IJCSS International Journal of Complex Systems in Science

Int. J. Complex Systems in Science
vol. **6(1)** (2016), pp. 69–77

Some rankings based on PageRank applied to the Valencia Metro

Francisco Pedroche[1,†], Miguel Romance[2] and Regino Criado[2]

[1] *Institut de Matemàtica Multidisciplinària, Universitat Politècnica de València*

[2] *Department of Applied Mathematics, Rey Juan Carlos University, & Center for Biomedical Technology (CTB), Technical University of Madrid, Spain*

Abstract. In this paper we apply a recent model of Multiplex PageRank to the multiplex network formed with the 9 lines of the metro of Valencia (Spain). We compute the PageRank vector following different approaches and we compare the results with those recently obtained for the Madrid metro system.

Keywords: Google matrix, PageRank, Multiplex networks
MSC 2000: 90B15 ,94C15, 15B51

† **Corresponding author:** pedroche@imm.upv.es
Received: July 27th, 2016
Published: November 30th, 2016

1. Introduction

The definition and computation of centrality measures in multiplex networks is a very active line of research [2], [3], [10]. One of its applications is focused on transportation systems [5], [1]. In this paper we deal with a metro system and we have three goals. First, we want to compare the results obtained by some recent centrality measures based on PageRank. Second, we would like to confirm if the obtained trends when using the Madrid metro system in [12] still hold when dealing with the Valencia metro system. Third, we shall identify some issues that should be taking into account when considering a metro system as a (multiplex) network. Note that our main objective is to compare some centrality measures in these graphs more than offering a reliable tool for these systems. To that end, one should take into account some other features (see, e.g., [8]).

2. Methodology

The Valencia metro and tram system [13] is formed by 9 line stations. The number of stations (stops) in each line is shown in table 1. All the lines are linear graphs except lines 4 and 6 (that have a cycle), and line 4 (that is a tree). The total number of stations is 134. For each line we construct an adjacency matrix of size 134 in the following form: we put the real nodes (stations) of the line and we add the rest of the stations by putting a loop in them (actually, we put a loop in all the stations). In this form we can consider the whole system as a multiplex. For example, line 1 (which is linear) has 40 stations that contribute in the adjacency matrix with 39 links. By putting a loop to all the elements (that is, by putting 1 in the diagonal of the adjacency matrix) we have that the number of nonzero elements of the adjacency matrix corresponding to line 1 is: 39*2 + 134=212. Each of these adjacency matrix represents a layer of the multiplex network.

Line #	1	2	3	4	5	6	7	8	9
Number of stations	40	33	27	33	18	19	16	4	22

Table 1: Number of stations in each line of the Valencia metro.

In each layer, the nodes that do not belong to a line are called *virtual* nodes, while the rest are called *real* nodes. One of the great advantages of our formulation is that we have a very easy tool to distinguish (to penalize, actually) these nodes. It suffices to assign to them a sufficiently small component of the personalization vector in the corresponding layer. We follow the same criterium as in [12] and we take the following value to the component of the personalization vector of any real node

$$v_{real} = \frac{\alpha}{max_r},$$

where α is the usual parameter of Google's PageRank, and max_r is the maximum number of real nodes over all the layers. In this example, $max_r = 40$ (see Table 1). Therefore, $v_{real} = 0.85/40 \approx 0.0212$ in the computations presented in this paper. The rest of the components of the personalization vector in each layer are scaled such as all the components of the personalization vector sum up to 1.

3. Ranking by degree

A first ranking of the lines can be done by considering the degree of each node, that is, the number of links of each station. To perform this computation we

Ranking	degree	Name of the station	Lines
1	5	Empalme	1, 2, 4
2	4	Àngel Guimerà	1, 2, 3, 5, 9
	4	Benimaclet	3, 4, 6, 9
4	3	Alameda	3, 5, 7, 9
	3	Colón	3, 5, 7, 9
	3	Dr. Lluch	4, 6
	3	Jesús	1, 2, 7
	3	La Cadena	4,6
	3	Les Arenes	4, 6
	3	Mediterrani	6
	3	Primat Reig	4, 6
	3	Rosas	3, 5, 9
	3	Torrent	1, 2, 7
	3	TVV	4
	3	Vicent Andrés Estellés	4

Table 2: Top 15 ranking by degree

can construct the graph formed by the union of the 9 graphs corresponding to the lines (we call this graph the *projected* graph). We convert the resulting adjacency matrix to a matrix of $(0, 1)$ elements. That is, we are not weighting by using the number of repeated links. For example, node 74 (Colon) and node 75 (Xàtiva) are connected by three lines that share the same tracks. As a result, the degree of station Colon is 3, although this station belongs to 4 lines (by the way, this is an important difference with the Madrid metro system: in Madrid the majority of the lines do not share the tracks with other lines). The top-15 ranking by the degree of the nodes in the projected graph is shown in Table 2. Note that nodes with the same degree are sorted alphabetically. In the whole system there are 13 stations with degree 3.

4. Ranking by PageRank of the projected graph

By considering the projected graph as before we can compute the usual PageRank. In this case we use an homogeneous distribution of the personalization vector, that is $\mathbf{v} = \mathbf{e}/n$. Note that all the nodes in the projected graph are *real* nodes and therefore we don't need to use virtual nodes. in Table 3 we show the resulting top-15 ranking . Only 107 iterations were needed to obtain convergence with a tolerance for the stopping criterium of $tol = 10^{-10}$ (the

Ranking	PageRank	Name of the station	Lines
1	0.01534094	Empalme	1, 2, 4
2	0.01222489	Benimaclet	3, 4, 6, 9
3	0.01210409	Àngel Guimerà	1, 2, 3, 5, 9
4	0.01123666	Rosas	3, 5, 9
5	0.01108987	TVV	4
6	0.01098181	Torrent	1, 2, 7
7	0.01089436	Vicent Andrés Estellés	4
8	0.00994722	Grau-Canyamelar	6, 8
9	0.00962004	Primat Reig	4, 6
10	0.00942137	Jesús	1, 2, 7
11	0.00926334	La Cadena	4, 6
12	0.00924681	Alameda	3, 5, 7, 9
13	0.00911324	Colón	3, 5, 7, 9
14	0.00891431	Mediterrani	6
15	0.00878211	Las Arenas	4, 6

Table 3: Top 15 ranking by usual PageRank of the projected graph

iterative method stops when the norm of the difference of two consecutive iterated vectors is lower than *tol*).

5. Ranking by multiplex PageRank

By using the multilayer model defined in [12] and the personalization vectors explained above we obtain the top-15 ranking shown in Table 4. 4922 iterations are needed in order to obtain convergence with the same stopping tolerance as before, $tol = 10^{-10}$. In this example, the test spent about 10 seconds of real clock time in a standard computer.

6. Ranking by average PageRank

To serve as a further comparison, in this section we calculate the usual PageRank of each layer. To that end we use the concept of *real* and *virtual* nodes explained above and the corresponding personalization vector. After that, we assign to each node the mean value of the 9 corresponding values of the PageRank in each layer for that node. The top-15 ranking of this average PageRank for each node is shown in Table 5.

Ranking	Multiplex PageRank	Name of the station	Lines
1	0.01011508	Àngel Guimerà	1, 2, 3, 5, 9
2	0.01007682	València Sud	1, 2, 7
3	0.01006990	Sant Isidre	1, 2, 7
4	0.01003807	Paiporta	1, 2, 7
5	0.01003776	Avinguda del Cid	3, 5, 9
6	0.01001567	Safranar	1, 2, 7
7	0.00998696	Nou d'Octubre	3, 5, 9
8	0.00997018	Xàtiva	3, 5, 9
9	0.00995061	Mislata	3, 5, 9
10	0.00994444	Picanya	1, 2, 7
11	0.00992008	Mislata-Almassil	3, 5, 9
12	0.00990124	Patraix	1, 2, 7
13	0.00989700	La Cadena	4, 6
14	0.00988808	Faitanar	3, 5, 9
15	0.00984703	Quart de Poblet	3, 5, 9

Table 4: Top 15 ranking by Multiplex PageRank

Ranking	Average PageRank	Name of the station	Lines
1	0.01410204	Àngel Guimerà	1, 2, 3, 5, 9
2	0.01192986	Benimaclet	3, 4, 6, 9
3	0.01168921	Alameda	3, 5, 7, 9
4	0.01165402	Colón	3, 5, 7, 9
5	0.01078847	Torrent	1, 2, 7
6	0.01060402	Empalme	1, 2, 4
7	0.01054702	Picanya	1, 2, 7
8	0.01042865	Paiporta	1, 2, 7
9	0.01038145	Rosas	3, 5, 9
10	0.01037067	València Sud	1, 2, 7
11	0.01034238	Sant Isidre	1, 2, 7
12	0.01032881	Safranar	1, 2, 7
13	0.01032279	Patraix	1, 2, 7
14	0.01032111	Jesús	1, 2, 7
15	0.01012576	Manises	3, 5, 9

Table 5: Top 15 ranking by average PageRank

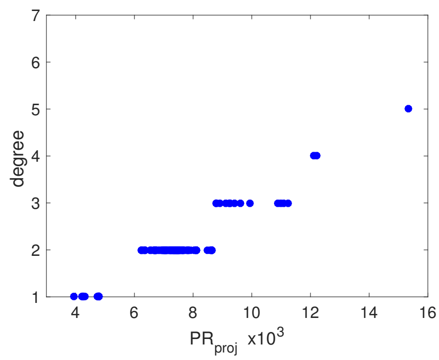

Figure 1: PageRank of the projected graph vs the degree of each node,

7. Comparisons

By using a linear regression we obtain that the values of the degree of each node correlate with the usual (projected) PageRank with a value of the squared coefficient of determination $r^2 = 0.866$. Furthermore, taking into account that both rankings have ties we also employ the *Kendall coefficient* τ *with penalty parameter* (see [6, 11, 4]) to perform this comparison. This value results to be $\tau = 0.680$ and hence there is a soft correlation between degree and PageRank, as it was expected.

By using linear regression we get that the values of the PageRank of the projected graph correlate with the multiplex PageRank with $r^2 = 0.124$. The value of the Kendall coefficient results to be $\tau = 0.081$.

Finally, by using a linear regression we obtain that the values of the average PageRank correlate with the multiplex PageRank with a value of $r^2 = 0.752$ and the value of the Kendall coefficient results to be $\tau = 0.660$.

We also compute the correlation of of the degree of each node with the multiplex PageRank, obtaining $r^2 = 0.234$ and $\tau = 0.576$.

In this results we see that the correlaton between multiplex PageRank and average PageRank is greater than in the case of the Madrid metro system. This could be due to the fact that in Valencia there are more lines than share the same graph and therefore, the mean of the PageRank in each layer is closer to the multiplex PageRank than in the case of Madrid in which each line is represented with a different graph.

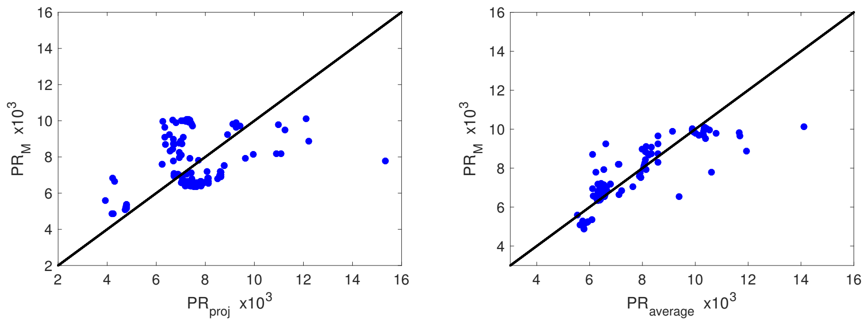

Figure 2: PageRank of the projected graph vs multiplex PageRank (left panel) and average PageRank vs multiplex PageRank (right panel)

8. Conclusions

We have shown that the basic principles outlined in [12] to apply the multiplex PageRank to a metro system (Madrid) still work to the valencian case. In particular, the criterium for assigning values for the personalization component of *virtual* nodes still work here and all the methodology can be adapted. Regarding the networks, we have noticed that there are some differences in the topological properties of the Madrid and Valencia metro systems. For example, in Madrid there's a cycle line, while in Valencia there is not. Other important difference is that in Valencia there a a great number of lines that share the same tracks. One issue that should be taken into account in future works is the convenience of taking into account the number of tracks that a line has (e.g, line 1 in Valencia metro has one track in some parts and two tracks in other parts). Other feature to be taken into consideration is that some lines are connected to others in their endings. As a consequence, the start and ending of a line do not have degree one.

Regarding the results we have obtained that there exists a soft correlation between degree and PageRank of the projected graph (like in the Madrid metro) but a greater correlation between multiplex PageRank and average PageRank than in the case of Madrid. We think this difference is due to the main feature of the valencian metro: some lines share the same tracks, and therefore the mean of the PageRanks of each layer is more similar to the resulting PageRank of the multiplex network than in the Madrid case.

Acknowledgements

This work has been partially supported by the project MTM2014-59906 (Spanish Ministry) and the grant *URJC-Grupo de Excelencia Investigadora* GARECOM (2014-2016).

References

[1] A. Aleta, S. Meloni, and Y. Moreno. *A multilayer perspective for the analysis of urban transportation systems* arXiv:1607.00072v1 (2016)

[2] F. Battiston, V. Nicosia, V., and V. Latora. *Structural measures for multiplex networks.* Physical Review E **89** (2014), 032804

[3] S. Boccaletti, G. Bianconi, R. Criado, C.I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendiña-Nadal, Z. Wang and M. Zanin, *The structure and dynamics of multilayer networks* Physics Reports **544**(1) (2014), 1-122.

[4] R. Criado, E. García, F. Pedroche and M. Romance, *A new method for comparing rankings through complex networks: Model and analysis of competitiveness of major European soccer leagues*, Chaos **23** (2013), 043114. doi:10.1063/1.4826446

[5] S. Derrible, *Network Centrality of Metro Systems*, PLoS ONE, **7**(7) (2012), e40575. doi:10.1371/journal.pone.0040575

[6] R. Fagin, R. Kumar, M. Mahdian, D. Sivakumar and E. Vee, *Comparing Partial Rankings*, SIAM J. Discrete Math., **20(3)** (2006), 628-648. doi:10.1137/05063088X

[7] E. García, F. Pedroche and M. Romance, *On the localization of the Personalized PageRank of Complex Networks*, Linear Algebra and its Applications **439**, 640 (2013).

[8] D. Gattuso, and E. Miriello. *Compared Analysis of Metro Networks Supported by Graph Theory.* Networks and Spatial Economics, 5: (2005) 395414.

[9] A. Halu, R. J. Mondragón, P. Panzarasa,and G. Bianconi, *Multiplex PageRank*, Plos One, Vol. 8, issue 10, 2013.

[10] M. Kivelä, A. Arenas, M Barthelemy, J.P. Gleeson, Y. Moreno and M.A. Porter, *Multilayer networks*, Journal of Complex Networks **2** (3) (2014), 203-271.

[11] F. Pedroche, R. Criado, E. García, M. Romance and V.E. Sánchez, *Comparing series of rankings with ties by using complex networks: An analysis of the Spanish stock market (IBEX-35 index)*, Networks Het.Media **10** (1) (2015), 101-125. doi:10.3934/nhm.2015.10.101

[12] F. Pedroche, M. Romance, R. Criado, *A biplex approach to PageRank centrality: From classic to multiplex networks*, Chaos: An Interdisciplinary Journal of Nonlinear Science . 26, 065301 (2016)

[13] `http://www.metrovalencia.es/`