Frontiers in Computing and Intelligent Systems
ISSN: 2832-6024 | Vol. 10, No. 1, 2024

Exploring the Influence from London Development on London Underground Mobility Through Network Analysis

Xinlei Shi

School of architecture, Xi'an University of Architecture and Technology, Xi'an Shaanxi, 710055, China

Abstract: As cities grows up, people continue to gather in cities, and human activities in cities have become more intense than ever before. These human activities will subtly reshape urban space, causing significant transformation of the urban physical environment and socio-economic environment. The method of big data provides a promising perspective for us to explore the impact of human activities on the urban environment. This study investigates the commuting patterns of London citizens in the subway and explores the impact of human activities on the importance of network nodes in subway stations. We use network analysis to characterize the node's importance through degree, closeness, and betweenness centrality, and then compare these centralities with the centrality with the human mobility as weight. The result shows that the node's importance based on centrality characteristics are largely reshaped by human mobility, which in turn drive the government to build new lines to hold the mobility flow. This research is helpful for providing insights for the London subway and urban construction.

Keywords: Network Analysis; London Underground Mobility; Urban Transportation Planning.

1. Introduction

As the population grows, trip frequency increases, and disposable income rises, the spatial structure of large cities is becoming increasingly complex, enabling residents to lead more diverse lifestyles [1]. These trends have significant and noticeable impacts on urban activities [2]. Urban sociology emphasizes that a city is not an autonomous entity with specific institutions and functions but rather a support and expression of human activities [3]. The emergence of subways has effectively facilitated residents' activities, providing convenience for their commuting [4].

Currently, the traditional method for describing human activity relies on surveys. However, this method is time-consuming, labor-intensive, and difficult to scale for large-scale investigations of human activities. Furthermore, the results collected from surveys are influenced by subjective perceptions, which can lead to biases in data quality. The advent of big data offers an opportunity to address these issues [5], [6]. Emerging big data sources, such as mobile signal data, public transportation card data, and social media data, can help provide an objective understanding of human activities [7]. Additionally, with the accelerated process of informatization, the integration of mobile devices with advanced method, such as network analysis allows us to sense how the urban development reshape the human mobility effectively [8], [9].

This article starts from urban transportation data and employs complex network analysis, using underground stations as nodes and constructing a weighted passenger flow network based on big data. It aims to explore the underground network in ideal situations, and when it encounters the real-world mobility pattern after urban development, how the node's characteristics transfer from the ideal situation. The research aims to utilize big data to explore the incompatible between the ideal construction and real-world mobility pattern after urban development. Our research is insightful for providing targeted suggestions for London transportation, encourage them to align the underground construction with real need from citizen's mobility, thereby improving the overall quality of life and satisfaction of residents.

This study uses the London underground as a case analysis and commuter data of urban residents between London underground stations, conducting the following analysis:

(1). Analyzed the degree centrality, betweenness centrality, and closeness centrality of the London underground transportation network.

(2). Compared the trends and influencing factors of the degree centrality, betweenness centrality, and closeness centrality of the London transportation network before and after population weighting.

(3). Provided certain feasible suggestions and recommendations for the current development status of the London underground.

2. Methodology

2.1. Data Process

The data used in this paper is the 2020 London underground commuting data, provided by Transport for London. It describes the origin-destination flows of the average number of people per minute between various stations within a subway network consisting of 438 nodes and 486 edges, along with the corresponding numbers and geographical coordinates of each station.

In the research presented in this paper, we first registered the subway flow data with the station data and conducted data cleaning to ensure the smooth progression of subsequent studies. Specifically, in terms of registration, we initially calculated the total outflow and inflow data for each station. Subsequently, we assigned these data to the locations of the subway stations to form a flow network. In terms of data cleaning, we detected subway station flow data with erroneous records and removed the relevant information of that station during the analysis process.

2.2. Network Analysis

To explore the matching relationship between the urban subway network structure and the mobility needs of London's citizens, this paper quantitatively calculates the degree centrality, betweenness centrality and closeness centrality of the London subway network. By computing the commuting flow-weighted centrality, we observe changes in the importance of the subway network influenced by human mobility.

2.2.1. Degree Centrality

In a network graph, degree centrality is measured by the total amount of direct links with the other nodes. Since as time goes by, the size of the network may vary, to decrease this possible size effect to degree centrality measurement. In this study, degree centrality represents the level of importance of a subway station as a node, as shown in Eq. (1).

$$C_i^D = \frac{\sum_{j=1}^{n} X_{ij}}{(n-1)(n-2)}, (i \neq j) \quad (1)$$

$\sum_{j=1}^{n} X_{ij}$ means the number of links directly connected with node i, and n means the total number of the nodes in an urban public transport network (UPTN).

2.2.2. Closeness Centrality

Closeness centrality is meant to measure one node to the others nodes' sum distances, if the length of node i's shortest paths with other nodes in the network is the shortest path, then node i has a high closeness centrality. It stands for the convenience and ease of connections between the focused node and the other nodes. For the London underground, stations with high closeness centrality have closer connections compared to those with lower closeness centrality. The fundamental formula is Eq. (2).

$$C_i^c = \frac{n-1}{\sum_{j=1}^{n-1} d(j,i)} (i \neq j) \quad (2)$$

where $d(j,i)$ is the shortest-path distance between i and j, and n-1 is the number of nodes reachable from i.

2.2.3. Betweenness Centrality

The betweenness centrality is a measure based on the shortest distance. The shortest path involved is the shortest actual route distance between the two nodes. In a UPTN, a node's centrality is positively correlated with the number of shortest paths passing through the node. The greater the number of shortest paths that pass through a node, the higher the betweenness of the node, and the more obvious it serves as the role of hubs or bridges in the UPTN. This indicator reflects the potential load on the node of the UPTN, which is important for measuring the traffic flow through each node. The betweenness indicator is expressed in Formula (3). In the research of this paper, betweenness centrality represents the importance of a node as a bridge connecting other nodes (or edges).

$$C_i^B = \sum_{s,t \in I} \frac{\sigma(s,t|i)}{\sigma(s,t)} \quad (3)$$

where I is the set of nodes, $\sigma(s,t)$ is the number of shortest (s,t)-paths, and $\sigma(s,t|i)$ is the number of those paths passing through some node i other than s, t. If s=t, $\sigma(s,t)$=1; and if i∈s, t, $\sigma(s,t|i)$=0.

2.2.4. Network Efficiency

The Network Efficiency function calculates the network efficiency of the UPTN, where N represents the total number of nodes in the graph. Network efficiency is a metric that measures the efficiency of information dissemination within a network. It is defined as the sum of the reciprocals of the shortest path lengths between all pairs of nodes, divided by the total number of possible node pairs. This metric reflects the speed of information propagation within the network. For this paper's purposes, a higher Network Efficiency indicates a stronger capacity for disseminating human flow.

The calculation formula is as formula (4):

$$E = \frac{\sum_{i \neq j} \frac{1}{d(i,j)}}{N(N-1)} \quad (4)$$

Where d(i,j) is the shortest path length between nodes i and j, and N is the total number of nodes in the network.

2.2.5. Largest Connected Component

LCC (Largest Connected Component) is a metric measuring the largest connected component's size relative to the overall network scale. A connected component is a set of nodes in a graph where any two nodes are connected by a path. The largest connected component is the largest such set in the graph. LCC calculates the proportion of the size of the largest connected component (LCC) in the network to the total number of nodes N. The size of the largest connected component refers to the largest subset of nodes in a network where every pair of nodes is mutually reachable.

The calculation formula is as formula (5):

$$LCC = \frac{|C_{max}|}{N} \quad (5)$$

Where LCC represents the proportion of the largest connected component. $|C_{max}|$ represents the size of the largest connected component in the network, i.e., the number of nodes it contains. N represents the total number of nodes in the network.

This indicator represents the proportion of the network that remains connected after the removal of any node or edge. A higher LCC value indicates stronger resistance of the network to disconnection.

3. Results

Figure 1 and Table 1 show three types of centralities in the London underground network and the top 10 stations with the highest centrality.

We applied three indicators of urban road network analysis to measure the centrality of various stations in London. Overall, the London underground transportation network covers a wide area with numerous lines and stations, especially in the downtown area, which connects multiple commercial and cultural centers, exhibiting a central radial distribution.

Degree centrality provides a straightforward indication of a station's importance within a transportation network. Nodes with higher degree centrality are typically hubs with significant traffic flow, capable of connecting multiple destinations, and they play a vital role in pinpointing key transportation nodes. According to Figure 1(a), the stations with the highest degree centrality are King's Cross St. Pancras and Baker Street, both scoring 0.0160, followed by Oxford Circus, Green Park, Bank, Earl's Court, and Waterloo, all scoring 0.0137. Among these, King's Cross St. Pancras and Waterloo serve as transportation hubs and crucial transfer points for connecting different railway stations, while stations like Oxford Circus represent bustling commercial centers.

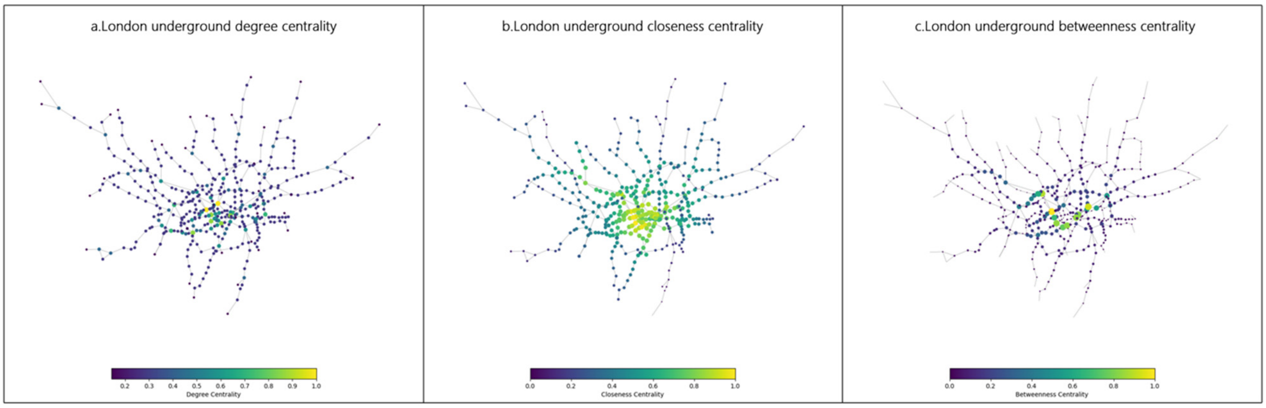

Fig 1. (a) degree centrality; (b) closeness centrality; (c) betweenness centrality

Table 1. The information about top 10 stations ranked in each of the three indicators

station name	degree	station name	closeness	station name	betweenness
King’s Cross St. Pancras	0.0160	Green Park	0.0948	Baker Street	36297.7758
Baker Street	0.0160	Bond Street	0.0937	Bethnal Green	33670.1083
Oxford Circus	0.0137	Westminster	0.0931	Finchley Road	32064.8004
Green Park	0.0137	Baker Street	0.0929	Bank	30443.4417
Bank	0.0137	Waterloo	0.0923	Green Park	30442.4381
Earl’s Court	0.0137	Bank	0.0920	Waterloo	30219.9000
Waterloo	0.0137	Oxford Circus	0.0916	Liverpool Street	29820.7417
Turnham Green	0.0114	Liverpool Street	0.0900	Westminster	27623.5417
Canning Town (DLR)	0.0114	Regent’s Park	0.0893	Bond Street	24635.6532
Liverpool Street	0.0114	Finchley Road	0.0892	West Hampstead	22536.6583

Closeness centrality reflects a node’s overall connectivity within a network, where nodes with higher closeness centrality are better at maintaining the network’s connectivity. In a transportation network, roads with high closeness centrality may be key points for optimizing traffic flow and reducing congestion. As shown in Figure 1(b), the closeness centrality of road traffic also exhibits a "center-edge" distribution pattern, where stations in the central region all have higher values, which then extend and diffuse outward, forming lower centrality values. Green Park, Bond Street, and Westminster are the top three stations with closeness centrality values of 0.0948, 0.0937, and 0.0931, respectively.

Betweenness centrality is based on the measurement of shortest paths. In a transportation network, a node’s centrality is positively correlated with the number of shortest paths passing through it. Roads or areas with higher betweenness centrality may be priorities for resource allocation and optimization. As shown in Figure 1(c), the distribution of betweenness centrality values also exhibits a strong alignment with the transportation lines (comparing it to the London underground map). Baker Street, Bethnal Green, and Finchley Road are the top three stations with the highest betweenness values, at 36297.7758, 33670.1083, and 32064.8004, respectively.

When considering the population mobility patterns of city residents, there will be some shifts in centrality, as shown in Figure 2 and Table 2.

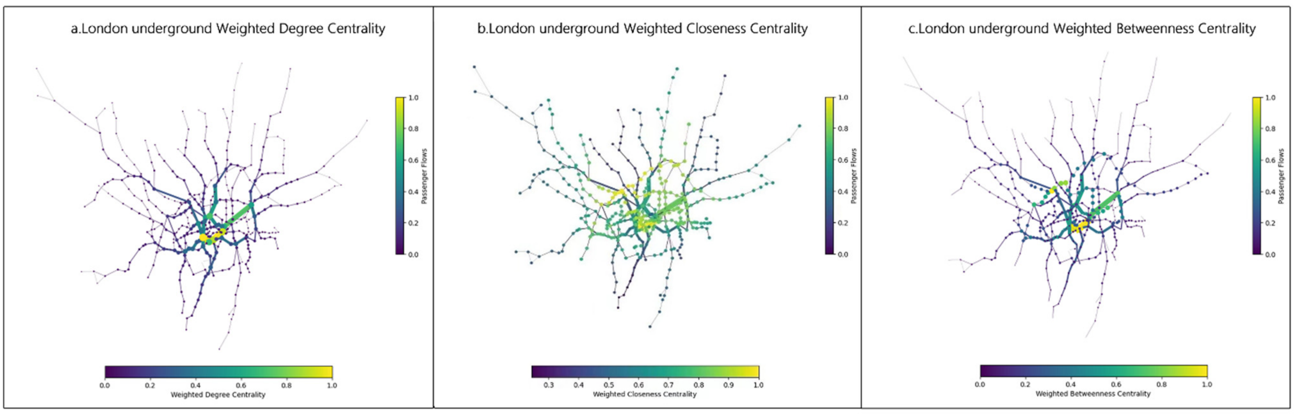

Fig 2. (a) degree centrality; (b) closeness centrality; (c) betweenness centrality

Table 2. The information about top 10 stations ranked in three indicators after weighed by flow

station name	degree	station name	closeness	station name	betweenness
Green Park	652550	West Hampstead	5.7515	West Hampstead	0.3490
Bank and Monument	601435	Finchley Road & Frognal	5.7139	Gospel Oak	0.2948
Waterloo	579510	Hampstead Heath	5.6820	Finchley Road & Frognal	0.2878
King's Cross St. Pancras	465741	Gospel Oak	5.6627	Hampstead Heath	0.2873
Westminster	445085	Brondesbury	5.6449	Stratford	0.2607
Liverpool Street	405495	Upper Holloway	5.6233	Mile End	0.2434
Euston	367930	Crouch Hill	5.5935	Willesden Junction	0.2292
Stratford	364617	Kentish Town West	5.5884	Whitechapel	0.2277
Baker Street	316087	Harringay Green Lanes	5.5674	Brondesbury	0.2054
Oxford Circus	295562	Brondesbury Park	5.5577	Brondesbury Park	0.2040

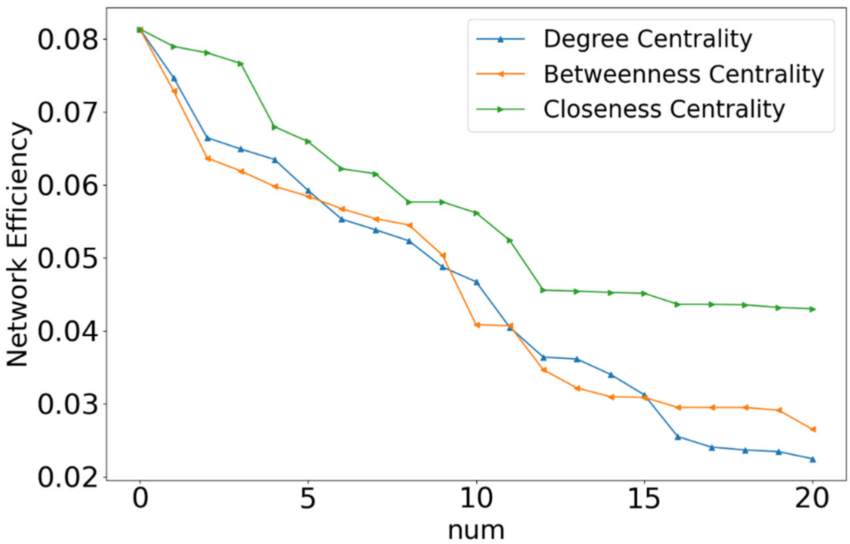

Fig 3. The network efficiency changes through removing the nodes according to degree centrality, closeness centrality, and betweenness centrality with ascending order

From the network efficiency analysis, the three centrality indicators all decrease with the number of deleted nodes. Closeness centrality has a smaller impact on network efficiency than degree centrality and betweenness centrality. The nature of association closeness, closeness A high point means that from a global perspective, it is always on the shortest path for other nodes. When nodes are deleted according to proximity, the loss of the deleted node is more easily borne by other nodes and has less impact on network efficiency. For the network, the impact of efficiency between centrality and degree centrality is more obvious. When the first two points are removed, the network efficiency drops significantly and levels off. This shows that the first two points bear more traffic transportation and passenger transfer responsibilities.

In the in-depth analysis of network efficiency, we observed a clear downward trend in three key centrality measures—closeness centrality, degree centrality, and betweenness centrality—as the number of removed nodes increased. Notably, closeness centrality exhibited a relatively milder impact on network efficiency compared to the other two metrics. This phenomenon can be reasonably explained by the intrinsic nature of closeness centrality: nodes with high closeness values typically occupy critical positions on the shortest paths between many other nodes within the global network structure. Therefore, when nodes are removed based on closeness centrality, their loss of function is more easily compensated by the reconfiguration of alternative paths in the network, mitigating the impact on overall network efficiency. In contrast, betweenness centrality and degree centrality have a more pronounced effect on network efficiency. Specifically, after removing the top two nodes with the highest betweenness or degree centrality, the network efficiency experienced a significant decline, followed by a gradual stabilization of this downward trend. This observation underscores the pivotal roles played by nodes identified by these two metrics. They not only serve as major channels for the flow of information or resources but also act as crucial intermediaries connecting many other nodes. Once these critical nodes fail, the network's connectivity and efficiency suffer considerable damage. In summary, by analyzing the impact of different centrality measures, we not only deepen our understanding of network vulnerability but also gain valuable insights for optimizing network design and enhancing its resilience.

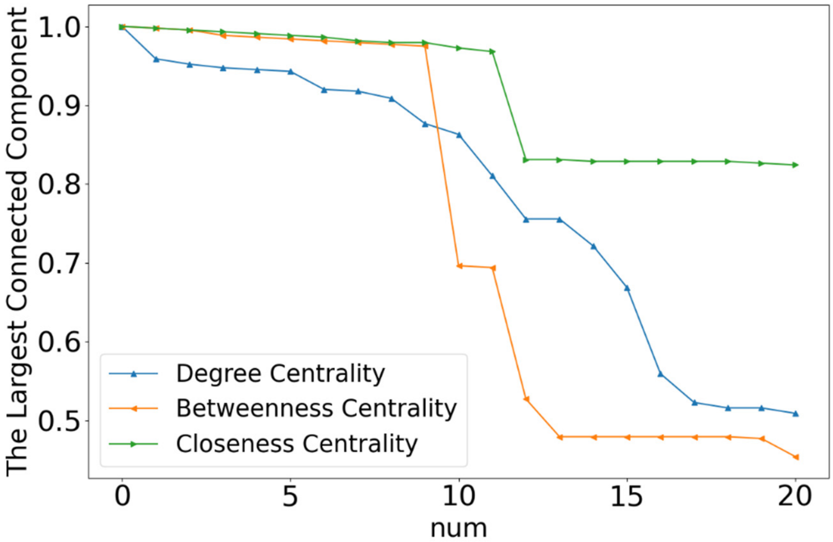

Fig 4. The largest connected component change through removing the nodes according to degree centrality, closeness centrality, and betweenness centrality with ascending order

At the beginning, the trends of closeness and betweenness centrality are very similar, showing little change with the removal of nodes. This indicates that regions with high values of these two centralities are in parts of the network with higher resilience. However, when a certain number of nodes are removed, the size of the LCC of the network decreases rapidly, suggesting that the network collapses into multiple isolated sub-networks. This indicates that the network is highly dependent on these nodes. In contrast, the degree centrality experiences a relatively stable change as nodes are removed.

4. Discussion

In terms of degree centrality, the previously highest-ranked King's Cross St. Pancras has dropped to fourth place, while Green Park has jumped from fourth to first. Oxford Circus has fallen from third to tenth place, while Waterloo has risen from sixth to third. Overall, the increase in passenger flow has further elevated the importance of subway stations that serve transfer functions, while weakening the significance of those that act as commercial hubs. For example, Stratford is a newly developed area. With the construction of railways in this region, Stratford has further assumed the responsibility of transporting residents from surrounding towns to London, which leads to a sharp increase in its passenger flow and results in a rise in its weighted degree centrality. Therefore, after incorporating passenger flow, we can observe the evolving role of Stratford in the urban development process. Regarding betweenness, there is a significant difference between the betweenness centrality with population weighting and the original betweenness centrality results. Specifically, the original betweenness centrality primarily revolves around the central urban area, represented by the commercial zone of Baker Street and the historic area of Westminster. The new betweenness centrality, however, extends more towards the periphery of the city, with an increase in betweenness order represented by West Hampstead and Gospel Oak. From the Figure (1) and (2), we can see that betweenness centrality decreases along the route from the city center to the north, as well as in the eastern London area, while it increases along the northeastern extension of the Overground line. This variation is primarily due to the construction of the Overground subway line, which stimulated population movement along the route and improved connectivity among stations in northern London, thereby enhancing the importance of nodes in that area. In contrast, the saturation of development in the central London area has led to a decrease in population activity within the central zone. For closeness, similar to betweenness, the construction of new subway lines in the northern areas has enhanced the connections among various stations in North London, leading to a collective increase in closeness centrality in that region. In contrast, the saturation of development in the central area has resulted in decreased population activity, and the new subway lines have alleviated passenger flow in the central region, causing a decline in the closeness centrality of subway stations in the city center.

Network efficiency and LCC reflect the resilience of the network from different perspectives. When nodes are removed, network efficiency indicates whether the remaining subway network can compensate for the decline in transmission efficiency caused by the removal of those nodes, while LCC shows whether the subway network remains interconnected after nodes are removed. Regarding network efficiency, degree centrality reflects the importance of nodes within local areas. When nodes are removed based on degree, it significantly impacts local network areas, resulting in a notable decline in network efficiency. In the case of betweenness, it primarily affects inter-regional paths and plays a role in connecting different network structures. Therefore, removing nodes based on betweenness will also lead to a decrease in the global efficiency of the network. Closeness focuses on the efficiency of passage between nodes in the network. When nodes are removed based on closeness, the flow that was handled by the removed node can be more easily shared by its surrounding nodes, leading to a minor impact of closeness on overall network efficiency. LCC is similar to network efficiency in that nodes with high degree centrality cause less global disruption but have a more significant localized impact. As a result, degree centrality continues to influence LCC, though changes are relatively small. Betweenness is crucial for the overall connectivity of the network, closely tied to its properties, as it focusses on the crucial edges that connect different clusters. Thus, the influence of removing nodes based on betweenness will significantly broke the clusters in the whole network, leading to rapid network fragmentation. While closeness centrality relies on the connectivity of neighboring nodes, it emphasizes the shortest path for one node with all the other nodes in the network. So, when removing nodes according to closeness centrality, the loss caused by node removing can be effectively support by neighboring nodes, reduce the impact on LCC. Especially for betweenness, the removal of the nodes at West Hampstead and Stratford significantly reduces the network's LCC, indicating that these two points are crucial hubs connecting sub-networks. In fact, Stratford serves as a transportation hub linking East and West London, while West Hampstead is a key transit point connecting Central London with the northern areas. Regarding closeness, the removal of the Victoria node also leads to a rapid decrease in the network's LCC, highlighting the importance of Victoria for connectivity in the local network. As a major transportation hub in Central London, Victoria plays a vital role in connecting various parts of the city center. Therefore, both network efficiency and LCC effectively help us identify these critical nodes within the network, playing an important role in enhancing the resilience of the network. Overall, the construction of the subway in the London area has effectively alleviated the pressures of resident activities in the city center, activated the station activity in the northern regions, and strengthened the overall resilience of the network.

5. Conclusion

This article analyzes the differences in degree, betweenness, and closeness centrality of the London underground under population mobility weighting compared to the original network. It examines the changes in usage patterns of the London underground resulting from the urban development in London, as well as the impact on network resilience. Overall, we find that the continuous development of suburban areas and adjacent towns has enhanced the importance of certain hubs connecting the suburbs with downtown, or linking nearby towns to London, such as Stratford and West Hampstead. Additionally, development in the northern regions of London has made subway stations in the north more active. In summary, the urban development of London has led to a more balanced overall growth, with population movement becoming more stable. At the same time, for key nodes like West Hampstead and Stratford, the London area should continue to closely monitor these points and enhance network resilience. The conclusions of this study hold significant importance for the ongoing development of the London underground.

References

[1] H. Wang et al., 'A multivariate hierarchical regionalization method to discovering spatiotemporal patterns', GIScience Remote Sens., vol. 60, no. 1, p. 2176704, Dec. 2023, doi: 10. 1080/ 15481603.2023.2176704.

[2] Z. Li et al., 'Exploring the association between multi-mode transport and the built environment: A comparative study of metro, bus, taxi, and shared bike use', Sustain. Cities Soc., vol. 114, p. 105789, Nov. 2024, doi: 10.1016/j.scs.2024.105789.

[3] X. Liu and H. Xia, 'Networking and sustainable development of urban spatial planning: Influence of rail transit', Sustain. Cities Soc., vol. 99, p. 104865, Dec. 2023, doi: 10.1016/ j.scs. 2023.104865.

[4] F. Ma, F. Ren, K. F. Yuen, Y. Guo, C. Zhao, and D. Guo, 'The spatial coupling effect between urban public transport and commercial complexes: A network centrality perspective', Sustain. Cities Soc., vol. 50, p. 101645, Oct. 2019, doi: 10.1016/j.scs.2019.101645.

[5] S. Porta, P. Crucitti, and V. Latora, 'The Network Analysis of Urban Streets: A Primal Approach', Environ. Plan. B Plan. Des., vol. 33, no. 5, pp. 705–725, Oct. 2006, doi: 10.1068/ b32045.

[6] G. Pflieger and C. Rozenblat, 'Introduction. Urban Networks and Network Theory: The City as the Connector of Multiple Networks', Urban Stud., vol. 47, no. 13, pp. 2723–2735, 2010, doi: 10.1177/0042098010377368.

[7] B. Schaller, 'The New Automobility: Lyft, Uber and the Future of American Cities', Jul. 2018, Accessed: Oct. 01, 2024. [Online]. Available: https://trid.trb.org/View/1527868.

[8] S. Chen and D. Zhuang, 'Evolution and Evaluation of the Guangzhou Metro Network Topology Based on an Integration of Complex Network Analysis and GIS', Sustainability, vol. 12, no. 2, Art. no. 2, Jan. 2020, doi: 10.3390/su12020538.

[9] M. Li, W. Yu, and J. Zhang, 'Clustering Analysis of Multilayer Complex Network of Nanjing Metro Based on Traffic Line and Passenger Flow Big Data', Sustainability, vol. 15, no. 12, Art. no. 12, Jan. 2023, doi: 10.3390/su15129409.