

Diversity of Building Blocks Unity of Construction Rule for Transportation Networks

Bangya Li^{1,a}, Liang Liu^{2*,b}

¹ School of Management, Shanghai University Shanghai 200444, China

² School of Management, Shanghai University Shanghai 200444, China

Abstract. The identification of local structure composition may help understand the network construction law from a bottom-up view. Focusing on local structure, the network subgraph approach was applied to systematically analyze 3- to 7-node subgraphs in several transportation networks in China. Results show that the number of subgraph and forms increase as the subgraph scale increase, but the rate are different, which indicates the diversity of building blocks in transportation networks. Moreover, both the network subgraph rank can be divided into two kinds, each corresponding to one of the transport geographical networks and transport service networks. The two kinds reflect the unity of construction rule of geographical network or service network due to the similarity of the cost, technical and space constraints. The methods and results can be used as methods and standards in the planning, evaluation and optimization of transportation networks.

1 Introduction

In the past two decades, great progress has been made in the study of the topology of research objects in various disciplines based on complex network theory[1], especially the common structure of networks spanning biological, technological and social systems[2]. Researchers have shown that the above-mentioned networks have common global structural characteristics, e.g., small-world or scale-free[3][4]. Recently, researches in statistical physics and molecular biology based on complex networks have further shown that networks with similar global structures may differ due to their own functional characteristics or generation mechanisms, reflecting extremely different local structures -- small-scale subgraphs frequently appearing in real networks[5]. In studying the transcriptional regulatory network of *Escherichia coli*, Milo found the small-scale subgraphs that appear in the real network are much higher than those in the random network with the same number of nodes and connections and defined such subgraphs as motifs and may have basic functional characteristics[6]. Moreover, different levels and sizes of subgraph distribution play an important role in characterizing network composition, reflecting the construction rules of complex network structure. The study of local network structure has attracted wide attention in the fields of natural, engineering and social research.

The transportation network (TN) is composed of a large number of stations and lines, which has the key support function for the sustainable development of economy and society[7]. Recently, many scholars have studied the global structure of transportation networks

based on complex network theory, and analyses show that the node degree of Indian railway geographic network is exponential distribution[8], the Chinese railway passenger network has the characteristics of drift power law distribution[9], and the distribution of Beijing bus network obeys the power-law distribution[10]. The subway engineering network has the characteristics of tree shape and small-world network[11], and its degree distribution follows the exponential form[12]. The national highway network of the United States is a random network with exponential distribution form[13], and so on. In general, the research on the global structure characteristics of transportation networks has been carried out. Results show that the TNs have the characteristics of small-world, and generally speaking, the transport geographical network has the characteristics of degree exponential distribution, and the transport service network has the characteristics of degree power-law distribution. The study effectively expands the cognition of the global structure of complex transport network.

However, little is known about the local structures of transportation network and their construction rules. Because of the different network size and connectivity, it seems difficult to compare such local structures and identify their construction rules in TNs. On the basis of the above research, we propose to build a network model based on some real transport geographical networks and transport service networks, and we aim to systematically investigate the basic local structures and their construction rules of such transportation networks based on complex network subgraph approach. The results may reveals, depicts and compares the building blocks and construction rule of different kinds of transportation networks.

^ae-mail: bangyali@163.com

^b* Corresponding author: e-mail: liangliu@shu.edu.cn

2 Data and methods

2.1 Transportation network data and model

In a complex network, nodes are used to represent individuals in the real system, and edges are used to represent the relationship between individuals. Two nodes connected by edges are regarded as adjacent. L model and P model are usually adopted in transportation network modeling[14]. The former regards transport stations as nodes, and if there are engineering lines connecting two adjacent nodes, an edge will be connected between them. The latter takes transport stations as nodes, and if there is a service line between any two nodes, an edge is connected between them. Model L generates a transport geographic network, while Model P generates a transport service network.

Transportation networks considered in this study include: China Expressway Network, China Railway Network, China National Road Network, Beijing Metro Network and Shanghai Metro Network, as well as China Air Passenger Network and China Railway Passenger Network. Among them, China's expressway network, railway network and national road network all take the city as the node, and the intercity roads or rails or other physical lines as the edges. Subway network takes subway

stations as nodes and tracks between stations as edges, that is, the above five networks generate transport geographic networks based on L model. China's air passenger transport network takes cities as its nodes and inter-city direct routes as its edges. China's railway passenger transport network takes cities as its nodes and inter-city direct passenger trains as its edges, the above two networks generate service networks based on the P model. Regardless of the distance of the lines and the direction of the connections, these networks are regarded as undirected and unweighted networks. Relevant data are from National Medium and Long-Term Railway Development Planning Diagram, National Highway Network Diagram, and Urban Subway Medium and Long-Term Planning Diagram, etc. The basic statistical datas of the network are shown in Table 1, N represents the number of stations in the transport network, M represents the number of edges in the transportation network, the average degree K refers to the average number of connections for all stations in the network. the average clustering coefficient C refers to average value of c (The clustering coefficient of one station is equal to the connections between its neighboring stations divided by all the possible connections between them), the average distance D refers to the average edges of the shortest path between any two stations in the network.

Table I. China's national transportation networks and their global characteristics

| Network | Number of nodes N | Number of edges M | The average degree K | The average Clustering coefficient C | The average distance D |
|---|-------------------|-------------------|----------------------|--------------------------------------|------------------------|
| Shanghai Metro Network (SH) | 212 | 235 | 2.226 | 0.002 | 13.203 |
| Beijing Subway Network (BJ) | 245 | 280 | 2.286 | 0.002 | 13.169 |
| China's expressway network (GS) | 259 | 380 | 2.934 | 0.026 | 9.150 |
| China Railway Network (TLP) | 427 | 533 | 2.496 | 0.042 | 15.777 |
| China National Highway Network (GD) | 1280 | 1456 | 2.275 | 0.003 | 25.313 |
| China Air Passenger Transport Network (HK) | 148 | 1065 | 14.392 | 0.695 | 2.135 |
| China Railway Passenger Transport Network (TLS) | 348 | 948 | 5.480 | 0.269 | 3.078 |

2.2 Subgraph search method

Subgraph search, that is, in the real network to search the specific scale of the subgraph, determine and classify the isomorphic subgraph. Because the number of subgraphs increases exponentially with the size of the network and subgraphs to be searched, its computational complexity is relatively high. Milo et al first adopted an exhaustive recursive search algorithm to search all subgraphs of a given scale[5]. In this algorithm, the input network is represented by an $N \times N$ adjacency matrix, and the corresponding derived subgraph is obtained by enumerating all the $n \times n$ submatrices in the input network. Only small-scale subgraphs can be found effectively, such as the subgraph of 3 nodes and 4 nodes.

Wernicke further studied and proposed ESU algorithm to find all subgraphs[15]. The algorithm starts from a node V in the input network and adds nodes with the following two properties to the set V_s , one is that their subscript must be greater than the subscript of V , and the other is that they do not belong to the adjacency point of a node in V_s . The

algorithm gradually increases the size of the subgraph from the subgraph of size 1. Although it is also an enumeration subgraph, it reduces the search space by providing constraints to ensure that a subgraph can be searched only once and no meaningless subgraph will be generated. However, due to the inherent complexity of the problem, the subgraphs that ESU searched cannot reach a large scale, and the computation speed is slow.

In order to improve the efficiency of subgraph search, Kashtan et al proposed a sampling algorithm to estimate the relative occurrence times of subgraphs [16]. The ESA algorithm randomly selects an edge and adds its associated vertex to the vertex set of subgraph $V_{subgraph}$. Then, the adjacency points of vertices in $V_{subgraph}$ are continuously added to the $V_{subgraph}$ until the subgraph reaches the specified scale and a sampled subgraph is obtained. This process is repeated until the number of subgraphs sampled reaches a predefined number. The ESA algorithm is independent of the network size and solves the problem that the number of searching subgraphs increases exponentially with the input network size and subgraph size. Theoretically, it can search larger subgraphs in the

network. In this study, sampling algorithm is used to determine all $n \times n$ ($n=3-7$) submatrices of the real network, and the number of subgraphs of the corresponding 3-7 nodes is obtained.

3 The structural diversity of building blocks in transportation network

3.1 Number of subgraph

The absolute number of 3- to 7-node subgraphs of the national transportation network is shown in Figure 1. Statistics show that the absolute number of the subgraphs of each transportation network grows exponentially with the size of the subgraphs to be searched, and the geographic network and service network have different growth rates. Taking the railway geographical network as an example, the absolute total number of 3- to 7-node subgraphs are 946, 1957, 4487, 10994 and 28107, respectively, with a difference of only two orders of magnitude. However, the number of 3- to 7-node subgraphs in the railway service network is 20474, 545637, 13909267, 323491265 and 683202238, respectively, with a difference of four orders of magnitude.

For the transport network belonging to the same geographic network or service network, the number of subgraphs of the same scale is positively correlated with the number of network nodes. For example, the number of nodes of Shanghai subway network, Beijing subway network, national railway network and national road network increases successively, so the number of subgraphs of corresponding network scales increases successively. It is note that the nodes number of expressway net is less than the national highway net, but the number of 6- and 7-node subgraphs is more than the national highway network, analysis shows that the number of subgraph have a great relationship with the average degree of the network, in general, The network with larger average degree has a relatively larger number of edges, It leads to a significant increase in each scale subgraphs, and behave more obviously as the subgraph size increase.

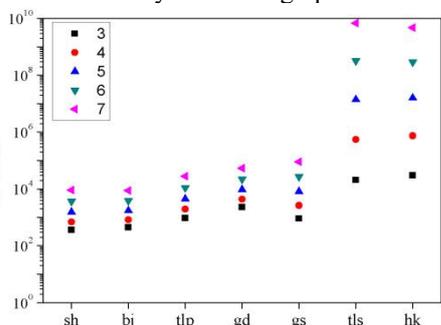


Figure 1. Absolute number of 3 to 7 nodes subgraphs in national transport networks

3.2 Subgraph form

Generally speaking, the possible subgraph forms in the network increase exponentially with the size of the subgraph to be searched. For example, there are 2 forms

of 3-node subgraphs, 6 forms of 4-node subgraphs and 28 forms of 5-node subgraphs. The number of subgraph forms of transport network is shown in Figure 2. Statistics show that all possible subgraph forms of each scale do not all appear in the transport network. Taking the railway geographical network as an example, there are 2, 5, 12, 31 and 80 types of forms in terms of 3- to 7-node subgraphs, respectively. There are 2, 6, 21, 112 and 853 subgraph forms of railway service network from 3- to 7-node subgraphs, and the number of subgraph forms of the two types has different growth rates. The results show that all the transport geographic network forms of 3- to 7-node subgraphs are contained in the transport service network, while the large quantum subgraph form of the transport service network does not appear in the transport geographic network.

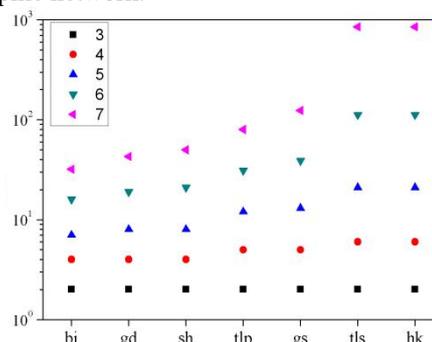


Figure 2. Number of categories of 3 to 7 nodes subgraphs in national transport networks

For the transport network belonging to the same geographic network or service network, the relationship between the number of subgraph forms of the same scale and the number of network nodes is not clear. For example, the number of nodes of Shanghai subway network, Beijing subway network, national railway network and national highway network increase successively, but the number of corresponding subgraph forms of Beijing subway network is less than that of Shanghai subway network, while the subgraph forms of national highway network are also less. The expressway network still has the most subgraph forms. It is preliminarily believed that this is related to the actual characteristics of the network. For example, the Beijing metro network has a radial and symmetrical ring structure, while the Shanghai metro network is more irregular, so it may lead to the differences in the form of subgraphs. The number of subgraph forms of railway service network and air service network is almost the same, which may represent all the local structure forms of transport service network. The number and types of the above subgraphs reflect the diversity of the basic structure of Transportation Networks from the local perspective, that is, the form and number of the specific subgraphs of different forms and scales of transport networks are different.

Transport geographic network and service network present different relative number order of subgraphs, mainly due to the difference of cost, technology and space constraints of transportation network planning and construction.

First, engineering cost constraint. Generally speaking, when the number of nodes of a local network is fixed, its connectivity or efficiency is proportional to the number of

network edges. However, when more edges are connected, the efficiency is improved, and the construction cost is higher. For example, the cost of subway network is positively correlated with the line length. Because the geographic network lines are physical lines, which are strongly restricted by the construction cost of the engineering, and the service network lines are abstract non-physical lines, which are less restricted by the construction cost of the engineering, the high cost and high connectivity mode of full connectivity is rarely seen in the geographic network, and more in the service network, but its number is also small.

Second, engineering technical constraint, as shown by the high density local structure, its connection construction has a certain engineering technical difficulty. For example, digging out lines from existing subway station platforms in use or taking space from the depth are often constrained by engineering technical capabilities. As a result, star and fully connected structures with high density and multiple edges appear less frequently in geographic networks, but can appear in large quantities in service networks without such engineering and technical constraint.

Third, the geographic space constraint, the performance is that the network occupies the space position, is subject to the space resource constraint, so the adjacent geographical location between the sites has a certain number of connections. For example, some stations in the core area of the subway network are limited to the number of connections due to the constraints of the surrounding space; adding an urban direct line to the expressway network is restricted by geographical factors, so it is impossible to increase and duplicate construction in large quantities. As a result, the star and fully connected structures are rare in geographic networks.

4 Uniformity of transport network construction rules

In this section, the sorting characteristics of the relative number of transportation networks subgraphs are considered. Select $n(=3\sim7)$ node size containing the

minimum edge $m(=n-1)$ of the three subgraph forms, as well as the maximum edge subgraph form. We take the 5-node subgraph as an example ($n=5$): the subgraph with minimum edge($m=4$) contains three typical subgraph forms of linear, Y-type and the star; the subgraph with maximum edge($m=10$) contains the fully connected subgraph form. Each subgraph has its own connotation in the transportation networks, the linear subgraph indicates that the stations in the transportation networks are successively connected but not annular; the star subgraph indicates that several unconnected sites are connected to a central site; the Y-type subgraph indicates that there are three unconnected sites connected with a central site, and one of the three sites leads to a number of successively connected sites; the fully connected subgraph indicate that all sites are connected to each other. The above subgraph types can be regarded as the typical local structure of the transportation network from the form and quantity.

Table 2 shows the relative number of linear, star, Y-type and fully connected four forms of the 3- to 7-node subgraphs of the transportation network, in which, 3-node subgraphs has only two forms, 4-node subgraphs has three forms, and 5- to 7-node subgraphs has four forms. The analysis shows that the relative number of four subgraph forms of each scale in seven transportation networks can be divided into two categories according to geographic network and service network. We take 5-node subgraph as an example: (1) in the transport geographic networks, Linear>Y-type>Star>Fully connected. The relative number of linear subgraph is the largest, accounting for about 50%, which is the main local structure of the network. Y-type subgraph is a secondary local structure, accounting for 35%-40%; the star subgraph accounts for less than 10%. The fully connected subgraph does not appear in the geographic network and is a unnecessary local structure. (2) in the transport service networks, Star>Y-type>Linear>Fully connected. The relative number of star subgraph is the largest, accounting for about 35%, which is the main local structure. The Y-type subgraph is a secondary local structure, accounting for about 20%. Linear subgraph accounts for about 3%, and fully connected subgraph rarely appears, which is a unnecessary local structure.

Table II. Relative number of typical subgraphs of Transportation Networks

| (n,m) | Subgraph type | graph | Shanghai Metro Network | Beijing Subway Network | China's expressway network | China Railway Network | China National Highway Network | China Railway Passenger Transport Network | China Air Passenger Transport Network |
|-------|-----------------|-------|------------------------|------------------------|----------------------------|-----------------------|--------------------------------|---|---------------------------------------|
| (3,2) | Linear | | 99.72% | 99.77% | 98.35% | 97.99 | 99.82% | 92.93% | 77.96% |
| (3,3) | Fully connected | | 0.28% | 0.23% | 1.65% | 2.01% | 0.18% | 7.07% | 22.04% |
| (4,3) | Linear | | 77.20% | 79.29% | 72.77% | 79.71% | 74.99% | 24.05% | 8.07% |
| (4,3) | Star | | 21.65% | 19.63% | 22.15% | 16.25% | 24.36% | 56.80% | 48.99% |
| (4,6) | Fully connected | | 0 | 0 | 0 | 0 | 0 | 0.36% | 4.46% |
| (5,4) | Linear | | 52.87% | 56.79% | 47.42% | 57.79% | 50.15% | 3.59% | 0.32% |
| (5,4) | Y-type | | 39.83% | 38.09% | 40.02% | 35.12% | 41.24% | 26.11% | 6.42% |
| (5,4) | Star | | 4.63% | 2.33% | 3.25% | 0.62% | 7.33% | 37.48% | 34.58% |

| | | | | | | | | | |
|--------|-----------------|---|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| (5,10) | Fully connected |  | 0 | 0 | 0 | 0 | 0 | 0.01% | 0.80% |
| (6,5) | Linear |  | 34.36% | 38.59% | 29.44% | 38.58% | 30.33% | 0.48% | 0.01% |
| (6,5) | Y-type |  | 25.52% | 25.89% | 23.54% | 23.32% | 26.29% | 2.41% | 0.12% |
| (6,5) | Star |  | 0.74% | 0.03% | 0.39% | 0 | 2.03% | 24.16% | 24.39% |
| (6,15) | Fully connected |  | 0 | 0 | 0 | 0 | 0 | 0.0003% | 0.13% |
| (7,6) | Linear |  | 21.66% | 25.34% | 17.64% | 24.43% | 17.37% | 0.06% | 0.0002% |
| (7,6) | Y-type |  | 16.47% | 17.11% | 14.28% | 14.49% | 14.14% | 0.29% | 0.002% |
| (7,6) | Star |  | 0.09% | 0 | 0.04% | 0 | 0.50% | 14.65% | 16.67% |
| (7,21) | Fully connected |  | 0 | 0 | 0 | 0 | 0 | 0.00001% | 0.002% |

Note: the boldface indicates that the subgraph form has the largest relative number among all the subgraph forms of the same size in the same transport network

In general, the typical subgraph forms and quantity characteristics of local transportation network structures can be considered as a compromise of constraints, such as cost, geography, technology and the efficiency of supply.

(1) the transport geographic network is highly constrained by the engineering cost, technology and space. In this kind of network, the linear subgraph with low connectivity constitutes the basic framework; the star and fully connected subgraphs with high connectivity are difficult to appear due to the constraints. Y-type subgraph with medium connectivity appears more often in the network because of the balance between constraint and efficiency, and this kind of subgraph improves system connectivity effectively without significantly increasing the engineering cost and construction complexity.

(2) the transport service network is an abstract network, almost unaffected by the space constraints, and less affected by the cost and technical constraints. In this kind of network, the Y-type subgraph with medium connectivity and star subgraph with high connectivity constitute the basic structure, the linear subgraph with low connectivity is relatively rare, and fully connected subgraph with high connectivity also appears in the network as a feature component that greatly improves network connectivity.

The above characteristics actually show the relative uniformity of the local structure ordering from the perspective of geographic network and service network.

5 conclusion

In this study, subgraph search method of the complex network was applied to the bottom-up identification of the national transportation network, to discuss the structural characteristics and their construction rules. It is clarified even if the TNs belong to generalized transportation system, the building blocks of them(subgraph form and quantity) have diversity characteristics. At the same time, the subgraph ordering of the TNs can be divided into geographic networks and service networks, which reflects the unity of network construction rules of geographic networks or service networks due to the similarity of cost, technology and geographical constraints.

Complex functions are often conceived in simple structures, and there are a few simple rules that generate

complex systems. Different transportation networks with different scales are complex and diverse in the form and quantity of specific subgraphs, the transport geographic networks or transport service networks are generated from the bottom up based on the relatively unified construction rules, reflecting the constraints of cost, geography and technology. This rule can be used as an important principle and standard for the planning, design, construction and optimization of TNs.

As a new research field, local structure analysis of complex networks has made great progress and been widely used in biology and physics[17]. The local structure of complex network can be used to analyze the structure and organization of key engineering network, and further study the structure form and construction rules of the building blocks, including power network, communication network, etc[18]. On the basis of in-depth understanding of the behavior and function of building blocks, it will help us understand the dynamic characteristics of higher scale engineering network structure.

Acknowledgment

This work is supported by the Natural Science Foundation of China (Nos. 71602107 and 71874123), the China Scholarship Council (No. 201806895024), and the Shanghai Leading Academic Discipline Project (No. B310).

References

1. Watts D J and Strogatz S H, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393(6684), 1998, pp. 440-442.
2. Albert-László Barabási, and Réka Albert, "Emergence of Scaling in Random Network," *Science*, vol. 286(5439), 1999, pp. 509-512.
3. Albert R, Barabási A L, "Statistical mechanics of complex networks," *Rev. Mod. Phys.*, vol. 74(1), 2002, pp. 47-97.
4. Newman M E J, "The structure and function of complex networks," *SIAM Review*, vol. 45(2), 2003,

- pp. 167-256.
5. Milo R, Shen-Orr S, Itzkovitz N, et al, "Network motifs:simple building blocks of complex networks," Science, vol. 298(5594), 2003, pp. 824-827.
 6. Shen-Orr S, Milo R, Mangan S, et al, "Network motifs in the transcriptional regulatory network of escherichia coli," Nature Genetics, vol. 31(1), 2002, pp. 64-68.
 7. Steven H. Strogatz, "Exploring complex networks," Nature, vol. 410(6825), 2001, pp. 268-276.
 8. Sen P, Dasgupta P, Chatterjee A, et al, "Small-world properties of the Indian railway network," Nature, vol. 67(3), 2003, 036106.
 9. Wang R, Tan J X, Wang X, et al, "Geographic coarse graining analysis of the railway network of China," Physica A, vol. 387(22), 2008, pp. 5639-5646.
 10. Wu J J, Gao Z Y, Sun H J, et al, "Urban transit as a scale-free network," Modern Physics Letters B, vol. 18(19), 2004, pp. 1043-1049.
 11. Panagiotis A., David F, "Large subway systems as complex networks," Physica A, vol. 367, 2006, pp. 553-558.
 12. Seaton K A, Hackett L M, "Station, Trains and Small-world Networks," Physical A, vol. 339(3/4), 2004, pp. 635-644.
 13. Barabási A L, Bonabeau E, "Scale-free networks," Sci Am, vol. 288(5), 2003, pp. 60-69.
 14. M. Kurant, P. Thiran, "Extraction and analysis of traffic and topologies of transportation networks," Phys. Rev. E, vol. 74(3), 2006, 036114.
 15. Wernicke S, "Efficient detection of network motifs," IEEE Trans on Computational Biology and Bioinformatics, vol.3(4), 2006, pp. 347-359.
 16. Kashtan N, Itzkovitz S, Milo R, Alon U, "Efficient sampling algorithm for estimating subgraph concentrations and detecting network motifs," Bioinformatics, vol. 20(11), 2004, pp. 1746-1758.
 17. Roy Soumen, "Systems biology beyond degree, hubs and scale-free networks: the case for multiple metrics in complex networks," Systems and synthetic biology, vol. 6(1-2), 2012, pp. 31-34.
 18. Casaca, Henriques, "Using Components of the Mahalanobis Squared Norm of the Residuals Vector for Quality Control in Local Engineering Networks," Survey Review, vol. 43(321), 2011, pp. 241-251.