airport in Fiumicino and direct regional lines to Ancona, Genoa and Pescara pass through Termini station.

Other important Rome railway yards are “Roma Tiburtina” and “Ostiense” railway stations, from where the main lines of Lazio regional railway as well as the train to “Lido di Ostia” pass through. Lines to Pisa, Viterbo and the Vatican City originate from Roma San Pietro station, not far from St. Peter’s Basilica in the Vatican City. Rome’s metro system is an underground network serving the city, made up of three lines.
The Roman transportation company owns about 3k buses, about 170 trams, about 100 metro trains.Furthermore, there is private traffic. The resident population in the municipality is about 2.5 M, territory is 1,308 square kilometers and car park is about 2 M Automobiles and Motorcycles over 400 k. There are remarkable problems on Rome transportation, which is characterized by surface transit heavily influenced by traffic, bad weather and maintenance works, as inferred from the explained data. We should also highlight that road network in Rome (ancient city) is strongly characterized by narrow streets and double-parked cars, which create traffic jam and bottlenecks that slow down the transport system (Pompei, 2016).

Section III

In this section we introduce the materials and methods and the data sources used for the model.

Materials and methods

Before carrying out the data analysis, information from different sources with different formats had to be collected in a database. The information are related to:

- Identification of areas to be analyzed
- Demand data such as movements among identified areas
- Supply data such as transit data stops and routes

Data sources:

Telecommunications Operator for matrices Origin / Destination relating to demand

- Mobility Agency Site of Rome for public transit data

The matrices acquisition has been carried out through an automatic system for the calculation of Matrices Origin/Destination concerning movements of people within geographical areas. Movement tracking has been possible through automatic and pseudo-anonymous processing of passive location data, collected by another system, which locates all users of that Operator, including roaming customers, by the time they interact with the network. More specifically, user location is detected by events such as switching on and off the terminal, location update procedure (periodically and aperiodically), return in the service coverage area, switch of the technology from 2G to 3G or 4G, incoming/outgoing call or instant messages.
For each day and for each area, MOD (Origin / Destination Matrix) system produces 24 files, one per hour. The collected data refer to a whole week and relate to about 2.5 M individual users.
The Origin / Destination Matrix has as many rows and columns as areas; each element of the matrix indicates the number of trips originated from the area represented on the row and destined to the area represented on the column. On the matrix diagonal there are the movements occurring within the same area.

Public transport dataset

The Mobility Agency of Rome offers the public transport dataset, in GTFS (General Transit Feed Specification) format, on Internet.

Sub-division of the analysis area

Once the area has been geographically defined, it is necessary to analyze the occurring movements. Therefore, the area has to be divided into sub-zones among which movements occur. In our case, the area is bounded by a rectangle whose lower right vertex has geographic coordinates latitude lat=12.3679 and longitude lon=41.7909, while upper right vertex has lat=12.6233 and lon=41.9957, approximately including the GRA (Ring Road around Rome). This area has been divided into 55 x 55 sub-areas with the same surface and 500 m side.

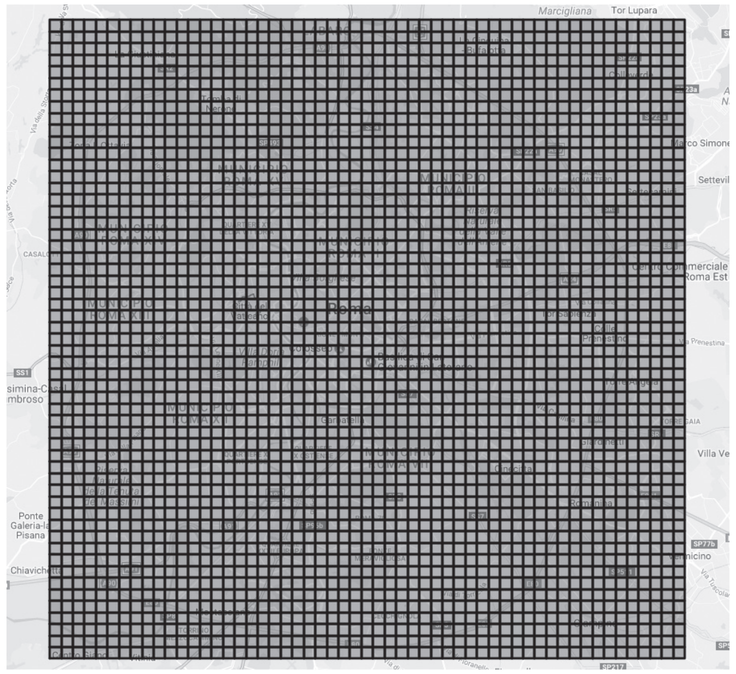

Figure 1. Rome, 55x55 zones

Section IV

Application of the optimization algorithm, Google APIs and simulation

We detail the procedure to optimize the transport network in Rome. First we applied the optimization algorithm. For this purpose, we selected the Ant Colony System based on maximum flow at minimum cost. The aim is to collect the maximum number of passengers by slightly varying the transport mean route (Yu Bin, et al. 2005). After that, we are able to simulate the new transport line based on the departure time according to mobility agency of the city and on the MOD matrix with data collected through the cellular system. Finally, data are further integrated with the information of user traffic based on the Google API in order to define the better routes in relation with the traffic condition.

In particular, the Google Maps Distance Matrix API is a service that provides travel distance and time for a matrix of origins and destinations, based on the recommended route between start and end points.
The Google Maps Directions API is a service that calculates directions between locations. They allow to search for directions for several modes of transportation, including transit, driving, walking, or cycling.
The algorithm of Google API is coded as follows:

```
Start
Loop /* zone */
        Loop /* neighbour zones */
                Call Google API for the path (time) WITHOUT traffic
                Loop /* time X */
                        Analyze the behaviour at time X
                        Call Google API for the path (time) WITH traffic at time X
                        Log the time difference between WITH and WITHOUT traffic
                        time = time +1 hour
                Until time>24 hours
                Zone = zone + 1 (*next zone*)
        Until all neighbour zones are analized
Until all matrix zone are analyzed
```

In the following figure it is reported the flow chart of the algorithm.

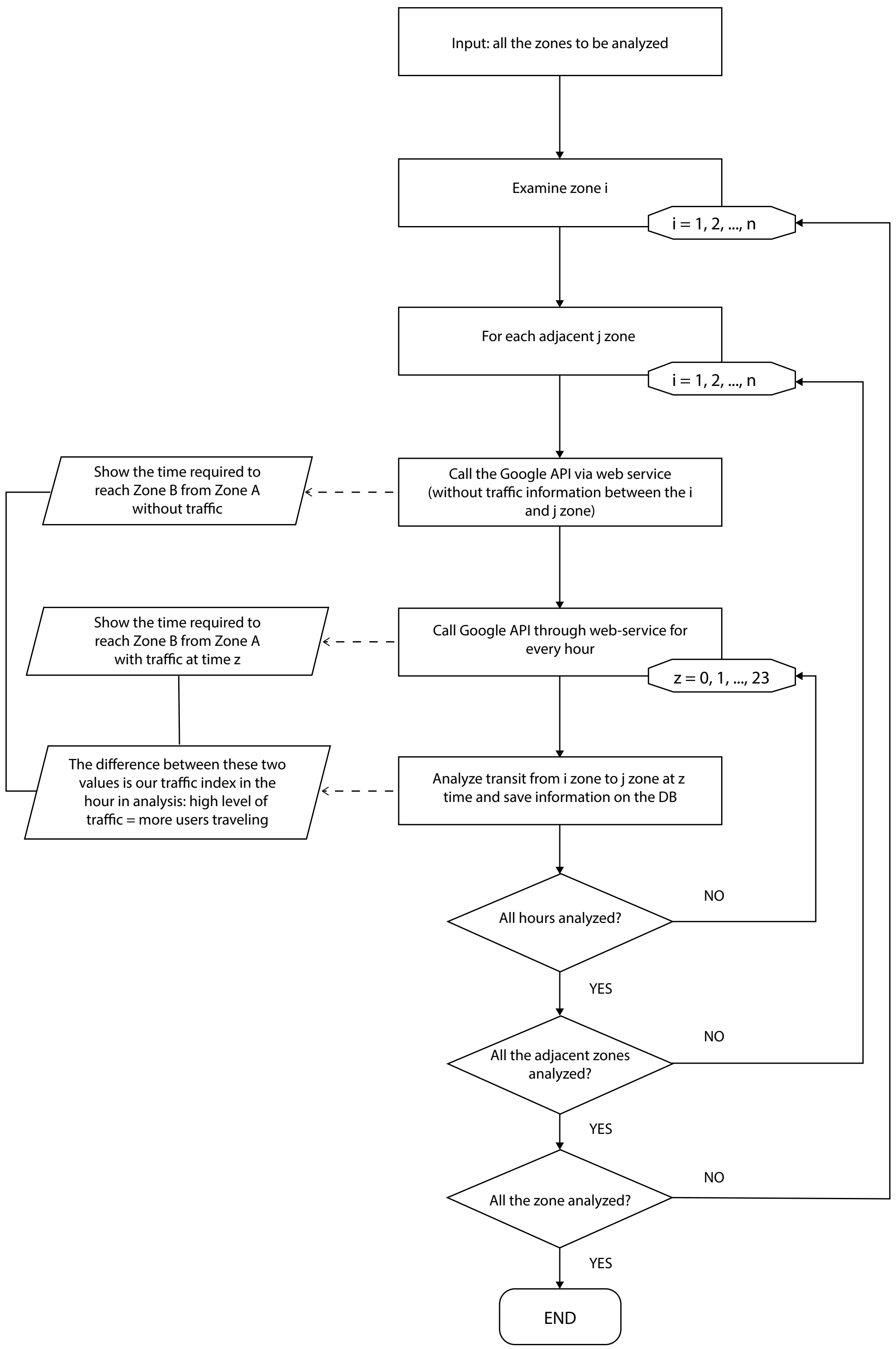

Figure 2. Flow graph of the Google API algorithm

An improvement of the algorithm

A further improvement (an example with only 3 selected lines) in the optimization process can be performed by analyzing the system behavior after deleting one or more transport lines due to malfunctions of transport means or drivers unavailability on the low loaded lines.

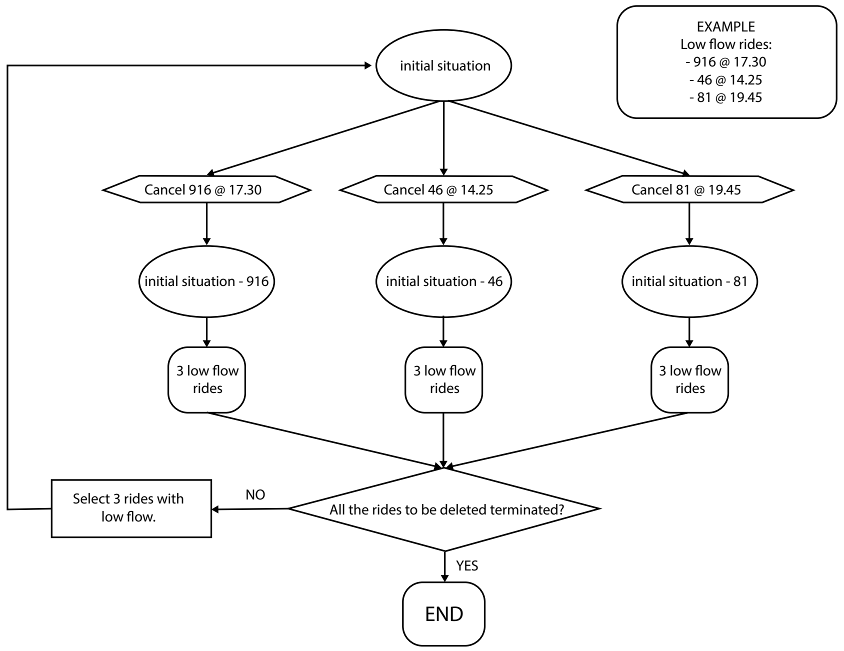

Figure 3. Flow graph to cancel a bus ride

In the flow graph we start from the simulations described above, we examine the n bus travel that have certain performance parameters (in the example there are 3 bus travel with less potential users). The number of bus travel to be analyzed n, in our case 3, is also configurable. The main cycle, analyzes the 3 most unsuccessful bus travel (those that may be eligible to be eliminated), the cycle completes 3 new simulations in each of which one of the 3 candidate lines is deleted, at the end of each simulation they are identified again the 3 new bus travel with lower turnout. The triplet of new bus travel with a smaller total affluence determines the branch of the tree that will be further refined by a new iteration. In each iteration a branch of the tree is expanded with the 3 slower runs until the target reaches (eliminating the required X rides). The application of this part of the algorithm (rides cancellation) is useful to the manager but does not affect the results that will be explained in the next section. It's an optimization in order to manage particular situations like strikes or drivers unavailability.

Results

The routes proposed by this model are longer than the original ones. We expected it because of the feature of our study. We did not want an algorithm for searching shortest path but something asymptotically approaching maximum flow – minimum cost benchmark. Therefore, it was foreseeable that the maximum flow was not achieved with the basically linear routes as the existing ones. Another feature we expected from new paths was that they were limited in length, being ants leaded by the virtual pheromone, which tends to favor the shortest routes, without considering the fact that ants are virtually "killed" in the event of lost. In order to check the length of the lines we carried out a comparison between ACS path and original ones. The examination of average lengths confirms this slight tendency to ACS lines to be longer than the original ones with the average number of nodes 26 versus 20 of the original routes. Considering 0.4 km the size of the areas and the chessboard movements, we conclude that the ACS paths measure on average about 11 km versus 8.0 km of the original routes. The increase is slightly less than 30%. We will now analyze the output of the simulation considering the same original timetable but with lines redesigned through ACS. The comparison of simulated and the original data are shown in the table below. We see the time slot division and the absolute and percentage differences. The data related to the requests are the same while those relating to passengers highlight that the new tracks are more effective. The original lines carry 22.7% of the potential demand, while the redesigned one 34.1%, achieving a difference of 2 million passengers per day, which corresponds to 150.28%.

H	Request	Final	Transp	Difference
1	239.376	162.451	76.925	28.326
2	232.449	186.966	45.483	21.492
3	155.415	128.377	27.038	12.918
4	234.360	198.495	35.865	16.133
5	171.192	141.096	30.096	14.242
6	211.821	136.712	75.109	31.112
7	393.024	253.172	139.852	51.452
8	886.587	591.382	295.205	91.819
9	1.225.332	819.294	406.038	127.565
10	1.233.795	804.145	429.650	138.080
11	1.193.826	769.734	424.092	137.917
12	1.220.847	785.766	435.081	146.145
13	1.233.009	795.142	437.867	144.889
14	1.299.210	849.153	450.057	152.121
15	1.196.361	782.431	413.930	137.722
16	1.188.357	780.062	408.295	134.077
17	1.160.328	759.879	400.449	129.935
18	1.151.448	759.849	391.599	126.351
19	1.083.366	709.902	373.464	122.493
20	923.106	600.121	322.985	108.179
21	707.700	451.577	256.123	86.683
22	490.767	309.050	181.717	63.454
23	346.488	216.466	130.022	46.497
24	121.320	75.365	45.955	15.628
	18.299.484	12.066.587	6.232.897	2.085.230

Yearly req.	6.679.311.660
Yearly trans.	1.751.755.702
Transported %	34,1%

Table 1 – Original lines data vs ANT

We can already say that an effective 11% of length increasing has been fully rewarded by an increase of more than 50% of passengers. It is interesting to underline that the maximum increments (close to 200%) have occurred precisely in hours with lower efficiency by the original lines.

Conclusions

The goal of this paperwork is to illustrate the theoretical basis, implementations and optimization results of Rome public transport network. This article aims to change the current paths of transport lines in order to increase the flow of passengers without rising travelled kilometers.
As in all the network improvement processes is important to decide in which aspects we wish to boost the performance, on the one hand because it affects the methodology to be applied, on the other because the obtained results shall be assessed with appropriate metrics. Parameters such as transit times, location of bus stops, types of means, maintenance periods are just some of the variables that may be subject to optimization.
The algorithm used is a maximum flow – minimum cost.
The source data, except the OD matrices from the telecommunication world, came from the database that the public agency "Rome's Mobility Services" provides on its website and Google Maps.
Google Maps bases its traffic views and faster-route recommendations on two different kinds of information: historical data about the average time it takes to travel a particular section of road at specific times on specific days and real-time data sent by sensors and smartphones that report how fast cars are moving right then.
The assessment of the study was carried out through a 24-hour travel simulation, both along the original paths and the route optimized with the algorithm.
At the simulation level, the results show that the lines calculated with the algorithm allow the increasing of passengers transported by more than 50% at the same frequency of travel, restricting the increase in length of the paths and obtaining a passenger / km ratio higher than one of current lines. Furthermore, the vehicles running through optimized paths have obtained a greater average filling ensuring a number of passengers served at each step higher than 20% compared to current lines.
Other applications are tourist flow optimization, less-known site promotion and local economy improvement. Through this method, we can get the position of tourists by their mobile terminal, localizing each smartphone in an area of the city. Furthermore, we can associate the tourist flow to the position of the monuments and other important attractions or events. In this way, local administrations can better face high-density tourist areas and optimize tourism facilities.

References

Longo Michela, Hossain Chowdhury Akram, Roscia Mariacristina (2013), *Smart Mobility for Green University Campus*, IEEE

Mirri Silvia, Prandi Catia, Salomoni Paola, Callegati Franco, Melis Andrea, Prandini Marco (2016), *CrowdSensing for Smart Mobility through a Service-Oriented Architecture*, "Hindawi" https://www.hindawi.com/journals/misy/2016/2821680/ref/

Pompei Fabio (2016), *Smart Mobility in Rome*, CESC

Pompei Fabio (2017), *Ant Colony Optimisation and geolocation technologies for the transportation assignment problem - Use of ACO and geolocation technologies for performance improvement and test of Local Public Transport*, IEEE Computer Society Signature Conference on Computers, Software and Applications

Pompei Fabio (2017), *Geolocation for LPT - Use of geolocation technologies for performance improvement and test of Local Public Transport,* EEEIC 2017 – IEEE 17th annual conference of the International Conference on Environmental and Electrical Engineering

Pompei Fabio, Caruso M.C., Giuliano Romeo, Mazzenga Fabio (2017), *Mobility management for Smart Sightseeing*, IEEE Automotive Conference of Electrical and Electronic Technologies for Automotive

Tokody Dániel, Schuster György, Holicza Peter (2016), *The smart mobility aspects of Intelligent RailWay,* 11th IEEE International Symposium on Applied Computational Intelligence and Informatics

Yu Bin, Zhongzhen Yang, Chuntian Cheng, Chong Liu (2005), *Optimizing bus transit network with parallel ant colony algorithm*, Proceedings of the Eastern Asia Society for Transportation Studies, V. 5, pp. 374 - 389
http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.538.7440&rep=rep1&type=pdf

Sintesi

Lo sviluppo di un modello di ottimizzazione per la rete di transito degli autobus basata sulla rete stradale e sull'ODM zonale (Origin Destination Matrix) ha come obiettivo il miglioramento massimo di un servizio pubblico. Le informazioni sulla fonte sono state ottenute tramite i dati telefonici registrati degli utenti in transito, raccolti da un operatore italiano di telecomunicazioni. Il modello utilizza le API di Google ed è testato con i dati del sondaggio della città di Roma. I risultati mostrano che è possibile ottenere una rete di bus migliorativa con meno trasferimenti e tempo di viaggio.