


 首都经济贸易大学出版社基金资助

基于模块和鲁棒性的 复杂网络结构和 功能特性研究及 协同优化

JIYU MOKUAI HE LUBANGXING DE
FUZA WANGLUO JIEGOU HE
GONGNENG TEXING YANJIU JI XIETONG
YOUHUA

马丽丽
郑志明
张占利
姜鑫 等 © 著

 首都经济贸易大学出版基金资助

基于模块和鲁棒性的 复杂网络结构和 功能特性研究及 协同优化

JIYU MOKUAI HE LUBANGXING DE
FUZA WANGLUO JIEGOU HE
GONGNENG TEXING YANJIU JI XIETONG
YOUHUA

马丽丽
郑志明
张占利
姜鑫 等 © 著

 首都经济贸易大学出版社
Capital University of Economics and Business Press

· 北 京 ·

图书在版编目 (CIP) 数据

基于模块和鲁棒性的复杂网络结构和功能特性研究及协同优化/马丽丽等著. —北京: 首都经济贸易大学出版社, 2017. 1

ISBN 978 - 7 - 5638 - 2519 - 6

I. ①基… II. ①马… III. ①计算机网络—网络结构—研究 IV. ①TP393.02

中国版本图书馆 CIP 数据核字 (2016) 第 144178 号

基于模块和鲁棒性的复杂网络结构和功能特性研究及协同优化
马丽丽 郑志明 张占利 姜鑫 等著

责任编辑 薛晓红

封面设计  砚祥志远·激光照排
TEL: 010-65976003

出版发行 首都经济贸易大学出版社

地 址 北京市朝阳区红庙 (邮编 100026)

电 话 (010) 65976483 65065761 65071505 (传真)

网 址 <http://www.sjmcb.com>

E-mail publish@cueb.edu.cn

经 销 全国新华书店

照 排 首都经济贸易大学出版社激光照排服务部

印 刷 北京玺诚印务有限公司

开 本 710 毫米 × 1000 毫米 1/16

字 数 154 千字

印 张 8.75

版 次 2017 年 1 月第 1 版 2017 年 4 月第 2 次印刷

书 号 ISBN 978 - 7 - 5638 - 2519 - 6/TP · 42

定 价 22.00 元

图书印装若有质量问题, 本社负责调换
版权所有 侵权必究

主要作者简介

马丽丽，女，汉族，1982年11月出生于河北省秦皇岛市。

2002—2006年就读于大连理工大学数学科学学院信息与计算科学专业，并于2006年6月获得理学学士学位。2006年9月被保送到北京航空航天大学数学与系统科学学院基础数学专业直接攻读博士学位，师从郑志明教授，北京航空航天大学“数学、信息与行为”教育部重点实验室成员，空天系统大型计算与信息处理中的前沿数学问题研究教育部创新团队成员。2012年7月开始入职首都经济贸易大学统计学院。

在本科与研究生期间获得国家级和校级奖项10项。2013年参与多元统计分析学科建设申请并成功获批。2014年主持申请国家自然科学基金青年基金项目“非结构依赖性路由策略构建的潜在几何度量空间方法”（No 11401396）并成功获批。

责任编辑：薛晓红

封面设计： 砚祥志远·激光照排
TEL: 010-65976003

摘 要

近年来，复杂网络的研究由单纯的结构研究发展为动力学过程、功能特性的研究，以及三者随机、动态、非线性的演化过程中呈现出的协同作用、交互影响。随着研究的深入，人们发现现实网络的形态虽然千差万别，大多数却展现出令人惊奇的相似性，包括结构上的小世界、无标度、高聚类、模块等特性，功能方面的导航性以及对随机攻击的鲁棒性和对目的攻击的脆弱性等。复杂网络领域未来的一个研究重点是探究众多现实网络是如何如此高效地实现其功能的，而结构与功能的相互影响又是如何推动着网络的演化发展的，这是复杂网络研究方面极具挑战性的方向，也是复杂网络研究的终极目标。本书正是从这一点出发，重点针对能够体现结构和功能关联性的网络模块结构和网络鲁棒性进行研究。

首先，我们首次提出了基于节点自身相似性来探索网络模块结构形成的机制，以揭示模块结构之所以成为现实世界网络共有的结构特性的原因。该机制借助了复杂网络隐度量空间的思想，利用节点在隐度量空间中的性质构造是基于节点的潜在相似性，来探测节点在隐空间中模块结构的算法。通过将模块结构与相应可视网络中的模块结构进行对比我们发现，两种模块结构的 modularity 或 fitness 度量标准的数值是很接近的，特别是后者，两种模块结构的 fitness 值以 90% 以上的概率误差不超过 0.0002，更以 100% 的概率误差不超过 0.0005，且在 fitness 度量标准下，两种模块划分的节点匹配程度均高于 95%。我们的研究结果说明，节点潜在相似性很大程度上是现实网络模块共性形成的原因，即由于节点之间始终会存在或高或低的相似性，随着功能的实现和不断的自我演化，现实网络终将演化出具有模块结构的形态。通过分别考察隐度量空间模型中的聚类强度参数与相应可视网络模块结构的 modularity 和 fitness 度量标准之间的关系，我们发现，强聚类的网络往往倾向于具有较为明显的模块结构。另外，在网络模块结构探测算法方面，针对聚类网络，基于代数学中群在集合上作用的思想，我们提出了一种高效的模块结构划分算法——GAS 算法。它以网络最本质的代数表示—— r -轮换作为算



法的核心。在算法执行之前，我们提出网络的“模块可分度”的概念，以衡量网络模块划分的必要性，而网络的潜在模块结构会在算法执行的过程中随着置换群在集合上作用的轨道的分离而被探测出来。在经典的聚类网络以及 LFR benchmark 模型上的数值试验验证了我们所提出的 GAS 算法的高效性和准确性。进一步，在复杂网络模块结构的应用方面，我们首次基于模块结构的理念提出了反映城市交通状况（拥塞或畅通）的模型。该模型是利用城市浮动车数据系统返回的实时车辆信息建立起的以浮动车辆为节点的网络，将该模型应用于北京城市交通我们发现，在不同时间点北京城市交通网具有高斯度分布的形式。通过考察不同时刻北京城市交通网模块结构的 fitness 数值我们发现，相对于交通畅通状况，交通拥塞发生时交通网络会呈现出更为明显的模块结构。我们的研究结果说明，网络模块结构是探索交通拥塞问题的一个非常可行的方案。

其次，观察到网络用户需求在现实世界网络功能实现中发挥的关键作用，我们首次将用户需求这一关键因素引入复杂网络鲁棒性研究中，从保证网络功能实现的角度考察复杂网络的抗目的攻击性。基于目的攻击的原理是打击网络中少量重要性较高的节点，即目的攻击与网络节点重要程度或节点中心性密切相关，我们首次提出了一个融合网络结构、动力学过程以及网络用户需求三大网络基本要素的新的中心性定义方法。对该中心性的评估结果表明，其可以恰当地反映不同节点对具有不同需求的网络用户的传播过程发挥的不同作用，亦可以给出对不同用户而言节点重要程度的正确的排序。我们以数值模拟的手段考察了不同网络族中节点的该中心性的分布性质。进一步，由于对网络进行的目的攻击往往是基于节点的重要性排序，我们考察了在该中心性定义下不同类网络中节点的重要性排序，数值模拟结果显示，对不同用户而言，节点的重要性排序可能是完全不同的，则对不同网络用户而言，支持其预期传输过程的关键节点可能是完全不同的。我们同时在人工构造的网络和现实网络中验证了这一点，证明了同一个目的攻击对具有不同需求的网络用户带来的致命程度大小可能完全不同，因此，从保证网络在被攻击之后仍可以最大程度地提供给用户进行用户可接受的传输过程以实现网络功能的角度而言，我们可以通过用户在用户可接受的范围内适当调节其传输需求的方法来降低目的攻击对于用户传输造成的致命性。另外，针对与网络鲁棒性密切相关的节点中心性，我们根据网络中扩散现象（信息传输、病毒扩散等）发



生的特点，基于其动态生成函数提出了反映节点排序的动力学中心性的概念，并借助动力学熵来衡量节点的动力学中心性与网络拓扑结构之间的关系，从而从某种程度上揭示结构与功能的相互影响。

关键词：网络模块结构，隐度量空间，交通拥塞，节点中心性，网络鲁棒性

Abstract

Recently, rather than only researching the structural topology of complex networks, more and more research has been done on the dynamical processes and functions, and their interaction in the random, dynamic and nonlinear evolution of complex networks. Research results show that many networks have wonderful similar features, such as small-world features, scale-free features, high clustering and community structures, navigability of complex networks, network robustness for random attacks and fragility for targeted attacks. But why real-world networks can realize their functions with such high efficiency and how is the interaction between network structure and network function promoting the evolution of the network? It is a most important and challenging task in the future research of complex networks, and it is also the ultimate aim of the research of complex networks. Based on this task, in this book we deeply study network community structure and network robustness, which are two hot topics in the research of complex networks and quite correlated both with network structure and network function.

First, we propose a node-similarity based mechanism to study the formation of communities to explore why most real-world networks in various fields share a universal topological property as community structure. In this mechanism, we apply the concept of hidden metric spaces of complex networks. Based on the features of network nodes in the hidden metric space, we construct an algorithm to detect hidden communities of nodes in the space. By comparing this community structure with that in the corresponding observed network, we find that the values of modularity or fitness of these two kinds of community structures are very close, especially for network fitness. The difference between the values of network fitness for these two community structures are no larger than 0.0002 with the probability of more than 90%, no larger than 0.0005 with the probability of 100%, and the segments of nodes for these two kinds of community structures detected based on the



fitness parameter are matching with the probability of more than 95%. Our research results imply that the potential similarity between nodes is largely a latent cause of network communities. That is to say, because of the high or low similarity existing between network nodes, communities would finally arise in most real-world networks with the realization and evolution of network functions. By studying the relationship between the clustering strength parameter in the hidden metric space model and network modularity or network fitness, we find that strongly clustered networks generally tend to have clear community structures. In addition, a well targeted algorithm (GAS algorithm) for detecting communities in high clustered networks by presenting group action technology on community division is proposed. During the processing of this algorithm, the underlying community structure of a clustered network emerges simultaneously as the corresponding partition of orbits by the permutation groups acting on node set are achieved. As the derivation of the orbit partition, an algebraic structure r -cycle can be considered as the origin of the community. To be a priori estimation for the community structure of the algorithm, the community separability is introduced to indicate whether a network has distinct community structure. By executing the algorithm on several typical networks and the LFR benchmark, it shows that this GAS algorithm can detect communities accurately and effectively in high clustered networks. It is suggested that techniques in the viewpoint of algebra can have a fresh light on detecting communities in complex networks. Furthermore, for the first time, we propose a new network model to characterize the states of city traffics (congestion or free) based on the concept of network community structure. The model is constructed based on the real-time data from the Floating Car Data system of a city and with the floating cars as the network nodes. By applying this model in the traffic of Beijing city, we find that the traffic network of this city always has a Gaussian degree distribution at different times. After computing the fitness values of the community structures of the Beijing traffic network at different times, we also find that the network would have clearer community structures when the traffic is congested than when the traffic is free. All these results suggest that it is a feasible mechanism to explore the congestion problems of cities by applying network community structures.



Second, observing the critical roles of client demands playing in realizing the functions of the real-world networks, we introduce the client demand into the research of network robustness to study the realization of network functions after targeted attacks on the networks. To reveal the correlation between the demands of network clients and the safety and robustness of real-world networks, we propose a new definition of node centrality by incorporating not only the topology and dynamics of the network but also the demands of network clients. We evaluate this centrality as a proper index to quantify the contributions of different nodes to ensuring acceptable and successful transportation for different network clients, and to give correct orders to nodes by distinguishing node importance to different network clients. We study the distribution of this centrality for kinds of network ensembles by simulations. Furthermore, since targeted attacks on networks are generally based upon the importance order of network nodes, we research node importance order in different networks according to this new definition of node centrality. Simulation results show that node importance order in a network may quite different for different network clients. That is to say, the vital nodes in the network for the acceptable and successful transportation of different clients may completely different. We clarify this both in artificial networks and in real-world networks. Then the extent of the damage of a targeted attack on a network may quite different for different clients. To realize the transmission functions of a network to the full for network clients after the network is attacked, we could reduce the damage for the clients by adjusting client demands within their acceptable range. In addition, according to the features of the diffusion processes on complex networks, the dynamical generating function of the process is proposed, which deeply reflects the basic characteristic of the process and is mutually decided with the dynamical process. Based on the analysis of the dynamical generating function, dynamical centrality is introduced for each node, which determines the relative importance of nodes and exhibits the capability that a node collects and communicates information with its neighbor environment over the network in the diffusion process.

Keywords: Network community structure, Hidden metric space, Traffic congestion, Node centrality, Network robustness

目 录

第一章 绪论	1
1.1 研究背景及意义	3
1.2 国内外研究进展及趋势	5
1.3 本书的主要思想与工作	8
1.4 本书的组织结构	12
第二章 网络隐度量空间中节点模块结构的探索	15
2.1 引言	17
2.2 网络模块结构概述	20
2.3 复杂网络的隐度量空间概述	25
2.4 一个基于网络隐度量空间来探索网络模块结构的机制	29
2.5 总结和讨论	34
第三章 基于代数学的复杂网络模块划分算法	35
3.1 引言	37
3.2 预备知识	38
3.3 网络模块划分的 GAS (Group Action on Set) 算法	43
3.4 GAS 算法实现及其在网络上的数值试验	46
第四章 基于新建交通网络模型模块结构的交通拥塞分析	55
4.1 引言	57
4.2 交通网络模型	60
4.3 利用网络模块理论探索城市交通拥塞问题	63
4.4 总结与讨论	65



第五章	有向加权网络中基于网络用户需求的节点中心性	67
5.1	引言	69
5.2	有向加权网络中的 R - 用户子图	72
5.3	用以度量节点对用户贡献大小的用户需求中心性	74
5.4	有向加权网络中用户需求中心性的数值模拟结果	76
5.5	总结和讨论	83
第六章	用户需求对有向加权无标度网络中目的攻击破坏性的影响	85
6.1	引言	87
6.2	人工网络中用户需求对目的攻击破坏性影响的数值模拟结果	88
6.3	现实网络实例中用户需求对目的攻击破坏性的影响的数值模拟结果	90
6.4	结论	92
第七章	复杂网络上基于扩散过程的节点动力学中心性研究	93
7.1	复杂网络上的扩散过程	95
7.2	扩散过程的动态生成函数	97
7.3	动力学中心性与动力学熵	100
第八章	结论	105
8.1	结论	107
8.2	展望	109
参考文献		111
致谢		126



第一章

绪 论



1.1 研究背景及意义

近年来,以互联网为代表的计算机和信息工程技术的迅猛发展使人类社会大步迈入了网络时代,各类复杂信息和通信网络构成了现代社会的重要信息基础设施,所处理的数据量已达到万亿至兆亿量级,并且仍在不断增长,每时每刻都有千万至数亿的进程同时运行,对海量信息进行处理和传输,为人类生产和生活提供着各方面的服务。从因特网到万维网,从电力网到交通网,从生物体内的神经网络到新陈代谢网,从科研合作网到各种政治、经济、社会关系网,超大规模复杂信息系统的涌现及应用在极大地推动人类社会信息化进程的同时,也给人类社会的发展带来了新的挑战:传染病和计算机病毒的快速传播以及大面积停电事故时有发生,城市交通严重堵塞在大中型城市更是每天上演。因此,人类社会的日益网络化要求人类对各种人工及自然复杂网络的结构、性质和行为有更清楚的认识并能驾驭,更充分地发挥大型复杂信息系统在现代社会发展中的功能,这是我国建设和谐信息化社会的重大战略需求,对增强我国信息产业的核心竞争力具有基础性、前瞻性和全局性的深远战略意义。

在复杂网络领域的研究中,受到因特网、社会网络、生物网络等网络形式系统经验研究的启发^[1-4],研究者们提出了许多技术和模型,可以帮助我们理解和预测这些复杂系统的行为,包括小世界效应、度分布、聚类系数以及网络相关等概念,随机图模型、网络生长和偏好连接的模型,以及发生在网络上的动态过程^[5-9]。大量科学研究表明,现实世界中许许多多的网络都是具有小世界 (small-world)^[10]或无标度 (scale-free)^[11]等共同属性特征的复杂网络。复杂网络是大规模信息网络系统物理层面的表征形式,关于复杂网络的研究是复杂性理论研究的一部分,它是在传统的图论和组合学上的新发展,已经涉及物理学、数学、信息科学、计算科学、管理科学、系统科学、社会科学、金融经济科学等诸多科学领域,在交通、能源、通信、电子甚至医学等行业都发挥出愈来愈重要的作用。

复杂网络简而言之即为呈现高度复杂性的网络,其复杂性主要体现在结构复杂性、节点多样性、连接多样性、网络演化、动力学行为复杂性以



及多重复杂性因素相互融合^[12-13]。现实世界复杂网络节点数目巨大，网络结构呈现出多种不同特征；网络中的节点可以代表任何事物，如人际关系网中节点代表单独个体，万维网中节点表示不同网页等；现实世界网络节点之间的连接可能存在方向性且权重往往存在差异性；现实世界复杂网络中节点或连接随时产生与消失，导致网络结构在网络演化过程中不断发生着变化；网络功能的实现离不开网络平台上各种复杂的动力学过程，如能量的传输、信息包的传递以及病毒的传播等。上述多重复杂性分别存在的同时又相互影响，导致更为难以预料的结果，例如，设计一个电力供应网络需要考虑此网络的进化过程，其进化过程决定网络的拓扑结构，当两个节点之间频繁进行能量传输时，他们之间的连接权重会随之增加，通过不断的学习与记忆逐步改善供应网络的性能。目前，关于复杂网络及其功能特性的研究涉及计算机网络与信息科学、非平衡统计物理学、耗散结构与协同学、临界现象与自组织临界性、相变理论、量子统计与场论、数理统计及图论等诸多学科，主要问题集中在复杂网络的统计拓扑特征、网络演化及其建模、复杂网络的动力学行为与功能机制等方面。网络拓扑结构与网络动力学行为及功能实现相辅相成、密不可分。拓扑结构是动力学行为及网络功能实现的基础物理平台，对复杂网络结构特征尺度的分析，是建立高效、和谐的通信硬件环境、优化网络功能的一个核心问题，而功能的需求又影响结构的变化，这两个方面共同作用于网络上的物理传播过程，进而构成整个网络科学研究的系统框架。

随着社会经济与技术的快速发展，人类社会已经全面迈入网络时代，自然界和人类社会的诸多领域均可以用复杂网络的形式来描述，例如全球因特网、描述社会个体之间关系的社会网络以及交通网络等（如图 1.1 所示），复杂网络理论在这些领域的广泛应用可以有效地解决如因特网中数据包的传递，社会网中信息的交流、流言的传播、舆论的扩散以及城市交通拥塞等众多现实问题。正因如此，复杂网络的研究受到包括数学、物理学、计算机科学、生物医学以及经济学、管理学和社会学等不同领域的广泛关注。对综合了随机、动态、非线性三大特征且日趋复杂的网络结构和动力学行为的深入分析研究，以及针对社会需求而对网络功能的高度优化，已经成为复杂网络研究的基本核心问题。