

文章编号:1672-9331(2018)03-0035-05

# 基于 Newman 快速算法的航运网络社团结构

蒋 鹤<sup>1</sup>, 范小晶<sup>2</sup>, 封学军<sup>1</sup>, 蒋柳鹏<sup>1</sup>

(1.河海大学 港口海岸与近海工程学院, 江苏 南京 210098;

2.中交上海航道勘察设计研究院有限公司, 上海 200120)

**摘 要:** 为了研究航运网络的网络结构及其相关特性,以航运网络的社团结构为对象研究,构建了基于模块度的 Newman 快速算法。选择由 453 个港口和 3 444 条边构成的“二十一世纪海上丝绸之路”(简称“海丝之路”)无向无权航运网络为案例,从航运网络连接性视角,采用 Newman 快速算法对“海丝之路”航运网络进行计算,得出该航运网络的社团结构性质:网络中只有一个由 173 个港口节点构成的核心社团,且为领导者社团,度值接近于幂律分布,其他均为小型社团,表明“海丝之路”集装箱航运网络为巨型社团结构,且具有核心节点。

**关键词:** “二十一世纪海上丝绸之路”航运网络; 复杂网络; 社团发现; 网络结构; Newman 快速算法; 模块度

**中图分类号:** U659

**文献标识码:** A

## Community structure of shipping network based on Newman fast algorithm

JIANG He<sup>1</sup>, FAN Xiao-jing<sup>2</sup>, FENG Xue-jun<sup>1</sup>, JIANG Liu-peng<sup>1</sup>

(1.College of Harbour Coastal and Offshore Engineering, Hohai University, Nanjing 210098, China;

2.Shanghai Waterway Engineering Design and Consulting Co. Ltd., Shanghai 200120, China)

**Abstract:** To study the network structure and related properties of shipping network, this paper researches the community structure of shipping network. The Newman fast algorithm based on modularity is constructed. And "The twenty-first Century Maritime Silk Road" shipping network is chosen as the case, which is unweighted and undirected shipping network, and composed of 453 ports and 3 444 edges. From the perspective of shipping network connectivity, the Newman fast algorithm is used to calculate "The twenty-first Century Maritime Silk Road" shipping network. The structural properties of this shipping network can be obtained. There is only one core community in this shipping network, which is leader community, and consists of 173 ports. The degree follows the power-law distribution. Other communities are non-core communities. The "The twenty-first Century Maritime Silk Road" container shipping network owns huge community structure with the core nodes.

**Key words:** "The twenty-first Century Maritime Silk Road" shipping network; complex network; community structure detection; network structure; Newman fast algorithm; modularity

收稿日期:2018-04-27

基金项目:国家自然科学基金资助项目(41401120)

通讯作者:蒋柳鹏(1984-),男,江苏泰州人,河海大学讲师,博士,主要从事水运经济与规划、复杂网络等方面的研究。E-mail:jsjlp@hhu.edu.cn

随着全球经济一体化进程的快速推进,航运网络发挥着越来越大的作用。航运网络是一类典型的复杂网络。

复杂网络理论作为复杂性理论研究的前沿,日益受到学者们的重视。从系统整体的角度去认识复杂网络的特性正成为研究的热点。复杂网络理论应用于港口航运系统的研究主要集中于以下几个方面:①关于航运网络的复杂性特征方面的研究。陈芙英<sup>[1]</sup>以海上丝绸之路东南亚航运网络为对象,验证了海上丝绸之路东南亚航运网络符合复杂网络特征;李电生<sup>[2]</sup>以中国城市港口航运网为研究对象,得出中国港口航运网具有“小世界”、无序性等特点。②关于航运网络的结构特征的研究。Wilmsmeier等<sup>[3]</sup>通过对拉丁美洲和加勒比地区港口的空间演化进行分析,得出集装箱航运网络中港口的空间布局及其特征;蒋柳鹏、蹇令香等<sup>[4,5]</sup>以港口群或航线的视角来对港口系统的发展规律进行分析,得出港口的空间结构和演化趋势。③从网络视角分析航运网络结构。Ducruet<sup>[6]</sup>以全球海运流为研究对象,得出网络增长的集聚特性和运输量分布特点;Ducruet和Wang等<sup>[7]</sup>以船舶的动态为视角对中国的全球航运连通性进行分析,强调了技术、经济和政治因素对改变内外部连通性分布的影响。

综上所述,目前已有的研究成果主要集中于航运网络的特征。对航运网络社团结构的研究成果较少,使得针对航运网络的一些独特属性无法在网络模型中体现出来。针对现有研究的不足,作者从复杂网络的视角,依托社团发现方法,对航运网络的结构开展研究。

## 1 航运网络要素及数据采集

### 1.1 复杂网络的一般复杂性特征

真实复杂网络中的拓扑结构既不是完全规则的,也不是随机生成的,而是由局部密集的子图与部分随机连接而形成的复杂的网络结构。随着复杂网络理论的突破性发展,网络分析的方法体系不断充实和完善,目前主流的复杂网络拓扑性质及其特征条件如表1所示。通过分析网络的主要特征,如:度分布、平均路径长度、集聚系数等,就可以判断该网络的主要特性。

表1 常用的网络拓扑性质及其主要特征条件

Table 1 The common topological properties of network and its main characteristic condition

序号	复杂网络拓扑性质	特征条件
1	小世界特性	平均路径长度小、集聚系数高
2	无标度特性	度分布符合幂律分布
3	随机特性	度分布为泊松分布

在上述复杂网络各个拓扑性质对应的特征条件中,节点的度值是与该节点直接相连的边数,反映节点在复杂网络中的联通程度。通过计算网络中节点的度值,可以了解网络中所有节点的度值分布情况。平均路径长度是网络中任意两个节点之间相互连接需要经过的最少边数的平均值。对所有节点对之间的距离求平均值,即可得到网络的平均路径长度。一般情况下,复杂网络中的平均路径长度比较小。集聚系数表示与复杂网络中某个节点相连的各个节点之间相互连通的概率。某个节点的集聚系数等于与该节点相连的所有节点之间(除去该节点)相连的边的数目占可能存在的最大边数的比例。

### 1.2 航运网络构成要素的分析与采集

航运网络主要由港口和航线构成,港口为网络中的节点,连接港口与港口的航线为边。航运网络可以定义为 $G(V, E)$ ,其中, $V = \{v_i | i = 1, 2, \dots, n\}$ 为节点集合; $n$ 为港口的数量; $E = \{e_{ij} | i, j = 1, 2, \dots, m, i \neq j\}$ 为航运网络的边; $m$ 为航线的数量。该网络不考虑边的方向和权,为无向无权网络。

文中数据来源于中国国际海运网([www.shippingchina.com](http://www.shippingchina.com))的集装箱班轮船期数据。选取研究范围内的港口,边的范围为研究范围内所有的班轮航线。采用基于JAVA的爬虫软件“八爪鱼”数据采集器,应用其内部自定义采集网站页面信息的功能,采集中国国际海运网“货代船期”一栏中装运港和对应的卸货港之间的信息,得到研究范围内港口和航线的原始数据。在对原始数据进行去重、修正的基础上,将其转化为网络的度矩阵。部分矩阵结构示意图如表2所示。

### 1.3 航运网络的社团结构

社团是具有某些共同特征的单元相聚而成的集合。复杂网络中的结点集合可以根据其特征属

性划分为不同的子集。社团发现的主要目标就是依靠网络拓扑性质发掘出具有相似性质的节点组。分析复杂网络中的社团结构有助于深入理解其内部的组织结构和每个社团对整个网络的功能和作用。航运网络作为复杂网络中的一种,也具有社团结构。航运网络除了具有常见复杂网络的拓扑性质之外,还具有航运网络特有的性质与特点,这些性质与特点受空间距离、政治、文化、贸易往来等因素影响。对航运网络进行社团结构的划分,分析航运网络内部的结构特点,可以为港口之间的贸易分析提供新的视角,方便后续的政策制定、贸易策略的调整以及资源的合理配置。

表 2 船期数据结构

Table 2 The data structure of sailing date

	新加坡	香港	深圳	宁波	上海	巴生	...
新加坡	0	1	1	1	1	1	...
香港	1	0	1	1	1	1	...
深圳	1	1	0	1	1	1	...
宁波	1	1	1	0	1	1	...
上海	1	1	1	1	0	1	...
巴生	1	1	1	1	1	0	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

注:“0”表示两港之间无航线;“1”表示两港之间存在航线。

## 2 基于模块度的航运网络社团划分方法

### 2.1 社团结构的衡量标准——模块度

本世纪初,Girvan 和 Newman 发现并研究了复杂网络的社团结构特性,使得社团结构成为复杂网络的重要特征之一。其作为表征复杂网络特性的重要性日渐受到重视。目前,普遍接受关于社团结构的衡量标准是由 Newman 和 Girvan 定义的模块度函数<sup>[8]</sup>。模块度是指网络中连接社团结构内部节点的边所占的比例与任意连接该网络中节点的边的比例的差的期望。可用  $Q$  函数定量描述社团划分的模块化水平。

假设网络中已经划分出社团结果,该划分结果包含  $k$  个社团。定义  $k \times k$  维的对称矩阵  $e = (e_{ij})$ ,其中,矩阵元素  $e_{ij}$  表示第  $i$  个社团和第  $j$  个社团之间连接的边数与网络总边数的比例。这里的网络总边数指网络中包含的所有边的总数。设

$\text{Tre} = \sum_i e_{ii}$  为矩阵中对角线上各元素之和,其表示网络中社团内部节点之间相连的边数在网络总的边数中所占的比例。设  $a_i$  为每行或者每列中各元素之和,其表示与第  $i$  个社团中的节点相连的边在所有边中所占的比例,则  $Q$  函数可以表达为:

$$Q = \sum_i (e_{ii} - a_i^2) = \text{Tre} - \|e^2\|。 \quad (1)$$

式中:  $\|e^2\|$  表示矩阵  $e^2$  中所有元素之和。

若社团内部边的比例不大于任意连接节点的边的比例的期望值,则  $Q \leq 0$ 。 $Q$  的上限为  $Q = 1$ 。 $Q$  越接近 1,说明社团结构越明显,相应的社团结构划分良好。而在实际的网络中, $Q$  值通常在 0.3~0.7 之间。在划分社团的过程中,计算每一种划分对应的模块度值( $Q$  值),并找出最大  $Q$  值所对应的一种或者几种划分,即为最优的社团结构划分方式。

### 2.2 基于模块度的 Newman 快速算法

经典的 GN 算法由 Girvan 和 Newman 两位学者提出。GN 算法属于一种分裂社团的算法,该算法是基于模块度优化的社团发现算法,根据边介数从大到小的顺序,将边从网络中移除,直至整个网络被划分为各个社团。但由于计算边介数的时间复杂度非常高,每一次的网络划分都需要计算边介数,大大增加了算法的复杂度,因此仅仅局限于研究中等规模的复杂网络。为了改进 GN 算法时间复杂度高的问题,Newman 在 GN 算法的基础上提出 Newman 快速算法<sup>[9]</sup>。Newman 快速算法是一种寻找模块度值最大的贪婪算法,同时也属于往网络中添加边的凝聚算法。该算法的核心思想是一种自底向上的层次聚类过程,该方法通过最大化模块化率来检测社团结构,大大降低了时间复杂度,可以用于分析节点数达几百万的复杂网络。具体算法过程如下。

1) 首先把含有  $n$  个的节点复杂网络看做  $n$  个社团,即初始的情况为:

$$e_{ii} = \begin{cases} 1/2m; & \text{节点 } i \text{ 和 } j \text{ 之间有连接} \\ 0; & \text{其他} \end{cases}。 \quad (2)$$

$$a_i = k_i/2m。 \quad (3)$$

式中: $k_i$  为节点  $i$  的度; $m$  为该复杂网络包含的总边数。

2) 根据贪婪算法的原理,沿着使  $Q$  增加或者减小速度最快的方向,依次合并两个有边连接的

社团,并计算合并后的  $Q$  值增量:

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j). \quad (4)$$

合并以后,对相应的元素更新,并将与  $i$  和  $j$  社团的行和列相加。

3) 重复上一步,直到每个节点都合并成为一个社团为止。

### 3 案例分析

选取“二十一世纪海上丝绸之路”(简称“海丝之路”)航运网络为研究对象,时间范围为 2017 年 5 月至 2017 年 9 月,共有 453 个港口,3 444 条边。港口的地理空间分布如图 1 所示。

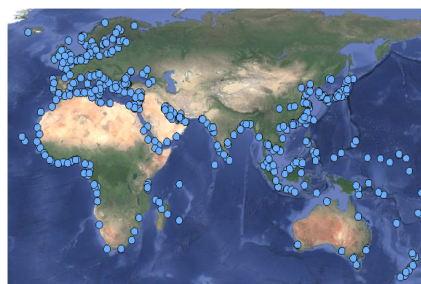


图 1 港口地理分布图

Fig. 1 Geographical distribution of ports

采用 2.2 节中的社团结构划分方法,将由 453 个节点和相互连接的边构成的复杂网络划分成 225 个社团。将这些团根据其社团内部的节点数量及其拓扑性质,分为三大类,社团划分结果见表 3。

表 3 社团划分结果

Table 3 The results for community division

社团类别序号	该类所含社团数量	该大类所含港口数量	每个社团内部所含节点数量/数量范围	节点度值范围
第一类	1	173	173	8~207
第二类	25	81	2~7	1~17
第三类	199	199	1	1~27

通过表 3 可知,“海丝之路”航运网络为巨型社团结构,有且只有一个巨型社团,即第一类社团,第一类社团由 173 个港口节点组成。“海丝之路”航运网络中度值排名前 20% 的港口均在该社团,如:新加坡港(度值 207)、香港港(度值 193)、

深圳港(度值 187)、宁波港(度值 185)、上海港(度值 185)等。第一类社团内部港口地理空间分布如图 2 所示。节点的度值频数分布如图 3 所示。节点累计概率占总体的比例图如图 4 所示。

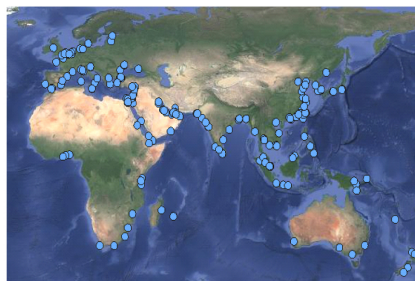


图 2 第一类社团地理空间分布图

Fig. 2 Geographical distribution of community 1

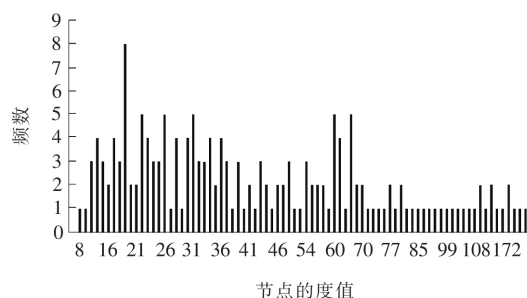


图 3 第一类社团的节点度分布

Fig. 3 The distribution of the degree for the first community

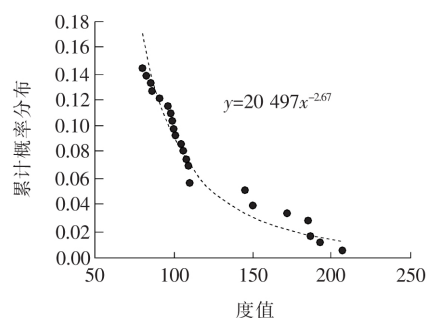


图 4 第一类社团度值排名前 22 的节点累计概率占总体的比例图

Fig. 4 The proportion diagram of cumulative probability of the nodes in the top 22 for the first community

从图 4 可知,第一类社团中的度值接近于幂律分布,且由于领导者社团内部往往存在几个度值较大的节点,可判定在该社团为领导者社团,即第一类社团中具有核心港口。根据帕累托法则,将该社团内部的核心港口初步判定为度值排名前

34 位的港口。核心港口的相关信息如表 4 所示。网络中第二类社团普遍为小社团,度值不高,而且有明显的地理分布。社团内部部分港口地理空间分布如图 5 所示。从图 5 可以看出,该类社团中

的 81 个港口中有 29 个港口分布于欧洲大陆,可见欧洲国家虽然在欧盟大背景的影响下各个成员国之间贸易往来比较密切,但仍存在独立于核心港口社团之外的小社团。

表 4 核心港口相关信息  
Table 4 The corresponding information of core port

英文名称	中文名称	所在大洲	所在国家	度值	英文名称	中文名称	所在大洲	所在国家	度值
SINGAPORE	新加坡	亚	新加坡	207	GUANGZHOU	广州	亚	中国	99
HONG KONG	香港	亚	中国	193	ROTTERDAM	鹿特丹	欧	荷兰	98
SHENZHEN	深圳	亚	中国	187	VALLETTA	瓦莱塔	欧	马耳他	96
NINGBO	宁波	亚	中国	185	GENOVA	热那亚	欧	意大利	91
SHANGHAI	上海	亚	中国	185	PORT SAID	塞得港	非	埃及	86
KELANG	巴生	亚	马来西亚	172	NHAVA SHEVA	那瓦西瓦	亚	印度	85
XIAMEN	厦门	亚	中国	150	ISTANBUL	伊斯坦布尔	亚	土耳其	82
BUSAN	釜山	亚	韩国	145	BARCELONA	巴塞罗那	欧	西班牙	80
QINGDAO	青岛	亚	中国	145	LE HAVRE	勒阿佛尔	欧	法国	79
JEBEL ALI	杰贝阿里	亚	阿联酋	110	PIRAEUS	比雷埃夫斯	欧	希腊	79
KAOHSIUNG	高雄	亚	中国台湾	109	ANTWERP	安特卫普	欧	比利时	78
TANJUNG PELEPAS	丹戎帕拉帕斯港	亚	马来西亚	109	DAMMAM	达曼	亚	沙特阿拉伯	77
JEDDAH	吉达	亚	沙特阿拉伯	108	VALENCIA	瓦伦西亚	欧	西班牙	77
HAMBURG	汉堡	欧	德国	106	DUBAI	迪拜	亚	阿联酋	75
DALIAN	大连	亚	中国	104	MANILA	马尼拉	亚	菲律宾	74
TIANJIN	天津	亚	中国	101	FUZHOU	福州	亚	中国	73
COLOMBO	科伦坡	亚	斯里兰卡	100	AQABA	亚喀巴	亚	约旦	72

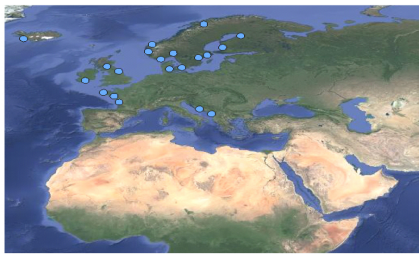


图 5 第二类社团部分港口地理空间分布图  
Fig. 5 Geographical distribution of the second community

4 结论

作者以航运网络的社团结构为对象研究,构建了基于模块度的 Newman 快速算法。选择“海丝之路”集装箱航运网络为案例,将其沿线 453 个港口与航线组成的航运网络进行社团划分,并将划分后的社团按照社团内部的拓扑性质分类研

究,得出如下结论。  
1) “海丝之路”集装箱航运网络具有一个巨型的第一类社团,而且该社团属于领导者社团,并初步判定该社团中具有 34 个核心港口。  
2) “海丝之路”集装箱航运网络具有社团结构特性,为巨型社团结构,且具有核心节点。  
3) 本研究可为我国“海丝之路”集装箱航运网络的发展提供技术支撑。  
研究核心节点与整个航运网络之间的互动机制,将是下一步的研究方向。

〔参考文献〕

[1] 陈芙英,胡志华.海上丝绸之路东南亚航运网络的复杂性分析[J].上海大学学报:自然科学版,2016,22(6):804—812.

(下转至第 93 页)

- 统谐波检测方法[J].电力系统保护与控制,2015(3):84—90.
- XIAO Zhu—li, GONG Ren—xi, CHEN Shuang, et al. Detection of harmonics in power system based on modified S—transform[J]. Power System Protection and Control, 2015(3):84—90.
- [8] 吕帅,张靠社.基于小波变换去噪预处理的 EMD 谐波检测方法[J].电网与清洁能源,2016,32(6):58—61.
- LV Shuai, ZHANG Kao—she. EMD harmonic detection method of denoising preprocessing based on wavelet transform[J]. Power System and Clean Energy, 2016, 32(6):58—61.
- [9] Smith J S. The local mean decomposition and its application to EEG perception data[J]. Journal of the Royal Society Interface, 2005, 2(5):443—454.
- [10] 焦李成,杨淑媛,刘芳,等.神经网络七十年:回顾与展望[J].计算机学报,2016,39(8):1697—1716.
- JIAO Li—cheng, YANG Shu—yuan, LIU Fang, et al. Seventy years beyond neural networks: retrospect and prospect[J]. Chinese Journal of Computers, 2016, 39(8):1697—1716.
- [11] 宋海军,黄传金,刘超宏,等.基于改进 LMD 的电能质量扰动检测新方法[J].中国电机工程学报,2016,28(8):74—78.
- SONG Hai—jun, HUANG Chuan—jin, LIU Chao—hong, et al. A new power quality disturbance detection method based on the improved LMD[J]. Proceedings of the CSEE, 2016, 28(8):74—78.
- [12] 杜培伟,黄纯,王美,等.基于改进 LMD 的谐波扰动检测新方法[J].电力系统及其自动化学报,2016,28(8):74—78.
- DU Pei—wei, HUANG Chun, WANG Mei, et al. New method of harmonic disturbance detection based on improved LMD[J]. Proceedings of the CSU—EPSA, 2016, 28(8):74—78.
- [13] 任达千,杨世锡,吴昭同,等.LMD 时频分析方法的端点效应在旋转机械故障诊断中的影响[J].中国机械工程,2012,23(8):951—956.
- REN Da—qian, YANG Shi—xi, WU Zhao—tong, et al. Research on end effect of LMD based time—frequency analysis in rotating machinery fault diagnosis[J]. China Mechanical Engineering, 2012, 23(8):951—956.
- 
- (上接第 39 页)
- CHEN Fu—ying, HU Zhi—hua. Analysis of shipping network of Southeast—Asian route along maritime silk road based on complex networks[J]. Shanghai University: Natural Science, 2016, 22(6):804—812.
- [2] 李电生,张腾飞,钟丹阳.基于港口航运网拓扑结构的港口层次划分:以中国城市港口航运网为例[J].中国海洋大学学报:社会科学版,2017(2):85—90.
- LI Dian—sheng, ZHANG Teng—fei, ZHONG Dan—yang. Port hierarchy partition based on the topological structure of port shipping network: an empirical study on ports of Chinese cities and shipping network[J]. Ocean University of China Social Science, 2017(2):85—90.
- [3] Wilmsmeier G, Monios J, Pérez—Salas G. Port system evolution: the case of Latin America and the Caribbean[J]. Journal of Transport Geography, 2014, 39(39):208—221.
- [4] 蒋柳鹏,封学军,王伟.“港口—产业—城市”复合系统协调度模型[J].水利经济,2011,29(1):11—14.
- JIANG Liu—peng, FENG Xue—jun, WANG Wei. "Port—Industry—City" composite system coordination model[J]. Economics of Water Resources, 2011, 29(1):11—14.
- [5] 蹇令香,李东兵,刘玲玲.我国集装箱港口体系演进规律研究[J].经济地理,2012,32(12):91—96.
- JIAN Ling—xiang, LI Dong—bing, LIU Ling—ling. Research on the evolution law of China container port system[J]. Economic Geography, 2012, 32(12):91—96.
- [6] Ducruet C. Multilayer dynamics of complex spatial networks: the case of global maritime flows(1977—2008)[J]. Journal of Transport Geography, 2017(60):47—58.
- [7] 迪克吕特,王列辉.中国的全球航运连通性:当代内部和外部的动态(1890—2016)[J].中国地理科学,2018(2):202—216.
- Ducruet C, Wang L. China's global shipping connectivity: internal and external dynamics in the contemporary era(1890—2016)[J]. Chinese Geographical Science, 2018(2):202—216.
- [8] Newman M E J. Detecting community structure in networks[J]. European Physical Journal B, 2004, 38(2):321—330.
- [9] Newman M E J. Fast algorithm for detecting community structure in networks[J]. Physical Review E Statistical Nonlinear and Soft Matter Physics, 2004, 69(6):066133.