ISSN: 2664-9640

Research on the Division of Container Shipping Network Community Based on Louvain Algorithm

Shoupeng Xie [1], Yan Zhang [2] and Cheng Zhang [2, *]

[1] China Harbour Engineering Company Ltd, Beijing 100027, China;

[2] College of Harbour Coastal and Offshore Engineering, Hohai University, Nanjing, Jiangsu 210098, China.

* Corresponding Author

Abstract

In order to explore the network structure and characteristics of the container shipping network, this paper builds a container shipping network model, and uses the Louvain algorithm to divide the community structure of the network. The container shipping of countries along the "21st Century Maritime Silk Road" composed of 453 ports and 3444 sides was selected as a case to construct an undirected shipping network with weights. Based on the Louvain algorithm in the community discovery method, the undirected contains distance rights. The "21st Century Maritime Silk Road" port shipping network is divided into 12 communities. By comparing the results of the division with the geographical location-based port group division method, it proves the rationality of the weighted network division results and the effectiveness of the community discovery algorithm in the shipping network.

Keywords

Container shipping network, Community discovery, Complex network, Louvain algorithm.

1. Introduction

With the continuous acceleration of global economic integration, trade exchanges between countries in the world are getting closer and closer. As a low-cost and large-volume transportation method, maritime transport plays a vital role in international trade. International maritime transport The structure and operation mode of the network (hereinafter referred to as the shipping network) has a profound impact on the pattern of international trade[1]. Because the shipping network has the characteristics of large volume and many elements, the existing research on shipping network focuses on the macro level, and the research results are difficult to systematically guide the development of the local shipping network from the micro level. Therefore, in order to combine the macro-level research and micro-level research of the shipping network to explore the relationship between different levels of shipping networks, there is an urgent need for the shipping network Carry out the division of the scientific system.

There are a large number of nodes (ports) and connections (routes) in the shipping network. The connections between each node are intricate, that is, a network that is not completely connected, or a network that is not randomly connected. Due to factors such as the geographic location of the port, the shipping network is both It shows a certain regularity and a certain randomness, which belongs to a complex network system. In recent years, complex network theory has been applied to the field of shipping research. For example, Pablo Kaluza et al.[2] analyzed and compared the route networks of three typical ships: bulk carrier, container ship, and oil tanker, and found that the route networks of the three types of ships have different routes. César Ducruet et al.[3] took the global container liner route network from 1996 to 2006 as the research object, analyzed the centrality of the network, and proved that the global container liner Robustness of the route network; Tian Wei[4] analyzed the shipping network of Maersk Line and proved the small world and scale-free characteristics of the network, and pointed out the difference between this network and other complex networks; Xiong Wenhai[5]From the perspective of the structure and function of the shipping network, the global shipping network was studied, and the structure and evolution mechanism of the shipping network were analyzed; Du Chao et al.[8] analyzed the China container shipping network, and a large number of studies have proved the complex network The feasibility of the theory in container shipping network analysis[9]-[13], and the shipping network satisfies the properties of "small world network" and "scale-free network"[2] [3].

However, the existing research results mainly focus on the topological properties and statistical indicators of the shipping network. The research object is the macroscopic properties of the network. There are few research results on the microstructure of the shipping network community structure, which makes some unique properties of the shipping network impossible. Reflected in the network model. In addition, compared to point-to-point transportation methods such as tankers and bulk shipping, container shipping networks have higher complexity, and container shipping has fixed schedules and routes, and its structure is more regular. Therefore, this article will use container Shipping is a representative shipping system as the research object. In summary, from the perspective of complex networks, this article conducts research on the community structure of container shipping networks based on the community discovery model.

2. Complex network model of container shipping

2.1. The basic elements of the complex network of container shipping

For the container shipping network, the nodes of the network are ports, and the sides are the routes between ports. The two together form the basic structure of the network. Then the container shipping network can be defined as: $G(V,E)$. Among them, $v=\{V_i|i=1,2,\cdots,n\}$ is the node of the network, that is, the port, where n is the number of ports, $E=\{e_{i,j}|i,j=1,2,\cdots,n,i\neq j\}$, which means the edge of the network, that is, the route between ports. Considering that the container liner routes are mostly circular or round-trip forms, and our research object is the connectivity of the system and the community structure, therefore, the container shipping network is regarded as an undirected network. In addition, the length of the route between the two ports reflects the relationship between the nodes to a certain extent. Therefore, the route distance between the ports can be used as the weight of the edge to construct a complex network model with undirected rights.

2.2. Construction of a complex network of container shipping

According to the basic components of the complex network of container shipping, the basic data for constructing the network is the shipping schedule data. This article selects the container liner schedule of China International Shipping Network as the basic data, and obtains the inter-port container liner schedule within the research scope. Store the acquired raw data on the basis of preprocessing. The basic data is divided into two parts: port information and liner route information. The port information data elements include the Chinese and English name of the port, the latitude and longitude of the continent, country, and geographic location of the port. The container liner route information elements include the names of the loading and unloading ports of each liner route. The network adjacency matrix after sorting is shown in Table 1. The length of the route between ports is the weight of the edge in the network, denoted as $\omega_{i,j}$.

ISSN: 2664-9640

Table 1 Network adjacency matrix

	Shanghai	Shenzhen	Port Klang	Hamburger	…
Shanghai	0	1	1	1	…
Shenzhen	1	0	1	1	…
Port Klang	1	1	0	1	…
Hamburger	1	1	1	0	…
…	…	…	…	…	…

2.3. Community structure of container shipping network

A community is a collection of units with certain common characteristics. The node set in a complex network can be divided into different subsets according to their characteristic attributes. The main goal of community discovery is to find nodes with similar properties by relying on the nature of the network topology. group. Analyzing the community structure in a complex network helps to understand its internal organizational structure and the function and effect of each community on the entire network. As a kind of complex network, shipping network also has a community structure[3]. As a typical transportation network, the formation of the container shipping network is affected by many factors such as topological advantages and geographic limitations. For example, the topological advantage promotes the port nodes in the network to preferentially form the edge of the central port node, while geographical restrictions. The factor is often in the consideration of freight, which prompts the port nodes to give priority to forming a connection with a shorter transportation distance. In addition, factors such as social economy and urban development are inevitably intertwined in the development and evolution of the container shipping network, and manifest as certain network characteristics.

With the in-depth research on the methods of community discovery in complex networks, many algorithms dedicated to dividing the network structure have been developed. Common community discovery methods include: community discovery methods based on the idea of splitting and cohesion algorithms, and community discovery based on graph partitioning. Methods, community discovery methods based on local information, etc. Based on the characteristics of undirected rights in the container shipping network, this paper uses the Louvain algorithm to divide the undirected container shipping networks with rights into communities in order to understand the results of the container shipping network.

3. Algorithm of shipping network community division based on Louvain algorithm

3.1. The modularity of community structure

There is no universally recognized definition of the community structure of complex networks. Intuitively, the network is composed of multiple “communities”. These communities have close internal connections and relatively few connections between communities. The schematic diagram is shown in Figure 1. The current definitions of associations mainly include: "local associations" and "global associations", "strong associations" and "weak associations", "leader associations" and "self-organizing associations", etc. However, in the study of the division of associations, these definitions of associations cannot have a clear standard to quantify the division of associations, nor can they measure the effect of the division of associations.

At present, it is generally accepted that the measure of community structure is the modularity function defined by Newman and Girvan[15]. Modularity refers to the difference between the

ISSN: 2664-9640

proportion of edges connecting the internal nodes of the community structure and the expected value of the proportion of the edges connecting other nodes in the network and the internal nodes of the community structure subtracted from the expected value. The Q function can be used to quantitatively describe the modularity level of community division. The basic calculation formula of modularity is as follows:

$$Q = \frac{1}{2m}\sum i,j\left[A_{ij} - \frac{k_i k_j}{2m}\right]\delta(C_i, C_j) \tag{1}$$

$$\delta(u,v) = \begin{cases} 1 & if(u == v) \\ 0 & else \end{cases} \tag{2}$$

In the formula, A_{ij} is the element in the adjacency matrix in the network. If there is an edge between v_i and v_j, then $A_{ij} = 1$, otherwise A_{ij}=0. k_i is the degree value of node i, $\frac{k_j}{2m}$ is the probability of edge connection between node j and other nodes in the network, so $\frac{k_i k_j}{2m}$ is the theoretical number of edges between node i and j.

In order to facilitate the analysis of the value of modularity and its physical meaning, formula (1) and (2) are equivalently processed. Assuming that the result of the community divided in the network contains k communities, define a $k \times k$-dimensional symmetric matrix $e = (e_{ij})$. Among them, e_{ij} represents the ratio of the number of edges connected between the i-th community and the j-th community to the total number of edges in the network, $Tre = \sum_i e_{ii}$ is the sum of the elements on the diagonal of the matrix, which represents the inside of the community in the network The proportion of the number of edges connected between nodes in the total number of edges in the network, a_i is the sum of the elements in each row or column, which represents the number of edges connected to the nodes in the i community among all the edges Accounted for. Therefore, the Q function can also be expressed as:

$$Q = \sum_i (e_{ii} - a_i^2) = Tre - \| e^2 \| \tag{3}$$

Among them, $\| e^2 \|$ represents the sum of all elements in the matrix e^2. If the proportion of edges within the community is less than or equal to the expected value of the proportion of edges connecting any nodes, then $Q \leq 0$. The upper limit of Q is 1. The closer to the upper limit, the more obvious the community structure, and the corresponding community structure is well divided. In the actual network, the Q value is usually between 0.3 and 0.7. In the process of dividing communities, calculate the modularity value ($Q\ value$) corresponding to each division, and find one or more divisions corresponding to the largest Q value, which is the optimal division of the community structure.

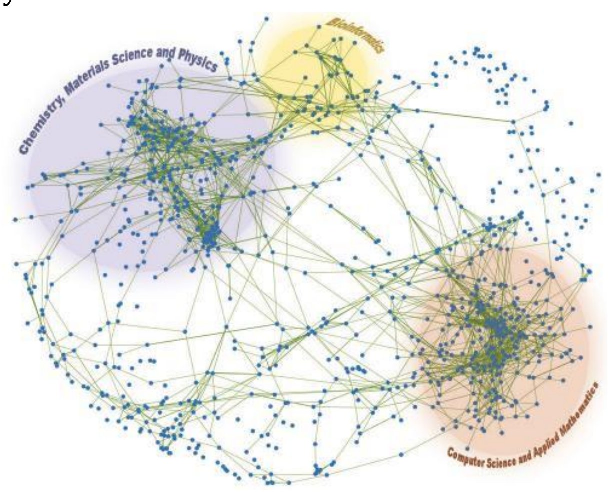

Figure 1: Schematic diagram of network community structure

ISSN: 2664-9640

3.2. Container shipping network partition algorithm based on Louvain algorithm

Louvain algorithm is a community discovery algorithm based on modularity proposed by Yale University scholar Vincent on the basis of GN algorithm[16]. The Louvain algorithm can accurately divide the hierarchical structure of the community, maximize the modularity of the entire network, avoid the omission of small community structures, and can be used to divide the weighted network. In order to make the modularity index usable in Louvain's algorithm, the formula (1) needs to be transformed as follows:

$$Q = \frac{1}{2m} \sum C \left[\sum in - \frac{(\sum tot)^2}{2m} \right] \tag{4}$$

Among them, $\sum in$ represents the sum of the weight values of the internal edges of the community C, and $\sum tot$ represents the sum of the weight values of the edges between the internal nodes of the community C and all other nodes in the network.

After further simplification of the above formula:

$$Q = \sum C \left[e_C - {a_C}^2 \right] \tag{5}$$

Therefore, the modularity can be regarded as the difference between the sum of the weights of the edges in the community and the sum of the weights of the edges between the nodes in the community and all other nodes in the network. Since the port shipping network studied in this paper is an undirected network, the calculation of modularity can be understood as the sum of the degree values of the internal edges of the community minus the total degree value of all port nodes in the community.

The Louvain algorithm is divided into two stages: In the first stage, each node in a network with n nodes is assigned to a community with a different number. Therefore, in the initial stage, each node forms a community by itself. For any community node i, consider adding it to another community node j and calculate the incremental change of the modularity Q:

$$\Delta Q = \left[\frac{\sum in + 2k_{i,in}}{2m} - \left(\frac{\sum tot + k_i}{2m}\right)^2 \right] - \left[\frac{\sum in}{2m} - \left(\frac{\sum tot}{2m}\right)^2 - \left(\frac{k_i}{2m}\right)^2 \right] \tag{6}$$

If the incremental change is greater than 0, put it on the node with the largest change in the corresponding modularity; otherwise, the node will keep the original community number unchanged. Repeat the process until the numbers of all nodes no longer change, that is, the modularity value of the network no longer changes. In the second stage, each community with a different number calculated in the first stage is regarded as a new node, and a new network is constructed with the new node unit and its edges. Then the new network formed will repeat the process of the first stage until the maximum modularity value is found; that is, as the iteration continues, the modularity value of the network no longer increases. The specific steps of the algorithm are as follows:

1. Treat each node in the network as a community;

2. For each node i, try to join it to the community of other nodes connected to node i in turn, and calculate the modularity increment in the network before and after joining. If the modularity increment is greater than 0, write down the number of the community where the neighbor node is located, and add node i to the community; otherwise, node i will remain unchanged in the original community;

3. Repeat the above two steps until the community numbers of all nodes in the network no longer change (the nodes no longer change);

4. Regard each community calculated by the above steps as a new node, the weight of the edge between nodes in the original community is converted to the weight of the new node, and the

weight of the edge between the original community is converted to the weight of the new node. Border rights

5. Repeat the above steps until the modularity value of the community in the entire network no longer changes.

The Louvain algorithm addresses the shortcomings of some community discovery algorithms that omit small communities, and proposes different levels of community merging calculations. The first stage is to initially merge closely connected nodes, and then based on the calculation results, the first stage is divided into communities as The nodes of the new network construct a new network. In the second stage, the new network is merged and divided into larger communities, so as to avoid missing some small communities, and the result of the division is closer to the actual situation of the network.

4. Case study

4.1. "21st Century Maritime Silk Road" container shipping network

The "21st Century Maritime Silk Road" is an important part of China's "One Belt, One Road" initiative. Its core is a shipping network composed of ports and routes. With the gradual advancement of the strategy, it will become an important channel for communication between China and countries along the route. This article selects the container liner network of countries along the "21st Century Maritime Silk Road" as the research object.

The geographical scope along the "21st Century Maritime Silk Road" mainly covers economic regions such as ASEAN, the Middle East, South Asia, North Africa, and Europe. The container shipping network of the "21st Century Maritime Silk Road" selected in this article is from May 2017 to September 2017. There are 453 ports in 123 countries and regions on four continents.

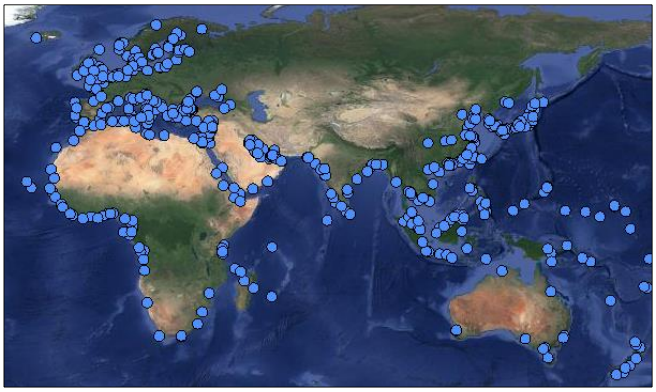

Figure 2: "21st Century Maritime Silk Road" Network Port Distribution Map

Collect the liner routes between the ports in the research area, and get 7038 connections between the port nodes in the network after de-duplication and correction. Considering the geographical location of the port, the distribution of the container shipping network of the "21st Century Maritime Silk Road" is shown as follows. It will be calculated that the average degree value of the network is 23.21, and the degree value distribution diagram of port nodes is shown in Figure 2. Its degree distribution obeys the power law distribution, which is a scale-free network. The smallest port node degree value is 1, and the largest port node degree value is 207 (Singapore Port). The average agglomeration coefficient of the network is equal to 0.578, which shows the small-world characteristics of the container shipping network of the "21st Century Maritime Silk Road".

4.2. Analysis of community structure

Using the Louvain algorithm-based weighted network community structure division method in Section 2.2, the weighted "21st Century Maritime Silk Road" port shipping network composed

ISSN: 2664-9640

of 453 nodes and their interconnected edges with distance weights is divided into 12 The distribution of communities and communities is shown in Figure 3. The port nodes of the same color in the topology diagram are in a unified community. It can be seen directly from the figure that the port nodes of the same color are clustered together. Regardless of whether the port nodes in the network are core port nodes close to the network center or general port nodes with a smaller degree, as long as they are port nodes in a community, the connection between them is basically better than that of other non-same communities. The ties should be close. The modularity value Q of the network is 0.727, which is a larger value among the modularity values of general complex networks, which further shows that the internal relations of the container shipping network of the "21st Century Maritime Silk Road" are closer, and the port community's internal resistance to external interference strong ability.

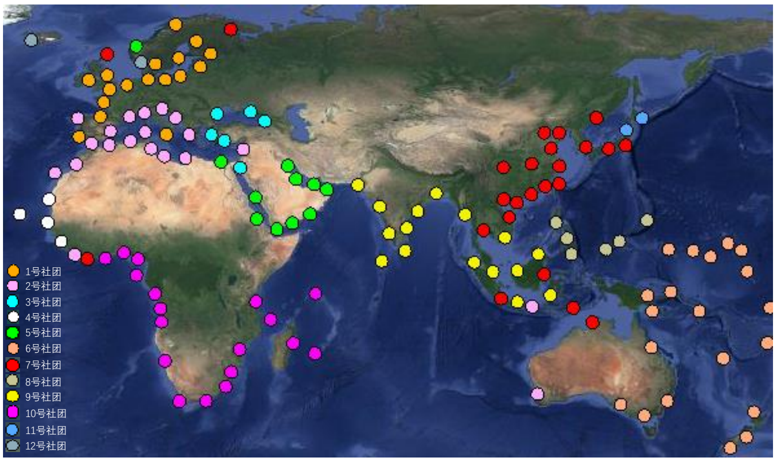

Figure 3: Schematic diagram of the distribution of associations

The 12 societies are represented by numbers 1 to 12 respectively. Among them, the internal port nodes of the No. 1 community are mainly distributed in Western Europe and Northern Europe and basically include the main ports in Western Europe and Northern Europe; the internal port nodes of the No. 2 community are mainly distributed in southern Europe and North Africa, and the Mediterranean coast; the internal ports of the No. 3 and 5 communities The nodes are mainly distributed in West Asia, the internal port nodes of the No. 3 community are mainly distributed in the Black Sea-Aegean Sea-the southeastern coast of the Mediterranean Sea, and the internal port nodes of the No. 5 community are mainly distributed on the Red Sea coast-the coast of the Arabian Peninsula-the coast of the Persian Gulf; the interior of the No. 4 community The port nodes are mainly distributed in West Africa; the internal port nodes of the No. 6 community are mainly distributed in the South Pacific-the west coast of Australia; the internal port nodes of the No. 7 community are mainly distributed on the coast of Northeast Asia-East Asia (China)-Southeast Asia; the internal port nodes of the No. 8 community Mainly distributed in the Philippines and nearby waters; No. 9 community internal port nodes are mainly distributed along the coast of South Asia-Southeast Asia; No. 10 community internal port nodes are mainly distributed along the coast of West Africa-South Africa; No. 11 community internal port nodes are distributed in Japan; No. 12 community internal port nodes The port nodes are distributed in Norway and Iceland.

Figure 4: No.1 community

Figure 5: No.2 community

Figure 6: No.3 community

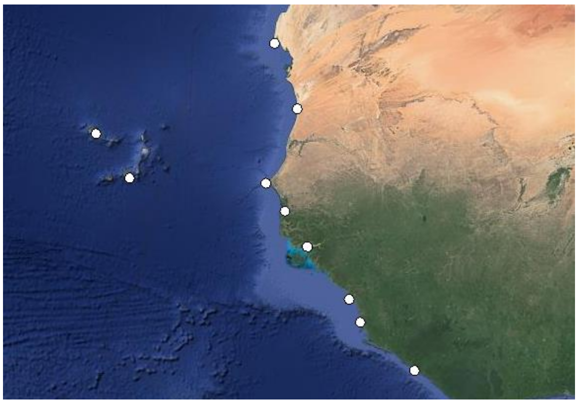

Figure 7: No.4 community

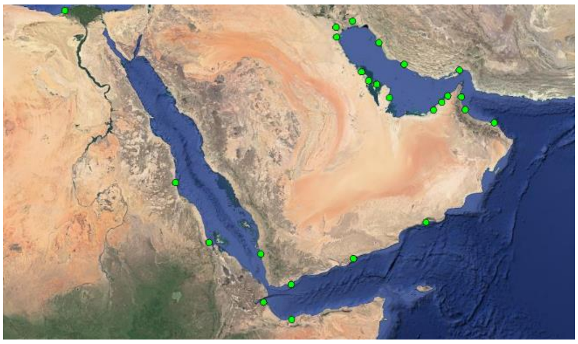

Figure 8: No.5 community

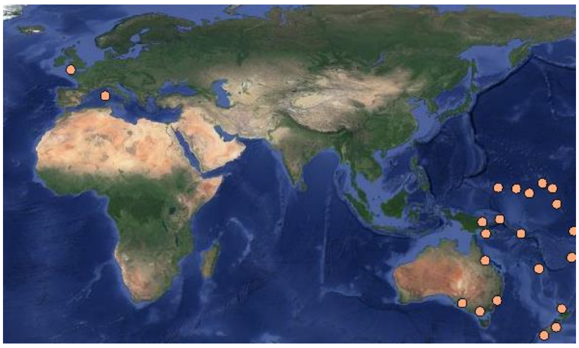

Figure 9: No.6 community

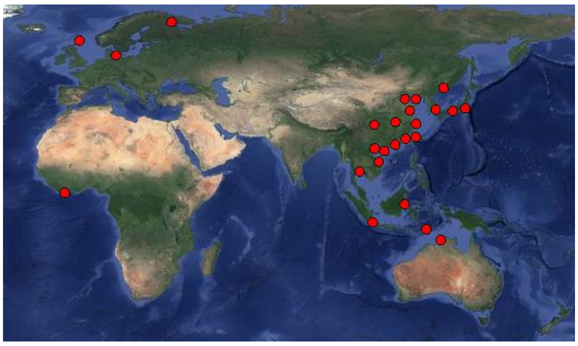

Figure 10: No.7 community

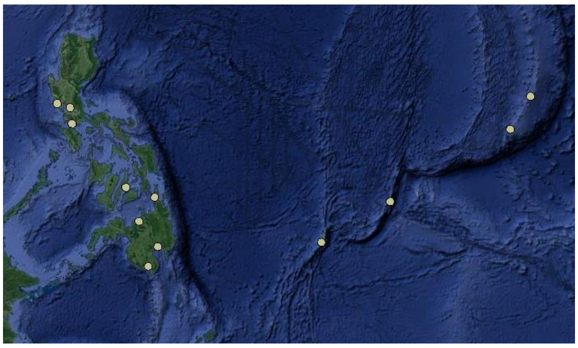

Figure 11: No.8 community

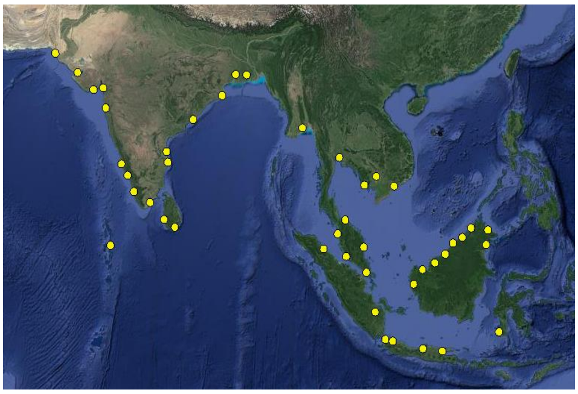

Figure 12: No.9 community

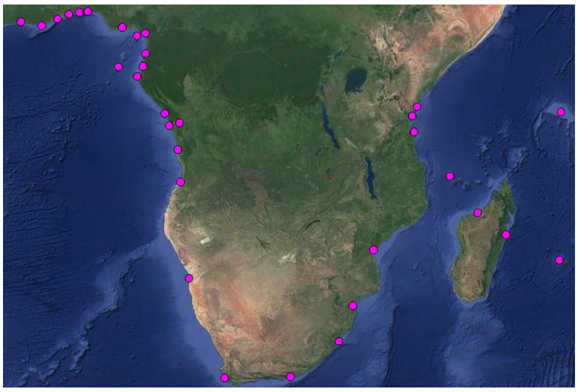

Figure 13: No.10 community

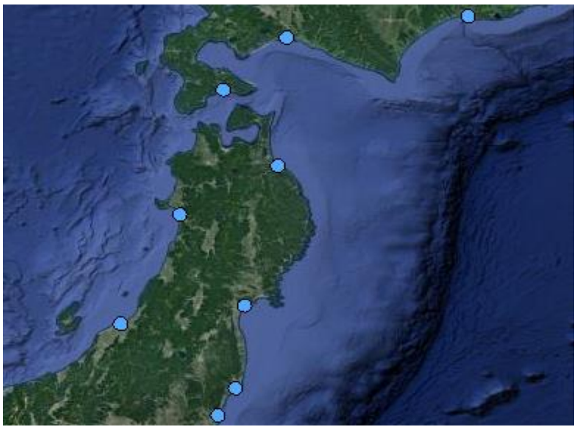

Figure 14: No.11 community

Figure 15: No.12 community

Combining the number of port nodes in each community, the distribution of geographic spatial locations, and the distribution of port node degrees within the community, the community and the entire "Sea Silk Road" port shipping network are systematically divided. Different from the traditional geographical divisions such as "Southeast Asia", "West Asia", "North Africa" and other geographical meanings, this article discusses the 414 of the eight communities 1, 2, 3, 5, 6, 7, 9, and 10. The division of port nodes in the port combines the geographical location and connectivity factors between port nodes, and the geographical location of the port nodes of the "Sea Silk Road" port shipping network is re-divided from the perspective of the network system. In terms of the number of nodes and their geographic spatial distribution, the port nodes within the communities 1, 2, 3, 5, 6, 7, 9, and 10 almost cover all levels of ports in the region where the internal port nodes of the community are located. From the perspective of the degree value range of internal port nodes in the community, the eight internal port nodes in the above-mentioned community cover the core port nodes in the region where they are located.

In the “Sea Silk Road” port shipping network studied in this article, in order to eliminate the interference of actual geographical factors, only connectivity and route distance are considered in the division of communities, and the information of the country or region where the port node is located is not calibrated in the input program data. However, the results of the division of associations show that the ports of most countries and regions are still divided into the same associations. As a result, the scientific quantitative analysis of the “Marine Silk Road” port and shipping network from the perspective of network connectivity is truly irrelevant. The division of interference from geographical factors such as countries and regions. However, the result of the division is in harmony with the traditional geographical division of the country and region where the port node is located, which also proves the rationality of the result of the weighted network division.

ISSN: 2664-9640

5. Conclusion

This paper uses the Louvain algorithm to divide the undirected "21st Century Maritime Silk Road" container shipping network with distance rights into 12 communities, and scientifically quantitatively analyzes the "21st Century Maritime Silk Road" container shipping network from the perspective of network connectivity. Division not restricted by geographical factors. The port nodes of each community are analyzed in combination with their geographic spatial distribution, and it is found that the topological division result is compatible with the traditional geographical division of the port, which also proves the rationality of the weighted network division result.

References

[1] Yan Zhanglong. *Research on Topology and Evolution of World Shipping Network* (MA, Dalian Maritime University, China 2013).

[2] Kaluza P, Kölzsch A, Gastner M T, et al: The complex network of global cargo ship movements, Journal of the Royal Society Interface, Vol.7 (2010) No.48, p.1093-1103.

[3] Ducruet C, Notteboom T: The worldwide maritime network of container shipping: spatial structure and regional dynamics, Global Networks, Vol.12 (2012) No.3, p.395-423.

[4] Tian Wei. *Research on the Network Effect of Container Ports* (Ph.D., Dalian University of Technology, China 2008).

[5] Xiong Wenhai. *Research on the Structural Characteristics and Dynamic Behavior of the World Shipping Network* (Ph.D., Qingdao University, China 2009).

[6] Wu Di. *Research on the Vulnerability and Risk Control of the Maritime Silk Road Container Shipping Network* (Ph.D., Dalian Maritime University, China 2018).

[7] Ducruet C, Notteboom T: The Worldwide Maritime Network of Container Shipping: Spatial Structure and Regional Dynamics, Global Networks, Vol.12 (2012) No.3, p.395-423.

[8] Han Zenglin, An Xiaopeng, Wang Li, Wang Chengjin, Wang Lihua, Li Yajun: The layout and optimization of China's international container transportation network, Acta Geographica Sinica, Vol.04 (2002), p.479-488.

[9] Du Chao, Wang Jiao'e, Mo Huihui: Research on the Spatial Pattern and Complexity of China's Container Shipping Network, Resources and Environment in the Yangtze River Basin, Vol.25 (2015), p. 190-198

[10] . Kam-Fung Cheung,Michael G.H. Bell,Jing-Jing Pan,Supun Perera: An eigenvector centrality analysis of world container shipping network connectivity, Transportation Research Part E, (2020), p. 140.

[11] César Ducruet: The geography of maritime networks: A critical review, Journal of Transport Geography, (2020), p. 88.

[12] Zhang Jinwei. *Analysis on the Characteristics and Robustness of the Maritime Silk Road Liner Shipping Network* (MA, Dalian Maritime University, China 2020).

[13] Feng Xuejun, Zhang Cheng, Jiang Liupeng, Zhang Yan, Jiang He: "Maritime Silk Road" Container Shipping Network Routing Strategy Research, Complex systems and complexity science, Vol.14, (2017), No.04, p. 58-65.

[14] Jiang He, Fan Xiaojing, Feng Xuejun, Jiang Liupeng: Shipping Network Community Structure Based on Newman Fast Algorithm, Journal of Changsha University of Science and Technology (Natural Science Edition), Vol. 15, (2018), No. 03, p. 35-39+93.

[15] Newman M E J, Fast Algorithm for Detecting Community Structure in Networks, Physical review E, Vol. 69,(2004), No. 6.

[16] Blondel VD, Guillaume JL, Lambiotte. R, Lefebvre, E, Fast unfolding of communities in large networks, Computer Science, (2008).