NETWORKS AND HETEROGENEOUS MEDIA
©American Institute of Mathematical Sciences
Volume **7**, Number **3**, September **2012**
doi:10.3934/nhm.2012.7.415
pp. **415–428**

STRUCTURAL PROPERTIES OF URBAN BUS AND SUBWAY NETWORKS OF MADRID

Mary Luz Mouronte

Departamento de Ingeniería Telemática,
Escuela Politécnica Superior, Universidad Carlos III de Madrid
Ave. Universidad, 30, Edif. Torres Quevedo, Leganés, 28911 Madrid, Spain

Rosa María Benito

Grupo de Sistemas Complejos and Departamento de Fisica y Mecanica,
E.T.S.I. Agrónomos, Universidad Politécnica de Madrid, 28040 Madrid, Spain

Abstract. The goal of this research is to estimate different parameters in the urban bus and the subway networks of Madrid. The obtained results will allow learning more about both types of networks: modularity, most important stops, sensitivity in the district networks (districts with highest and lowest sensitivity), bus line concentration by detected communities, communication capacity for these networks (districts with the greatest and less number of inner and external communications), and relation between network and dweller density by district. This study can help to improve the transport networks: reducing the district sensitivity, adding new stops or routes, etc.

1. Introduction.

The public transportation system of a big city yields social, economic and environmental benefits. It avoids injuries and fatalities caused by car accidents, is less stressful for passengers and improves the environment reducing pollution and road congestion.

Madrid city has a population of 3,254,950 on a extension of 60,683 hectares, and a high developed by public transportation network. It would be possible to improve it by means of a better planning: identification of critical stops, optimization of existing routes (changing trajectories, creating or removing stops).

This paper analizes the characteristics of urban bus and subway networks of Madrid such as stops, routes and their densities. These two transport networks are abstracted in a graph where a set of structural parameters [7], [10] (shortest distance between nodes, betweeness, detection of clusters and robustness) is analyzed both in the entire city and its districts. As a result, a useful help based on network science is provided to learn and improve the features of the transport infrastructure.

There are several works about the transport network characteristics. Majima et al. [6] propose an algorithm to optimize waterbus and bus network to reduce traffic congestion and increase the redudancy of transport system under disaster circumstances. The aforementioned paper investigates five transport networks, one railway, three subways and one hypothetical waterbus lines in Japan and their combinations from the perspective of complex networks. Some of the most important

---

2000 *Mathematics Subject Classification.* Primary: 05C82, 68R10; Secondary: 93A10.
Key words and phrases. Graph, network science, betweenness, robustness, communities.

known results about efficiency, vulnerability and cost for complex networks are reviewed from a mathematical point of view by Criado et al. [2]. Moreover, it includes new results that expand the domain of the theory to the realm of directed networks. This mathematical framework is subsequently used to carry out a comparative study of those performance measures over a significant sample of subway networks worldwide. Criado et al. [3] analized some relationships between the spectral radius and the vulnerability of a network, and some estimations for efficiency and dynamical importance are also given. Kwona and Jung [5] investigate the express bus flow in Korea and its network topology. By using a gravity type model, they find that the bus flow between cities depends on the square root of the product of the population size of both cities. On the other hand, the total bus flow of a city depends on only its population size.

The rest of the paper is organized as follows: Section 2 describes the subway and bus 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.

2. **Subway and urban bus networks of Madrid.** Madrid has an urban bus network with 204 lines and 4,455 stops and, a subway network with 16 lines and 272 stops. It has 21 districts that are administrative regions into which subdivide the city to distribute and manage the exercise of civil or political rights, public functions, and services. All districts of Madrid City, their position in the city map and its number of stops are shown in Figure 1.

As can we see in Figure 1 there is high heterogeneity in the number of stops for district. Figure 2 shows that the districts have quite different population densities and an unequally distribution of stops. The density of stops does not change proportionally to the density of dwellers. The density of bus-stops is greater than the subways one in all districts, reaching the biggest difference in districts 4 (30.68 $Stops/Km^2$) and 7 (30.17 $Stops/Km^2$) and the smallest one in district 8 (1.45 $Stops/Km^2$). Table 1 depicts the number of bus and subway stops by district.

3. Analysis method.

3.1. **Design of network.** In our work, the subway and urban bus networks of Madrid are abstracted in a graph $G = (E; V)$, where E is the set of nodes corresponding to the stops and V is the set of links between them, that connect the consecutive stops.

3.2. **Structural parameters calculation.** In order to characterize both networks we have carried out and study of different topological properties in the graphs: average distances between nodes, links, betweenness, robustness and communities structure [7], [10].

3.2.1. *General characteristics.* The average distance among reachable pairs$< l >$ is 9.797401905 in the whole bus network and 1.779177619 in the subway network, there is a greater number of intermediate stops in the bus network. The total number of links is 307 in the subway network and 10,079 in the bus network. The values of these parameters for all districts are shown in Tables 2 and 3. It can be noticed that in most districts the urban bus network has more internal than external links, with a significant difference in the districts 13 (60.14 % more inner links), 10 (58.82 % more inner links) and 11 (58.12 % more inner links). The opposite occurs in the subway network, with the most relevant discrepancies happening in districts 3

Table 1. Number of stops in the subway and bus networks for different districts in Madrid

District	Subway stops	Bus stops
1	15	168
2	13	171
3	7	142
4	15	180
5	22	263
6	10	115
7	12	153
8	14	368
9	11	329
10	9	293
11	13	283
12	6	177
13	6	277
14	5	149
15	15	259
16	16	278
17	3	186
18	6	173
19	10	102
20	55	262
21	9	127

(66.66 % more external links), 7 (66.66 % more external links) and 12 (66.66 % more external links). The districts are better connected by means of bus network. Figure 3 shows the percentage of external links in the urban bus and subway networks by district. This percentage allows knowing which is the communication degree outwards in a district. It would be possible to detect the district with poorer communications (district 13 in the urban bus network and district 16 in the subway network) .

3.2.2. *Betweenness.* The betweenness b_i of a node i in a network is defined by the expression:

$$b_i = \Sigma_{n \neq i \neq l; n \neq l} \frac{g_{nl}(i)/g_{nl}}{(N-1)(N-2)} \tag{1}$$

where g_{nl} is the number of the shortest paths from node n to node l, $g_{nl}(i)$ is the number of the shortest paths from n to l that pass through node i and N is the number of nodes in the network.

The value of betweenness of a node allow us to measure the relative importance of a stop in the network. High betweenness indicates that the stop has a higher than average likelihood of being on the shortest path from one node to another.

The betweenness is calculated for subway and bus networks (a ranking for each network is showed in Table 4 and Table 5 respectively). In bus network ranking 1 =0.10894763256031012) and in subway network (ranking 1 = 0.25444397705160271).

TABLE 2. Link properties of the bus network of Madrid by district

District	$< l >$	Inner links	external links	% Inner links	% external links
1	9.25997	299	313	48.86	51.14
2	7.90711	283	226	55.60	44.40
3	10.96680	227	161	58.51	41.49
4	9.38603	281	233	54.67	45.33
5	7.67011	403	279	59.09	40.91
6	5.18269	158	131	54.67	45.33
7	4.29790	198	174	53.23	46.77
8	10.94646	606	166	78.50	21.50
9	12.25799	489	194	71.60	28.40
10	14.48633	482	125	79.41	20.59
11	9.63926	472	125	79.06	20.94
12	9.44822	327	119	73.32	26.68
13	10.98836	466	116	80.07	19.93
14	9.33035	192	77	71.38	28.62
15	13.75677	345	182	65.46	34.54
16	10.81246	433	142	75.30	24.70
17	9.60398	285	101	73.83	26.17
18	13.36202	254	71	78.15	21.85
19	6.36318	141	69	67.14	32.86
20	9.69083	372	132	73.81	26.19
21	10.38862	167	63	72.61	27.39

The districts having stops with betweenness values greater than 0.1 in the subway network are 1, 4, 5, 6, 7, 8, 9, 12 and 15, while in the bus network only district 1 reaches that value. Although all nodes in both network have low betweenness, which means there are not crucial nodes to maintain the connections between nodes. The subway network is more susceptible to disruption than the bus network because the betweenness in its nodes is higher, it has higher number of essential nodes.

The betweenness allow us to identify the most important stops in the network. Some actions could be necessary if a failure occurs in these nodes, safeguard mechanisms should be applied in these stops such as carefully planning maintenance work, avoiding traffic lights malfunction, improving the traffic flow in the vicinity or adding other stops in the surrounding area.

3.2.3. *Robustness.* Several researches have estimated the efficiency and vulnerability in networks [1], [2], [3]. We analyze the robustness of transport networks by calculating the value of $< l >$ when three stops with the highest betweenness are removed. We observe in our analysis that the distribution of distances changes drastically in subway networks but not in the bus network, that means that the first one is most robust than the latest one.

We also estimate the sensitivity [1] in$< l >$ as $\zeta_{3l} = | < l > /N - < l^{'} > /N^{'}|$ where $< l^{'} >$ and $N^{'}$ correspond to the values of the average path length and size of networks when three stops with high betweenness are removed. This calculation will help to know the robustness of the networks.

Table 3. Link properties of the subway network of Madrid by district

District	$< l >$	Inner links	external links	% Inner links	% external links
1	1.79167	10	31	24.39	75.61
2	1.90909	5	23	17.86	82.14
3	1.25000	3	15	16.67	83.33
4	2.00000	8	34	19.05	80.95
5	1.46154	8	51	13.56	86.44
6	1.83333	6	22	21.43	78.57
7	1.97917	16	30	16.67	83.33
8	2.15385	10	18	35.71	64.29
9	1.50000	5	22	18.52	81.48
10	1.57143	4	12	25.00	75.00
11	1.76923	7	21	25.00	75.00
12	1.33333	2	10	16.67	83.33
13	2.25000	4	8	33.33	66.67
14	1.66667	3	7	30.00	70.00
15	1.25000	6	30	16.67	83.33
16	2.76744	12	20	37.50	62.50
17	1.33333	2	4	33.33	66.67
18	1.25000	3	8	27.27	72.73
19	2.66667	6	13	31.58	68.42
20	2.07042	32	76	29.63	70.37
21	1.55556	15	12	29.41	70.59

Table 4. Ranking of betweenness in the subway network of Madrid. The stops are represented by their latitude and their longitude (N,W).

Ranking	Subway stop coordinates (N,W)	District
1	40.454262,−3.692436	6
2	40.508562,−3.669477	8
3	40.429113,−3.702199	7
4	40.475615,−3.664284	15
5	40.505234,−3.695805	8
6	40.501579,−3.695204	8
7	40.46889,−3.6763	5
8	40.495428,−3.692887	8
9	40.480186,−3.686256	8
10	40.458458,−3.689904	5
...	...	...
272	40.410098,−3.67424	9

We simulate that an important failure occurs (i.e. three important stops are disabled). ζ_{3l} in the overall bus network is 0.000047611 and in the whole subway network 0.006086097 so it can be stated that the latter network is more sensible to failures, althouh both are quite robust.

Table 5. Ranking of betweenness in the bus network of Madrid. The stops are represented by their latitude and their longitude (N,W).

Ranking	Stop coordinates (N,W)	District
1	40.417131, −3.696851	1
2	40.4171,−3.69834	9
3	40.41709,−3.702693	1
4	40.427725,−3.695397	7
5	40.425062,−3.690462	4
6	40.420227,−3.699195	1
7	40.39951,−3.670549	3
8	40.389312,−3.760071	10
9	40.413365,−3.692822	3
10	40.432788,−3.607872	20
...	...	...
4455	40.370678,−3.645852	18

Table 6. Values of sensitivity in $< l >$ with the three higher betweenness nodes removed (ζ_{3l}) in the bus network of Madrid by district

District	ζ_{3l}
1	0.00087282
2	0.00029139
3	0.00097980
4	0.00044224
5	0.00028929
6	0.00041859
7	0.00004497
8	0.00056029
9	0.00097588
10	0.00082296
11	0.00043239
12	0.00030114
13	0.00049322
14	0.00020509
15	0.00137391
16	0.00022167
17	0.00024932
18	0.00146463
19	0.00047618
20	0.00062220
21	0.00111088

Table 6 depicts ζ_{3l} for the bus network in all districts. The most and least sensitive districts are 18 and 7, respectively.

3.2.4. *Communities.* We have carried out a study of the community structure in the bus and subway networks by measuring the similarities between nodes by means of Walktrap Algorithm [4], [9]. The community calculation allows knowing if the nodes of the network can be grouped into sets of nodes such that each set of nodes is densely connected internally.

This method is based on random walks on G to identify communities. At each step in the random walk, the walker is at a node and moves to another node chosen randomly and uniformly from its neighbors. The sequence of visited nodes is a Markov chain where the states are the nodes of G. An adjacency matrix of N x N dimension $A(G)$ can be built as a bidimensional representation of the relationships between stops, where $A_{ij} = 1$ when a connection between nodes i and j exists and $A_{ij} = 0$ otherwise. At each step the transition probability from node i to node j is $P_{ij} = \frac{Aij}{ki}$, it is an element of the transition matrix P for the random walk. We also compute $D^{-1}A$ where D is the diagonal matrix of the degrees ($\forall_i, D_{ii} = k_i$ and $D_{ij} = 0$ where $i \neq j$). The random walk process is driven by powers of P: the probability of going from i to j in a random walk of length t is $(P^t)_{ij}$ we will denote simply as P^t_{ij} . All the transition probabilities related to node i are contained in the i^{th} row of P_t denoted as $P^t_{i\bullet}$. We then define an inter-node distance measure:

$$s_{ij} = \sqrt{\sum_{q=1}^{n} \frac{(P^t_{iq} - P^t_{jq})^2}{k_q}} = \parallel D^{1/2} P^t_{i\bullet} - D^{1/2} P^t_{j\bullet} \parallel \tag{2}$$

where $\parallel \bullet \parallel$ is the Euclidean norm of R^n . This distance can also be generalized as a distance between communities: s_{CiCj} or as a distance between a community and a node: s_{Cij}

We then use this distance measure in our algorithm. The algorithm uses an agglomerative approach, beginning with one partition for each node ($|\rho| = n$) . We first compute the distances for all adjacent communities (or nodes in the first step). At each step α , two communities are chosen based on the minimization of the mean σ_α of the squared distances between each node and its community.

$$\sigma_\alpha = \frac{1}{n} \Sigma_{C_i \in \rho_\alpha} \Sigma_{i \in C_i} s^2_{iC_i} \tag{3}$$

Instead of directly calculating this quantity first we calculate the variations $\Delta\sigma_\alpha$. Due to the fact that the algorithm uses a Euclidean distance, we can efficiently calculate these variations as

$$\Delta\sigma(C_1, C_2) = \frac{1}{n} \frac{|C_1||C_2|}{|C_1| + |C_2|} s^2_{C_1 C_2} \tag{4}$$

The community merge with the lowest $\triangle\sigma$ is gotten. The transition probability matrix is accordingly updated.

$$P^t_{(C_1 \cup C_2)\bullet} = \frac{|C_1| P^t_{C_1\bullet} + |C_2| P^t_{C_2\bullet}}{|C_1| + |C_2|} \tag{5}$$

and the process is repeated again updating the values of s and $\Delta\sigma$ then performing the next merge. After $n-1$ steps, we get one partition that includes all the nodes of the network $\rho_n = \{N\}$. The algorithm creates a sequence of partitions $(\rho_\alpha)_{1 \leq \alpha \leq n}$. Finally, we use modularity to select the best partition of the network, calculating Q_{ρ_α} for each partition and selecting the partition that maximizes modularity [8].

Modularity Q is defined as the fraction of links within communities minus the expected value of the same quantity for a random network. Let A_{ij} be an element of the networks adjacency matrix and suppose the nodes are divided into communities such that node i belongs to community C_i. Then Q can be calculated as follows:

$$Q = \frac{1}{2m}\Sigma_{ij}\{A_{ij} - \frac{k_i k_j}{2m}\}\delta_{C_i C_j} \tag{6}$$

where the $\delta_{C_i C_j}$ function is 1 if $C_i = C_j$ and 0 otherwise, m is the number of links in the graph, and k_i is the degree of node i. The sum of the term $\frac{k_i k_j}{2m}$ over all node pairs in a community represents the expected fraction of links within that community in an equivalent random network where node degree values are preserved.

The highest modularity in the subway network is 0.722057, with 19 communities, while in the bus network is 0.879664, with 40 communities. By means of community estimation we can know if there is a relationship between identified communities in the whole network and districts in the city. Figure 6 displays the number of detected communities by district, there are several communities by district and the elements in each community belong to different districts.

We also analyze the singularity (S) of a transport line in a district. S is defined as $\frac{\eta}{\tau^2}p_{ij}$, where η is the number of detected communities in the network, τ is the number of communities where the transport line appears and p_{ij} is the probabilty that the transport line i is present at the community j. A bus line which is present in few communities has high singularity but if it exists in many communities has low singularity. In all districts there is a partition on a few bus lines by communities and some of those lines are more significant than others. In the figures 7 and 8 S is plotted as a function of the community (j) for several districts for the subway and urban bus networks.

4. Conclusions.

Subway and urban bus networks of Madrid have many stops and routes servicing a large number of users, so careful planning is needed to ensure their smooth operation. Therefore, a study to know and to help to improve these networks is developed, performing calculations of different topological parameters in the whole Madrid network and its districts (with very different densities of inhabitants and stops). The capital of Spain has an urban bus network with 204 lines and 4,455 stops and a subway network with 16 lines and 272 stops (fewer lines and stops in the latter, as expected). In most districts the urban bus network has more internal than external links (e.g. district 5 with 403 inner links and 279 external links), but in the subway network is just the opposite (e.g. district 5 with 8 inner links and 51 external links). The analysis of the average distance between reachable pairs of stops shows that it is shorter within the subway network, i.e. there is a greater number of intermediate stops in the bus network.

Furthermore, low betweenness values suggest that there are not crucial nodes to maintain the node connections in both networks, although the betweenness in the bus network is higher. From a topological point of view, these stops connect regions of high clustering (containing hubs). The betweenness in the bus network is lower than the betweenness in the subway network. Thereby, several safeguard measures should be taken in the stops with the highest values. With respect to the sensititivity in both networks, the subway network has high sensitivity after the elimination of nodes with highest betweenness whereas the bus network is fairly insensitive, although it has concrete districts with high sensitivity (such as 15 and 21). The bus network has a high level of robustness.

Finally, after examining the existence of structures by district in these networks, it can be stated that there are several communities by district and the elements in each community belong to different districts. Besides, in all districts there is a partition on a few bus lines by communities and some of them are more significant than others.

Acknowledgments. Support from MICINN- Spain under contracts No. MTM2009 - 14621, and i-MATH CSD2006-32, is gratefully acknowledged. We want to express our appreciation to Emilio Garcia for his assistance.

REFERENCES

[1] J. P. Cardenas, et al., *The effect of the complex topology on the robustness of spanish SDH network*, in "Fifth International Conference on Networking and Services," IEEE Xplore, (2007), 2289–2301.

[2] R. Criado, et al., *Efficiency, vulnerability and cost: An overview with applications to subway networks worldwide*, Int. Journal of Bif. And Chaos, **17** (2007), 2289–2301.

[3] R. Criado, et al., *Understanding complex networks through the study of their critical nodes: Efficiency, vulnerability and dynamical importance*, in "International Conference on Modelling and Computation on Complex Networks and Related Topics Net-Works 2007," (2007), 23–30.

[4] B. Fields, et al, *Analysis and exploitation of musician social networks for recommendation and discovery*, IEEE Transactions on Multimedia, **13** (2011), 674–686.

[5] O. Kwona and W. S. Jung, *Intercity express bus flow in Korea and its network analysis*, Physica A: Statistical Mechanics and its Applications, **391** (2012), 4261–4265.

[6] T. Majima, M. Katuhara and K. Takadama, *Analysis on transport networks of railway, subway and waterbus in Japan*, Emergent Intelligence of Networked Agents Studies in Computational Intelligence, **56** (2007), 99–113.

[7] M. E. J. Newman, *The structure and function of complex networks*, SIAM Review, **45** (2003), 167–256.

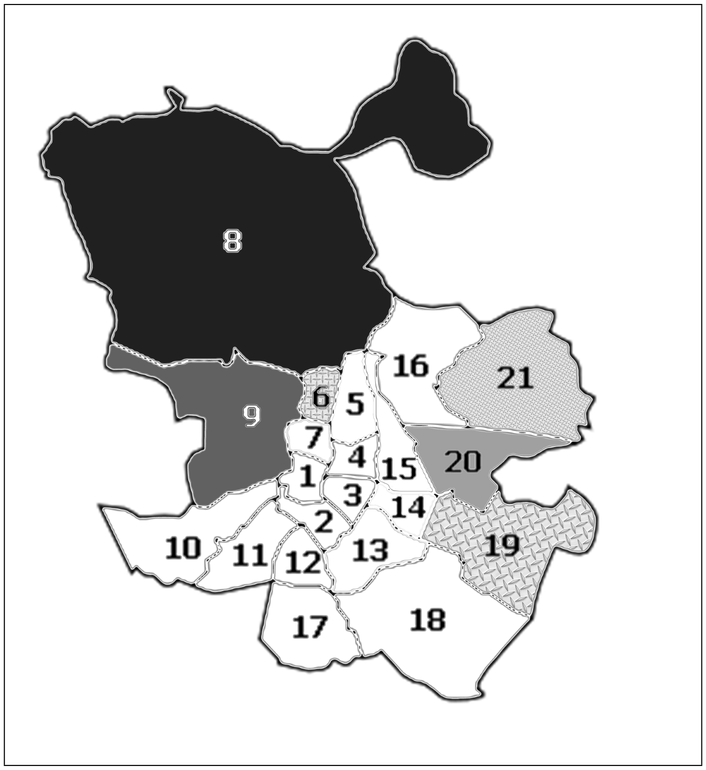
[8] M. E. J. Newman, *Modularity and community structure in networks*, Proc. Natl. Acad. Sci. USA, **103** (2006), 8577–8582.

[9] P. Pons and M. Latapy, *Computing communities in large networks using random walks*, J. Graph Algorithms Appl., **10** (2006), 191–218.

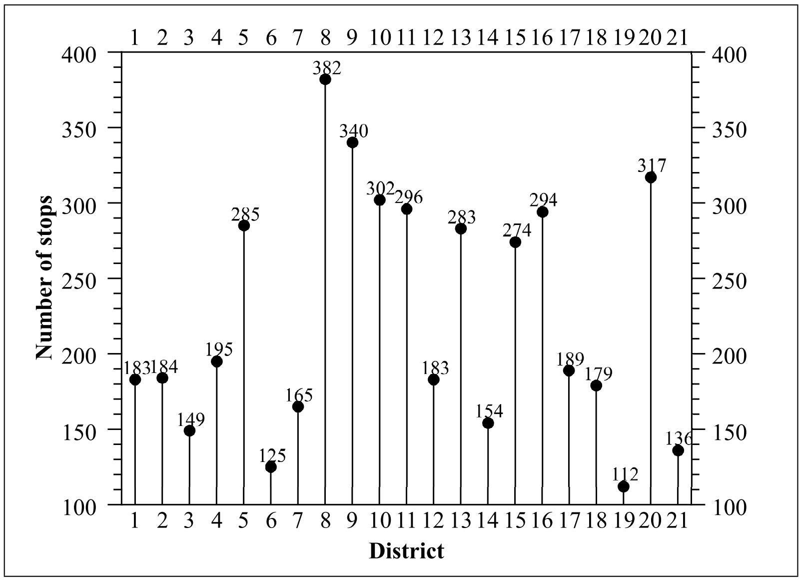
[10] D. Watts and S. Strogatz, *Collective dynamics of 'small-world' networks*, Nature, **393** (1998), 440–442.

Received December 2011; revised June 2012.

E-mail address: `mmouront@it.uc3m.es`
E-mail address: `rosamaria.benito@upm.es`



(a)



(b)

Figure 1. (a) District location inside Madrid city; in plain and meshed grey tones the districts with the highest and lowest number of stops respectively. (b) Number of bus and subway networks by district.

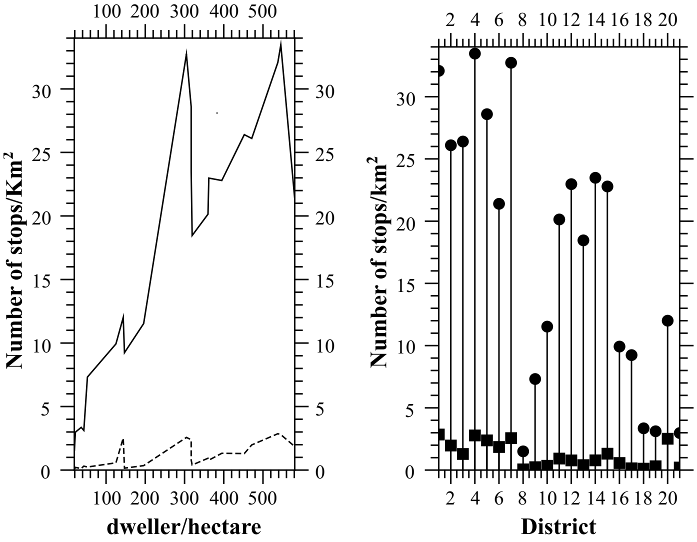


FIGURE 2. Density of urban bus (solid line) and subway (dashed line) networks plotted as a function of the dweller area (left side). Density of urban bus (circles) and subway (squares) networks by district (right side)

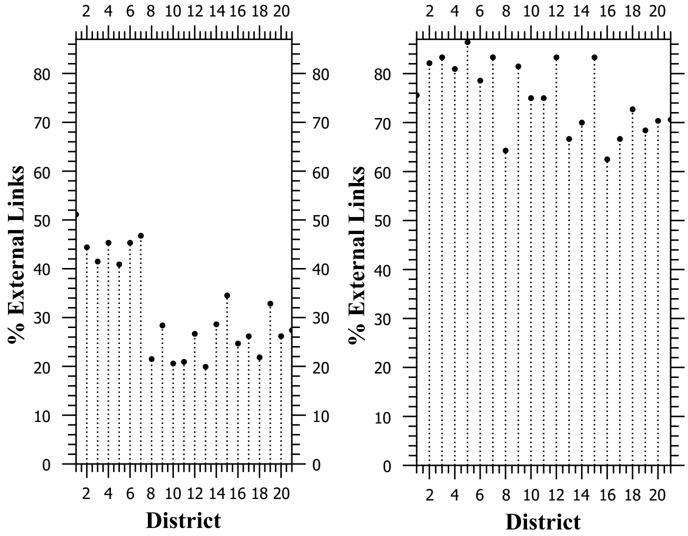


FIGURE 3. Percentage of external links in the urban bus (left) and in the subway (right) networks by district

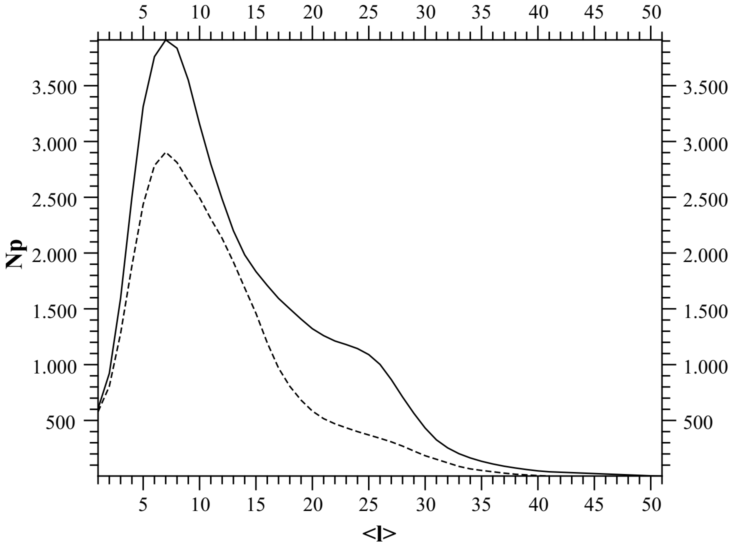


FIGURE 4. Distribution of the number of pairs of nodes Np separated by the shortest distance in the original network (solid line) and in the same network but with the three highest betweenness nodes removed (dashed line) in the subway network

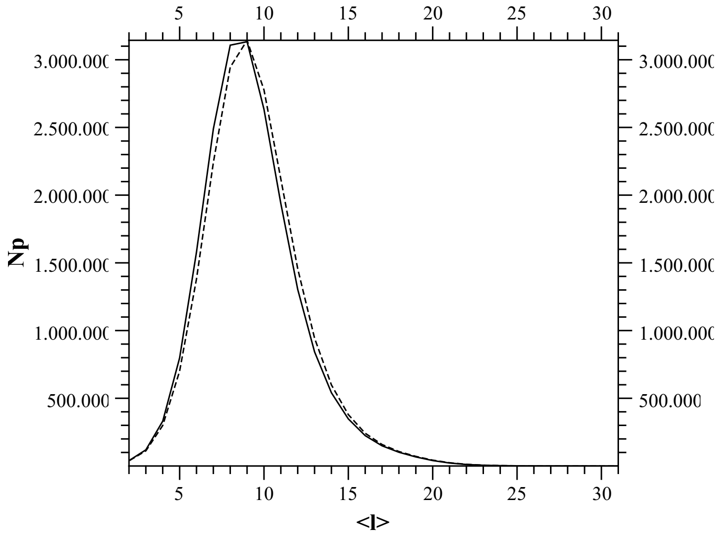


FIGURE 5. Same as Figure 4 for the bus network

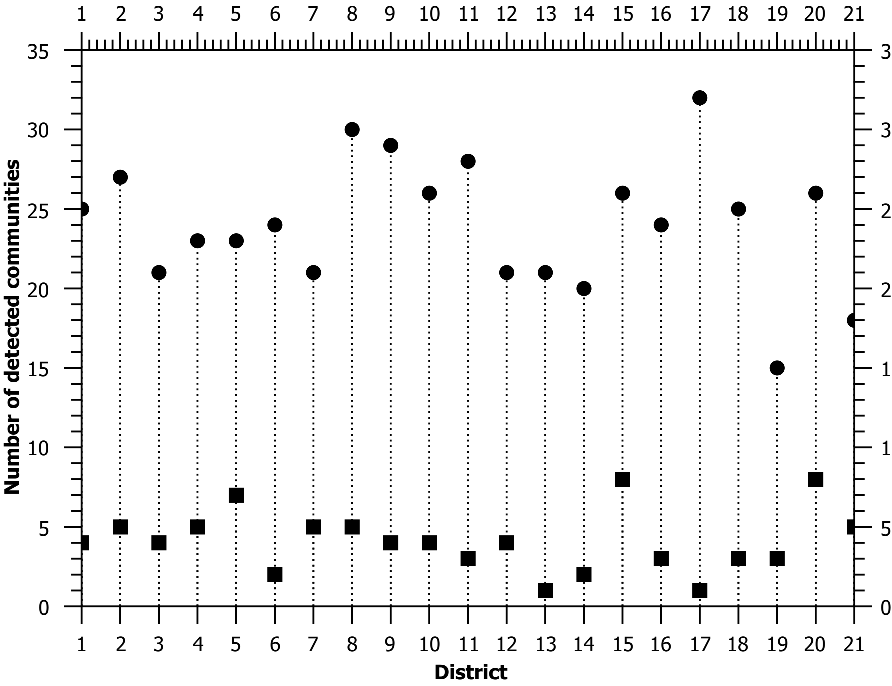


FIGURE 6. Number of detected communities by district of Madrid city. Subway network (squares) and bus network (circles)

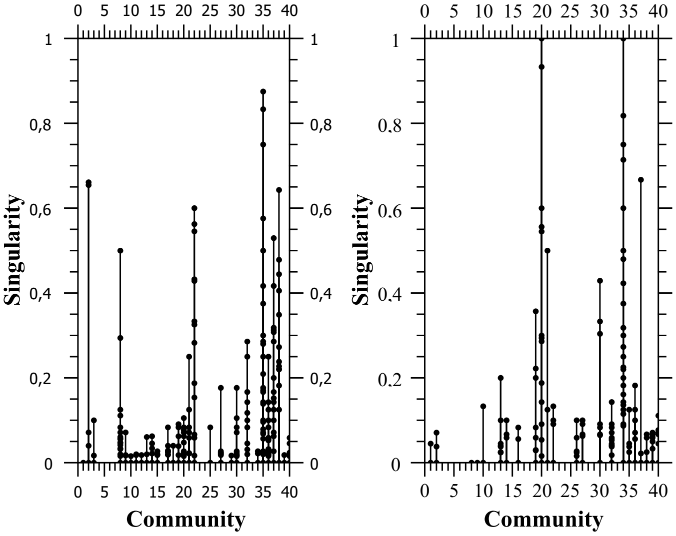


FIGURE 7. Singularity of subway lines in districts 9 (left) and 5 (right) as a function of the communities.

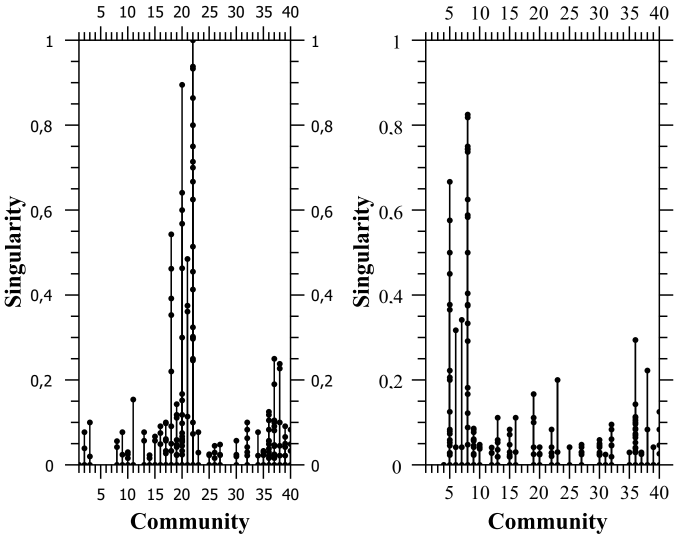


Figure 8. Singularity of urban bus lines in districts 8 (left) and 10 (right) as a function of the communities.