NETWORKS AND HETEROGENEOUS MEDIA doi:10.3934/nhm.2015.10.127
©American Institute of Mathematical Sciences
Volume **10**, Number **1**, March **2015** pp. **127–148**

STRUCTURAL ANALYSIS AND TRAFFIC FLOW IN THE TRANSPORT NETWORKS OF MADRID

Mary Luz Mouronte [1,2]

[1] Grupo de Sistemas Complejos. Universidad Politécnica de Madrid
[2] Escuela Técnica Superior de Ingeniería y Sistemas de Telecomunicación
Carretera de Valencia km. 7, 28031 Madrid, Spain

Rosa María Benito [1,3]

[1] Grupo de Sistemas Complejos. Universidad Politécnica de Madrid
[3] Escuela Técnica Superior de Ingenieros Agrónomos
Avenida Complutense 3, 28040 Madrid, Spain

Abstract. As the framework to characterize the subway and urban bus networks of Madrid city three topological spaces: geographical stop space, transfer space and route space, are considered. We show that the subway network exhibits better structural parameters than the urban bus network, with higher performance since in average a stop is reachable passing through less number of stops and carrying out less number of transfers between lines. We have found that the cumulative degree distributions of the subway and urban bus networks correspond to an exponential function, while the degree-degree correlations present a power law distributions in both transport systems. The relationship between transport flows and population are also studied at the city level by analyzing the flow between all the district (administrative areas) of Madrid. We prove that these flows can be described by a Gravity Model which takes into account the population from the origin and destination districts as well as the number of sections of a transport line that passes through two different districts.

1. **Introduction.** Urban transport systems (TS) are responsible for the mobility of many passengers in a city. This system is also responsible of negative traffic impact both in the citizens and the environment. The improvement of the energy efficiency of the TS can be achieved by improving the urban public transport network. A city public TS that is appealing, is a requirement so that people be able to reduce the usage of their own vehicles. As public TS become more complex, an analysis of their network features can be of substantial help for planners since an increased knowledge of the features of a transport network (TN) is required in order to identify possible actions to be undertaken within the network.

The goal of this study is to characterize the urban bus and subway networks of Madrid by applying tools of network science [1]. It allows to compare both TN and to identify some common features.

The main public TS in Madrid is the urban bus and subway networks, which are highly developed. Madrid is the third-most populated municipality in the European Union after London and Berlin, and has a population of $3,254,950$ in an area of

2010 *Mathematics Subject Classification.* 00A99.
Key words and phrases. Network science, topological analysis, traffic flow, subway network, urban bus network.

$60,683$ hectares. Madrid is also subdivided into 21 districts, administrative regions that distribute and manage the exercise of civil or political rights, public functions, and services.

Madrid subway network consists of 16 lines and 322 stops. The first line began operations in 1919 and had 8 stops along 3.5 km, now, this network has a length of 283 km and it is the sixth largest subway network in the world after London, New York, Moscow, Seoul and Shanghai. The network transported 635 million passengers in 2011. The Madrid urban bus network consists of 230 lines with $4,635$ stops and its current length is $4,074$ km. Around 423 million passengers travelled on the network in 2011. This network is a large, reliable and ever-growing network.

This paper uses the network theory [1] to investigate urban bus and subway networks in Madrid in three topological spaces: Space L (Geographical Stop Space), Space P (Transfer Space) and Space R (Route Space) which, allow us to describe both TS from a structural point of view [13]. There are several pieces of research that have studied the structural characteristics of TN in different countries. Y.Z. Chen et al. [5], present empirical results for the urban bus TN of four major cities in China. They analyse the distribution of the number of nodes that are connected by a link when the stops belong to a common line, as well as the distribution of the bus lines that a stop joins. The authors show that both distribution present exponential function forms. B. Jiang [7], derives a topological pattern of urban street networks using a large sample, 40 U.S. cities and a few more from elsewhere of different sizes. It was found that all the topologies of urban street networks based on street-street intersections demonstrated a small-world structure, and a scale-free property for both street length and street intersections. M. Ke et al [8], shown that the urban public bus networks of Beijing and Chengdu in China exhibit small-world behavior and are hierarchically organized. C. von Ferber et al. [21], show some properties of public transport network of 14 large cities, which are analysed in different spacial representations. The studied features are: connectivity degrees, clustering coefficients and path length distributions, betweenness centralities and finally, the harness distribution (number of sequences of consecutive stations that are serviced by a specific number of parallel routes). The authors also formulated a model based on simple growth rules, which supports the statistics properties of the analysed public transport network. The novelty of our invesigation, with respect to the analysis of the structural properties in TN, is that it analyses the urban bus and subway networks of the Madrid metropolis by using the aforementioned three topological spaces. Similar results for the cummulative degree distribution to those shown in the papers [5] for urban bus transport network in chinese cities and in [21] for different countries have been obtained. However, in [5] only urban bus networks are studied and, in [21] all PTN analysed are either operated by a single operator or by a small number of operators with a coordinated schedule (e.g. expressed by a central website), rather than dividing these centrally organized networks into subnetworks of different transport means like bus and subway or in an urban and a sub-urban part the study treats each full PTN as an entity. In contrast, our investigation analyse two PTN operated by different companies that correspond to different transport means. The TN assortativity feature, which has been obtained calculating the correlation between the node degrees of neighboring nodes in terms of the correlation normalized coefficient (Pearson correlation coefficient) [14], and the distribution of the nearest network degree, which exhibits a similar behaviour in both TN, are also estimated. Adittionally, we analyse the efficiency and the redundancy level in these networks.

The function that regulates vehicle movement between two areas of Madrid (districts) and their inhabitants is also calculated. A similar study has been carried out previously for the intercity express bus flow between two korean cities [9], the intercity bus flow was described using a Gravity Model [20] that took into account the population of the origin and destination cities and the distance between them. However our work focus to characterize the urban bus and subway train flows between two districts within the city of Madrid. It analyses a different type of traffic that takes place inside a city and it is analyzed at a smaller geographical scale. In this case, in applying the gravity model we consider the population of origin and destination districts as well as the number of sections of the transport lines that passes through them. The behaviour of the traffic flow is modeled taking into account the total number of vehicles during the five working days of a week. There are several pieces of research that have studied the dynamical characteristics of TN in different countries. S. Ondŏs et al. [15], use the daily schedules for ordinary workdays, weekend days, and school vacation workdays of public TS in Bratislava, to pool them into a representative day, but useful as an analytical generalization. The authors shown that the TS behaves systematically throughout the daily cycle in response to changing demand. The daytime rhythm which the authors call "the urban heartbeat", can be clearly identified in the network structure. H. Soh et al. [17], studied the Singapore rail and bus transportation systems, using topological and dynamical analysis. The review of the weighted eigenvector centrality highlighted an important difference in traffic flows for both networks during weekdays and weekends, suggesting the importance of adding a temporal perspective missing from many previous studies. However, we model the traffic flow of urban buses and subway trains between two districts considering the strength of their relationship, which is determined by the number of line sections that passes through them.

The rest of the paper is organized as follows: Section 2 presents the topological analysis of Madrid subway and urban bus networks, in which several statistical features are investigated. In Section 3 we characterize the vehicle flow between districts by a Gravity Model and study how the changes of topology impact over the relationship between the population of one district and its transport flow. Finally, in Section 4 we end with some conclusions.

2. **Topological analysis.**

2.1. **Building networks.** Although a long distance may not exist between two stops on one TN, the displacement between them can take a significant time due to the fact that a bus line may make many loops or because might be necessary to make several transfers between lines before reaching the destination stop. These facts can be analysed by studying two different representations of a TN: Space L (Geographical Stop Space) and Space P (Transfer Space).

In Space L, each node represents one stop (bus shelter or subway station), and a link between two nodes means that the two stops are consecutive in the same line. The distance between two nodes is measured by the total number of stops (nodes) passed on the shortest path between them. However, distance measured in this way does not take into account the need to make transfers during a trip. This characteristic can be studied by taking into account Space P, where each node represents one stop, and one link joins a pair of stops when one line provides a direct service between them, that is, both stops are linked if they belong to the same bus or subway line. In Space P the shortest path magnitude is the minimum number of lines that is required to be used to reach a destination stop, the distance between nodes represents the number of necessary transfers between lines plus one. Finally, it is interesting to know whether the passengers can directly transfer between two lines without involving additional lines. This feature can be analysed by considering Space R (Route Space). In this space, nodes are defined as lines and common stops determine the links. Here a link means that passengers can directly transfer between two lines.

We represent Madrid urban bus and subway networks in three topological spaces: Space P, Space L and Space R. In these spaces, both TN are mapped in a graph $G' = (E'; V')$, in which E' is the set of nodes and V' is the set of links between them. An adjacency matrix of N x N dimension $A(G')$ can be constructed as a bidimensional representation of the relationships between nodes, where $Aij = 1$ when a connection between nodes i and j exists and $Aij = 0$ in the other case. N is the number of nodes in E'.

An scheme of the construction of the three topological spaces is illustrated in Figure: 1.

2.2. **Degree distributions.** The degree of a node i is the number of nodes directly connected to it:

$$K_i = \Sigma_j A_{ij} \tag{1}$$

The meaning of K_i is different in each topological space. In Space L, K_i is the number of stops (i.e. bus shelter or subway station) with direct connections with stop i, this parameter represents the number of directions one passenger can take from a given stop i. In Space P, K_i is the total number of existing paths, without transfer, from stop i, that is, the number of stops that can be accessed without switching between lines. In Space R, K_i is the number of stops that link line i with other lines.

With respect to the urban bus network of Madrid in Space L, the stop with the highest connectivity is Plaza de Cibeles which has 13 lines, there is 1 stop that has 12 lines, 2 stops which have 2 lines and 5 stops that have 10 lines. It is also detected that most stops have less than 6 lines (96.63%) and only 9 stops have more than 9 lines. With respect to the subway network, the stops that have more lines are Alonso Martínez and Avenida de América which have 7 lines; there are 5 stops which have 6 lines; 97.83% of stops have less than 6 lines and only 7 stops have more than 5 lines.

Taking into account the networks corresponding to Space R, we can get information of the lines with the largest number of connections with other lines. The results are depicted in Table 1 and Table 2 for the urban bus and subway networks respectively. It can be noticed that the lines most connected are lines 10 and 103 for the subway and urban bus networks respectively.

The degree distribution of G', $P(K)$, determines the probability that a randomly chosen node will have degree K. In Figure 2 the cumulative degree distributions $CP(K)$ for both urban TN in the Space L, P and R are shown. They can be in all cases well fitted to an exponential function defined as:

$$CP(K) = b_1 \exp \frac{-K}{\alpha} \tag{2}$$

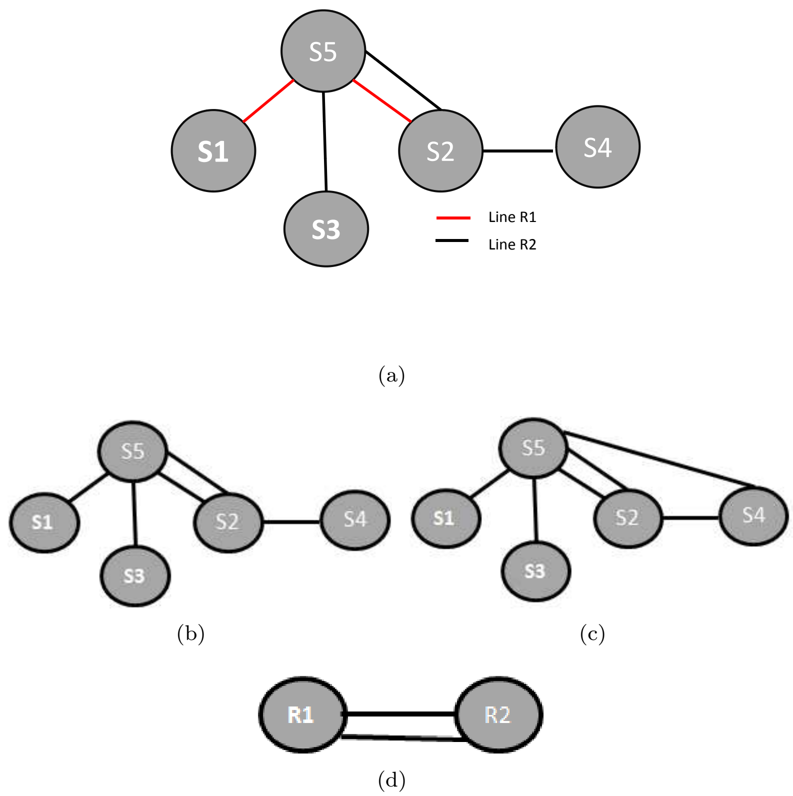

FIGURE 1. (a) An example of **U**rban **T**ransport **N**etwork (UTN) consisting of two lines (R_1 (red) and R_2 (black)) and five stops (S_1, S_2, S_3, S_4 and S_5). R_1: $S_1 \to S_5 \to S_2$, R_2: $S_3 \to S_5 \to S_2 \to S_4$. (b) Network visualisation in Space L. (c) Network visualisation in Space P. (d) Network visualisation in Space R.

TABLE 1. Madrid urban bus network lines with the largest number of connections to other lines (lines with the largest degree and the value of degree in Space R).

Line	Degree
103	150
116	132
150	124
N24	122
132	110
137	104
172L	102
71	102

Table 2. Madrid subway network lines with larger number of connections to other lines (lines with larger degree and the value of degree in Space R).

Line	Degree
10	13
4	10
1	9
2	9
5	9
6	9
7	9
9	9

Using natural logarithms:

$$ln(CP(K)) = \frac{-K}{\alpha} + ln(b_1) \tag{3}$$

multiple linear regression (MLR) is applied to model the relationship between $ln(CP(K))$ and K, by fitting a linear equation to the observed data:

$$ln(CP(K)) = n_1 K + n_2 \tag{4}$$

The values of the fitted parameters (n_1, n_2) together with the determination coefficient (DC) and the mean squared error (MSE) for all the networks considered are given in Table 3.

We can establish that urban bus and subway networks are single scale networks [3], [10], [11], [22] for all spaces. This distribution, which is characterised by a connectivity distribution with a fast decaying tail, decreases exponentially much faster than a power law distribution which characterises scale-free networks. The absence of a power-law distribution in these networks can be explained as result of the limited capacity of the nodes (bus shelter, subway station or line). Physical costs of adding links between nodes also limit the number of possible links attaching to a given node. Both causes shorten a power-law regime, since a single node cannot acquire the number of links necessary to reach a scale free network. Amaral et al. [3] also infered that physical constraints would prevent the formation of scale free networks in air traffic networks.

In the urban bus and subway networks of Madrid exist a small number of nodes with high degree. The urban bus network, for example, has 0.02% of its nodes with the highest degree $K = 13$ and $K = 450$ in Space L and P respectively, and 0.04% of its nodes with the highest degree $K = 150$ in Space R. The subway network, for example, has 0.80% of its nodes with the highest degree $K = 7$, 0.4% of its nodes with the highest connectivity $K = 99$ and 5.88% of its nodes with the highest $K = 13$ in Spaces L, P and R respectively. This distribution implies that these TN can withstand random attacks of several nodes (stops or lines) but are vulnerable in the case of directed attacks towards particular nodes. A random attack on a single or on a very few nodes will, in general, not bring down the network.

Although a directed attack towards a node (bus shelter, subway station or line) with high connectivity can be very relevant [18], [2], [6] and [25]. Nodes with large numbers of links give shelter to one relatively higher number of passengers. In order to the functioning of a node be suitable, its number of passengers (load) must be less than those for which the node was planned (its capacity). If a node has problems (i.e. subway station that has a power failure, subway line halts due to signal failures, bus line suspended due to a demonstration, bus shelter with access problems, etc.) its passengers should be directed to other nodes, causing a redistribution of them in the nodes (e.g. a problem in a shelter implies that the passengers should move to another stop, or, a failure in a line implies that the passengers should switch to another line). If the failing node deals with a small amount of passanger, there will be little effect on the network because the amount of passangers that needs to be redistributed is small. This is the common situation of random failure of nodes. However, if the node with problems contains a large amount of passengers, the consequences could be relevant because this number of passengers would redistribute themselves and it is possible that for some nodes (bus shelter, subway station or lines), the new number of passengers could exceed their capacity. If these nodes also fail, it will cause a further redistribution of passengers, and so on.

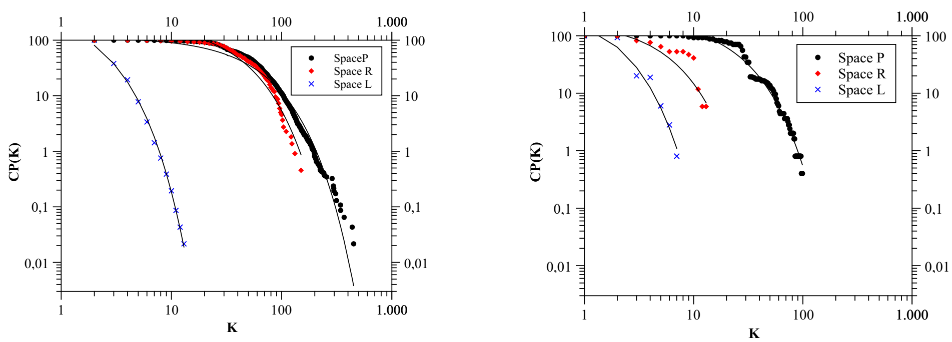

FIGURE 2. Left side: Cumulative degree distribution $CP(K)$ in the Spaces L, R and P for the urban bus network. Right side: $CP(K)$ in the Spaces L, R and P for the subway network.

TABLE 3. Parameters (n_1, n_2) of the fit of $CP(K)$ to an exponential funtion (equation 4), and the correspondent determination coefficient (DC) and mean squared error (MSE) for both TS networks.

	Space	n_1	n_2	DC	MSE
Subway network	L	-0.809950819	5.766582843	0.964370312	0.368414537
	P	-0.059522044	5.268155424	0,979913554	0.246030897
	R	-0.253785351	5.221173662	0.836015539	0,457193529
Urban bus network	L	-0.759935391	5.910276419	0.997937408	0.130647084
	P	-0.022880518	4.727338432	0.968773483	0.337766063
	R	-0.036563518	5.339890129	0.946964537	0.315484262

The mean degree of a network $(< K >)$ is the average of the degrees of all nodes in a network.

$$< K >= \Sigma_i \frac{K_i}{N} \tag{5}$$

TABLE 4. Topological properties of the Madrid subway networks in the three considered spaces.

	M Ns ϱ	16 322 0.84736842	
Space	L	P	R
$<K>$	2.42	29.39	6.00
K_{max}	7	99	13
$<Knn>$	3.17	36.36	7.77
Knn_{Max}	4.1	45.66	10.25
Knn_{Min}	2.23	15.06	1.67
$<l>$	10.19	2.26	1.62
$EGlob$	0.08140987	0.29839786	0.63888889
rD	0.270223	0.092046	−0.154867

TABLE 5. Topological properties of the Madrid urban bus networks in the three considered spaces.

	M Ns ϱ	230 4,635 0.41140757	
Space	L	P	R
$<K>$	2.71	43.86	23.41
K_{max}	13	450	150
$<Knn>$	3.66	98.63	29.78
Knn_{Max}	4.54	182.19	37.06
Knn_{Min}	2.33	55.20	8
$<l>$	18.77	3.20	2.33
$EGlob$	0.064930854	0.33642985	0.48056732
rD	0.199994	0.034606	0.232538

This magnitude is not representative, since the degree distribution is skewed. In this case, a more appropriate magnitude is the Maximum Degree of the network (K_{Max}), which is the highest degree of its nodes.

$<K>$ and K_{Max} for the subway and urban bus networks are shown in Tables 4 and 5 respectively. These networks verify:

$$K_{Max}(L) \leq K_{Max}(R) \leq K_{Max}(P)$$

and

$$<K>(L) \leq <K>(R) \leq <K>(P)$$

These facts are due to the different connectivity of the topological spaces. The average ($<K>$) and the maximum value (K_{Max}) of K are smaller in the subway network.

2.3. **Nearest neighbour degree.** The nearest neighbour degree $Knn(K)$ is defined as:

$$K_{nn}(K) = \Sigma_{K'=0}^{\infty} K' P(K'/K) \tag{6}$$

where $P(K'/K)$ is the conditional probability that a link belonging to a node with degree K links to a node with degree K'. Statistical variations in K_{nn} can be supressed by obtaining the corresponding cumulative distribution, CK_{nn}.

In both networks $CKnn$ can be fitted by means of a power law, such as:

$$CK_{nn} = b_2 m_1^K \tag{7}$$

Using natural logarithms:

$$ln(CKnn) = ln(m_1)K + ln(b_2) \tag{8}$$

Where b_2 and m_1 are fitted parameters.

Analogously, MLR method is applied to model the relationship between the cumulative neighbour degree distribution with K.

Table 6 shows the values of the parameters: m_1, b_2, DC and MSE for Spaces L, P and R in the urban bus and subway networks. As it can be seen, all the cumulative distributions can be well fitted to power law. Figure 3 shows $CKnn(K' \geq K)$ as a function of K.

TABLE 6. Fitted parameters of the cumulative distribution of the nearest neighbour degree $CKnn(K` > K)$ given by equation 8. The determination coefficient (DC) and mean squared error (MSE) are also given.

	Space	$ln(m_1)$	$ln(b_2)$	DC	MSE
Subway network	L	0.765634513	35.46252791	0.882592399	0,23049165
	P	0.964464371	3470.429487	0.904636214	0.272055784
	R	0.829266757	108,2103956	0.908903839	0.257374745
Urban bus network	L	0.840552334	76.47270121	0.884550132	0.255249454
	P	0.98786125	28109.14075	0.977398745	0.151441537
	R	0.954207558	2800.009178	0.885119331	0.303441252

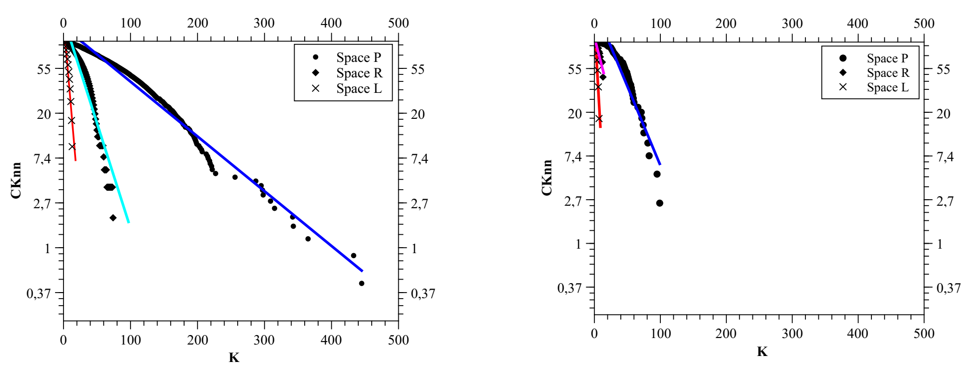

FIGURE 3. Left side: Cummulative nearest neighbours degree distributions CK_{nn} for the urban bus network in Spaces L, R and P. Right side: Cummulative nearest neigbours degree distributions CK_{nn} for the subway network in Spaces L, R and P.

In both networks the values of $< Knn >$, Knn_{Min} and Knn_{Max} are larger in Space P than in Space L, which is due to the fact that Space P is more dense than Space L. Space R presents an intermediate position between Spaces L and P, since this space has a density mid-way between the densities of the other two spaces. These facts are also supported by the values of the coefficient m_1 in all Spaces $m_1(L) \leq m_1(R) \leq m_1(P)$, as can be observed in Figure 3.

In Space L, $< Knn >$, Knn_{Max} and Knn_{Min} are similar in both TN, that is, the average nearest neighbour degree of one stop is nearly identical in both TN. However in the Spaces P and R these parameters are very different, as can be observed in Table 4 and Table 5.

2.4. **Network efficiency.** The network efficiency allows evaluate the operation of a TS. We study the network efficiency in the urban bus and subway networks of Madrid by analysing two topological parameters: the average shortest path of the network ($< l >$) and the Global Network Efficiency (*EGlob*).

The average shortest path, $< l >$, is defined as:

$$< l >= \frac{1}{N}\Sigma_i < l(i) >= \frac{1}{N^2}\Sigma_{i,j} l(i,j) \tag{9}$$

Where:

$$< l(i) >= \frac{1}{N}\Sigma_j l(i,j) \tag{10}$$

and $l(i,j)$ symbolises the distance between i and j nodes, i.e., the number of links for the shortest path between them.

In Space L, $< l >$ means the number of stops required by passengers to reach the destination stop (bus shelter or subway station), on average.

In Space P, $< l >$, is the minimum number of lines that is required to be used to reach a destination stop. We can define the *Transfer Capacity* as $< l >$ - 1, this parameter is a relevant indicator in order to estimate the adequacy of a TN. A passenger wishes to reach the destination stop through the least number of transfers. Often, passengers cannot reach the destination without a transfer on a long distance trip. In general the minimum number of transfers should not be more than two, otherwise it could be considered that the passanger trip is bothersome.

From Tables 7 and 8, we can observe that the most percentage of nodes are reachable through one, two or three transfers in both TS. 99.77% of the paired nodes are reachable within four transfers and the average minimum number of transfers is 2.20 in the urban bus network. In the subway network, 100% of the nodes are reachable within three transfers and the average minimum number of transfers is 1.26. The higher the transfer number, the worse the performance for one TS. The Global Network Efficiency (*EGlob*) is a measure of the performance of the network under the supposition that the efficiency for sending information between two nodes i and j is proportional to their reciprocal distance $l(i,j)$. Accordingly [4], *EGlob* is defined as:

$$EGlob = \frac{\Sigma_{i \neq j \in G'} l(i,j)^{-1}}{N(N-1)} \qquad 0 \leq EGlob \leq 1 \tag{11}$$

EGlob quantifies the efficiency of communication between all pairs of nodes on the network (stops or lines), under the assumption that communication flows along the shortest paths available. Considering just one pair of nodes, if a link joins these two nodes, the path between them has length 1 and so communication is maximally efficient for that pair, therefore that path has an efficiency of 1. If the shortest

TABLE 7. Percentage of reachable nodes according to the number of transfers in the urban bus network

%	Number of transfers
1.19	0
15.08	1
50.34	2
29.76	3
3.40	4
0.23	≥ 4

TABLE 8. Percentage of reachable nodes according to the number of transfers in the subway network

%	Number of transfers
11.85	0
54.09	1
30.43	2
3.63	3

path between a pair of nodes, i and j, has length $l(i,j)$, then its efficiency is $\frac{1}{l(i,j)}$. The mean value of that efficiency of each pair, taken over all pairs of nodes in the network, is $EGlob$. Only a completely connected network (where all possible links are present) has a global efficiency of 1. All other networks have an $EGlob \leq 1$.

$EGlob(L)$ in the subway network is bigger than in the urban bus network. This states that passengers, on average, will reach one destination stop by subway more easily (the mean steps to get from one stop to another is smaller). $EGlob(L)$ in the subway network is similar to the obtained $EGlob(L)$ in the subway network of Seoul [4].

$EGlob(P)$ in the subway network is bigger than in the urban bus network. This means that on average less transfers are required in a trip by subway.

Likewise, $EGlob(R)$ in the subway network is bigger than in the urban bus network, one passenger arrives to one line, on average, passing through less stops in the subway network.

2.5. **Mean service efficiency.** The mean service efficiency of a network, ϱ, is defined as the ratio between the total number of stops of the network and the product of the total number of lines multiplied by the average number of stops per line:

$$\varrho = \frac{Ns}{M\phi} \quad 0 \leq \varrho \leq 1 \tag{12}$$

Where Ns is the total number of stops in the networks, M is the number of lines, and ϕ is the mean number of stops per line, being $\phi \leq Ns$.

According to the equation 12 high levels of the redundancy in the network imply low ϱ. Taking into account that to provide high redundancy in a TN implies to carry out large investments in infrastructures, the transport companies should get a good balance between these two magnitudes during the TN planning. The TN planning should consider that the passengers wish to get from node i to other node j fast but they will not choose an unreliable albeit fast TS. (e.g. in the case of one line is down the passengers should have enough alternate lines to get their destination).

The value of ϱ is 0.84736842 and 0.41140757 in the subway and urban bus networks respectively. The subway network has a larger (ϱ), that is, a smaller redundancy, than the urban bus network.

There are several pieces of investigations that analyse the service efficiency all of them use different methods to the those used in our research, for instance, Takeuci et al. [19] discuss the situation and reason which make the public subsidy inevitable, and propose a method of practicing subsidies. They introduce and define a measurement named the route-potential, which can distinguish the bus route which may be subsidized, this calculation is applied in each bus route in Nagoya and usefulness of this measurement is confirmed. The authors clasify a bus route into a competitive and sustainable business course and a civil minimum course giving priority to mobility of citizens by using route potential as index. Mizokami et al. [12] evaluates bus routes in Kunamoto area by two different aspects, productive efficiency and the possibility of boosting the potential demand. The bus riding survey was performed in order to investigate the actual condition of bus utilization and propose a rational method of reorganization on bus network on Kumamoto urban area. The author also use a behavioral intention (BI) method to forecast the Kumamoto-Dentetsu railway, in this method, the behavioral intention is directly measured from targeted people to imagine actual behavior in new traffic environment. BI method is based on the attitude theory, therefore, all factors which affect traffic behavior do not need to be specified and converted into quantity to get included in function model.

2.6. **Correlations.** In this section the degree$-$degree correlation between connected nodes of the analysed TN is estimated for the three studied topological spaces. This correlation can be estimated calculating the normalized correlation coefficient r_D, which is defined as [14]:

$$r_D = \frac{1}{\sigma_D(q)^2} \Sigma_{K',K} K'K\{e_D(K',K) - q_D(K')q_D(K)\} \tag{13}$$

where $P_D(K)$ and $q_D(K)$ are the normalized degree distribution and the normalized remaining degree distribution respectively. The remaining degree distribution refers to the probability that following a randomly chosen link, the remaining degree of the reached node is K.

$e_D(K',K)$ is the joint probability that the two nodes on each side of a randomly chosen link have K' and K remaining degrees, respectively.

$$q_D(K) = \frac{(K+1)P_D(K+1)}{\Sigma_{K'} K' P_D(K')} \tag{14}$$

$$\sigma_D(q)^2 = \Sigma_K K^2 q_D(K) - < \Sigma_K K q_D(K) >^2 \tag{15}$$

In an assortative network, in terms of their degree-degree correlation behavior, high-degree nodes interact with high-degree nodes, in a neutral network the nodes connect to each other with the expected random probabilities and finally, in a disassortative network high-degree nodes tend to avoid linking to each other.

The values of r_D are shown in Tables 4 and 5 in Spaces L, P and R for the subway and urban bus networks respectively. The urban bus network is assortative

($r_D \geq 1$) for the three studied spaces, this means that one line or one stop preferably links to another one with many connections; therefore a district with many bus lines preferentially attaches to a district with a large number of bus lines and one line with a large number of stops tends to connect to another one with high number of stops. On the contrary, the subway network is assortative ($r_D \geq 1$) for Spaces L and P, but dissassortative ($r_D \leq 1$) for Space R. These facts are based on the size of the three spaces, that is, Space R for the subway network, which has only 16 lines, is the smallest space. These results are in agreement with the study of Sienkiewicz and Holyst [16] related to 22 public transport networks in Poland, which establishes that when the number of nodes is larger than 500 in some of those polish public transport networks, the network is usually assortative. Zhang et al. also validated this property in the bus network of Beijing [26].

3. **Urban transport flow.** Transport flow is defined as the movement of vehicles carrying passengers between different geographic destinations or within a certain region. The distribution of transportation flows in a network defines the rational variation of passenger flow in a TN that meets the needs for the transportation of passengers. Therefore, the understanding of the mathematical function and the factors that control the transportation flows is highly relevant.

In this section we will characterise the transport flow in Madrid. To this end we will consider Madrid is divided into 21 administrative areas and study the flow of urban buses and subway trains from one district to another. In Figure 4 the total flow of vehicles from each district of Madrid is represented. It can be noticed that the highest flow correspond to districts 13 and 15 with 1,210,082 and 1,177.725 vehicles (total number of urban buses and subway trains) for the 5 working days of a week respectively.

Figure 4 shows the 21 district locations inside Madrid. Tables 9 and 10 and Figures 5 and 6 present some data related to transport flow in Madrid. Tables 9 and 10 show the highest and smallest values of the vehicle flow between districts for the urban bus and subway networks respectively. Figures 5 and 6 present the different flows between districts for both TS.

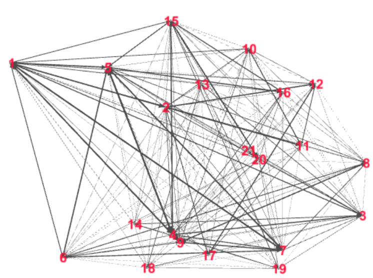

FIGURE 5. Visualization of buses flow between districts (darker links means higher flow).

Origin District	Destination District	Buses Flow
5	4	29,185.3
4	5	25,629.6
7	1	25,119.6
..	..	..
..	..	..
..	..	..
10	16	93
9	12	59.7
6	13	21.9

TABLE 9. Ranking of outward bound bus flow by district in the 5 working days of a week.

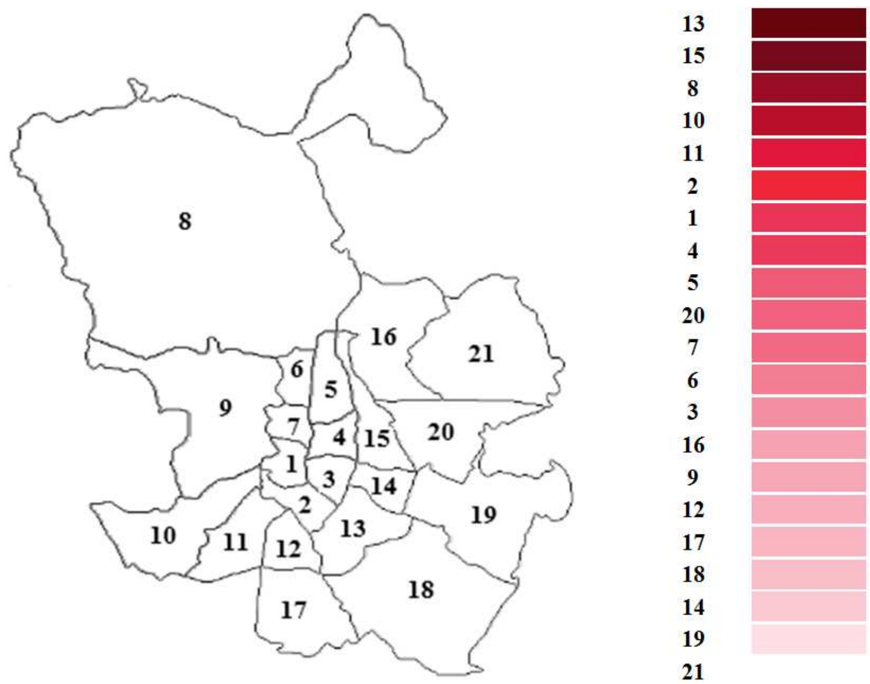

FIGURE 4. Geographic representation of the vehicle flow (total number of urban buses and subway trains) corresponding to different districts inside Madrid (darker red tones corresponding to higher flows).

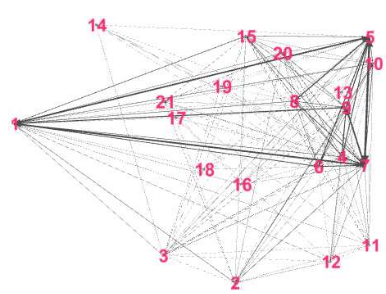

FIGURE 6. Visualization of subway train flow between districts (darker links means higher flow).

Origin District	Destination District	Subway Trains Flow
5	7	16,106
7	5	16,106
7	1	10,229.2
1	7	10,229.2
9	7	11,769
7	9	11,769
..	..	..
..	..	..
..	..	..
8	19	186.5
6	19	186.5
7	19	186.5
5	19	186.5
4	19	186.5
19	4	186.5
19	5	186.5
19	7	186.5
19	6	186.5
19	8	186.5

TABLE 10. Ranking of outward bound subway train flow in the 5 working days of a week.

In the following sections we study the characteristics of traffic flow in Madrid subway and urban bus networks. Our analysis uses the data about vehicle movements during the 5 working days of a week. This information is available on the web sites of the **E**mpresa **M**unicipal **T**ransportes (EMT) [23] and **M**etro **M**adrid (MM) [24].

3.1. **Urban bus flow.** In order to study the urban bus flow, we build a graph $G'' = (E''; V'')$, in which E'' is the set of nodes and V'' is the set of links between them. A node represents a district and a link symbolises that at least one line stretch of a bus line which goes from one district to another exists. A line stretch is a section of a transport line that passes through two different districts. The beginning and the end of a line stretch are the two consecutive stops in the line, which are located in different districts. In Figure 7 an schematic example of the construction of G'' is shown.

G'' represents a weighted networks whose nodes are the districts and the weighted links identify the existing interaction between two nodes and its strength. The weight of a link w_{ij} between two districts i, j is the total number of line stretches between them. The cost of a link between two districts i and j (C_{ij}), can be established as the total number of line stretches between i and j divided by the maximum number of existing line stretches between two any districts in the TN:

$$C_{ij} = w_{ij}/Maximun(w_{ij} \,{}_{\triangle_{ij\epsilon E''}}) \tag{16}$$

The relationship between the population of a district and the number of urban

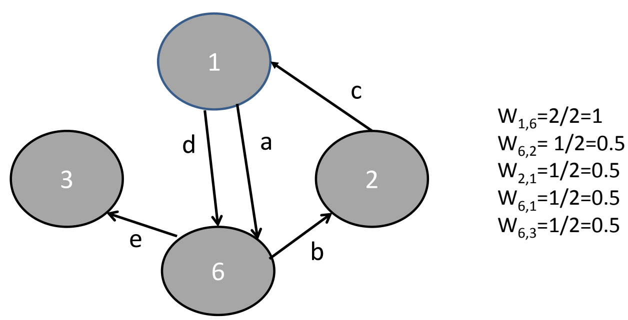

FIGURE 7. Schemes of the construction of G'' for the analysis of traffic flow between districts (nodes). In this example, traffic between all districts is shown for one line. The movement of vehicles happens between districts $1 \to 6, 6 \to 2, 2 \to 1, 1 \to 6, 6 \to 3$. The line stretches (a, b, c, d, e) between districts are shown.

buses during the five working days of a week departing from each district can be defined as:

$$T_{ij} = f(P_i, P_j, C_{ij}) \tag{17}$$

Where T_{ij} is the traffic from i to j district, that is the number of vehicles that leaves the district i and arrives to district j. P_i and P_j are the population of districts i and j, that is, the number of people living in the districts i and j respectively.

We found that T_{ij} can be well described by the Gravity Model given by the following equation:

$$T_{ij} = W \frac{P_i^{g_1} P_j^{g_2}}{C_{ij}^{g_3}} \tag{18}$$

We apply the MLR method to model the existing relationship between $log(P_i)$, $log(P_j)$, $log(C_{ij})$, and $log(T_{ij})$ by fitting a linear equation to the empirical observed data. The values fitted by the linear equation are denoted as: $log(\hat{T_{ij}})$.

$$log(\hat{T_{ij}}) = log(W) + g_1 log(P_i) + g_2 log(P_j) + g_3 log(C_{ij}) \tag{19}$$

For the available information, $DC = 0.964131014$ and $MSE = 0.112516415$. Therefore, $log(T_{ij})$ can be well fitted by means of a linear equation, where: $log(W) = 5.253284138$, $g_1 = 0.991209323$, $g_2 = -0.098554038$ and $g_3 = -0.013005331$.

Figure 8 shows the normalized values of $\hat{T_{ij}}$ ($\hat{T_{ijN}}$) as function of the normalized values of the empirical data T_{ij} (T_{ijN}). The normalization of $\hat{T_{ij}}$ and T_{ij} have been done with respect to the maximum values of $\hat{T_{ij}}$ and T_{ij} respectively. It can be noticed that the data are well clustered along the straight line:

$$\hat{T_{ijN}} = -0.035661556\ T_{ijN}\ -0.463841329.$$

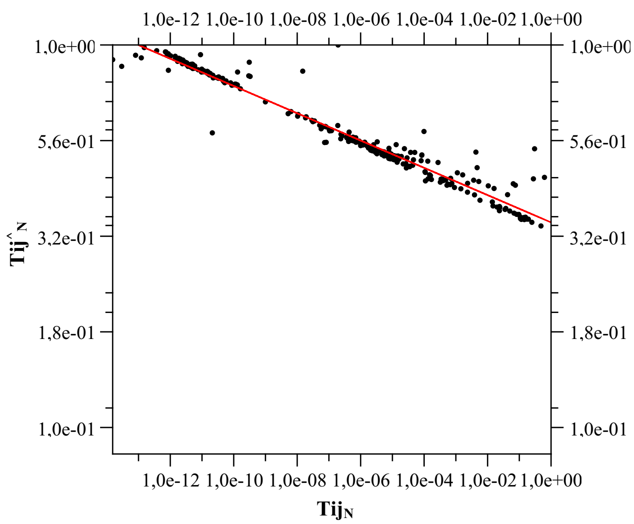

FIGURE 8. For the urban bus network, normalized values of $\hat{T_{ij}}$ ($\hat{T_{ijN}}$) with respect to the maximum $\hat{T_{ij}}$, as function of the normalized values of T_{ij} (T_{ijN}) with respect to the maximum T_{ij}.

We apply the Gravity Model with the fitted data given above, between all the pairs of i and j districts (where i, j vary between 1 and 21). The calculation of the total number of buses T_i for the district i is carried out by summing T_{ij} over j, $T_i = \Sigma_{j=1}^{j=21} T_{ij}$. We show the total number of buses T_i for a district i depends on its population P_i, with a relationship:

$$T_i \sim P_i^h \tag{20}$$

Using logarithms:

$$log(T_i) \sim hlog(P_i) \tag{21}$$

by applying the MLR method, we obtain that the value of h is 1.273672338 and the value of DC is 0.890794517. The results are shown in Figure 9 that depicts the relation between the number of buses that leaves the district i (T_i) and its population (P_i). We observe the data can be well fitted to the equation 20.

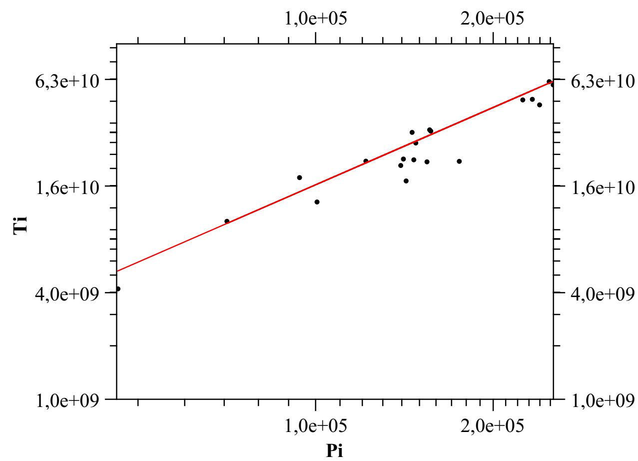

FIGURE 9. Correlation of the number of buses that leaves the district i (T_i) as function of its population (P_i). The red line correspond to the fitting to $log(T_i) \sim hlog(P_i)$, where $h = 1.273672338$ and $DC = 0.890794517$.

Next, we apply the Gravity Model to study how the changes of topology impact over the proportionality between the population of one district and its urban bus flow. Two methods to modify the structure of the network were used:

- Method 1: two links are randomly chosen from the network, one connecting districts i and j, and other joining q and o districts; next the partner of each district is changed, by modifying the original links $i-j$ and $q-o$ to $i-o$ and $j-q$. This method maintains the degree of each node. We repeat this process a significant number of times (2,000 times).
- Method 2: one link is randomly eliminated from the network and then this link is connected between two randomly chosen districts. This action is done 2,000 times. This algorithm does not conserve the original degree distribution.

A total of 506 networks were generated with each method.

Table 11 shows the obtained results by applying the MLR method to model the existing relationship between $log(T_i)$ and $log(P_i)$ in the aforementioned 506 networks, where $< h >$ represents the average value of h, σ_h is the typical deviation of h, $< DC >$ represents the average value of DC and σ_{DC} is the typical deviation of DC.

It can be noticed that the average exponent $< h >$ is equivalent for both types of generated random networks and it has also a similar value to the obtained exponent for the real urban bus network.

TABLE 11. Exponent h of the relationship between the total number of buses, T_i , and the population, P_i of the district i given by the equation 20, in the random networks. For these networks, $< h >$ represents the average value of h, σ_h is the typical deviation of h, $< DC >$ represents the average value of DC and σ_{DC} is the typical deviation of DC.

Method used to generate the random network	$< h >$	σ_h	$< DC >$	σ_{DC}
Method 1	1.274	0.002	0.944	0.001
Method 2	1.2699	0.0006	0.9403	0.0003

3.2. **Subway flow.** We apply the same methods described in the previous section to analyse the subway flow between districts in Madrid city. We fit the empirical observed data to the equation 19 and obtain that $DC = 0.948760437$ and $MSE = 0.11282044$. Therefore, we can conclude that $log(T_{ij})$ can be well fitted by a linear equation where: $log(W) = 2.740974027$, $g_1 = 1.122587533$, $g_2 = 0.137863317$ and $g_3 = 0.137863317$.

Figure 10 shows the normalized values of $\hat{T_{ij}}$ as function of the normalized values of T_{ij}. It can be noticed that the data are well clustered along the straight line:

$$\hat{T_{ijN}}=0.143663494\ T_{ijN}+8.649260843.$$

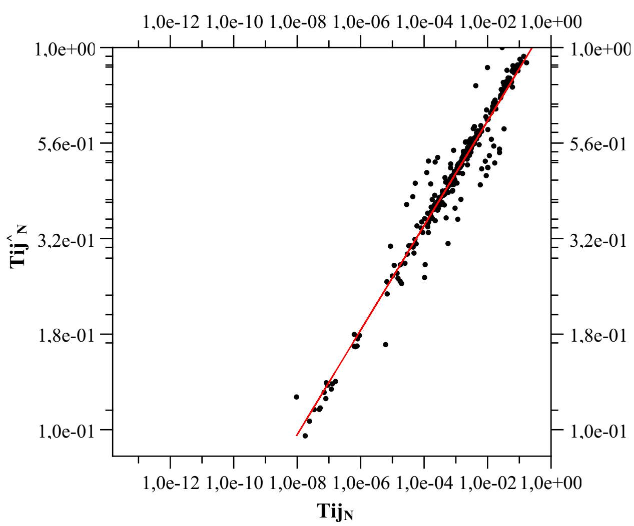

FIGURE 10. For the subway network, normalized values of $\hat{T_{ij}}$ ($\hat{T_{ijN}}$) with respect to the maximum $\hat{T_{ij}}$, as function of the normalized values of T_{ij} (T_{ijN}) with respect to the maximum T_{ij}.

Similar to the case of the urban bus network, the total number of subway trains T_i of a district i depends on its population P_i following the relationship given by equations 20 and 21.

The values of the exponent h that is 1.417400199, with $DC = 0.907918443$ was gotten from MLR method. The corresponding results are plot in Figure 11, that shows the relation between the number of subway trains that leaves the district i (T_i) and its population (P_i).

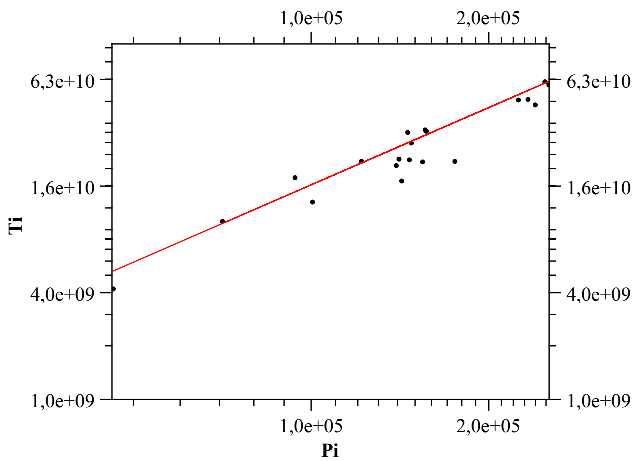

FIGURE 11. Correlation of the number of subway trains that leaves the district i (T_i) as function of its population (P_i). The red line correspond to the fitting to $log(T_i) \sim hlog(P_i)$, where $h = 1.417400199$ and $DC = 0.907918443$.

We also study how the changes of topology impact over the proportionally between the population of one district and its subway train flow. Methods 1 and 2, which were described in the previous section, are applied.

Table 12 shows the obtained results by means of MLR method to describe the existing relationship between $log(T_i)$ and $log(P_i)$ in the 506 generated networks.

TABLE 12. Parameters of the relationship between the total number of subway trains, T_i , and the population, P_i, for a district i in the random networks generated by two methods. For these networks, $< h >$ represents the average value of h, σ_h is the typical deviation of h, $< DC >$ represents the average value of DC and σ_{DC} is the typical deviation of DC.

Method used to generate the random network	$< h >$	σ_h	$< DC >$	σ_{DC}
Method 1	1.239	0.008	0.825	0.002
Method 2	1.247	0.006	0.826	0.002

Analogously to the case of the urban bus flow, we observed that the exponent h is similar for both types of generated random networks. However, the value of h is lower in these networks than in the real subway network.

4. **Conclusion.** In this research, we have investigated several structural parameters and traffic flows of Madrid urban bus and subway networks.

In order to detect the fundamental characteristics of these networks, three topological spaces were analysed: Space L (Geographical Stop Space), Space P (Transfer

Space) and Space R (Route Space). The interest of the aforementioned spaces lies in the fact that they allow us to know specific static structural aspects of both TN, such as: directions one passenger can take from one given stop, number of stops passed on the shortest path between two given stops, needed transfers during a trip, whether one passenger can transfer directly between two lines or must pass through other additional ones, etc.

We find that, on average, pasengers can reach a destination in less steps, that is, passing through less number of stops travelling by subway (10.19 stops) than taking the bus (18.77 stops). The required number of transfers to reach a destination stop is also, on average, less in the subway network (2.26) than in the urban bus network (3.20). Likewise, the average number of necessary line steps to reach a destination line from an origin line is less in the subway network (1.62 lines). In the subway network, the Mean Service Efficiency, which was defined as the ratio between the total number of stops and the product of the total number of lines multiplied by the average number of stops per line, is the double of its value in the urban bus network. We can conclude, the subway network exhibits better topological parameters and therefore, some measures to improve the urban bus network could be studied, such as: modifying some existing stops and lines to reduce the average path length, or, increasing stop redundancy.

The maximum number of lines that one stop has is higher in the urban bus network (13 lines), than in the subway network (7 lines). The largest number of existing paths (without transfer) from one stop is 99 and 450 in the subway and urban bus network respectively. The largest number of stops that link two lines is 13 in the subway network and 150 in the urban bus network.

We also show that the cumulative degree distribution of the urban bus and subway networks can be described by an exponential function in all studied spaces, while the degree-degree correlations correspond to a power law ($CK_{nn} \sim m_1^K$) in both TS. Taking into account the cumulative degree distribution of both networks, we can establish they are single-scale, which implies these TN can withstand the random attacks of several nodes (stops or lines) but it is vulnerable in the case of direct attacks towards particular nodes.

Considering the assortativity property, we detect that one stop with high connectivity preferably links to another one with many connections in both TS. However, a district with many bus lines preferentially attaches to a district with a large number of bus lines, whereas the opposite occurs in the subway network, where a district with a few links preferentially links to another one with a large number of lines.

Regarding the traffic flow between districts, it has been adequately described by a Gravity Model taking into account the population from the origin and destination districts as well as the total number of line stretches between them. We have got similar results to those obtained for the intercities Korean express bus system [9] and detected similar behaviour in two types of urban TN considered in Madrid city. It should be remarked that the results found are on a smaller geographic scale since we have studied the vehicle flow between two districts (existing administrative areas inside a city) in the same city. In [9] the authors also described the intercity bus flow using a gravity model. They estabished that the total bus flow of a city depended only on its population and we show that the total vehicle flow (buses and subway trains) for a district depend exclusively on the district population.

In the future, we will analyse more deeply the vulnerability of Madrid subway and urban bus networks using different existing methods. One of them is the generalised concept of percolation through which resilience is evaluated. This procedure is based on the calculation of the giant component size (largest connected cluster) once an arbitrary problem in a node or set of nodes has ocurred.

Acknowledgments. This research was supported by the Ministry of Economy and Competitiveness-Spain under Grant No. MTM2012-39101.

REFERENCES

[1] R. Albert and A. L. Barabási, Statistical mechanics of complex networks, *Reviews of Modern Physics*, **74** (2002), 47–97.
[2] R. Albert, H. Jeong and A. Barabasi, Error and attack tolerance of complex networks, *Nature*, **406** (2000), 378–382.
[3] L. A. N. Amaral, A. Scala, M. Barthélémy and H. E. Stanley, Classes of small-world networks, *Proceedings of the National Academy of the United States of America*, **97** (2000), 11149–11152.
[4] K. H. Chang, K. Kim, H. Oshima and S. M. Yoon, Subway networks in cities, *Journal of the Korean Physical Society*, **48** (2006), S143–S145.
[5] Y. Z. Chen, N. Li and D. R. He, A study on some urban bus transport networks, *Physica A*, **376** (2007), 747–754.
[6] J. Hao, J. Yin and B. Zhang, Structural fault tolerance of scale-free networks, *Tsinghua Science & Technology*, **12** (2007), 246–249.
[7] B. Jiang, A topological pattern of urban street networks: Universality and peculiarity, *Physica A*, **384** (2007), 647–655.
[8] M. Ke, et al., Power law and small world properties in a comparison of traffic city networks, *Chinese Science Vulletin*, **56** (2011), 3731–3735.
[9] O. Kwon, Intercity express bus flow in Korea and its network analysis, *Physica A*, **391** (2012), 4261–4265.
[10] V. Latora and M. Marchiori, Is the Boston subway a smallworld network?, *Physica A*, **314** (2002), 109–113.
[11] G. Mao and N. Zhang, A Multilevel simplification algorithm for computing the average shortest-path length of scale-free complex network, *Journal of Applied Mathematics*, (2014), Art. ID 154172, 6 pp.
[12] S. Mizokami, R. Kakimoto and J. Hashimoto, A method of line characteristic evaluation and network reorganization planning of bus systems, *Journal of Japan Society of Civil Engineers*, **793** (1995), 27–39.
[13] M. E. J. Newman, The structure and function of complex networks, *SIAM Review*, **45** (2003), 167–256.
[14] M. E. J. Newman, Assortative mixing in networks, *Physical Review Letters*, **89** (2002), 208701–208705.
[15] S. Ondǒs, I. Paulovičováa, L. Belušák and D. Husendová, Urban heartbeats (daily cycle of public transport intensity), in *GIS Ostrava 2014 - Geoinformatics for Intelligent Transportation*, 2014, 747–754.
[16] J. Sienkiewicz and J. A. Holyst, Statistical analysis of 22 public transport networks in Poland, *Physica Review E*, **72** (2005), 046127.
[17] H. Soh, et al., Weighted complex network analysis of travel routes on the Singapore public transportation system, *Physica A*, **389** (2010), 5852–5863.
[18] Z. Su, et al., *Robustness of Interrelated Traffic Networks to Cascading Failures,* Scientific Reports, 2014.
[19] D. Takeuchi and K. Yamada, Theory of public subsidies for city bus and development of route-potential as a measurement for that decision making, *Journal of Infrastructure Planning and Management,* **1991** (1995), 183–192.
[20] J. Tinbergen, *Shaping the World Economy: Suggestions for an International Economic Policy,* Twentieth Century Fund, New York, 1962.
[21] C. von Ferber, T. Holovatch, Y. Holovatch and V. Palchykov, Public transport networks: empirical analysis and modeling, *The European Physical Journal B*, **68** (2009), 261–275.

[22] D. J. Watts and S. H. Strogatz, Collective dynamics of 'small-world' networks, *Nature*, **393** (1998), 440–442.

[23] Web site of the **E**mpresa **M**unicipal **T**ransportes (EMT), 2014. Available from: www.emtmadrid.es/.

[24] Web site of the **M**etro **M**adrid (MM), 2014. Available from: www.metromadrid.es/.

[25] P. Zhang et al., *The Robustness of Interdependent Transportation Networks Under Targeted Attack,* EPL (Europhysics Letters), 2013.

[26] H. Zhang, P. Zhao, J. Gao and X. Yao, The analysis of the properties of bus network topology in Beijing basing on complex networks, *Mathematical Problems in Engineering*, **2013** (2013), Article ID 694956, 6 pages.

Received July 2014; revised November 2014.

E-mail address: mouronte.lopez@upm.es
E-mail address: rosamaria.benito@upm.es