The Twentieth International Multi-Conference on Computing in the Global Information Technology
ICCGI 2025

An Exploratory Modelling of the Spanish Railway System Through Graphs

Mary Luz Mouronte López
Universidad Francisco de Vitoria
Higher Polytechnic School
maryluz.mouronte@ufv.es

Goals

- A preliminary analysis of the Spanish railway network is carried out in this research using graphs.
- In this preliminary study, some topological properties of the Spanish railway system were analysed (See Methodology Section for details). The suitability of the Erdös Rényi model to reproduce the statistical properties of the network was also studied.
  As is well known, the Erdös Rényi model allows the generation of random graphs.

Background

- Network Theory has been utilised in the analysis of social networks (Barabasi, 2003), economic and financial systems (Kenett & Havlin, 2015), multigraph networks studies in different types of systems (Shafie, 2015), (Sumit et al., 2020), (Wang et al., 2020).
- There are also works related to the study of railway systems through Network Theory in several countries (Hernández & Flores, 2022), (Wang et al., 2020), (Ko & Prokhorchenko, 2013). There are analyses assessing the impact of new transportation modes on the use of the railway system (Wiseman, 2019).
- There is research that studies the expansion of the Spanish railway system and certain indicators referring to improvements made to its infrastructure (Esteban-Oliver & Martí-Henneberg, 2022), (Esteban-Oliver & Martí-Henneberg, 2024). The risk of accidents in the network has also been examined (Mateos et al., 2016).

Information about the data used

The data sources used in the study were: Secretaría de Estado de Digitalización e Inteligencia Artificial del Ministerio para la Transformación Digital y de la Función Pública https://avance.digital.gob.es/es-es/Paginas/index.aspx. Horarios de alta velocidad, larga distancia y media distancia. https://datos.gob.es/es/catalogo/ea0003337- horarios-de-alta-velocidad-larga-distancia-y-media-distancia1. License: Creative Commons Attribution 4.0. License: https://creativecommons.org/licenses/by/4.0/deed.es_ES. Date of creation May 21, 2019 - 0:00 (UTC+02:00). Renfe Data. Horarios de alta velocidad, larga distancia y media distancia. https://data.renfe.com/es/dataset/horarios-de-alta-velocidad-larga-distancia-y-media-distancia License: Creative Commons Attribution 4.0. License: https://creativecommons.org/licenses/by/4.0/ Date of creation November 29, 2018. Last update of data November 8, 2024. Metadata last updated November 29, 2018.

Methodology

- Python (Python, s. fa.), (Python, s. fb.) language was used to process the data and to build a graph in which the nodes are the stations and the links are the existing connections between them.
- Python package NetworkX (NetworkX developers, 2024), (Hagberg, et al., 2008) was utilised, which provides several algorithms to study networks. Specifically, we used those that made possible to calculate the following topological attributes of the railways system:
  - Betweenness,
  - Clustering coefficient,
  - Degree,
  - Eigenvector centrality,
  - Neighbors,
  - Pagerank,
  - Mean shortest path length,
  - Degree assortativity coefficient,
  - Number of connected components,
  - Global efficiency,
  - Local efficiency,
  - Number of bridges,
  - Number of maximal cliques,
  - Average communicability,
  - Number of nodes and links

  The algorithm that implement the Erdös Rényi model was also utilised.

  (see the References Section for more information on the research describing the methods).
- Gephi tool (Bastian et al., 2009) was used to plot the railways system graph.
- R language (R Core Team, 2021) was utilized to plot the distribution of some topological parameters (betweeness centrality, clustering coefficient, degree, eigenvector centrality, neighbors and pagerank) by quartiles.

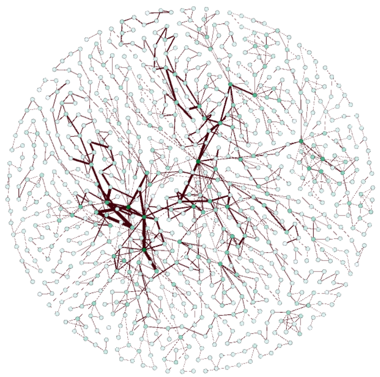
  Note: See R and Python packages used in the References Section.

Results

Some topological attributes of the Spanish railway system

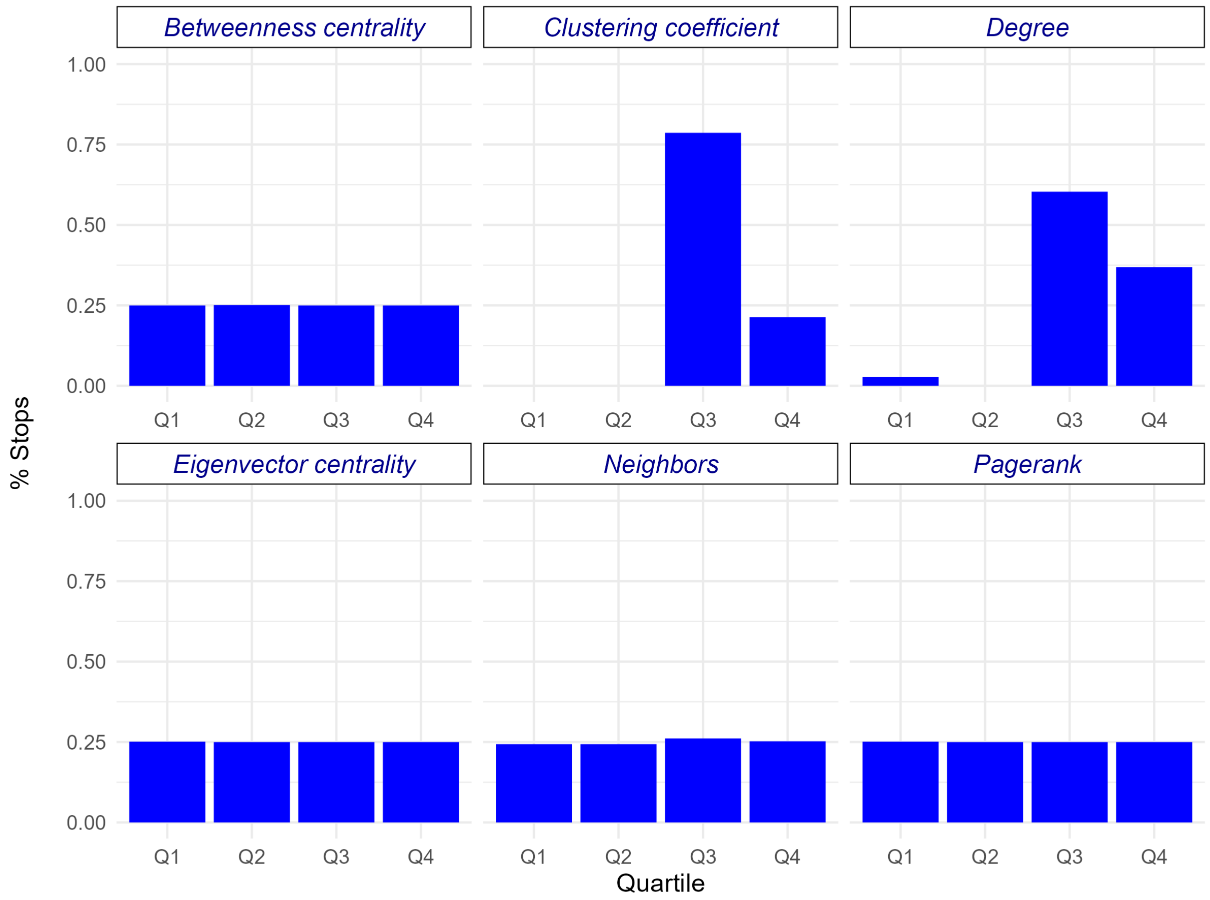
	Maximum	Minimum	Mean	Standard Deviation	Median
Betweenness	0.578	0	0.010	0.033	0.002
Clustering coefficient	1	0	0.2467	0.3565	0
Degree	18	1	2.959	1.978	2
Eigenvector centrality	0.405	0	0.010	0.034	0
Neighbors	15	1.5	3.8901	1.95501	3.5
Pagerank	0.0059	0.0004	0.0013	0.0006	0.0011

Mean shortest path length:9.2490
Degree assortativity coefficient: 0.2473
Number of connected components: 1
Global efficiency: 0.1327
Local efficiency: 0.2656
Number of bridges: 74
Number of maximal cliques: 804
Average communicability: 2.3046 (±0.9598)
Number of nodes: 790
Number of links: 1,161



Graph of the analysed network

Results (II)



Distribution of betweeness centrality, clustering coefficient, degree, eigenvector centrality, neighbors and pagerank by quartile

5 stations with the higher degree:

- Estación de tren Zaragoza-Delicias
- Estación de tren Madrid-Chamartín
- Estación de tren Córdoba
- Estación de tren Madrid-Puerta de Atocha
- Estación de tren Ourense

5 stations with the higher betweenness:

- Estación de tren Pozuelo
- Estación de tren Penyaflor
- Estación de tren Espinosa de Villagonzalo (apd)
- Estación de tren El Romeral
- Estación de tren Puerto Escandón

5 stations with the higher eigenvector centrality:

- Estación de tren Zaragoza-Delicias
- Estación de tren Madrid-Chamartín
- Estación de tren Córdoba
- Estación de tren Madrid-Puerta de Atocha
- Estación de tren Ourense

116 stations show a clustering value equal to 1

5 stations with the higher number of Neighbors:

- Estación de tren Toledo
- Estación de tren Guadalajara – Yebes
- Estación de tren Villanueva de Córdoba-Los Pedroches
- Estación de tren L'Hospitalet de Llobregat
- Estación de tren El Prat Aeroport

5 station with the higher pagerank:

- Estación de tren Zaragoza-Delicias
- Estación de tren Madrid-Chamartín
- Estación de tren Córdoba
- Estación de tren Madrid-Puerta de Atocha
- Estación de tren Ourense

Results (III)

Erdös Rényi model with probability=0.1 (Number of nodes: 790, Number of links: 31,101)

	Maximum	Minimum	Mean	Standard Deviation	Median
Betweenness	0.0019	0.0005	0.0011	0.0002	0.0011
Clustering coefficient	0.1195	0.0833	0.0999	0.0056	0.0998
Degree	104	52	78.7367	8.1651	79
Eigenvector centrality	0.0475	0.0229	0.03538	0.0037	0.0354
Neighbours	82.8333	76.9500	79.5870	0.8735	79.5910
Pagerank	0.0016	0.0009	0.0013	0.0001	0.0013

Mean shortest path length: 1.9005
Degree assortativity coefficient: -0.0060
Number of connected components: 1
Global efficiency: 0.5498
Number of bridges: 0
Average communicability: 2.3046 (± 0.9598)

Results (IV)

Erdös Rényi model with probability=0.5 (Number of nodes 790, Number of links 156,115)

	Maximum	Minimum	Mean	Standard Deviation	Median
Betweenness	0.0008	0.0005	0.0006	0.0006	4.7702×10^{-05}
Clustering coefficient	0.5067	0.4955	0.5010	0.0015	0.5010
Degree	449	348	395.2278	14.6015	395
Eigenvector centrality	0.0404	0.0313	0.0356	0.0013	0.0356
Neighbors	397.3906	393.9976	395.7713	0.5221	395.7765
Pagerank	0.0014	0.0011	0.0013	4.2121×10^{-05}	0.0013

Mean shortest path length: 1.4991
Degree assortativity coefficient: -0.0086
Number of connected components: 1
Global efficiency: 0.7505
Number of bridges:0
Average communicability: 2.305 (± 0.9598)

Conclusions

- Although further analysis is needed to fully characterize the network, the following conclusions were obtained:

  - The nodes with the highest degree centrality are those that exhibit the highest eigenvector and pagerank centralities.
  - The stations with the highest betweenness are not among those with the highest degree centrality. However, these stations, being relevant interconnection points, should be specially protected.
  - None of the stations with the highest degree centrality are among those with the highest average degree neighborhoods. Because of this, a failure at these nodes could have a higher impact.
  - The network presents a degree assortativity coefficient with a low positive value, which indicates that there is a certain tendency of the nodes to connect others with analogous degrees.
  - Efficiencies show a value close to 0.5, which is a medium magnitude.
  - Most nodes exhibit an average clustering and an average degree centrality in the third quartile.
  - There are 804 subgraphs in which all stations are linked with all other. This is a magnitude 1.77 % higher than the number of nodes of the network.
  - A model other than Erdös Rényi model seems to be necessary to reproduce the statistical properties of the network.

- A comparison of the topological properties shown by this network with those exhibited by other railway systems could be of interest in order to detect characteristics that may be intrinsic to railway systems.

- A study relating topological parameters to other travel data (times, number of passengers, etc.) could help to understand the influence that network topology may have on travel characteristics.

Next Steps

- Study of models that could reproduce the main statistical properties of the network.
- Analysis of railway networks in other countries in order to find commonalities and differences between them.

References

- Barabasi, A.-L. (2003). Linked: How Everything Is Connected to Everything Else and What It Means for Business, Science, and Everyday Life. Plume Editions. New York.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: an open source software for exploring and manipulating networks. International AAAI Conference on Weblogs and Social Media. https://doi.org/10.1609/icwsm.v3i1.13937
- Berman, A. & Plemmons, R.J. (1994). Nonnegative Matrices in the Mathematical Sciences. Classics in Applied Mathematics, SIAM, Philadelphia. https://doi.org/10.1137/1.9781611971262
- Brandes, U. (2001). A Faster Algorithm for Betweenness Centrality. *Journal of Mathematical Sociology, 25*(2),163-177. https://doi.org/10.1080/0022250X.2001.9990249
- Brandes, U. (2008). On Variants of Shortest-Path Betweenness Centrality and their Generic Computation. *Social Networks*, 30(2), 136-145. https://doi.org/10.1016/j.socnet.2007.11.001
- Brandes, U. & Pich, C. (2007). Centrality Estimation in Large Networks. *International Journal of Bifurcation and Chaos, 17*(7), 2303-2318. https://dx.doi.org/10.1142/S0218127407018403
- Berge, C. (1958). Théorie des graphes et ses applications. Dunod, Paris, France.
- Bonacich, P. (1972). Technique for analyzing overlapping memberships. *Sociological Methodology, 4*, 176–185. https://doi.org/10.2307/270732
- Bron, C. & Kerbosch, J. (1973). Algorithm 457: finding all cliques of an undirected graph". *Communications of the Association for Computing Machinery (ACM), 16*(9), 575–577. http://portal.acm.org/citation.cfm?doid=362342.362367
- Cazals, F. & Karande, C. (2008). A note on the problem of reporting maximal cliques. *Theoretical Computer Science*, *407*(1–3), 564–568 https://doi.org/10.1016/j.tcs.2008.05.010
- copy Python package. copy — Shallow and deep copy operations. https://docs.python.org/es/3.9/library/copy.html
- Cover, T. M. & Thomas, J. A. (2006). Elements of Information Theory (Wiley Series in Telecommunications and Signal Processing). Wiley-Interscience, USA.
- Costantini, G. & Perugini, M. (2014) Generalization of Clustering Coefficients to Signed Correlation Networks*, PLoS ONE, 9*(2), 1-10. https://doi.org/10.1371/journal.pone.0088669
- Data.table package R. https://cran.r-project.org/web/packages/data.table/index.html
- Erdős, P. & Rényi, A. (1959). On Random Graphs. *Publicationes Mathematicae Debrecen, 6*, 290-297.
- Esteban-Oliver, G. & Martí-Henneberg, J. (2022). The Expansion of the Spanish Railway Network (1848–1941): An Analysis through the Evolution of Its Companies. *Revista De Historia Industrial — Industrial History Review* 31(85), 87-144. https://doi.org/10.1344/rhiihr.37875
- Esteban-Oliver, G. & Martí-Henneberg, J. (2024). The Spanish railway network, 1848–2023. *Revista de Historia Económica / Journal of Iberian and Latin American Economic History, 42*(1), 153-169. https://doi.org/10.1017/S021261092400003X
- Estrada, E. & Hatano, N. (2008). Communicability in complex networks. *Physical Review E, 77*(3), 1-20. https://doi.org/ 10.1103/PhysRevE.77.036111
- Fagiolo, G. (2007). Clustering in complex directed networks. *Physical Review E, 76*(2), 1-16. https//doi.org/10.1103/PhysRevE.76.026107
- Foster, J. G., Foster, D. V., Grassberger, P., & Paczuski, M. (2010). Edge direction and the structure of networks. *Proceedings of the National Academy of Sciences of U S A (PNAS),107*(24), 10815–10820. https://doi.org/10.1073/pnas.0912671107
- Freeman, L. C. (1977). A set of measures of centrality based on betweenness. *Sociometry, 40*(1), 35-41. https://doi.org/10.2307/3033543

References (II)

- Gephi (w. d.). Gephi makes graphs handy. The Open Graph Viz Platform. https://gephi.org/
- Ggplot2 package R. https://cran.r-project.org/web/packages/ggplot2/index.html
- Gilbert, E. N. (1959). Random Graphs. *The Annals of Mathematical Statistics, 30*(4), 1141-1144. https://doi.org/10.1214/aoms/117770609
- Hagberg, A. A., Schult, D. A., & Swart, P. J. (2008). Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008), G. Varoquaux, T. Vaught, & J. Millman (Eds), SciPy Conference – Pasadena, CA, August 19-24, 2008, 11–15. G. Varoquaux, T. Vaught, & J. Millman. https://doi.org/10.25080/HXDH2061
- Harris, C.R., Millman, K.J., & van der Walt, S.J. et al. (2020). Array programming with NumPy. *Nature, 585*, 357–362. https://doi.org/10.1038/s41586-020-2649-2
- Hernández, S. & Flores, I. (2022). Analysis of railway systems applying complex networks: The case of Mexican Railway System. *Ingeniería Investigación y Tecnología, 23*, 1-10. https://doi.org/10.22201/fi.25940732e.2022.23.3.017
- Kendall, M. G. (1955). Further contributions to the theory of paired comparisons. *Biometrics, 11*(1), 43–62. https://doi.org/10.2307/3001479
- Kenett, D.& Havlin, S. (2015). Network science: A useful tool in economics and finance. *Mind & Society, 14*, 155-167. https://doi.org/10.1007/s11299-015-0167-y
- Ko, T. & Prokhorchenko, A. (2013). Investigation into Train Flow System on Ukraine's Railways with Methods of Complex Network Analysis. *American Journal of Industrial Engineering, 1*, 41-45. https://doi.org/10.12691/ajie-1-3-1
- Landau, E. (1895). Zur relativen Wertbemessung der Turnierresultate. Deutsches Wochenschach, 11, 366-369.
- Langville, A. N. & Meyer, C. D. (2005). A survey of eigenvector methods of Web information retrieval. *Society for Industrial and Applied Mathematics (SIAM) Review, 47*(1), 135–161. https://doi.org/10.1137/S0036144503424786
- Latora, V. & Marchiori, M. (2001). Efficient behavior of small-world networks. *Physical Review Letters 87*(19), 1-4 https://doi.org/10.1103/PhysRevLett.87.198701
- Zeng, X., Lu, J., Kerre, E. E. , Martínez, L., & Koehl, L. (2016). Proceedings of the 12th International FLINS Conference (FLINS 2016). Uncertainty Modeling in Knowledge Engineering and Decision Making, Roubaix, France, 24 – 26 August 201. World Scientific. https://doi.org/10.1142/10207
- math Python package. *math — Mathematical functions*. https://docs.python.org/3.9/library/math.html
- McKinney, W. (2010). Data Structures for Statistical Computing in Python. Proceedings of the 9th Python in Science Conference (SciPy 2010), Austin, 28 June-3 July 2010, 56-61. https://doi.org/10.25080/Majora-92bf1922-00a
- Mateos, A., Jiménez-Martín, A. & Pichardo, H. (2012). Risk Assessment of the Spanish National Railway System. Proceedings of the 12th International FLINS Conference (FLINS 2016). Uncertainty Modeling in Knowledge Engineering and Decision Making, Roubaix, France, 24 – 26 August 2016, 10, 1099-1104. https://doi.org/10.1142/9789814417747_0176
- Math package *math — Mathematical functions*. https://docs.python.org/3.9/library/math.html
- NetworkX developers (2024). *Networkx. Network Analysis in Python*. https://networkx.org/
- Newman, M. E. J. (2003). Mixing patterns in networks. *Physical Review E, 67*, 1-14. https://doi.org/10.1103/PhysRevE.67.026126
- Onnela, J.P. , Saramäki, J., Kertész, J., & Kaski, K. (2005). Intensity and coherence of motifs in weighted complex networks. *Physical review E: Statistical, nonlinear, and soft matter physics, 71*(6), 1-5. https://doi.org/10.1103/PhysRevE.71.065103
- Numpy Team (w.d.). *Numpy. The fundamental package for scientific computing with Python*. https://numpy.org/

References (III)

- operator Python package. *operator — Standard operators as functions.* https://docs.python.org/3.9/library/operator.html
- os Python package. *os — Miscellaneous operating system interfaces.* https://docs.python.org/3.9/library/os.html
- os.path Python package. *os.path — Common pathname manipulations.* https://docs.python.org/3.9/library/os.path.html
- Page, L., Brin, S., Motwani, R., & Winograd, T. (1999). *The PageRank citation ranking: Bringing order to the Web*. http://ilpubs.stanford.edu:8090/422/1/1999-66.pdf
- Pandas development team (2020). *pandas-dev/pandas: Pandas (1.4.4) [Computer software]. Zenodo.* https://doi.org/10.5281/zenodo.3509134
- Python (s. fa.). *Python.* https://www.python.org/
- Python (s. fb.). *The Python Language Reference*. https://docs.python.org/3.9/reference/index.html
- R Core Team (2021). *R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.* https://www.R-project.org/
- Saramäki, J., Kivelä, M., Onnela, J. P., Kaski, K., & Kertész, J. (2007) Generalizations of the clustering coefficient to weighted complex networks. *Physical review E: Statistical, nonlinear, and soft matter physics, 75*(2), 1-4. https://doi.org/10.1103/PhysRevE.75.027105
- Shafie, T. (2015). A Multigraph Approach to Social Network Analysis. *Journal of Social Structure, 16,* 1-21. https://doi.org/10.21307/joss-2019-011
- shutill Python package. *shutil — High-level file operations*. https://docs.python.org/es/3.9/library/shutil.html¶
- Sumit J., Ahmed S. M., & Sudipta S. (2020). Network formation with multigraphs and strategic complementarities. *Journal of Economic Theory, 188, 1-35.* https://doi.org/10.1016/j.jet.2020.105033
- Tomita, E., Tanaka, A., & Takahashi, H. (2006). The worst-case time complexity for generating all maximal cliques and computational experiments. *Theoretical Computer Science, 363*(1), 28-42. https://doi.org/10.1016/j.tcs.2006.06.015
- Von Mises, R. & Pollaczek-Geiringer, H. (1929). Praktische Verfahren der Gleichungsauflösung. *ZAMM - Journal of Applied Mathematics and Mechanics / Zeitschrift für Angewandte Mathematik und Mechanik, 9, 152-164.* https://doi.org/10.1002/zamm.19290090206
- Van der Walt , S. & Millman, J. (Eds.). (2010). *Proceedings of the 9th Python in Science Conference* (SciPy 2010)*, June 28 - July 3, 2010, Austin, Texas.* Van der Walt , S. & Millman, J. *https://doi.org/10.25080/Majora-92bf1922-012*
- Varoquaux, G., Vaught, T., & Millman, J. (Eds.). (2008). *Proceedings of the 7th Python in Science Conference* (SciPy2008)*, August 19-24, 2008*, Pasadena, CA. Varoquaux, G., Vaught, T., & Millman, J. https://doi.org/10.25080/PFVC8793
- Wang, W., Cai, K., Du, W., Wu, X., Tong, L., Zhu, X., & Cao, X. (2020). Analysis of the Chinese railway system as a complex network. *Chaos, Solitons & Fractals, 130,* https://doi.org/10.1016/j.chaos.2019.109408
- Wang, Y., Zhu, D., Yin, G. , Huang, Z., & Liu, Y. (2020). A unified spatial multigraph analysis for public transport performance. *Scientific Reports, 10, 1-9.* https://doi.org/10.1038/s41598-020-65175-x
- Wei, T.-H. (1952). The Algebraic Foundations of Ranking Theory. *[Doctoral thesis, University of. Cambridge].*
- Wiseman Y. (2019). Driverless Cars will Make Passenger Rail Obsolete. *IEEE Technology and Society, 38*(2), 22-27. https://doi.org/10.1109/MTS.2019.2913067

Thank you very much!!