Motif analysis of urban rail transit network

Yunfang Ma *, Jose M Sallan, Oriol Lordan

Department of Management, Universitat Politècnica de Catalunya, Barcelona, Spain

ARTICLE INFO

Article history:
Received 11 January 2023
Received in revised form 27 April 2023
Available online 6 July 2023

MSC:
00-01
99-00

Keywords:
Rail transit network
Topology
Network motif
Motif detection
Motif decomposition

ABSTRACT

The connectivity of the rail transit stations is an effective way to evaluate the economic value of the station. In order to explore the influence of the local structure in the complex system of urban rail transit, this paper constructs complex network models based on the Beijing rail transit real network by using the space P and space L methods. The topological structure and global characteristics of the line and station networks are analyzed, and all motifs from 3 to 8 nodes of both networks are obtained using the motif detection algorithm. A subgraph decomposition algorithm for complex network motifs is designed based on five typical subgraphs. The results show that both networks of Beijing rail transit have different typical numbers and distributions of subgraphs, with Y-shaped subgraphs and line subgraphs being the most common for high-node and low-node motifs, respectively. This research proposes a way to assess the connectivity of the rail transit system and has significant reference value for optimizing network functions in the design and planning of rail transit networks. The findings also contribute to the ongoing discussions on network reliability and resilience, as the subgraphs identified in this study could potentially have implications for the network's performance under different scenarios.

© 2023 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Introduction

Rail transit is critical to developing and achieving a wide range of social, economic, and environmental objectives [1]. Urban rail transit is a typical complex system, and the characteristics of its complex network are of great significance for the research and resolution of urban transportation-related problems and the rational planning of transportation networks [2]. The interaction of many stations and lines compose the urban rail transit network, which plays a vital role in the circulation of people and logistics [3]. The primary function of the transportation network is to transport passengers or materials quickly and efficiently. Its network design should establish service routes between nodes as much as possible to achieve high connectivity between networks, but at the same time, it is necessary to avoid waste of resources.

The urban rail transit network is complex because of its components, such as station connection, route planning, and passenger flow loading [4–6]. However, the complex network's components still have a certain hierarchy level, and there are also nonlinear interactions between the components. The complex system gradually integrates from low to high [7]. The planning and construction of transportation infrastructure is an essential indicator of economic development and urbanization. For the local structure of a complex network, the more edges, the higher the functional connectivity. To uncover the structural design principles, [8] defined network motifs as patterns of interconnections occurring in complex networks at numbers significantly higher than those in randomized networks. As the essential element structure of a complex network, the network motif constitutes the entire network from bottom to top, representing frequently occurring

* Corresponding author.
E-mail address: yunfang.ma@upc.edu (Y. Ma).
functional subgraphs in real networks. Therefore, the entire network's structure and function are profoundly significant from the rail transit network's motif level.

Establishing a scientific analysis model is the first step in examining the rail transit network's characteristics. The modeling methods mainly include Space P and Space L [9,10]. Space-wised modeling methods have gradually become mainstream, which have laid a good foundation for exploring the rail transit system. Heretofore, most studies analyze the network's local and overall characteristics from indexes based on complex network modeling, including degree, shortest path length, degree distribution, clustering coefficient, etc. [10]. Network motifs have relatively extensive research in systems biology and statistical physics [11–13], but they still need to be developed in management science and engineering applications [14]. The existing research on the rail transit network motif has realized the gradual improvement from 3 to 8 nodes, and the subgraph is described from the bottom up by the three-node motifs [15,16]. However, such studies remain narrow in focus dealing only with the three-node subgraph, which is a common subgraph in all subgraphs and has the same structure and functional connotation, not enough to describe the local network's actual function, but very little attention has been paid to the motif's structure in the rail transit network.

Although numerous studies have focused on rail transit networks, there is a lack of research on the rail transit network motif. The contribution of this study is threefold. Firstly, using lines and stations as the penetration point, explore the structure of the rail transit network from the macro and micro perspective to see whether it has achieved the optimal allocation of resources.; secondly, from one station to another, sometimes the time is the same, but there can be different paths. If it does not improve travel efficiency from a global perspective, is it necessary to exist for different paths? The internal structure of such different paths belongs to the network's local structure; therefore, one of the goals of this paper is to explore the internal structure of the rail transit network from the perspective of network motifs. Thirdly, we aim to design a motif decomposition algorithm to describe the function of the network motif so that we can know how the network motif's structure affects the network's operation.

This paper takes Beijing's rail transit network as the research object and conducts modeling and analysis from station circulation and line connection. We discovered two networks' motifs based on a motif detection algorithm. We designed the motif decomposition algorithm to explore characteristics of the local structure in the complex rail transit system based on five typical subgraphs: star subgraphs (Q1), line subgraphs (Q2), Y-shaped subgraphs (Q3), ring subgraphs (Q4), and fully connected subgraphs (Q5). The proposed algorithm can target all motifs from 3 to 8 nodes to realize the complex rail transit network's local structure and function. The motif analysis mechanism of different nodes is established, which realizes the study of the primitive characteristics of the Beijing rail transit network, reveals the regularity of the network structure, and further reveals and understands the network motif research and the transportation network structure.

2. Literature review

The most straightforward representation of the rail transit network is that the nodes represent the stations and the links represent the physical connections [17,10] describes the design and implementation of Space-P and Space-L methods in Shenzhen metro and proposes that the Space-L model has higher centrality, while the Space-P model has a more robust overall anti-attack capability. [18] obtains the topological network by Space-P and develops the improved local-world evolving rail transit network model to reflect the transportation network's real characteristics. Therefore, Space-wised methods are effective ways for building rail transit network. Besides, topology analysis is a mathematical tool used to study the properties of space and objects, and it can be applied to rail transit networks in several ways [19,20] gives an account of the connection between complex network topology, traffic behavior, and maximal connected component. [21] attempts to show the correlation between the network's topological properties and traffic volume. [22] discusses the case of Boston and Vienna urban rail transit networks on the topological characteristics from clustering coefficient, path length, and average vertex degree. [23] contests the claim that the star structure is the optimal traffic network structure when considering the cost of traffic congestion. [24] argues that the urban rail transit network with preferential connection characteristics can develop into a scale-free network, and the node distribution follows the power law. [25] investigates the factors that determine most subway networks are scale-free by topological analysis of 33 subway networks. Comprehensively, topology analysis can provide valuable insights into the structure and behavior of rail transit networks, enabling transit planners and operators to make informed decisions about network design, operation, and maintenance.

Network motif has been proposed for almost two decades and used in multiple areas, but awareness of motif in management science is just in recent ten years. [26] proposes two motif-based extraction methods for extracting the functional backbones of complex networks based on the higher-order organization of salient motifs. [27] considers the implications of network motifs on China's Airline Network and argues that adjusting the number of proper network motifs is helpful to optimize the overall structure of airline networks, which is profitable for air transport sustainable development. [28] focuses on the motif-based analysis of network resilience and reliability under various intentional attacks, revealing networks' local dynamics and vulnerability. [29] identifies the terror hubs and vulnerable motifs of complex networks and finds that star structures frequently occur in the network, meaning one source could attack many targets and vice versa. [30] provides an overview of the overview of the subgraph structure in the air traffic network and proposes that lower-connected subgraph, medium-connected subgraph structures, and higher-connected subgraph structures are critical to improving the overall capacity of the network. [30] pays attention to the riders' travel behavior
Y. Ma, J.M. Sallan and O. Lordan Physica A 625 (2023) 129016

using temporal motifs. The variation in temporal motifs across travelers from different public transportation modes reveals the difference in the travel behavior of users, which has verified the importance of the motif in public transportation networks. [31] characterizes the urban transportation networks by using network motifs, but the weakness is that it only simply describes and compares the frequency and distributions of different sizes of the motifs. The application of network motif analysis in rail transit is still at an early stage, but it has great potential to improve our understanding of network behavior, optimize network design, and enhance system resilience.

Overall, the reported studies provide essential insights into rail transit networks. According to previous studies, network motif analysis is a practical and helpful way to uncover the local structure and display the stations’ connectivity preference, evaluating structural rationality. A systematic understanding of how network motifs contribute to rail transit is still lacking, which could reveal the connection details of the local rail transit network. Applying motif analysis, we can better understand the existing rail transit networks, which can help optimize the layout of the rail transit networks.

3. Data and methods

3.1. Data description

This paper takes Beijing rail transit as an example for the research. Up to December 2021, the Beijing Subway consists of 24 lines, including 19 rapid transit lines, two airport rail links, one maglev line, and two light rail lines [32]. If the stations are not double-counted, there are 359 operating stations. Tongyunmen station in line 6, Gaojiayuan station, Taoranqiao station in line 14, Suzhou Jie station in line 16, and Laoguanli station in line Yizhuang T1 suspended activation before 2021. We will not consider all of these stations. Finally, there are 24 lines and 354 stations for Beijing rail transit. The Beijing rail transit line is shown in Fig. 1.

3.2. Space P and Space L

Space-P connects nodes if there is at least one route between them, and Space-L connects if consecutive stops are on a given route. Many studies employ either Space-L or Space-P to study the rail transit network [10,18,33]. The space-P method can intuitively represent travel routes under different modes of transportation, which is more consistent with people’s cognitive habits, and it can efficiently perform path planning and path search, such as determining the shortest and fastest paths [34]. In contrast, Space L can intuitively represent the physical characteristics of the rail transit network, such as the length, and the number of stations, which is more consistent with reality, and can conveniently perform network topology analysis, such as determining the shortest path and minimum spanning tree [10]. We will use Space P to construct a Beijing rail transit (BRT) line network and Space L to construct a Beijing rail transit (BRT) station network.

BRT line network is composed of all transit lines. We construct an adjacency matrix model based on the Space P method regarding the Beijing rail transit line as an independent node. If there is a service line between two nodes, connect the edge to get the BRT line network graph. For example, regarding line 1 as node 1 and line 3 as node 2, we get an adjacency matrix of order 24×24 as follows:

\[ C_{ij} = \begin{cases} 1, & v_i, v_j \text{ connected} \\ 0, & v_i, v_j \text{ disconnected} \end{cases} \]

When \( C_{ij} = 1 \), line \( i \) and line \( j \) intersect, which means line \( i \) can transfer to line \( j \) directly. When \( C_{ij} = 0 \), two lines are not connected.

BRT station network is composed of all transit stations. To regard each station as a network node, connect the edge if there is a line between two nodes. Based on Space L, we construct the BRT station network. We get an adjacency matrix \( D \) of order 354×354.

\[ D_{ij} = \begin{cases} 1, & v_i, v_j \text{ connected} \\ 0, & v_i, v_j \text{ disconnected} \end{cases} \]

The matrix has the following features:

(1) The constructed graph is undirected since the stations are interconnected. That is, the adjacency matrix is symmetric, and the main diagonal is 0;

\[ D_{ii} = 0 \]

(2) The degree of any vertex \( v_i \) is the number of all non-zero elements in column \( i \) (or row \( i \));

\[ K_i = \sum D_{ii} \]

(3) We do not consider the line distance, so the construction is a non-weighted network.
3.3. Topology properties

The properties of complex networks that do not depend on the specific positions of nodes and the specific shapes of edges are called the topological properties of complex networks, which determine the functions, dynamic properties, and evolution characteristics of complex networks [35]. In this study, we will use five different measures to analyze the topology properties of the BRT line network and BRT station network: degree, clustering coefficient, network diameter, average shortest path length, and modularity.

Degree $k$: The degree of a node $i$ in the BRT network is the number of edges connected with node $i$, that is, the number of stations connected to another station in the station network or the number of lines connected to any other line in the line network:

$$k_i = \sum_{j=1}^{n} a_{ij}$$

Clustering coefficient $C_i$: Reflecting the level of clustering degree of Beijing’s entire rail transit network, it is the ratio of the average degree to the stations’ scale or lines in the network.

The clustering coefficient $C_i$ of node $i$:

$$C_i = \frac{2M_i}{k_i(k_i - 1)}$$

$M_i$ – The actual number of edges of the sub-network composed of node $i$ and $k_i$ adjacent nodes.
The average clustering coefficient $C$ of the network:

$$
C = \frac{1}{N} \sum C_i / N
$$

$N$ – All vertices in the network, that is, the total number of lines or stations of the BRT network.

Network diameter $D$: The longest path lengths between two nodes. The distance between nodes refers to the shortest distance between the two most distant nodes in the network.

Average shortest path length $l$: The average number of steps along the shortest paths for all possible pairs of network nodes. It measures the efficiency of information or mass transport on a network.

The average shortest path length $l_i$ of node $i$:

$$
l_i = \frac{1}{N(N + 1)/2} \sum d_{ij}
$$

$d_{ij}$ – The distance between node $i$ and node $j$

Modularity: It measures the strength of a network's division into communities. Here, all nodes we focus on belong to the same community. Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules. Ranges $[0, 1]$, when the modularity tends to be 0, the network does not exit the communities, but when the modularity tends to be 1, it means the communities' structures are much clearer.

3.4. Network motif detection

As many small-scale subgraphs with the same structure appear in the network, motifs partially portray the specific pattern of the interconnection of real networks and directly affect the structure and function of the network organization. Therefore, the study of the motifs can intuitively analyze the network's microstructure [15,36]. As the smallest research unit, motifs as the “building blocks” of the network, composing the overall network from bottom to top. The network motifs’ detection includes three steps: generation of random networks, subgraph search, and motif evaluation [8].

3.4.1. Generation of random networks

Network motifs of a network are those that appear more frequently than in a random network. Then, we need to generate random networks of similar structures. The generated random network needs similar statistical properties to the real network, such as degree distribution. This paper selected the degree distribution as the construction benchmark, one of the complex network's most important statistical properties. There are three commonly used algorithms for constructing a random network based on the degree sequence: the switching algorithm, the matching algorithm, and the go-with-the-winners algorithm [37]. Since the switching algorithm is faster and more accurate than the matching algorithm and the go-with-the-winners algorithm, this paper applies the switching algorithm to generate random networks and constructs 1000 random networks for the BRT line network and the BRT station network, which has the same degree distribution as the real network.

3.4.2. Subgraph search

There are two subgraph search strategies: full enumeration and sampling [38]. Full enumeration can enumerate all subgraphs of the selected size in the original network and the random networks, while the sampling method can set the probability of the edge occurrence of the subgraph. The enumeration method is more accurate, but for large networks, the more subgraph nodes, the slower the speed. The random sampling algorithm is more friendly to detecting large network subgraphs, but the result is not as accurate as the enumeration. The BRT line network contains 24 nodes, and the BRT station network contains 354 nodes, which are not super-large networks, so to ensure the accuracy of the results, we will use the full enumeration method to retrieve subgraphs.

3.4.3. Motif evaluation

As a qualitative measure of statistical significance, the $Z$ score [8] evaluates the motif’s importance. The higher the $Z$ score, the more critical the motif is in the network.

$$
Z_i = \frac{(C_{\text{real}} - C_{\text{randi}})}{\text{std}(\sigma_{\text{randi}})}
$$

$C_{\text{real}}$ – Concentration of subgraph $g_i$ in real networks, that is, percentage of the number of subgraphs in a real network to all subgraphs in the same proportion;

$C_{\text{randi}}$ – Average concentration of $g_i$ in randomized networks, that is, the average percentage of the number of subgraphs in a randomized network to all subgraphs in the same proportion;

$\sigma_{\text{randi}}$ – The standard deviation of the average concentration of the subgraph in a randomized network.

If the subgraph $g_i$ reaches statistical significance, then it is a motif. Statistical significance needs to meet the following conditions:

(1) $Z > 0$;

(2) The probability $P$-value that the number of occurrences of the subgraph in the corresponding randomized network is greater than or equal to the number of occurrences in the real network needs to be very small and reach a certain
threshold. The smaller the value, the more critical the motif is in the network. Here, the $P$-value of all motifs is $P < 0.01$, as determined by comparison to 1000 randomized networks.

\[
P(N_{\text{randi}} \geq N_{\text{reali}}) < 0.01
\]

$N_{\text{reali}}$ – The absolute number of the subgraph $g_i$ occurrences in the real network;

$N_{\text{reali}}$ – The absolute number of average occurrences of the subgraph $g_i$ in the randomized network.

(3) The absolute number of the subgraph $g_i$ occurrences in the real network is not less than the lower limit $U$. Here, $U = 5$ means subgraphs need to appear independently in the real network at least five times, and subgraphs appearing less than five times are ignored.

\[
N_{\text{reali}} \geq U
\]

(4) The absolute number of the subgraph $g_i$ in the real network is significantly higher than in the randomized network.

\[
N_{\text{reali}} - N_{\text{randi}} > 0.1N_{\text{randi}}
\]

The motifs for which the four conditions above hold can be called motifs of the BRT line or station networks.

### 3.5. Motif decomposition

#### 3.5.1. Typical subgraphs

The global structure of the complex network is realized through a layered combination of small-scale motifs. Therefore, as the essential component of the complex network, the motif is decomposed layer by layer from the macroscopic rail transit network to the microscopic motif component, analyzing the complex network’s combination mechanism from bottom to top. The key to the connectivity of the motif function is the degree of the vertex. The greater the degree, the higher the motif’s connectivity for a given number of nodes.

For transportation networks, independent and combined analysis of network subgraphs can clarify the network’s local structural characteristics, thereby revealing the construction method and mechanism of the transportation network [39,40]. Typical local subgraphs of transportation networks include star subgraphs ($Q_1$), line subgraphs ($Q_2$), Y-shaped subgraphs ($Q_3$), ring subgraphs ($Q_4$), and fully connected subgraphs ($Q_5$) [41,42]. Those subgraphs have a specific impact on the functionality of line and station networks.

In the rail transit station network, the star subgraph ($Q_1$) represents that several disconnected stations are connected to a central station. The line subgraph ($Q_2$) represents the sequential connection of stations but does not form a loop. The Y-shaped subgraph ($Q_3$) means that three unconnected stations connect to a central station, and one of the three stations can lead to several connected stations in turn. The ring subgraph ($Q_4$) represents that several stations are connected in sequence and form a ring. The fully connected subgraph ($Q_5$) denotes that any two stations are connected in pairs, from one station can directly transfer to any other station in the graph.

In the rail transit line network, the star subgraph ($Q_1$) means several unconnected lines can transfer through another central line. The line subgraph ($Q_2$) represents several lines connected in sequence; one line can be transferred to another, but not every two lines in the subgraph are interconnected. The Y-shaped subgraph ($Q_3$) means that three unconnected lines connect to another central line, and one of the three lines can gradually transfer to several other lines. The ring subgraph ($Q_4$) represents that passengers can gradually transfer from the original line and return to the original starting line. The fully connected subgraph ($Q_5$) means every two lines are connected in pairs, and passengers can transfer directly from one line to another. Take the 5-node subgraph as an example; typical subgraphs of a transportation network are shown in Fig. 2.

### 3.5.2. Motif decomposition algorithm

One of the graph representations is a recursively constructed graph class, defined by a set of primitive or base graphs, in addition to one or more operations (called composition rules) that compose larger graphs from smaller subgraphs. Each operation involves fusing specific vertices from each subgraph or adding new edges between specific vertices from each subgraph [43]. The design strategy of the graph decomposition algorithm usually adopts the method of
dynamic programming. First, solve the problem on the base graphs defined for the given class, and then combine the solutions for subgraphs into a solution for a larger graph formed by the specific composition rules that govern the construction of members in the class [43]. The decomposition algorithm currently includes tree decomposition [44], path decomposition [45], and branch decomposition [46]. Based on the design strategy of the graph decomposition algorithm and these three decomposition algorithms, this study proposes a motif decomposition algorithm based on typical subgraphs. The difference is that a typical subgraph, namely the base graph, is defined and given at the initial stage of decomposition, and the motif is decomposed on this basis.

The network’s complexity lies in its elementary particles and many different combinations of elementary particles, which lies in the interaction of constituent particles [47]. Although the combination can show different possibilities and diversity, for the system, the form of the global structure is realized layer by layer through the combination of small-scale subgraphs. As the essential component of the network structure, the motif fundamentally affects the entire complex network's structure and function [48]. Three-node subgraphs to eight-node subgraphs are the basis of the BRT line and station network [49]. Since the number of different motifs differs, the subgraphs are classified and analyzed according to different nodes. The 3~5 nodes subgraphs have few nodes, and the number is small, so it is classified as a low-node subgraph, while the 6~8 nodes subgraphs have more nodes and a larger number, classified as a high-node subgraph.

The three-node and four-node motifs are simple in structure and have clear functional connotations. For example, a three-node loop subgraph represents three nodes connected; a four-node fully connected subgraph means four nodes connect in pairs. However, for motifs with 5 to 8 nodes, the types of motifs increase, and the structure is more complex as the number of nodes increases. It is necessary to decompose the obtained subgraph based on the typical subgraph and get the combined form of the 5~8 node subgraph to analyze these high node motifs in-depth. The decomposed subgraphs are mutually independent; the retrieved subgraphs do not contain each other. The search process starts with fully connected subgraphs (Q5), followed by ring subgraphs (Q4), star subgraphs (Q3), line subgraphs (Q2), and Y-shaped subgraphs (Q3) successively to ensure that the subgraphs are independent of each other. The decomposition of the subgraph is carried out in the following steps:

1. Define types of typical subgraphs as \(Q(i = 1 \sim 5)\), where star subgraph is \(Q_1\), line subgraph is \(Q_2\), Y-shaped subgraph is \(Q_3\), the ring subgraph is \(Q_4\), and fully connected subgraph is \(Q_5\);

2. \((n, m)\) represents nodes and edges. For n-node motifs, search for typical subgraphs based on n-nodes;

3. Let \(m = n(n - 1)/2\), scilicet, \((n, m) = (n, n(n - 1))/2\) which represents that there is an edge connection between any two points in the subgraph to search for subgraph \(Q_5\), namely fully connected subgraphs, record the result, go to the next step;

4. Let \(m = n\), scilicet, \((n, m) = (n, n)\), to search for ring subgraphs, namely \(i = 4\), record the result, go to the next step;

5. Select the subgraph form with the fewest edges under the scale of \(n\) nodes, namely \((n, m) = (n, n - 1)\), according to the types of \(Q_3\), to search for star subgraphs, line subgraphs, and Y-shaped subgraphs in turn from \(i = 1 \sim 3\), record the result, go to the next step;

6. Let \(n' = n - 1\), follow the above search method to search for the motif again until \(n' = 3\) to realize the complete decomposition of the subgraph.

7. The final decomposed subgraph combination formula is:

\[
Q_{\text{motifID}} = \sum_{i=1}^{5} AQ_i \quad (A \text{ is constant})
\]

Take the subgraph 5-1 (Fig. 3) in the BRT line network as an example,

When \(n = 5\),

When \((n, m) = (n, m = n(n - 1))/2 = (5, 10)\), no fully connected subgraphs exist;

When \((n, m) = (n, m = n) = (5, 5)\), no ring subgraphs exist;

When \((n, m) = (n, m = n - 1) = (5, 4)\), there is one line subgraph \(Q_2\), two Y-shaped subgraph \(Q_3\);

When \(n = 4\),

When \((n, m) = (n, m = n(n - 1))/2 = (4, 6)\), no independent fully connected subgraphs exist;

When \((n, m) = (n, m = n) = (4, 4)\), no independent ring subgraphs exist;

When \((n, m) = (n, m = n - 1) = (4, 3)\), no independent typical-subgraphs exist;

When \(n = 3\),

When \((n, m) = (n, m = n) = (3, 3)\), there is one ring subgraph \(Q_4\);

When \((n, m) = (n, m = n - 1) = (3, 2)\), no independent typical subgraphs exist;

Fig. 3. Subgraph 5-1 in the BRT line network.
Table 1

<table>
<thead>
<tr>
<th></th>
<th>N</th>
<th>E</th>
<th>(\bar{k})</th>
<th>C</th>
<th>D</th>
<th>l</th>
<th>M</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRT line network</td>
<td>24</td>
<td>58</td>
<td>4.833</td>
<td>0.441</td>
<td>5</td>
<td>2.246</td>
<td>0.205</td>
</tr>
<tr>
<td>BRT station network</td>
<td>354</td>
<td>397</td>
<td>2.243</td>
<td>0.008</td>
<td>55</td>
<td>17.659</td>
<td>0.829</td>
</tr>
</tbody>
</table>

Finally, the subgraph combination formula of subgraph 5-1 is:

\[
Q_{5-1} = Q_2 + 2Q_3 + Q_4
\]

The final result represents that subgraph 5-1 is combined by a line subgraph, a Y-shaped subgraph, and a ring subgraph. There is no independent star subgraph and fully connected subgraph. Nevertheless, a Y-shaped subgraph has five nodes, which means that the subgraph has a central line connecting multiple unconnected lines. According to this algorithm, the combination formula of each node motif could be obtained in turn.

Besides, define the subgraph combination strength \(R_{motifID}\) as the evaluation index of the subgraph combination obtained by decomposition, which is used to describe the way and degree of the typical subgraphs combined into n-node subgraphs. Namely,

\[
R_{motifID} = C_{reali} \sum_{i=1}^{5} \frac{a_i}{A_i}
\]

where,

- \(a_i\) – the number of typical subgraphs contained in a specific subgraph;
- \(A_i\) – the sum of all the corresponding typical subgraphs;
- \(C_{reali}\) – Concentration of subgraph in real networks.

When \(i = 1\), it represents the line subgraphs' subgraph combination strength of the evaluated motif; \(i\) from 2 to 5 represent the star subgraph, the Y-shaped subgraph, the ring subgraph, and the fully connected subgraph orderly.

The number and types of motifs with 6 to 8 nodes are significantly more than those with 3~5 nodes for the BRT line and station networks. We proposed the relative combination strength to evaluate the subgraphs. The relative subgraph combination strength \(|R_{motifID}|\) defined,

\[
|R_{motifID}| = |C_{reali}| \sum_{i=1}^{5} \frac{a_i}{A_i}
\]

where, \(|C_{reali}|\) is the Relative concentration.

\[
|C_{reali}| = \frac{N_{reali}}{\sum_{i} N_{reali}}
\]

4. Results

4.1. Topology of BRT networks

This study constructed the BRT line network from a macro perspective and the BRT station network from a micro perspective. The Beijing Subway is the urban rail transit system with the second largest operating mileage and the largest passenger volume in the world, so the Beijing Subway is typical in terms of subway scale and passenger flow [50,51].

Due to Beijing’s vast area, passengers usually need to go through line transfers when traveling. The choice of lines is crucial for passengers. Therefore, building a network with lines as nodes is significant to passenger travel. The station network uses the station as a node and describes passenger travel point-to-point so that it can pay more attention to the passenger’s itinerary. Combining the line network with the station network, we can see the rationality of the subway line setting and the relationship between the station setting and the line connection. Based on the adjacency matrix, we can get the topological graph of the Beijing rail transit line and station network, as shown in Fig. 4.

The Beijing rail transit line can show the situation of the line network, and the topological graph can intuitively show the overall line network of Beijing rail transit and the connection between various stations. For example, we can see a closed-loop connection in the topology subgraph. The Universal Resort Station connects this partial subgraph. Universal Resort Station belongs to the subway transfer station and terminal station of the Batong Line and Line 7 of the Beijing Metro. Since it belongs to the terminal station of the two lines, a ring connection has appeared. It also shows excellent connectivity and a greater possibility of going to other stations and lines. The main topological properties of the BRT line network and station network are shown in Table 1.
The BRT line network contains 58 edges, and the maximum degree is 13, which is line 10, which means that it can transfer from line 10 once to 13 of the other 23 lines, followed by line 4, with a degree of 10. Both Line 10 and Line 4 are loop lines. The minimum degree is 1: Line 16, Daxing Line, S1 Line, Yanfang Line, Xijiao Line, Daxing airport express, and Yizhuang T1 line, which means these five lines can only transfer to another. Moreover, the network diameter is 5; it can transfer to another line up to 5 times from one line. The average degree is 4.833; that is, one line is connected to 4.833 lines on average, representing the transfer ability between lines and the BRT network’s reachability. The average shortest path length between two pairs of nodes in the entire network is 2.246; an average of 2.246 lines can reach from one line to another. For the line network, the modularity coefficient is relatively small. Due to the continuous expansion of Beijing rail transit, the clustering coefficient has declined compared to 2019; Daxing Airport Express’s continuous expansion is one reason.

In the BRT station network, the maximum degree is five; those are Xizhimen station and Sanyuanqiao station, which can transfer to another five stations from both Xizhimen station and Sanyuanqiao station. The reachability depends on the location of these two lines. Xizhimen Metro Station is located on loop Line 2, while Sanyuanqiao Metro Station connects to the two terminal stations of Beijing Airport; both are the main transfer stations in Beijing. The minimum degree is 1, which can only transfer to another station from one station. These stations are usually the start or end of lines. For example, Anheqiao North station is the beginning of line four, and Tiantongyuan North station is the beginning of line five. The BRT station network’s diameter is 55; it can transfer up to 55 times from one station to another. The network’s average degree is 2.243; one station connects to 2.243 stations on average. The average shortest path length between two node pairs in the entire network is 17.659, which means it can reach another station by passing through 17.039 stations on average. Furthermore, the modularity is 0.829, which means that the BRT station network’s community structure is clear, and the stations are relatively concentrated.

4.2. Network motifs of BRT networks

It can be seen from the global characteristics of the rail transit network that there are different types of ring networks, star networks, and other typical subgraphs in the entire network. These typical subgraphs affect the rail transit operation. However, from the perspective of global characteristics, it is impossible to fully obtain the specific local characteristics of the entire network. In this case, the network motif can help us discover the network’s structure.

Because the BRT network has a loop radial symmetrical structure, it has a specific influence on the number and form of the BRT network’s subgraphs, which increase exponentially with the size of the subgraphs to be searched. According to the network motifs detection algorithm. The absolute number and the types of subgraphs of the BRT line network and the BRT station network are shown in Table 2.

The results show that the BRT line network contains 3 to 8 node motifs, while the BRT station network does not include the three-node motif. For the BRT station network, the number of each node motif is relatively minor. There is
Table 3
The BRT line network 3–5 nodes motifs.

<table>
<thead>
<tr>
<th>Motif</th>
<th>Legend</th>
<th>Subgraph combination formula</th>
<th>$C_{real}$%</th>
<th>Abs</th>
<th>Z</th>
<th>R%</th>
</tr>
</thead>
<tbody>
<tr>
<td>3-1</td>
<td>Q₂₋₁ = Q₄</td>
<td>19.08</td>
<td>50</td>
<td>3.9</td>
<td>19.08</td>
<td></td>
</tr>
<tr>
<td>4-1</td>
<td>Q₄₋₁ = Q₁ + 2Q₂ + Q₄</td>
<td>29.08</td>
<td>308</td>
<td>8.87</td>
<td>48.47</td>
<td></td>
</tr>
<tr>
<td>4-2</td>
<td>Q₄₋₂ = 2Q₁ + 2Q₄</td>
<td>10.86</td>
<td>115</td>
<td>2.21</td>
<td>14.48</td>
<td></td>
</tr>
<tr>
<td>5-1</td>
<td>Q₅₋₁ = Q₃ + 2Q₅ + Q₄</td>
<td>13.80</td>
<td>481</td>
<td>11.35</td>
<td>1.88</td>
<td></td>
</tr>
<tr>
<td>5-2</td>
<td>Q₅₋₂ = Q₁ + 2Q₃ + 3Q₄ + 3Q₄</td>
<td>9.58</td>
<td>334</td>
<td>9.51</td>
<td>4.67</td>
<td></td>
</tr>
<tr>
<td>5-3</td>
<td>Q₅₋₃ = 6Q₂ + 2Q₃ + 3Q₄</td>
<td>2.84</td>
<td>99</td>
<td>5.41</td>
<td>1.05</td>
<td></td>
</tr>
<tr>
<td>5-4</td>
<td>Q₅₋₄ = Q₁ + 6Q₂ + 4Q₃ + 5Q₄</td>
<td>3.85</td>
<td>134</td>
<td>4.67</td>
<td>2.78</td>
<td></td>
</tr>
<tr>
<td>5-5</td>
<td>Q₅₋₅ = Q₁ + 6Q₂ + 4Q₃ + Q₄</td>
<td>3.47</td>
<td>121</td>
<td>4.19</td>
<td>5.20</td>
<td></td>
</tr>
<tr>
<td>5-6</td>
<td>Q₅₋₆ = 2Q₂ + Q₃ + Q₄</td>
<td>5.65</td>
<td>197</td>
<td>3.66</td>
<td>0.76</td>
<td></td>
</tr>
<tr>
<td>5-7</td>
<td>Q₅₋₇ = Q₁ + 4Q₂ + 4Q₃ + 2Q₄</td>
<td>0.60</td>
<td>21</td>
<td>2.85</td>
<td>0.32</td>
<td></td>
</tr>
<tr>
<td>5-8</td>
<td>Q₅₋₈ = Q₁ + 8Q₂ + 12Q₃ + 7Q₄</td>
<td>1.41</td>
<td>49</td>
<td>2.34</td>
<td>1.58</td>
<td></td>
</tr>
</tbody>
</table>

one type of four-node motif, four types of five-node motifs, nine types of six-node motifs, 25 types of seven-node motifs, and eight-node motifs, including 58 types in the BRT station network. The absolute number and the form of subgraphs of the BRT line network are higher than those of the BRT station network.

The BRT line network contains all the motifs of the four-node and five-node of the BRT station network. The line network’s averagedegree is 4.833, and the averagedegree of the station network is 2.243. A network with a larger averagedegree means that the number of edges of each vertex is relatively large, and it will also cause a significant increase in the number of subgraphs of different sizes.

4.3. Motif decomposition results

4.3.1. Low-node motif decomposition

A network does not have all kinds of motifs. We obtain the BRT line network 3–5 node motifs through the motif decomposition algorithm proposed in the previous chapter, as shown in Table 3. The subgraphs are ordered by Z value from high to low. The bigger the Z value, the higher the importance of the motif. There is only one typical ring three-node motif and two four-node motifs containing the typical star subgraph (Q₁) and ring subgraph (Q₄). Motif 4-1 is not a typical subgraph that can be decomposed into combinations of star subgraph (Q₁), line subgraph (Q₂), and ring subgraph (Q₄), representing the three lines that connect in pairs, and the fourth line only connects to one of them. Motif 4-2 is composed of star subgraphs (Q₁) and ring subgraphs (Q₄), which are close to fully connected subgraphs (Q₅) and have good connectivity.

In the BRT line network, the concentration of the 3-node motif is 19.08%. There are fifty 3-node motifs in the real network. The structure of the 3-node motif is relatively simple, but it has a high concentration and affects the network a lot. The concentration of motif 4-1 is 29.08%, the real network contains 384 of the same motifs, which is higher than the concentration of motif 4-2, whose number is 115 in the line network. Furthermore, the subgraph combination strength of motif 4-1 is much higher than other motifs. The higher the combination strength is, the more stable the motif. The main reason is that motif 4-1 has higher concentration.

On the one hand, it represents the importance of the subgraph, and on the other hand, it also represents that the subgraph’s connectivity and utilization are much higher than that of other subgraphs. It shows that one line can quickly transfer to the other three lines, which can significantly improve the entire network’s operating efficiency, and it is also the mode that will be given priority in the construction of the transportation network. Although the lines’ excessive connection will make them fully connected, it will significantly increase costs. Considering the constraints of geospatial factors, it will also make it difficult to realize a fully connected subgraph.

There are 3485 five-node motifs of eight types in the BRT line network. In the real network, the concentration of motif 5-1 is 13.80%, that is, 481 subgraphs of the same type, the highest proportion of 5-node subgraph, followed by motif 5-4, whose subgraph concentration is 10.29%, that is, 329 subgraphs of the same type. After decomposition, the 5-node motif
contains a total of 95 typical subgraphs, of which five star subgraphs ($Q_1$), accounting for 5%; 35 line subgraphs ($Q_2$), accounting for 37%; 32 Y-shaped subgraphs ($Q_3$), accounting for 34%; 22 ring subgraphs ($Q_4$), accounting for 23%, and one fully connected subgraphs ($Q_5$), accounting for 1%. The results show that line subgraphs ($Q_2$) > Y-shaped subgraphs ($Q_3$) > ring subgraphs ($Q_4$) > star subgraphs ($Q_1$) > fully connected subgraphs ($Q_5$). The number of line subgraphs ($Q_2$) and Y-shaped subgraphs ($Q_3$) is relatively large, which is the main local structure of the network. Ring subgraphs ($Q_4$) are the secondary local structures of the network. The other two subgraphs account for less than 10%.

In the five-node subgraphs, motif 5-1 has the highest concentration and the Z-value, which show the importance of motif 5-1, while motif 5-5 has the highest subgraph combination strength is motif 5-5, though its concentration and Z-value is lower than motif 5-1. The main reason is that fully connected subgraphs ($Q_5$) greatly increase the importance of the motif 5-5. The combination of multiple typical subgraphs shows that the subgraph has good connectivity and interchangeability. In the real Beijing rail transit network, the connection among line 1, line 4, line 9, line 10, and line 2 belongs to motif 5-5. Line 1, Line 4, Line 9, and 10 interconnect in pairs, which are typical fully connected subgraphs ($Q_5$), and Line 1 also connects to Line 2, forming a connection-mode of a typical fully connected subgraph ($Q_5$) and a typical star subgraph ($Q_1$), as shown in Figs. 5 and 6. The pairwise interoperability of these four lines can greatly help passengers choose a better travel route.

The BRT station network 3 ~ 5-node motifs are as shown in Table 4. The types and number of motifs in the station network are significantly lower than those in the line network. The table shows that the station network does not contain three-node motifs, only one four-node motif, and four five-node motifs. In the station network, the concentration of motif 4-1 is 0.735%. The actual network contains 800 motifs of the same type. Although the concentration value is low, since the 4-node subgraph detected in the station network is much higher than the line network, the number of motifs of the same type of station network in the actual network is much higher than the number of line networks. In addition, the Z value of the motif 4-1 in the station network is larger. It shows the importance of the motif. It represents that the three stations connect in pairs, and the fourth station only connects to one of the stations. The connection mode is the same as the connection mode of motif 4-1 of the line network, but it represents a different meaning. The line network takes the line as the main body, while the station network takes the station as the main body.

There are 2301 subgraphs in the station network containing four types of five-node motifs. After deconstruction, the 5-node motif contains 16 typical subgraphs, including one star subgraph ($Q_1$), a 6% difference, five line subgraphs ($Q_2$), a 31% difference; six Y-shaped subgraphs ($Q_3$), a 38% difference; and four ring subgraphs ($Q_4$), a 25% difference; there are no fully connected motifs ($Q_5$) in the network. Therefore, from the perspective of the five-node motif, the line subgraphs ($Q_2$) and the Y-shaped ($Q_3$) subgraphs are the primary local structure, and the ring subgraphs ($Q_4$) and the star subgraphs ($Q_1$) are
the secondary local structure, while fully connected subgraphs (Q_5) are non-essential local structures. The concentration of motif 5-4 is the highest, and the subgraph combination strength is also the highest, indicating the high frequency and utilization.

4.3.2. High-node motif decomposition

High-node motifs are more complex than low-node motifs. We select the top 5 motifs with 6 to 8 node motif Z-value (sorted by the importance of motifs) in the BRT line and station networks for analysis. The combination of 6~8 node motifs in the BRT line and the station network are shown in Tables 5 and 6.

For six-node motifs, the line network subgraphs are 9395, including 42 types, while the number of station network subgraphs is 5284, including nine types of six-node motifs. The concentration ratios of the six-node motifs of the line network are 7.28% and 6.53%, respectively, while the concentration ratios of the top two Z-values are 0.14% and 0.06%. It shows that the concentration of the existing algorithm does not greatly influence Z-value; therefore, the importance of

---

**Table 4**
The BRT station network 3~5 nodes motifs.

<table>
<thead>
<tr>
<th>Motif</th>
<th>Legend</th>
<th>Subgraph combination formula</th>
<th>C_{real} %</th>
<th>Abs</th>
<th>Z</th>
<th>R%</th>
</tr>
</thead>
<tbody>
<tr>
<td>4-1</td>
<td><img src="4-1" alt="Image" /></td>
<td>$Q_{4-1} = Q_1 + 2Q_2 + Q_4$</td>
<td>0.74</td>
<td>800</td>
<td>24.34</td>
<td>2.21</td>
</tr>
<tr>
<td>5-1</td>
<td><img src="5-1" alt="Image" /></td>
<td>$Q_{5-1} = Q_1 + 2Q_3 + Q_4$</td>
<td>0.22</td>
<td>236</td>
<td>32.68</td>
<td>0.34</td>
</tr>
<tr>
<td>5-2</td>
<td><img src="5-2" alt="Image" /></td>
<td>$Q_{5-2} = Q_2 + 2Q_3 + Q_4$</td>
<td>0.35</td>
<td>379</td>
<td>30.30</td>
<td>0.27</td>
</tr>
<tr>
<td>5-3</td>
<td><img src="5-3" alt="Image" /></td>
<td>$Q_{5-3} = 2Q_2 + Q_3 + Q_4$</td>
<td>0.44</td>
<td>474</td>
<td>20.16</td>
<td>0.36</td>
</tr>
<tr>
<td>5-4</td>
<td><img src="5-4" alt="Image" /></td>
<td>$Q_{5-4} = 2Q_2 + 3Q_3 + Q_4$</td>
<td>1.17</td>
<td>1277</td>
<td>4.55</td>
<td>0.96</td>
</tr>
</tbody>
</table>

**Table 5**
The BRT line network 6~8 nodes motifs.

| Motif | Legend | Subgraph combination formula | C/ % | Abs | | C_{real} | | R_{motifs} |
|-------|--------|------------------------------|------|-----|---|------------|-----------|
| 6-1   | ![Image](6-1) | $Q_{6-1} = Q_1 + 3Q_2 + 7Q_3 + 4Q_4$ | 0.14 | 13  | 0.02 | 0.02 |
| 6-2   | ![Image](6-2) | $Q_{6-2} = 2Q_1 + 2Q_3 + 2Q_4$ | 0.06 | 6   | 0.01 | 0.01 |
| 6-3   | ![Image](6-3) | $Q_{6-3} = 3Q_1 + 20Q_2 + 5Q_3 + 4Q_4$ | 0.25 | 23  | 0.04 | 0.06 |
| 6-4   | ![Image](6-4) | $Q_{6-4} = 2Q_1 + 6Q_2 + 6Q_3 + Q_5$ | 2.90 | 272 | 0.51 | 0.85 |
| 6-5   | ![Image](6-5) | $Q_{6-5} = 3Q_1 + 4Q_2 + 4Q_3 + Q_4$ | 2.35 | 221 | 0.41 | 0.21 |
| 7-1   | ![Image](7-1) | $Q_{7-1} = 5Q_1 + 25Q_2 + 14Q_3 + 8Q_4$ | 0.04 | 907 | 0.11 | 0.10 |
| 7-2   | ![Image](7-2) | $Q_{7-2} = 5Q_1 + 7Q_2 + 18Q_3 + 15Q_4$ | 0.03 | 590 | 0.07 | 0.06 |
| 7-3   | ![Image](7-3) | $Q_{7-3} = 4Q_1 + 50Q_2 + 33Q_3 + 13Q_4 + 2Q_5$ | 0.06 | 1307| 0.17 | 0.40 |
| 7-4   | ![Image](7-4) | $Q_{7-4} = 3Q_1 + 5Q_2 + 7Q_3 + 6Q_4$ | 0.06 | 1307| 0.17 | 0.05 |
| 7-5   | ![Image](7-5) | $Q_{7-5} = 3Q_1 + 3Q_2 + 12Q_3 + 10Q_4$ | 0.18 | 3796| 0.48 | 0.26 |
| 8-1   | ![Image](8-1) | $Q_{8-1} = 3Q_1 + 20Q_2 + 14Q_3 + 7Q_4$ | 0.05 | 1799| 0.27 | 0.26 |
| 8-2   | ![Image](8-2) | $Q_{8-2} = 2Q_1 + 6Q_2 + 26Q_3 + 9Q_4$ | 0.05 | 2119| 0.32 | 0.23 |
| 8-3   | ![Image](8-3) | $Q_{8-3} = 2Q_1 + 7Q_2 + 7Q_3 + 20Q_4$ | 0.02 | 919 | 0.14 | 0.09 |
| 8-4   | ![Image](8-4) | $Q_{8-4} = 3Q_1 + 4Q_2 + 19Q_3 + 22Q_4$ | 0.02 | 919 | 0.14 | 0.11 |
| 8-5   | ![Image](8-5) | $Q_{8-5} = 6Q_1 + 3Q_2 + 12Q_3 + 18Q_4$ | 0.02 | 799 | 0.12 | 0.10 |
The BRT station network 6–8 nodes motifs.

| Motif | Legend | Subgraph combination formula | C/% | Abs | |C_reals| |R_motif|D| |
|---|---|---|---|---|---|---|---|
| 6-1 |  | \(Q_{6-1} = Q_1 + 2Q_3 + Q_4\) | 0.11 | 602 | 0.09 | 0.07 |
| 6-2 |  | \(Q_{6-2} = Q_2 + 4Q_2 + 2Q_3 + Q_4\) | 0.27 | 1400 | 0.20 | 0.25 |
| 6-3 |  | \(Q_{6-3} = 2Q_2 + Q_3 + Q_4\) | 0.32 | 1701 | 0.25 | 0.13 |
| 6-4 |  | \(Q_{6-4} = 2Q_2 + Q_3 + Q_4\) | 0.30 | 1601 | 0.23 | 0.13 |
| 6-5 |  | \(Q_{6-5} = Q_1 + Q_2 + 2Q_3 + Q_4\) | 0.30 | 1601 | 0.23 | 0.21 |
| 7-1 |  | \(Q_{7-1} = Q_1 + 6Q_3 + Q_4\) | 0.07 | 906 | 0.10 | 0.07 |
| 7-2 |  | \(Q_{7-2} = Q_1 + 2Q_2 + 3Q_3 + Q_4\) | 0.20 | 2500 | 0.28 | 0.23 |
| 7-3 |  | \(Q_{7-3} = Q_1 + 2Q_2 + 9Q_3 + Q_4\) | 0.05 | 600 | 0.07 | 0.08 |
| 7-4 |  | \(Q_{7-4} = Q_1 + Q_2 + 3Q_3 + Q_4\) | 0.17 | 2194 | 0.25 | 0.17 |
| 7-5 |  | \(Q_{7-5} = Q_1 + 2Q_2 + Q_4\) | 0.21 | 2704 | 0.30 | 0.21 |
| 8-1 |  | \(Q_{8-1} = Q_1 + 2Q_2 + 3Q_3 + Q_4\) | 0.12 | 3688 | 0.28 | 0.18 |
| 8-2 |  | \(Q_{8-2} = Q_1 + 5Q_2 + 10Q_3 + 2Q_4\) | 0.09 | 2893 | 0.22 | 0.29 |
| 8-3 |  | \(Q_{8-3} = Q_1 + 2Q_3 + Q_4\) | 0.05 | 1494 | 0.11 | 0.05 |
| 8-4 |  | \(Q_{8-4} = Q_1 + 5Q_3 + Q_4\) | 0.11 | 3401 | 0.26 | 0.14 |
| 8-5 |  | \(Q_{8-5} = Q_1 + 4Q_2 + 11Q_3 + Q_4\) | 0.05 | 1589 | 0.12 | 0.13 |

The combined strength of the sub-subgraphs is explained. For the line network, the motif 6-1 to the motif 6-5 all include typical star subgraphs \(Q_1\), line subgraphs \(Q_2\), and ring subgraphs \(Q_4\). It contains a total of 78 subgraphs, of which nine star subgraphs \(Q_1\), accounting for 12%, 35 line subgraphs \(Q_2\), accounting for 45%, 22 Y-shaped subgraphs \(Q_3\), accounting for 28%, and 11 ring subgraphs \(Q_4\), accounting for 14%, and a fully connected subgraph \(Q_5\), accounting for 1%. The line subgraph \(Q_2\) is the leading local structure, the Y-shaped subgraph \(Q_3\), the star subgraph \(Q_1\), and the ring subgraph \(Q_4\) are the secondary local structure, and the fully connected subgraph \(Q_5\) is the unnecessary local structure. Besides, the Z-value of motif 6-1 is the highest, but the subgraph concentration and subgraph combination strength of motif 6-4 are the largest. The strong combination of subgraphs of motif 6-4 is that motif 6-4 contains fully connected subgraphs \(Q_5\), further strengthening the importance of this motif type. The six-node subgraph of the station network contains a total of 25 typical subgraphs, of which three star subgraphs \(Q_1\), accounting for 12%, nine line subgraphs \(Q_2\), accounting for 36%, and eight Y-shaped subgraphs \(Q_3\), accounting for 32%, five ring subgraphs \(Q_4\), accounting for 20%, excluding fully connected subgraphs \(Q_5\). According to the proportion of the subgraphs, the typical line subgraphs \(Q_2\) and typical Y-shaped subgraphs \(Q_3\) of the station network are the necessary local structures, the typical star subgraphs \(Q_1\), and the ring subgraphs \(Q_4\) are the secondary local structures, and fully connected subgraphs \(Q_5\) are non-essential local structures. Motif 6-1 has the highest Z value, motif 6-3 has the highest concentration, and motif 6-2 has the most robust combination of subgraphs. The reason lies in the diversity of motif 6-3 subgraphs combinations.

For the 7-node motif, the number of 7-node subgraphs detected in the line network is 21,087, including 192 types of 7-node motifs. The concentration ranges from 0.21% to 0.05%. According to the decomposed motif combination, the 7-node motif of the line network contains a total of 246 typical subgraphs, of which 18 are star subgraphs \(Q_1\), accounting for 7%; 90 line subgraphs \(Q_2\), accounting for 37%, 84 Y-shaped subgraphs \(Q_3\) account for 34%, 52 ring subgraphs \(Q_4\) account for 21%, and two fully connected subgraphs \(Q_5\) account for 1%. According to the proportion, for the 7-node motif of the line network, the line subgraph \(Q_2\), the Y-shaped subgraph \(Q_3\), and the ring subgraph \(Q_4\) are the necessary local structures, and the star subgraph \(Q_1\) and the fully connected subgraph \(Q_5\) are the secondary local structures. The list shows that the highest Z value is motif 7-1, motif 7-5 has the highest concentration, and motif 7-3 has the highest subgraph combined strength. Different results show the importance of each type of motif at different levels. The number of 7-node subgraphs detected by the station network is 12,757, including 25 7-node motifs. The Z value ranges from 0.05%
to 0.21%. The station network 7-node motif selects the top five motifs sorted by the Z value for analysis and decomposes the subgraph according to the motif decomposition algorithm, including 38 typical subgraphs, five star subgraphs \((Q_1)\), accounting for 13%, seven line subgraphs \((Q_4)\), accounting for 18%, and 21 Y-shaped subgraphs \((Q_2)\), accounting for 55%, five ring subgraphs \((Q_5)\), accounting for 13%, excluding fully connected subgraphs \((Q_6)\). According to the proportion, it can be seen that the Y-shaped subgraph \((Q_2)\) is the necessary local structure, the star subgraph \((Q_1)\), the line subgraph \((Q_4)\), and the ring subgraph \((Q_5)\) is a secondary local structure, and the fully connected subgraph \((Q_6)\) is an unnecessary local structure. It can be seen from the motif legend that the 7-node motif of the line network is more complex than the 7-node motif of the station network. The reason is that the average degree of nodes in the line network is 4.833, while the average degree of nodes in the station network is 2.243. The average degree of the line network is equivalent to twice the average degree of nodes, which also makes its complexity much higher than the same type of motif for the station network.

For the 8-node motif, the number of 8-node subgraphs is 39,973 in the line network, including 801 types of 8-node. According to the results obtained by the decomposition algorithm, the 8-node motifs of the line network contains 210 typical subgraphs, 16 star subgraphs \((Q_1)\), accounting for 8%, 40 line subgraphs \((Q_4)\), accounting for 19%, and 78 Y-shaped subgraphs \((Q_2)\), accounting for 37%, 76 ring subgraphs \((Q_5)\), accounting for 36%, excluding fully connected subgraphs \((Q_6)\). According to the proportion, Y-shaped subgraphs \((Q_2)\) and ring subgraphs \((Q_5)\) are necessary local structures, star subgraphs \((Q_1)\) and line subgraphs \((Q_4)\) are secondary local structures, and fully connected subgraphs \((Q_6)\) are unnecessary. The number of 8-node subgraphs in the station network is 31,789, including 59 8-node motifs. Decomposing different types of motifs, the station network 8-node motifs contains a total of 53 typical subgraphs, of which five star subgraphs \((Q_1)\), accounting for 9%, 11 line subgraphs \((Q_4)\), accounting for 21%, and 31 Y-shaped subgraphs \((Q_2)\), accounting for 58%, and six ring subgraphs \((Q_5)\), accounting for 11%, which also do not include fully connected subgraphs \((Q_6)\). The Y-shaped subgraph \((Q_2)\) belongs to the necessary local structure, the line subgraph \((Q_4)\) belongs to the secondary local structure, and the fully connected subgraph \((Q_6)\) belongs to the non-essential local structure. Similarly, the 8-node motif structure of the station network is far less complex than the 8-node motif structure of the line network. The reason is mainly the degree of the node, which significantly affects the local structure.

### 4.3.3. Results

The horizontal and vertical comparisons are carried out on the typical subgraphs in the motifs of different nodes of the Beijing rail transit line and station networks. Horizontally, it compares the proportions of typical subgraphs of both rail transit networks’ same node motifs. At the same time, the vertical is a comparison of the proportions of different types of typical subgraphs of the same network. The results can be used to evaluate the current rail transit network better, find the balance point between the line network and the station network, and increase the station’s economic value to a greater extent.

Horizontally compare the proportions of the typical subgraphs of the BRT station network and the line network. Tables 3 to 6 shows the star subgraphs \((Q_1)\), line subgraphs \((Q_4)\), Y-shaped subgraphs \((Q_2)\), ring subgraphs \((Q_5)\), and fully connected subgraphs \((Q_6)\) in the line network and station network high-node motifs are respectively analyzed. The results are shown in Figs. 7(a), 7(b), 7(c), 7(d).

The connectivity of distinct typical subgraphs is diverse. Line subgraphs \((Q_4)\) have low connectivity, star subgraphs \((Q_1)\) and fully connected subgraphs \((Q_6)\) have high connectivity, and Y-shaped subgraphs \((Q_2)\) have medium connectivity. While the ring subgraph \((Q_5)\) lies in its structure’s particularity, the connectivity is lower than that of the Y-shaped but higher than the line subgraph. According to Fig. 7(a), for the 5-node motif, the line network has the highest proportion of line subgraphs \((Q_4)\), followed by Y-shaped subgraphs \((Q_2)\), and the station network has the highest proportion of Y-shaped subgraphs \((Q_2)\), followed by line subgraphs \((Q_4)\). It means that for combinations with fewer nodes in the line network, most of them are connected line by line, and the Y-shaped subgraph in the station network represents a central station between the nodes as a transfer station to ensure the connectivity of the line.

In addition, it can be also found that there is no full-connected typical subgraph in the 5-node station network. Therefore, the subgraph of the lower node of the station network is mainly Y-shaped, and line-shaped, ring-shaped and star-shaped are supplemented to form the actual station network; in the same-node line network, the proportions of star subgraphs \((Q_1)\) and ring subgraphs \((Q_5)\) are similar to those of the station network, including fully connected subgraphs \((Q_6)\), but the proportion is tiny.

Similarly, as shown in Fig. 7(b), for the 6-node motif, in the line network and the station network, the line subgraph \((Q_4)\) has the highest proportion in both networks, followed by the Y-shaped subgraph \((Q_2)\), and the ring subgraph \((Q_5)\) and the star subgraphs \((Q_1)\) are similar, while the proportion of fully connected subgraphs \((Q_6)\) is deficient in the line network, but does not exist in the station network. Therefore, for the 6-node motif, the line subgraph \((Q_4)\) and Y-shaped subgraphs \((Q_2)\) are the main in the two networks, and the star subgraph \((Q_1)\) and the ring subgraph \((Q_5)\) complement each other.

According to Fig. 7(c), in the 7-node motif, the Y-shaped subgraph presents a peak shape in the station network, representing that more than 50% of the 7-node structures have a central station connecting three unconnected stations, and connect to several other stations through one of these three stations in turn. The distribution of the remaining star, line, and ring subgraphs presents a relatively balanced pattern. Similarly, there is still no fully connected subgraph \((Q_6)\) structure in the 7-node motif. In the line network of the same node, the proportions of line subgraphs \((Q_4)\) and Y-shaped
subgraphs \((Q_3)\) are similar, but the fully connected subgraphs \((Q_5)\) account for only 1\%, and the proportions of each typical subgraph are relatively balanced.

In the typical 8-node subgraphs distribution in Fig. 7(d), the Y-shaped subgraphs \((Q_3)\) in the station network also show a peak shape, accounting for more than 50\%. Star subgraphs \((Q_1)\) and ring subgraphs \((Q_4)\) accounted for similarly. The line subgraphs \((Q_2)\) account for a higher proportion than the star subgraphs \((Q_1)\) and the ring subgraphs \((Q_4)\), but they are still much smaller than the Y-shaped subgraphs \((Q_3)\). The result shows the importance of the Y-shaped connection to the current station network.

The proportions of Y-shaped subgraphs \((Q_3)\) and ring subgraphs \((Q_4)\) are similar in the line network. Combining the two subgraphs can greatly improve the network’s connectivity, while the proportions of star subgraphs \((Q_1)\) and line subgraphs \((Q_2)\) almost coincide with the proportion of similar subgraphs in the station network. For the 8-node motif, neither of the two types of networks contain fully connected subgraphs \((Q_5)\). The reason is that the more nodes, the more difficult and costly to achieve full connectivity. It shows that Y-shaped subgraphs are the most important for maintaining high connectivity in station networks, while line subgraphs are dominant in line networks. Ring and star subgraphs complement the connectivity of both types of networks. The proportion of fully connected subgraphs is low in all cases due to the high cost of achieving full connectivity as the number of nodes increases. The analysis emphasizes the significance of connectivity and subgraph distribution in network design.

Longitudinally compare the proportions of typical subgraphs of the BRT line and station networks. The typical subgraph of 5~8 nodes of the line network is shown in Fig. 8(a). The line network contains all types of typical subgraphs. For different node motifs, overall, the typical line subgraphs \((Q_2)\), typical Y-shaped subgraphs \((Q_3)\), and typical ring subgraphs \((Q_4)\) are more evenly distributed; the typical star subgraphs \((Q_1)\) account for a relatively small proportion, and the fully connected typical subgraphs \((Q_5)\) account for the lowest proportion. Specifically, the typical line subgraphs \((Q_2)\) of 5-node, 6-node, and 7-node motifs in the line network account for the highest proportion, while 8-node motifs are relatively small; there is little difference in the proportional distribution of Y-shaped subgraphs \((Q_3)\) under different nodes; the 8-node motif has the highest proportion of ring subgraphs \((Q_4)\), which is in sharp contrast with the proportion of 6-node subgraphs. The proportion of typical subgraphs of 5~8 nodes in the station network is shown in Fig. 8(b). Whether for high-node or low-node motifs, the proportion of Y-shaped typical subgraphs \((Q_3)\) presents an absolute advantage. The proportions of line typical subgraphs \((Q_2)\) and typical ring subgraphs \((Q_4)\) are similar, and the proportion of typical star subgraphs \((Q_1)\) is relatively stable. Besides, there are no fully connected subgraphs \((Q_5)\) in the station network. It provides insights into the structural characteristics of the BRT line and station networks, which could be helpful for future planning and design of similar transportation systems.

In summary, the line typical subgraph \((Q_2)\) and Y-shaped typical subgraph \((Q_3)\) are the main local structures in the line network, the typical ring subgraph \((Q_4)\) is the secondary local structure. The typical star subgraph \((Q_1)\) is complementary to the network. Fully connected subgraphs \((Q_5)\) occupy a deficient proportion in the line network, but they play an essential role in the connectivity of the line network. Y-shaped typical subgraphs \((Q_3)\) play a prominent role in the station network,
the main local structure. Typical line (Q_2) and ring subgraphs (Q_4) are secondary local structures. The star subgraphs (Q_1) do not account for a high proportion but have a stable distribution in all motifs. The station network does not contain fully connected subgraphs (Q_5), which are significant for connectivity. Therefore, for the station network, the proportion of fully connected subgraphs (Q_5) can be appropriately increased within the range allowed by the cost, and the connectivity among stations can be improved.

The results provide valuable insights into the local structures of transportation networks, particularly the line and station networks, by examining the proportions of typical subgraphs with different sizes and motifs. These findings can be used to enhance the design and optimization of transportation networks for improved connectivity and efficiency. For example, increasing the proportion of fully connected subgraphs in the station network could lead to better connectivity among stations and enhance the overall efficiency of the transportation system. Moreover, the study's results could also be applied to other types of networks, such as social or communication networks, to identify and optimize the primary local structures for enhanced connectivity and efficiency.

5. Conclusions

The Beijing rail transit line network and the station network are based on the Beijing rail transit real network, using P-space and L-space modeling methods to construct models from the perspectives of lines and stations. The two networks' topological structure and global characteristics are obtained, respectively, and all motifs from 3 to 8 nodes of the two networks are obtained using the motif detection algorithm. Besides, this paper designs a motif decomposition algorithm for the motif based on five typical subgraphs: star subgraphs (Q_1), line subgraphs (Q_2), Y-shaped subgraphs (Q_3), ring subgraphs (Q_4), and fully connected subgraphs (Q_5). The motifs are decomposed and calculated separately for the two networks. The results show that the two networks show different numbers and distributions of typical subgraphs. Comprehensive analysis results, the conclusions obtained are as follows:

(1) Although the line subgraph has low connectivity for the station network, it mostly appears as the basic structure of the network, and the proportion of occurrence is relatively high. The emergence of ring subgraphs (Q_4) as auxiliary subgraphs improves the network’s connectivity to a certain extent. Although star subgraphs (Q_1) and fully connected subgraphs (Q_5) have high connectivity, they are difficult to achieve. The typical Y-shaped subgraphs (Q_3) with moderate connectivity appear the most because of their constraints on the entire network and a certain efficiency balance. Under the premise of not significantly increasing the construction cost and complexity of the project, the connectivity of the entire network can be effectively improved.

(2) Compared with the station network, the line network is an abstract network less constrained by space, cost, and technology. The Y-shaped subgraph (Q_3) with medium connectivity and the line subgraph (Q_2) with low connectivity constitute the network's basic framework. Ring subgraphs (Q_4) appear less. Both star subgraphs (Q_1) and fully-connected subgraphs (Q_5) with high connectivity appear in the line network, significantly improving the network's connectivity.

(3) The station and line networks show the highest proportion of the Y-shaped subgraph (Q_3) of the high-node motifs and the line subgraph (Q_2) of the low-node motifs. Although the two portray the rail transit network from different perspectives, they show similar results, fully illustrating the importance of the Y-shaped subgraph (Q_3) and the line subgraph (Q_2) for transportation planning.

The identification of typical subgraphs and their distribution in the station network and line network can help identify potential vulnerabilities in the network's structure. For example, the research found that Y-shaped subgraphs (Q_3) and line subgraphs (Q_2) are the most common motifs in both networks. These motifs may be critical for maintaining the network's connectivity and should be prioritized in planning and design decisions to enhance the network's reliability and resilience. Moreover, the subgraph decomposition algorithm proposed in this research can be used to identify the distribution of motifs at different network scales and locations. This information can be useful for network planners to make decisions on where to place critical infrastructure, such as power sources and signal systems, and to plan for redundant connections to maintain network functionality during failures or disruptions.
Although the station network and the line network are the modeling and analysis of the transportation network from two perspectives, the line network is a macroscopic display of the station network, and the station network is a microscopic depiction of the line network. The station network is more susceptible to the constraints of geographic space and construction cost during the construction process; The line network can assist the station network and realize the reasonable distribution of stations. The two complement each other and realize the optimal allocation of resources. The motif decomposition algorithm for complex network motifs proposed in this paper has specific reference significance for existing motif research and motif detection.

According to the research results, the following guidelines can be suggested for the design, planning, and expansion of the rail transit network:

1. Incorporate Y-shaped subgraphs ($Q_3$) and line subgraphs ($Q_2$) into the design of the rail transit network. These subgraphs are important for maintaining the efficiency and connectivity of the network.

2. Use ring subgraphs ($Q_4$) as auxiliary subgraphs to improve the connectivity of the network, especially for the station network.

3. Consider the cost and complexity of constructing high-node motifs like star subgraphs ($Q_1$) and fully-connected subgraphs ($Q_5$). While these subgraphs have high connectivity, they are difficult to achieve and may not be cost-effective.

4. Use the line network as a macroscopic display of the station network and the station network as a microscopic depiction of the line network. The two networks complement each other and can assist in the reasonable distribution of stations and optimal allocation of resources.

5. The proposed subgraph decomposition algorithm for complex network motifs can be used as a reference for existing motif research and motif detection.

In further research, we will consider improving the existing motif detection tools to realize the decomposition of large-scale quantitative motifs according to the proposed motif decomposition algorithm and building different weighted rail transit networks to see how weighted network motif affects the rail transit system.

CRediT authorship contribution statement

Yunfang Ma: Conceptualization, Methodology, Software, Writing – original draft. Jose M Sallan: Conceptualization, Writing – review & editing. Oriol Lordan: Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

Y. Ma, J.M. Sallan and O. Lordan  
Physica A 625 (2023) 129016


