

TRANSPORT FINDINGS

How is Intraday Metro Ridership related to Station Centrality in Athens, Greece?

Athanasios Kopsidas¹, Konstantinos Kepaptsoglou¹¹ School of Rural, Surveying and Geoinformatics Engineering, National Technical University of Athens

Keywords: metro ridership, station centrality, network theory, intraday correlation, clustering, Athens, Greece

<https://doi.org/10.32866/001c.75171>

Findings

In this study, intraday correlations between station centralities and ridership at stations of the Athens metro system in Greece are explored. An unweighted L-space representation of the physical metro network is developed, and degree, closeness and betweenness are selected as station centrality measures. Hourly smart-card data are used for representing passenger flows. For station classification, principal component analysis and k-means clustering are utilized. The findings suggest that centrality and ridership usually move in opposite directions, morning peak-hour boardings are completely uncorrelated with station centrality, and metro stations can be classified as ‘central destinations’, ‘averagely central origins’, and ‘underutilized peripheral stations’.

1. Questions

Network theory is valuable for analyzing public transport networks and a topic of growing attention. In this sense, centrality measures have been associated with ridership at stations of urban rail transit, with degree, betweenness and closeness as the most common ones (He, Zhao, and Tsui 2019; Dai et al. 2022). For instance, passenger flows are related to centralities of both physical and service-level networks in tram systems (Luo, Cats, and van Lint 2020), while in metro systems, centralities of the physical network and the network of public transport alternatives demonstrate linear correlations with passenger flows at stations (Kopsidas, Douvaras, and Kepaptsoglou 2023b). Although network centralities seem strongly related to total passenger flows in metro networks, a weaker relation is encountered when it comes to the boarding and alighting volumes, at least at daily level (Kopsidas, Douvaras, and Kepaptsoglou 2023a). But are correlations consistently low during the day, or are there time periods with specific patterns? And if so, what can be learnt from such correlations? In addition, are there specific groups of stations with similar departure/arrival patterns? As such, it is interesting to explore the intraday fluctuations of potential correlations between station centrality and boarding/alighting volumes and classify the stations of a real-world metro system in Athens, Greece.

2. Methods

The Athens metro system is modeled as a complex network and analyzed in the Gephi visualization platform (<https://gephi.org/>). The system consists of 3 lines and 62 stations, with a daily ridership of about 500,000 passengers. Its

Table 1. Methodology equations

Measure	Equation	Components
Degree centrality	$C_i^D = \frac{\sum_j^n e_{ij}}{(n-1)(n-2)}$	C_i^D : Degree centrality of node i e_{ij} : edge formed by nodes i and j n : total number of nodes
Closeness centrality	$C_i^C = \frac{n-1}{\sum_{j \neq i \in N} d_{ij}}$	C_i^C : Closeness centrality of node i d_{ij} : distance from any node j to node i
Betweenness centrality	$C_i^B = \frac{\sum_{s \neq i \neq t \in N} \frac{\sigma_{st}^i}{\sigma_{st}}}{(n-1)(n-2)}$	C_i^B : Betweenness centrality of node i σ_{st} : number of shortest paths between any nodes s and t σ_{st}^i : number of shortest paths passing through node i
Z-score	$Z_i = \frac{C_i - \mu}{\sigma}$	Z_i : Standardized centrality of a station i C_i : Centrality of a station i μ : Average centrality among stations σ : Standard deviation of centrality among stations
Pearson coefficient	$p_{C,R^t}^t = \frac{Cov(C, R^t)}{\sigma_C \sigma_{R^t}}$	p_{C,R^t}^t : Hourly Pearson correlation coefficient between station centrality (C) and hourly ridership (R^t), within timeslot t $Cov(C, R^t)$: Covariance of C, R^t σ_C, σ_{R^t} : Standard deviation of C and R^t , respectively

corresponding network comprises 62 nodes and 128 edges. Three centralities of the L-space physical network are calculated, i.e., degree, closeness and betweenness (equations in [Table 1](#)), and ridership data are coded in hourly-based variables. Unweighted centralities are used as the simplest way to incorporate the fundamental intertemporal design of the physical network, which can capture essential ridership information from a reverse engineering point of view. Ridership data account for the 01/24/23, a representative weekday for the system. Timeslots within 07.00-00.00 are considered, when the system is fully operational. A snapshot of the dataset is presented in [Table 2](#). Subsequently, the correlation between centralities and boardings/alightings is measured through Pearson coefficients. A Pearson coefficient is calculated for every available timeslot, correlating each time station centralities with their hourly ridership (equation in [Table 1](#)). The within-day trend of the coefficients is presented in line graphs in [Figure 1](#).

To classify the stations into groups of similar centrality and ridership patterns, a k-means cluster analysis is conducted. Prior to clustering, the data need undergo principal component analysis (PCA) twice, separately for ridership and centrality data, to achieve dimensionality reduction. The 3 highly correlated centrality measures and 34 collinear ridership variables need to be reduced in essentially fewer factors explaining most of the original variable variance. PCA is very efficient for this reason, since 76.54% of total variance regarding station centrality is captured by a single component. Accordingly, 74.72% and 18.50% of total ridership variance are captured by the first and second components, respectively. The components are extracted if their Eigenvalues are greater than 1. A three-variable k-means clustering is then conducted. A k=3 number of clusters is selected according to the Elbow method, which suggests selecting the k for which the within-cluster sum of squared errors begins to decrease at a decreasing rate ([Figure 2](#)). Silhouette scores are also reported in [Figure 2](#) (the higher the better), but not followed

Table 2. Dataset snapshot

ID	Station	Degree	Closeness	Betweenness	Board_7	...	Board_23	Alight_7	...	Alight_23
1	Acropoli	0.0011	0.1393	0.2557	59	...	164	311	...	34
2	Aghia Marina	0.0011	0.1093	0.1008	841	...	67	482	...	225
3	Aghia Paraskevi	0.0011	0.0737	0.0645	254	...	51	268	...	72
...
62	Viktoria	0.0011	0.1456	0.4033	689	...	145	645	...	158

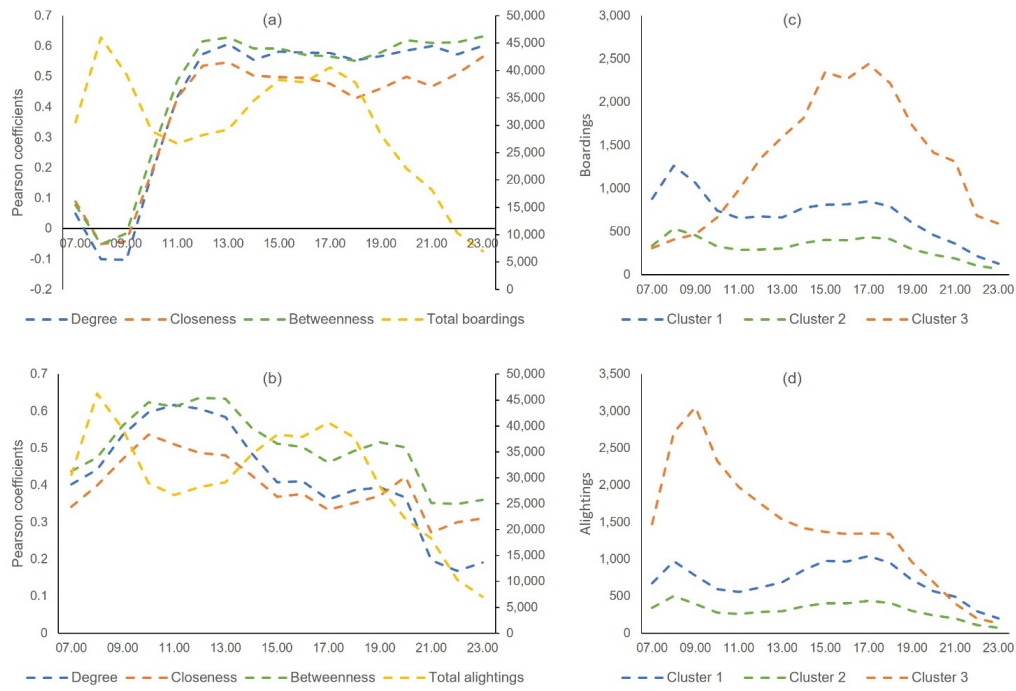


Figure 1. Hourly correlation coefficients between centralities and boardings (a) / alightings (b). Hourly distribution of within-cluster average boardings (c) and alightings (d)

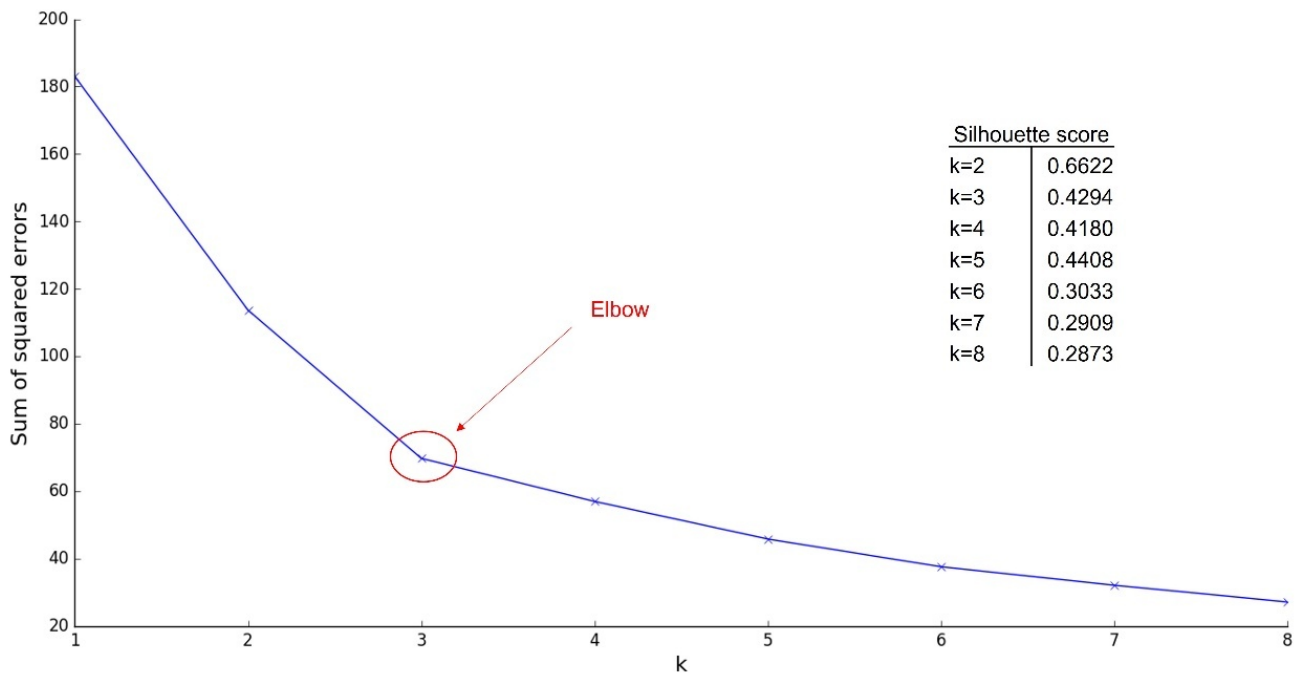


Figure 2. Clustering validation metrics

due to case-specific criteria. All variables are standardized to Z-scores (hence possible negative values) prior to PCA or clustering, as it is necessary to alleviate scale bias (standardization formula in [Table 1](#)). The Athens metro system and station classification are illustrated in [Figure 3](#), average centrality measures are presented in [Table 3](#), while within-cluster average ridership fluctuations are depicted in [Figure 1](#).

Table 3. Within-cluster centrality distribution

	Cluster 1		Cluster 2		Cluster 3	
	Mean	Range	Mean	Range	Mean	Range
Degree	-0.1081	[-1.7831, 3.2419]	-0.1898	[-1.7831, -0.1081]	3.2419	[3.2419, 3.2419]
Closeness	0.1917	[-1.4795, 1.5071]	-0.2231	[-1.7468, 1.5982]	1.8981	[1.8185, 2.0573]
Betweenness	0.1448	[-1.2005, 2.2111]	-0.2661	[-1.2005, 1.3039]	2.7676	[2.4176, 3.0033]



Figure 3. The Athens metro system with station clustering (image retrieved from Wikipedia, edited by the authors)

3. Findings

According to the results, morning boardings are completely uncorrelated with station centrality, but moderately high correlations exist during the rest of the day (within a range of 0.4 to 0.6). As for the alightings, considerable correlations are encountered at noon, while they drop quickly after 14.00. The most interesting finding is that the respective correlations seem to move in opposite directions with total boardings/alightings. That is, when travel demand surges (morning/evening peak-hours), the respective correlation drops, and vice versa. This is a crucial finding regarding disruption

management, since it suggests that excessive travel demand is not channeled mostly to central stations and therefore, it does not create additional traffic burdens to them. This way, the criticality of central stations, which is already high due to topology and high traffic volumes, does not further increase during peak-hours, but relatively to the other metro stations, it drops instead.

Correlation and cluster analysis suggest that boardings during morning peak-hours are mostly related to departures from low and average-centrality stations (residential areas), but morning alightings are mostly related to central stations, providing evidence of work-related mobility. In addition, evening peak-hour departures are related to central stations, contrary to arrivals, signaling commuting, as well. Cluster analysis also suggests the existence of three groups of stations with similar characteristics. The first (Cluster 3) mostly includes the most central and crowded stations, the second comprises peripheral stations with low traffic volumes (Cluster 2), and the third includes averagely central stations (Cluster 1). Interestingly, stations of Cluster 1 are the predominant origins of morning departures, although their total contribution to daily ridership is essentially lower than Cluster's 3. Driven by those characteristics, clusters 1-3 can be labeled as 'averagely central origins', 'underutilized peripheral stations' and 'central destinations'.

All in all, intraday analysis can reveal centrality-ridership correlations within specific timeslots. Such an analysis can highlight specific mobility patterns, and it can thus be a fast, economical and convenient alternative to extended Origin-Destination travel surveys, and a crucial supportive tool for public transport analysis.

Acknowledgments

This work was supported by the Basic Research Program, PEVE 2021, National Technical University of Athens.

Submitted: April 11, 2023 AEST. Accepted: May 05, 2023 AEST. Published: May 13, 2023 AEST.



This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-SA-4.0). View this license's legal deed at <https://creativecommons.org/licenses/by-sa/4.0> and legal code at <https://creativecommons.org/licenses/by-sa/4.0/legalcode> for more information.

REFERENCES

- Dai, Teqi, Tiantian Ding, Qingfang Liu, and Bingxin Liu. 2022. "Node Centrality Comparison between Bus Line and Passenger Flow Networks in Beijing." *Sustainability* 14 (22): 15454. <https://doi.org/10.3390/su142215454>.
- He, Yuxin, Yang Zhao, and Kwok-Leung Tsui. 2019. "Geographically Modeling and Understanding Factors Influencing Transit Ridership: An Empirical Study of Shenzhen Metro." *Applied Sciences* 9 (20): 4217. <https://doi.org/10.3390/app9204217>.
- Kopsidas, Athanasios, Aristeides Douvaras, and Konstantinos Kepaptsoglou. 2023a. "Exploring the Association between Network Centralities and Passenger Flows in Metro Systems," March. <https://doi.org/10.21203/rs.3.rs-2625640/v1>.
- . 2023b. "Extracting Metro Passenger Flow Predictors from Network's Complex Characteristics." In *Complex Networks and Their Applications XI: Proceedings of The Eleventh International Conference on Complex Networks and Their Applications: COMPLEX NETWORKS 2022—Volume 1*, 529–40. Cham: Springer International Publishing. https://doi.org/10.1007/978-3-031-21127-0_43.
- Luo, Ding, Oded Cats, and Hans van Lint. 2020. "Can Passenger Flow Distribution Be Estimated Solely Based on Network Properties in Public Transport Systems?" *Transportation* 47 (6): 2757–76. <https://doi.org/10.1007/s11116-019-09990-w>.