Topological Analysis of the Subway Network of Madrid

Mary Luz Mouronte
Departamento de Ingeniería y Arquitecturas Telemáticas.
Escuela Técnica Superior de Ingeniería y Sistemas deTelecomunicación
Universidad Politécnica de Madrid
Email: mouronte.lopez@upm.es

***Abstract*—In this paper, we study by means of Complexity Science the topological structure of the subway network of Madrid. Different statistic features are analyzed: Degree of a Node i (K_i), Degree Probability Distribution $P(K)$, Nearest Neighbour Degree $Knn(k)$, Clustering Coefficient, Average Path Length $< l >$, Mean Service Efficiency ρ, Global Network Efficiency $Eglob$, Correlation Coefficient r_D and robustness of the network. This analysis will allow to obtain a deeper knowledge of this network and it will also help to improve its management: insight about connectivity, most relevant stops, efficiency, vulnerability and way of growth.**

Keywords-Network Science; Transport Network; Statistical Analysis

I. INTRODUCTION

This paper calculates different mathematical parameters in the subway network of Madrid. Our research will allow to increase the current knowledge about this network. Structural properties of a subway network are very relevant for an effective transportation management in the urban cities. There are several works that analyze the characteristics of the transport networks:

Chen et al. [1], present an empirical investigation about the urban bus networks of four major cities in China: Hangzhou, Nanjing, Beijing and Shanghai.

Chen et al. [2], investigate the evolution of the dynamic properties in bus networks of Hangzhou, Nanjing, Beijing and Shanghai. New measurements of the average sum of the nearest-neighbors degree$-$degree correlation $D_{nn}(K)$ and the degree average edges among the nearest-neighbors $L(K)$ are proposed. The obtained results reflect that the considered transport network are organized randomly.

Chang et al. [3], study the subway network of Seoul, Tokyo, Boston and Beijing, by using the global and local efficiency. It is found that the Seoul subway network has a smaller global and local efficiency than the Tokyo network. The authors suggest that the Tokyo subway system is better for an overall distance trip but is weaker regarding incidents of disconnection. It is also shown for the subway networks of Boston, Seoul, Tokyo and Beijing, the global efficiency is inversely proportional to the length of the network. The Boston and Beijing local efficiencies are very low which means that these are somewhat deficient in some routes.

The rest of the paper is organized as follows: Section 2 describes the subway network of Madrid, in Section 3 the method of analysis and the results are presented, and finally in Section 4 we end with some conclusions.

II. THE SUBWAY NETWORK OF MADRID

Madrid is one of the most populous cities in the world. It has a population of 3,254,950 dwellers on an area of 60,683 hectares, and a high developed public transportation network. The subway network of Madrid is one of the largest subway networks in the world, rivalling other networks such as the Shanghai, Guangzhou, Beijing or Delhi. In 2007, this network became the second largest subway network in Europe after London. The subway network of Madrid has 16 routes and 322 stops. The first route of the network began its operation in 1919.

The subway network of Madrid is operated with more than 2,400 trains and its yearly ridership was 628 millions in 2012. This network has been tranformed by means of several improvement plans since 2011. Generally, the subway is open to the public from 6:00 AM to 1:30 AM every day of the year.

III. MATHEMATICAL ANALYSIS

We can map this network in three Topological Spaces: Space P, Space L and Space R. In these Spaces, the network is abstracted 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 built as a bidimensional representation of the relationships between nodes, where $Aij = 1$ when a connection between nodes i and j exists and $Aij = 0$ otherwise. N is the number of nodes in E.

In the Space L, one node represents one stop, and one link symbolizes an union between two nodes if one stop is the successor of the other on a subway route. Space L is named Stop Geographical Space. In the Space P, one node represents one stop, and one link joins a pair of stops if at least one route provides direct service. A link means that passengers can take at least one route for a direct travel between two stops. If passengers have to exchange routes then the pair of stops is linked by more than one link. Space P is called Subway-Transferring Space. In the Space R, nodes are defined as routes and common stops determine the links. Space R is called Route Space.

While carrying out a topological study of the network for these three Spaces, some parameters are estimated:

Degree of a Node K_i and Degree Probability Distribution

 ISBN: 978-1-61208-346-9

$P(K)$: The degree of a node i is the number of links connected to it:

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

Not all nodes in the subway network have the same number of links: how the degree is distributed among the nodes is an interesting property which can be analized by estimating the Degree Probability Distribution $P(K)$.

Nearest Neighbour Degree $K_{nn}(k)$, which is defined as:

$$K_{nn}(K) = \Sigma_{K'=0}^{\bowtie} K' p(K'/K) \tag{2}$$

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'. Therefore K_{nn} is the average degree of those nodes that are found by following the links originating from a node of degree K. The evolution of $K_{nn}(K)$ is related to the assortativity of the network, which indicates the tendency of a node of degree K to associate with a node of the same degree K. In an assortative network, $K_{nn}(K)$ increases with increasing K but, in a dissassortative network, $K_{nn}(K)$ decreases with increasing K while in a neutral network, $K_{nn}(K)$ does not depend on K.

Clustering Coefficient: given three actors i, j and w with mutual relations between j and i as well as between j and w, Clustering Coefficient is supposed to symbolize the likelihood i and w are also related. This parameter was used by Watts and Strogatz [4] for social networks analysis.

This concept can be explained by defining for $j \in V$,

$$m(j) = |\{i,w\} \in E : \{j,i\} \in E \quad and \quad \{j,w\} \in E\}| \tag{3}$$

and

$$t(j) = \frac{K_j(K_j - 1)}{2} \tag{4}$$

We named $m(j)$ the number of opposite links of j, and $t(j)$ the number of potential opposite links of j.

For a node j with $K_j \geqslant 2$, the Clustering-Coefficient is defined as:

$$C(j) = \frac{m(j)}{t(j)} \tag{5}$$

and Clustering-Coefficient of a graph $G = (V, E)$ is denoted as:

$$C(G) = \frac{1}{|V'|} \Sigma_{j \in V'} C(j) \tag{6}$$

where V' is the set of nodes i with $K_i \geqslant 2$.

The Clustering Coefficient of a node ranges between 0 and 1.

Average Path Length $< l >$, which is the average shortest path between all nodes of the network. We denote by $l(i, j)$ the distance between i and j, i.e., the number of links on a shortest path between them.

$$< l(i) > = \frac{1}{n} \Sigma_j l(i, j) \tag{7}$$

represents the average distance from i to all nodes. The average distance in G is defined as:

$$< l > = \frac{1}{n} \Sigma_i < l(i) > = \frac{1}{n^2} \Sigma_{i,j} l(i, j) \tag{8}$$

We denote by $D = max_{i,j} l(i, j)$ the diameter (D) of G, i.e., the largest distance between two nodes of the network.

Mean Service Efficiency (ρ) [9], which can be defined for the subway network as:

$$\rho = \frac{Ns}{M\phi} \quad 0 \leqslant \rho \leqslant 1 \tag{9}$$

Where Ns , M, and ϕ are the total of stops, the number of routes and the mean number of stops per route respectively. For a specific number of subway stops, a larger magnitude of ρ implies fewer subway routes that the transport company should maintain to satisfy the travel demand.

Global Network Efficiency (Eglob) [9], which may be described as:

$$E_{glob} = \frac{\Sigma_{i \neq j \in G} l_{ij}^{-1}}{Ns(Ns - 1)} \quad 0 \leqslant E_{glob} \leqslant 1 \tag{10}$$

Global efficiency is a measure of the performance of the network, under the assumption that the efficiency for sending information between two nodes i and j is proportional to the reciprocal of their distance $l(i, j)$.

Correlation Coefficient, the degree-degree correlation was analized as the correlation function between the remaining degrees [5] of the two nodes on each side of an link. Remaining degree means the degree of that nodes minus one. The normalized Correlation Coefficient is defined as:

$$r_D = \frac{1}{\sigma_D(q)^2} \Sigma_{u,v} uv \{e_D(u, v) - q_D(u) q_D(v)\} \tag{11}$$

where:

- $e_D(u, v)$ is the joint probability that the two vertices on each side of a randomly chosen link have u and v remaining degrees, respectively.
- $q_D(v)$ is the normalized distribution of the remaining degree [6].

$$q_D(v) = \frac{(v+1)P(v+1)}{\Sigma_u u P(u)} \tag{12}$$

$$\sigma_D(q)^2 = \Sigma_v v^2 q_D(v) - |\Sigma_v v q_D(v)|^2 \tag{13}$$

This quantitity was named by Newman [6] the Degree Assortative Coefficient. In an assortative network, r_D is positive but, in a dissortative network, r_D is negative while in a neutral network $r_D = 0$.

Some statistical parameters are available in Table I: Total of Stops (Ns); Number of Subway Routes (M); Clustering Coefficient ($< C >$), showing the subway routes density near each stop; Network Diameter (D), providing the maximum number of stops (or routes) on the shortest paths between any pair of stops (or routes); and finally, Average Shortest Path Length $< l >$, denoting the average number of stops (or routes) on all the shortest paths between any two stops (or routes).

Clustering Coefficient is considered to be a measure of the local connectivity of a graph. High clustering is associated with robustness of a network, that is resilience against random network damage. Considering this parameter the subway network shows moderate resilience.

A node with high K controls the traffic flow, acting as gatekeeper. A node with high k can also act as a link between two distant sectors of the network. The average path length ¡l¿ can be interpreted as a measure of efficiency in the flow

 ISBN: 978-1-61208-346-9

TABLE I: Empirical data corresponding to the Subway Network of Madrid

Space	Parameter	Value
	M	16
	Ns	322
Space L	$<k>$	2.42
	K_{max}	7
	$<C>$	0.01
	$<w>$	2.03
	w_{min}	2.00
	w_{max}	4.00
	D	30
	$<l>$	10.19
Space P	$<k>$	29.39
	K_{max}	99
	$<C>$	0.90
	$<w>$	2.08
	w_{min}	2.00
	w_{max}	6.00
	D	4
	$<l>$	2.26
Space R	$<k>$	6.00
	K_{max}	13
	$<C>$	0.65
	$<w>$	4.88
	w_{min}	1.00
	w_{max}	14.00
	D	4
	$<l>$	1.62

of the network.

Several researches have estimated the efficiency and vulnerability in networks [7][8]. We analyze the robustness of subway network by calculating the value of the average shortest path length ($<l'>$) and the distribution of the number of pairs of nodes Np separated by the shortest distance, in the original network and in the same network but with the highest K degree nodes removed for the tree Spaces. In the Space L, $<l'>= 10.93$; in the Space P, $<l'>= 2.27$ while in the Space R, $<l'>= 2.00$. In Figures 1, 2 and 3 we observe that the distribution of distances changes drastically in the Spaces L and R after the gatekeepers elimination. The network shows low robustness in both spaces (removal of a route or a stop elimination in a route occurs). This feauture is due to the current subway network design which, can be improved by means of optimization tasks. We can also notice that the most frequent short path length is 7 in the Space L (passangers should cross 7 stops without changing their route to get a destination in most cases), 2 in the Space P and 1 in the Space R (most routes are linked by a stop).

We also calculate $Eglob$ and ρ in the Space P since these parameters lack clear meaning in the other Spaces. $Eglob$ represents the total ability of the network to minimize the spatial resistance (or travel impediment), $E_{glob} = 0.298$ and $\rho = 0.847$.

The node degree and its distribution are very important properties for a network. From Figures 4, 5, 6 and Table I several conclusions can be obtained:

- In the Space L, we observe that the number of nodes

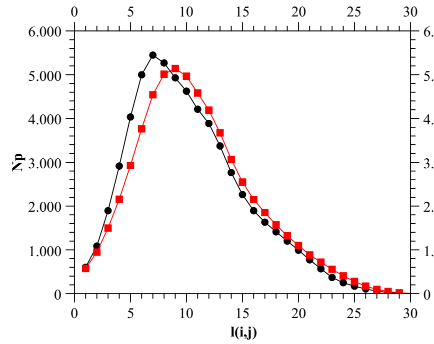

Figure 1: Np - $l(i,j)$ in the original network (black line) and in the same network but with the highest degree nodes removed (red line) in Space L

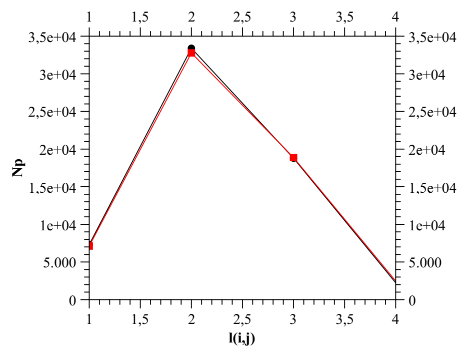

Figure 2: Np - $l(i,j)$ in the original network (black line) and in the same network but with the highest degree nodes removed (red line) in Space P

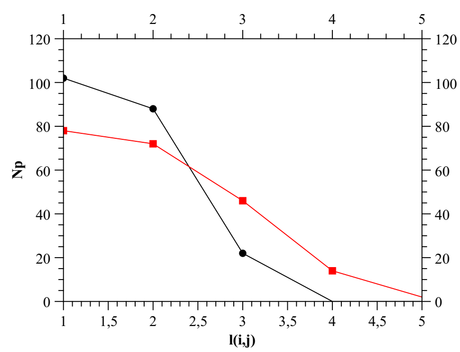

Figure 3: Np - $l(i,j)$ in the original network (black line) and in the same network but with the highest degree nodes removed (red line) in Space R

with degree $K = 2$ are the higher quantity, which means that a typical stop is directly connected to two other stops. In the Spaces P and R there are no nodes with a connectivity degree significantly different from the other nodes; the connectivity distribution is close to an uniform distribution. This happens because the company responsible for the urban transport in the city must ensure the uniform distribution of local equipment so that they are accessible to the entire population. In the Space R, we notice that routes with 9 common stops are the most frequent.

- The subway network constrained in different Geographical Spaces leads to different values of the node

Copyright (c) IARIA, 2014. ISBN: 978-1-61208-346-9

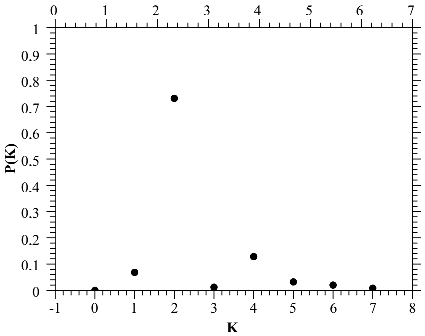

Figure 4: Degree Distribution in Space L

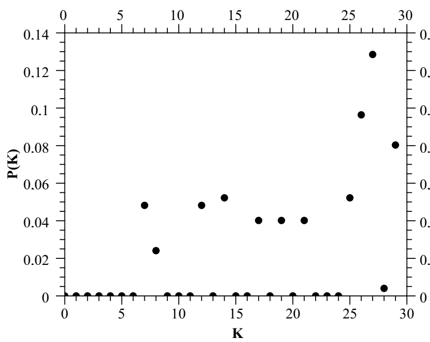

Figure 5: Degree Distribution in Space P

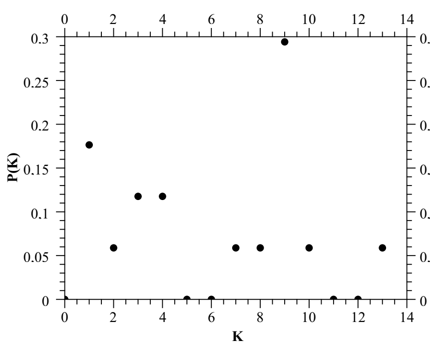

Figure 6: Degree Distribution in Space R

degree: low magnitudes in the Spaces L and R ($< K >$ is 2.42 and 6 respectively) but very high value in the Space P ($< K >$ is 29.39). This is because there are few common stops to different routes, although, from one origin stop many final stops can be reached. In the Space L, the most connected stops exist: Alonso Martínez and Avenida de América; in the Space P, the most connected node is also Alonso Martínez (this stop can be reached through most of the routes), finally, in the Space R, the most connected route is the route number 10. This happens because these elements are relevant communications centers in the city.

Regarding K_{nn} we can observe in Figures 7 and 8 that in some intervals it is difficult to establish whether the correlations are positive, negative or uncorrelated. The statistical variations in K_{nn} can be supressed by estimating its cumulated value. This magnitude decays with increasing K as it is showed in Figures 9 and 10; therefore we conclude that the network is assortative, that is the nodes in the subway network that have many connections tend to be connected to other nodes with many connections. This characteristic is also supported by the positive value of r_D in the spaces L and P (i.e., $r_D = 0.270223$ and $r_D = 0.092046$ respectively). This happens because during the tasks of design and planning to satisfy the traffic needs, was established that the new stops or new routes would be linked to other stops or routes that had similar connectivity in the network.

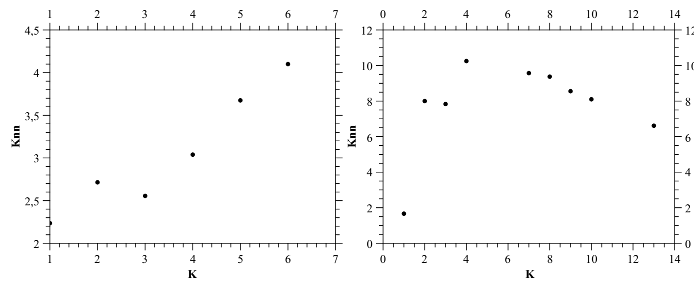

Figure 7: Left side: K_{nn}-K in Space L. Right side: K_{nn}-K in Space P

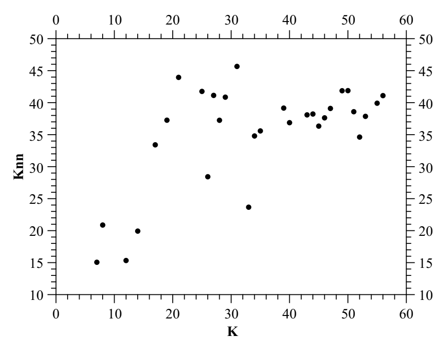

Figure 8: K_{nn} in Space R

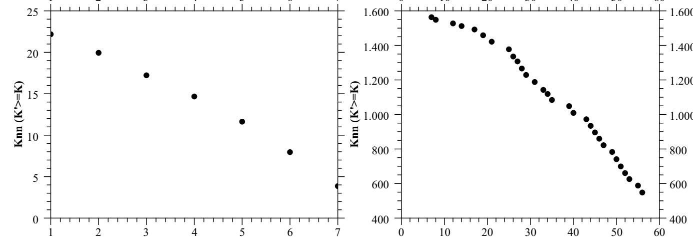

Figure 9: Left side: Cumulated K_{nn}-K in Space L. Right side: Cumulated K_{nn}-K in Space P

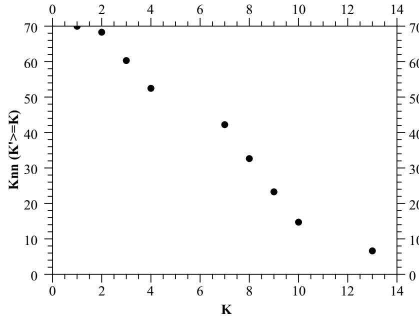

Figure 10: Cumulated K_{nn}-K in Space R

The interactions between nodes is higher in the Space R (nodes have a larger average weight). In the Spaces L and P the interaction magnitudes are very similar (2.03 and 2.08 respectively). Link weights in the network represent multiplicity of connections between stops (Spaces L and P) and between routes (Space R). Passengers have one route (one link in each direction) in terms of average to move from one stop to another.

IV. Conclusion

This research uses the Network Science as a mathematical method to translate networks into graphs, from which important properties are collected. The underlying structure of a network has relevant consequences for its performance.

The subway network of Madrid is studied in three topological Spaces: Space L (stop geographical space), Space P (subway-transferring space) and Space R (Route space). We can conclude:

The study of Space R allows to know the average number and maximum value of the subway routes that a stop joins. These magnitudes are 6 and 13, respectively. The number of shared stops by two specific routes is defined as the weight of the link joining them. The average weight of a link is 4.88. The Degree Probability Distribution shows that there are no routes with a connectivity degree significantly different from the other routes; the connectivity distribution is close to an uniform distribution. Due to assortativity property routes that have many connections tend to be connected to other routes with many connections.

The analysis of Space P allows to know more precisely the accessibility and convenience of the network. The degree of a node symbolizes the number of stops a passenger can go to directly without any change, while the distance between two nodes represents the shortest path between them. The average degree and maximum degree of a node is 29.39 and 99 respectively; the average distance between nodes is 2.26. The average clustering coefficient of a node is high, 0.9, which means that there is high probability that the neighbours of this node (all other nodes to which it is joined by an link) are also connected to each other. In this Space, the Degree Probability Distribution also shows that there are no stops with a connectivity degree significantly different from the other stops. Nodes with many links tend to join other nodes with many links.

In the Space L, one link between two stops exists if they are consecutive on at least one route. The average degree of a node is 2.42. In this Space, the network is assortative. The Degree Probability Distribution shows that many nodes have degree equal to 2.

Regarding robustness, we observe that the network is more robust in Space P than in the Spaces L and R. That is, if a failure occurs in one stop, those stops linked to it will be easily reachable by means of routes that provides a direct service between them, although the following stop on that route will be difficult to reach through one direct link; this network is more sensitive to problems in a route than in a stop.

Global efficiency can be an useful parameter for the assessment of the centrality before and after alterations to the network structure; these alterations can be caused by failures or planning changes.

Our future works will investigate deeper the vulnerability of subway network. We will also build a mathematical model that explains the growth of this network.

References

[1] Y. Z. Chen, N. Li, and D. R. He, ”A study on some urban bus transport networks”, Physica A, vol. 376, 2007, pp. 747−754.

[2] C. Y. Z. Chen, and N. Li, ”The randomly organized structure of urban ground bus-transport networks in China”, Physica A, vol. 386, 2007, pp. 388−396.

[3] K. H. Chang, K. Kim, H. Oshima, and S. M. Yoon, ”Subway networks in cities”, Journal of the Korean Physical Society, vol. 48, 2006, pp. S143−S145.

[4] D. J. Watts, and S. H. Strogatz, ”Collective dynamics of small-world networks”. Nature, vol. 393, 1998, pp. 440−442.

[5] M. E. J. Newman, ”Assortative mixing in networks”, Physical Review Letters, vol. 89, No. 20, 2002, pp. 208701−208705.

[6] M. E. J. Newman, ”Mixing patterns in networks”, Physical Review E, vol. 67, No. 2, 2003, pp. 026126−026139.

[7] J. P. Cardenas, Cardenas, R. M. Benito, M. L. Mouronte, and V. Feliu, ”The effect of the complex topology on the robustness of spanish SDH network”, in Fifth International Conference on Networking and Services,” IEEE Xplore, 2007, pp. 86−90.

[8] R. Criado, B. Hernández, and M. Romance, ”Efficiency, vulnerability and cost: An overview with applications to subway networks worldwide”, Int. Journal of Bif. And Chaos, vol. 17, 2007, pp. 2289−2301.

[9] Z. Zhen-Tao, Z. Jing, L. Ping, and C. Xing-Guang, ”An evolutionary model of urban bus transport network based on B-space”, Chinese Physics B, vol 17, No 8, 2008, pp. 2874−2880.