



华南理工大学  
South China University of Technology

# 博士学位论文

费德勒向量极值分析与复杂网络加边算法研究

作者姓名：李刚

学科专业：计算机应用技术

指导教师：郝志峰 教授

所在学院：计算机科学与工程学院

论文提交日期：2019年6月20日

# **Fiedler Vector Extremum Analysis and Research on Link Addition Algorithms in Complex Networks**

A Dissertation Submitted for the Degree of Doctor of Engineering

**Candidate: Li Gang**

**Supervisor: Professor Hao Zhifeng**

South China University of Technology  
Guangzhou, China

分类号: TP39

学校代号: 10561

学 号: 201210102137

## 华南理工大学博士学位论文

# 费德勒向量极值分析与复杂网络加边算法 研究

作者姓名: 李 刚

指导教师姓名、职称: 郝志峰 教授

申请学位级别: 博士

学科专业名称: 计算机应用技术

研究方向: 计算智能

论文提交日期: 2019年6月20日

论文答辩日期: 2019年5月29日

学位授予单位: 华南理工大学

学位授予日期: 年 月 日

答辩委员会成员:

主席: 吴定

委员: 杨晓伟 王甲海 方 方 徐雪如



# 摘要

随着计算机和互联网的快速发展，社交网络和计算机网络等实际网络的规模正在急剧增加，网络中不可避免的会有新的边（链接）增加，因此，加边扩容问题成为网络动态演化过程中结构变化的一个重要研究课题。加边扩容算法主要研究根据不同的优化目标设计最优的加边策略，大数据时代，随着网络规模增大，结构复杂性增加，完全基于网络结构的研究方法和结论显然不能满足现实应用中的需求，因此需结合网络的多种相关理论设计加边算法。本文分析了费德勒向量与图结构的关系，从代数的角度研究了网络的结构性质，并提出了快速有效的最大化代数连通度和传输容量的加边算法，另外，还结合社会科学中的结构洞理论提出了最小化网络平均最短路径的加边算法，本文旨在为大规模网络扩容提供高效可行的加边算法，同时为多角度研究网络结构提供思路。具体工作包括以下四个方面：

1、本文进一步研究了费德勒向量与图结构的关系。费德勒向量（Fiedler Vector）是图的拉普拉斯矩阵的第二小特征值对应的向量，其在图的分割，数据聚类，社交网络等领域有着广泛的应用，费德勒向量的重要意义在于其把图的代数特征（如拉普拉斯矩阵的特征值和特征向量）与图的拓扑结构特征（如图的连通性等）联系起来，使得人们可以通过特征向量分量的大小，正负等来研究图的结构，比如，费德勒向量分量的非负的值对应图顶点的导出子图是连通的。因此，费德勒向量常常被用于网络结构研究。本文的具体研究成果如下：

首先，本文定义了 $\alpha$ -匹配图，这种图的最小度点恰恰对应了费德勒向量分量的最大值。并且我们证明了树图是 $\alpha$ -匹配图。进一步的抽样统计实验表明大部分图都是 $\alpha$ -匹配图。

然后，本文给出了Fiedler-Breadth-First图的定义，这类图的特点是：如果以费德勒向量分量的最大值对应的顶点为根生成广度优先图，则费德勒向量分量的最小值一定在该图的最高层。为了验证该图类的存在性，本文找到并证明了一类这样的图。同样抽样统计实验表明大部分图都是Fiedler-Breadth-First图。

2、根据费德勒向量与图结构的关系设计了最大化代数连通度的算法。代数连通度是图拉普拉斯矩阵的第二小特征值，是多智能体等网络系统的基本性能指标。本文讨论了如何通过最大化代数连通度来增加网络的连通性和鲁棒性。目前已有的代数连通度最大化的有效算法需要直接计算它，这导致算法复杂度较高，难于应用在大规模的现实网络。本文在分析费德勒向量的基础上，提出了一种不需要计算代数连通度的启发式算法——最小度和最大距离（MDMD）算法。该算法在大型随机网络和现实自主

系统（AS）对等信息网络中进行了测试。仿真结果表明，该算法是有效的，与其他算法相比，有更短的运行时间。因此，它可以应用于非常大的网络，特别是大型稀疏网络。

3、研究了代数连通度与传输容量的关系，并提出了最大化网络传输容量加边算法。本文研究了如何通过加边来提高无标度网络的传输效率。在分析了代数连通度与传输容量之间相关性的基础上，提出了一种有效的加边策略：最大化代数连通度增量边(MACIE)算法。现有的方法大都基于拓扑结构指标，如网络的路径和度，而MACIE算法从网络的代数属性与结构指标关系角度研究和设计提高传输效率的算法，因此具有更短的运行时间。仿真结果表明，MACIE算法是有效的，其性能优于以往的减少结构孔（RSH）的策略。

4、应用社会学中结构洞理论设计了最小化平均最短路径长度算法。平均最短路径长度是网络的小世界特性的一个衡量指标，网络具有小世界特性，意味着平均最短路径较小。网络的动态演化过程，是一个小世界化的过程，因此如何加边，使得网络平均最短路径最小化有重要应用意义。本论文基于社会学中的结构洞理论设计了加边算法，该算法的核心思路是连接网络中结构洞数目最多的顶点。结构洞数目较多的顶点意味着有较多的路径通过，连接这样的点可以影响尽可能多的最短路径，因此可以有效减小平均最短路径长度。对比试验验证了我们算法的有效性。

**关键词：** 费德勒向量；代数连通度；网络优化；复杂网络；传输容量；平均路径长度

# Abstract

With the rapid development of computer and Internet, the scale of social networks and computer networks is increasing dramatically, and adding edges (links) to existing networks are somewhat inevitable. Therefore, link addition strategy has become an important research topic in the process of dynamic evolution of networks. Link-addition strategy mainly studies the optimal design of network according to different optimization objectives. In the era of big data, with the increase of network size and structure complexity, the research methods and conclusions based on network structure can not meet the needs of practical applications. Therefore, it is necessary to design link-addition algorithm with a variety of related theories of network. This thesis analyses the relationship between Fiedler vector and graph structure, studies the structural properties of networks from the algebraic point of view, and proposes a fast and effective link-addition algorithm to maximize algebraic connectivity and transmission capacity. In addition, a link-addition algorithm to minimize the average shortest path of networks is proposed based on the structural hole theory in social sciences. The purpose of this thesis is to provide an efficient and feasible Edge-Adding algorithm for large-scale network expansion, and to provide ideas for studying network structure from multiple perspectives. Generally speaking, the contributions of this thesis are four-fold.

1. In this thesis, we further study the relationship between Fiedler vector and graph structure. Fiedler vector is corresponding to the second smallest eigenvalue of Laplacian matrix of a graph. It is widely used in graph segmentation, data clustering, social network and other fields. The significance of Fiedler vector lies in its connection between the algebraic features of a graph (such as eigenvalues and eigenvectors of Laplacian matrix) and the topological structure of a graph (such as the connectivity of a graph). We can study the structure of graphs by the eigenvector components, positive and negative. For example, the induced subgraphs of vertices corresponding to the non-negative values of Fiedler vector components are connected. Therefore, Fiedler vectors are often used in network structural analysis. The specific research results of this paper are as follows:

Firstly, this paper defines  $\alpha$ -matching graph, whose minimum degree vertex corresponds to the maximum value of Fiedler vector. And we prove that the tree graph is a  $\alpha$ -matching graph. Further sampling experiments show that most of the graphs are

$\alpha$ -matching.

Then, the definition of *Fiedler-Breadth-First* graph is given. The characteristic of this kind of graph is that if the roots of the vertices corresponding to the maximum value of the Fiedler vector are used to generate the breadth-first graph, the minimum value of the Fiedler vector must be at the highest level of the graph. In order to verify the existence of this class of graphs, this paper finds and proves a class of such graphs. Sampling experiments also show that most of the graphs are *Fiedler-Breadth-First* graphs.

2. Based on the relationship between Fiedler vector and graph structure, an algorithm for maximizing algebraic connectivity is designed. Algebraic connectivity is the second smallest eigenvalue of the Laplacian matrix, and it is the basic performance index of multi-agent and other network systems. This paper discusses how to increase the connectivity and robustness of networks by maximizing algebraic connectivity. At present, the existing effective algebraic connectivity maximization algorithms need to calculate it directly, which results in high time complexity and is difficult to apply in large-scale real-world networks. Based on the analysis of Fiedler vector, a heuristic algorithm, Minimum Degree and Maximum Distance (MDMD), which does not need to calculate algebraic connectivity is proposed in this paper. The algorithm is tested in large random networks and real autonomous system (AS) peer-to-peer information networks. The simulation results show that the algorithm is effective and has shorter running time than other algorithms. Therefore, it can be applied to very large networks, especially large sparse networks.

3. The relationship between algebraic connectivity and transmission capacity is studied, and a link-addition algorithm for maximizing network traffic capacity is proposed. This paper studies how to improve the transmission efficiency of scale-free networks by adding edges. Based on the analysis of the correlation between algebraic connectivity and traffic capacity, an effective link-addition strategy is proposed: maximizing the incremental edge of algebraic connectivity (MACIE). Most of the existing methods are based on topological parameters, such as the path and degree of the network. MACIE algorithm studies and designs algorithms to improve transmission efficiency from the perspective of the relationship between algebraic attributes and structural parameters of the network, so it has shorter running time. The simulation results show that the MACIE algorithm is effective and its performance is superior to the previous reducing structural hole (RSH) strategy.



4. Based on the theory of structural holes in sociology, an algorithm for minimizing the average shortest path length is designed. The average shortest path length is a measure of the small-world characteristics of the network, which means that the average shortest path is smaller. The dynamic evolution process of network is a small-world process, hence how to add edges to minimize the average shortest path of network has important application significance. In this paper, based on the theory of structural holes in sociology, an link-addition algorithm is designed. The core idea of this algorithm is to connect the vertices with the largest number of structural holes in the network. A large number of vertices of structural holes means that more paths pass through. Connecting such vertices can affect as many shortest paths as possible, thus it can effectively reduce the average shortest path length. The validity of our algorithm is verified by comparative experiments.

**Key words:** Fiedler Vector; Algebraic Connectivity; Network Optimization; Complex Network; Traffic Capacity; Average Shortest Path Length

# 目录

摘    要	I
Abstract	III
目    录	VI
第一章 绪论	1
1.1 选题背景和意义	1
1.2 费德勒向量研究概述	2
1.2.1 费德勒向量理论研究概述	2
1.2.2 费德勒向量应用举例	4
1.3 复杂网络加边算法研究概述	5
1.3.1 问题描述	5
1.3.2 最大化代数连通度的加边算法	7
1.3.3 最大化网络传输容量的加边算法	7
1.3.4 最小化平均最短路径的加边算法	8
1.3.5 基于顶点中介中心性的加边算法	8
1.3.6 其他加边算法	9
1.4 本文主要工作	10
1.5 论文组织结构	11
第二章 数学基础与基础理论	13
2.1 图论基础	13
2.2 主要模型及参数	14
2.2.1 ER随机网络模型	14
2.2.2 BA无标度网络模型	15
2.2.3 流量模型	16
2.2.4 中介中心性 (Betweenness Centrality)	16
2.2.5 结构洞理论 (Structural Holes Theory)	18
2.2.6 相关系数	19
2.2.7 平均路径长度 (Average path length)	19
2.3 本章小结	20
第三章 费德勒向量的极值研究	21
3.1 相关符号和定义	21
3.2 费德勒向量的研究现状与成果	24

3.3	费德勒向量极值的研究现状与成果 . . . . .	28
3.4	本论文关于费德勒向量极值的主要理论成果 . . . . .	30
3.4.1	$\alpha$ -匹配图 . . . . .	31
3.4.2	<i>Fiedler-Breadth-First</i> 图 . . . . .	32
3.5	讨论 . . . . .	35
3.6	本章小结 . . . . .	35
第四章	基于费德勒向量极值的最大化代数连通度的加边算法 . . . . .	37
4.1	研究背景和意义 . . . . .	37
4.2	国内外研究现状及发展动态分析 . . . . .	38
4.2.1	问题的研究现状与成果 . . . . .	38
4.2.2	最大化代数连通度问题研究存在的问题 . . . . .	39
4.3	预备知识 . . . . .	39
4.3.1	问题定义 . . . . .	39
4.3.2	与本章相关的工作 . . . . .	40
4.4	最小度-最大距离 (MDMD) 算法 . . . . .	42
4.4.1	MDMD 算法 . . . . .	42
4.4.2	MDMD算法的理论分析 . . . . .	42
4.5	数值实验 . . . . .	44
4.6	本章小结 . . . . .	47
第五章	基于代数连通度的最大化网络传输容量的加边算法 . . . . .	49
5.1	研究背景和意义 . . . . .	49
5.2	国内外研究现状及发展动态分析 . . . . .	50
5.3	预备知识 . . . . .	50
5.4	与本章相关的工作 . . . . .	52
5.4.1	软方法 . . . . .	52
5.4.2	减小结构洞策略 (RSH) . . . . .	53
5.5	MACIE算法 . . . . .	53
5.6	数值实验 . . . . .	56
5.7	本章小结 . . . . .	59
第六章	基于结构洞理论的最小化平均最短路径的加边算法 . . . . .	61
6.1	研究背景和意义 . . . . .	61
6.2	国内外研究现状及发展动态分析 . . . . .	62
6.2.1	问题的研究现状与成果 . . . . .	62

6.2.2 预备知识 . . . . .	63
6.3 相关工作 . . . . .	64
6.4 基于结构洞的平均最短路径优化算法 . . . . .	68
6.5 数值实验 . . . . .	70
6.6 本章小结 . . . . .	74
<b>第七章 总结与展望 . . . . .</b>	<b>77</b>
7.1 工作小结 . . . . .	77
7.2 未来工作展望 . . . . .	79
<b>参考文献 . . . . .</b>	<b>81</b>
<b>攻读博士学位期间取得的研究成果 . . . . .</b>	<b>93</b>
<b>致谢 . . . . .</b>	<b>95</b>

# 第一章 绪论

## 1.1 选题背景和意义

当今社会，随着硬件水平的发展和人类需求的提高，人们对网络的依赖性越来越高。从出生起，每个人都被纳入了各种网络之中。伴随着人的“社会性”首先就有了亲戚朋友的社会关系网，与其他人沟通联系就用到了通讯网络，出门坐车就需要交通网络，购物又可以用到互联网，商品到客户手中依赖于物流网络，科学数据的采集和应用离不开传感器网络，等等。因此可以说网络是无处不在的，甚至可以说人的“网络性”和“社会性”是同等重要的。所以，对各种网络性质和网络模型的研究一直备受关注。为此更是催生了一门新的学科：复杂网络<sup>[1-3]</sup>。该领域融合了图论，计算机，数学，社会科学，物理学、生物学等多个学科的理论。同时，复杂网络的应用也覆盖了人们生活的方方面面。例如，对于社会网络<sup>[4,5]</sup>（包括，朋友关系网，科学家合著网，科学引文网等）的研究，可以进行进一步的挖掘社会中人们之间的关联性，用于对信息传播的控制，对恐怖袭击的预测等。对于电力网络<sup>[6]</sup>，交通网络<sup>[7,8]</sup>的研究，可以优化他们的拓扑结构，发现潜在的缺陷。对流行病等网络<sup>[9]</sup>的研究可以有效防控病毒的传播。而学者们对复杂网络的研究也是多角度的，比如，对复杂网络模型<sup>[10,11]</sup>的研究，对复杂网络上的动力学<sup>[12]</sup>的研究，对复杂网络同步性<sup>[13]</sup>的研究。

网络的加边扩容问题是复杂网络研究的热点问题，有着重要的应用意义，现实生活中的网络系统大多是动态变化的，这就涉及到网络的加边扩容问题，如何控制增加的边才能使得新的网络性能更优，如何最大可能的预测即将增加的不同人之的关联关系，网络拓扑中的微小变化（通过添加一些边）如何减轻其社区结构，从而产生更有效的信息传播过程，如何通过添加一些边使得网络拓扑中的微小变化可能增加图的导电性，并显著提高由更快的随机游动驱动的基于图采样的模拟的性能等<sup>[14]</sup>，这些都吸引了大量学者的关注。另外，平均最短路径是刻画网络小世界属性的一个指标，研究如何向网络中添加一些边使得网络的平均最短路径长度最小的问题，可以进一步理解动态网络的小世界演化进程。向网络中添加适当的边，还可以提高网络的高同步（Synchronizability）性<sup>[13,15]</sup>，有助于节点在较短的时间内达成一致（Consensus），从而提高网络的持久寿命。同样，加边策略的选择也有助于提高网络的鲁棒性<sup>[16-18]</sup>和结构可控性（Structural controllability）<sup>[19]</sup>，社交网络中的社团发现问题也可以通过加边策略和思路来解决<sup>[20,21]</sup>。因此，针对不同的优化目标指标，设计有效的加边策略，提高网络的性能是当今十分热门的话题。

对网络加边扩容，需要研究网络的结构特征，而费德勒向量（Fiedler Vector）

<sup>[22]</sup>是从代数学角度研究复杂网络结构特征的重要工具，其定义为图的拉普拉斯矩阵（Laplacian matrix）的第二小特征值对应的特征向量。它在以图为模型的各种网络系统中有广泛的应用。例如：图的分割问题<sup>[23]</sup>，社交网络的社团分析问题<sup>[24-27]</sup>，图像分割问题<sup>[28, 29]</sup>，数据的谱聚类问题<sup>[30, 31]</sup>，多智能体网络的一致性问题<sup>[32]</sup>，网络鲁棒性优化问题<sup>[33]</sup>，等等。但是从1975年费德勒向量被提出以来，其理论的研究文献相对较少。因此，费德勒向量相关的理论研究迫在眉睫。一旦理论研究有所突破，必然会带动其在网络应用领域的再一次的蓬勃发展。

本论文讨论了费德勒向量分量与图的拓扑结构之间的对应关系，同时针对不同的优化目标指标，提出了不同的复杂网络的加边算法。本论文的研究有着重要的应用和理论意义：

1) 费德勒向量的研究成果对进一步理解网络的结构具有理论意义。图的费德勒向量和图的拓扑结构是两个复杂网络的重要研究方向，一旦两者的对应关系能够建立起来，将会促进两个方向的更深层次的融合：一方面可以通过求解网络的费德勒向量，分析其分量值，来研究网络的结构特征，如上文提到的图像分割，谱聚类，社交网络的社团发现等问题，另一方面，也可以通过图的拓扑结构特征，比如，图的顶点的度，路径长度等来分析费德勒向量的值，进而在应用的过程中，可以不用求解费德勒向量的值，减少计算消耗。

2) 对复杂网络加边问题的研究具有现实应用意义。实际生活中许多网络都会面临加边扩容的问题，因此需要人们研究如何加边才能使得网络性能更好，比如交通网络<sup>[34]</sup>，航班网络<sup>[35]</sup>，无线传感器网络<sup>[36]</sup>等网络，需要考虑添加能够最大可能增加传输容量，提高传输效率的边。对于社交网络，可以考虑向网络中添加一些边，进而优化其连通性并提高其进行社会化过程的能力<sup>[14]</sup>。还可以通过添加边来最大程度减小网络中点对点的信息传递的延迟<sup>[37, 38]</sup>。

## 1.2 费德勒向量研究概述

本节将对费德勒向量的理论发展和应用进行概述。其中费德勒向量分量值与图的拓扑结构特征的关系尤其值得关注，任何关于他们关系的理论进展都具有重大的应用价值。

### 1.2.1 费德勒向量理论研究概述

费德勒向量于1975年由M. Fiedler首次提出，其重要意义在于虽然它是图对应的拉普拉斯矩阵的特征向量，但可以用于刻画图的拓扑结构特征，因此被广泛的研究。由于其意义深远，人们用M. Fiedler的名字命名。图1-1显示的是一个费德勒向量为 $X=(0.5138, 0.4383, 0.3145, 0.1283, -0.0531, -0.0689, -0.2901, -0.3209, -0.2450, -0.4170)$

的图。M. Fiedler在文章<sup>[22]</sup>提出的一个重要结论，图中对应费德勒向量分量值大于等于（或小于等于）零的点的导出子图是相互连通的。该结论建立了图的代数性质和结构性质之间的关系，因此对于图的代数性质研究和拓扑结构性质研究具有双重意义。

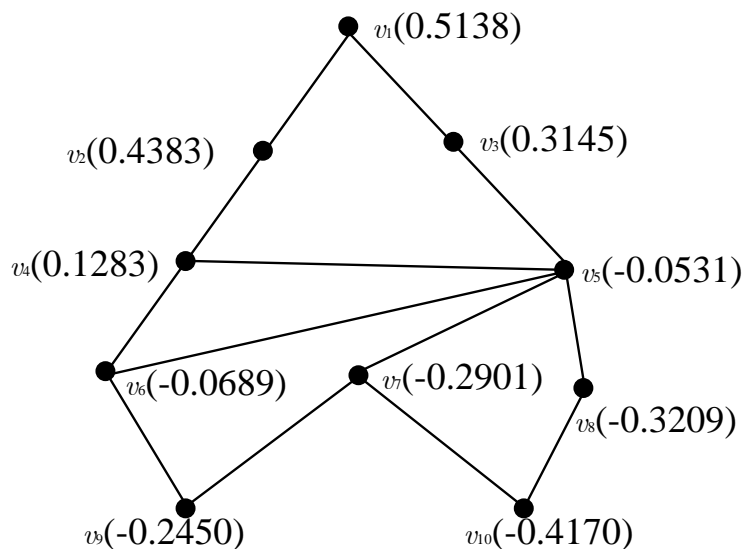


图 1-1: 费德勒向量为 $X=(0.5138, 0.4383, 0.3145, 0.1283, -0.0531, -0.0689, -0.2901, -0.3209, -0.2450, -0.4170)$ 的图。

费德勒向量提出以来，有许多关于拉普拉斯矩阵的特征值的研究成果，尤其是对于费德勒向量对应的特征值（定义为代数连通度）的研究<sup>[39-44]</sup>。然而，关于费德勒向量极值及其与底层图的拓扑性质关系的理论文章相对较少。

1974年，J.Rauch在研究热方程解的极值是提出假设：对于初始条件的某种通用选择，热方程解的极值在所考虑的区域边界处得到，详细内容参考<sup>[45]</sup>。这篇文章，引起了人们对费德勒向量极值的关注，但不幸的是这个猜想后来被证明在某些领域并不成立<sup>[46]</sup>。接下来，2011年Moo K. Chung等人对于树图，提出了相似的猜想：费德勒向量的极值可以在通过最长路径连接的两个顶点上求出，换句话说，树的费德勒向量分量的最大值与最小值对应的点是该树所有点中距离最远的两个点。但是，很快关于该猜想的反例被提出了<sup>[47,48]</sup>。Evans 等人<sup>[47]</sup>提出了一种费德勒玫瑰（Fiedler rose）图，不具有上述的特点。J.Lefèvre 等人<sup>[48]</sup>定义了Fiedler extrema diameter（FED）性质，来研究费德勒向量的极值。如果图的费德勒向量只有两个位于最长路径端点的极值，则称该图具有FED 性质。同时J.Lefèvre等人提出树图的极值点位于叶子顶点。近期，H. Gernandt等人<sup>[49]</sup>进一步研究了费德勒向量的极值问题，并研究了使得在最长路径的端点可以找到极值项的图类。

另外，2014年J. Urschel<sup>[50]</sup>等对费德勒向量分量与图的拓扑结构关系进一步研究，得出了新的结论：任意无向无权重的简单图都存在一个费德勒向量，使得费德勒向量

分量的非负值对应的点和负值对应的点分别连通。这个结论改进了M. Fiedler原有的结论。2017年, M. Begué<sup>[51]</sup>得出了另外一个结论: 具有正费德勒向量分量的点与具有负费德勒向量的点之间的距离不大于2。

从费德勒向量理论的发展过程来看, 都是以刻画图的结构特性为立足点。建立图的代数性质和结构性质之间的关系是图论领域中一种很自然的想法。有了这种关系, 人们就可以根据图的代数性质来研究结构性质, 反之亦然。本文的立足点也是如此, 我们进一步探讨了费德勒向量分量与图的结构的关系, 并且根据这些关系设计了多种算法, 使得费德勒向量得到更多的应用。

### 1.2.2 费德勒向量应用举例

下面我们简介一下费德勒向量的相关应用。正如前文提到的, M. Fiedler证明

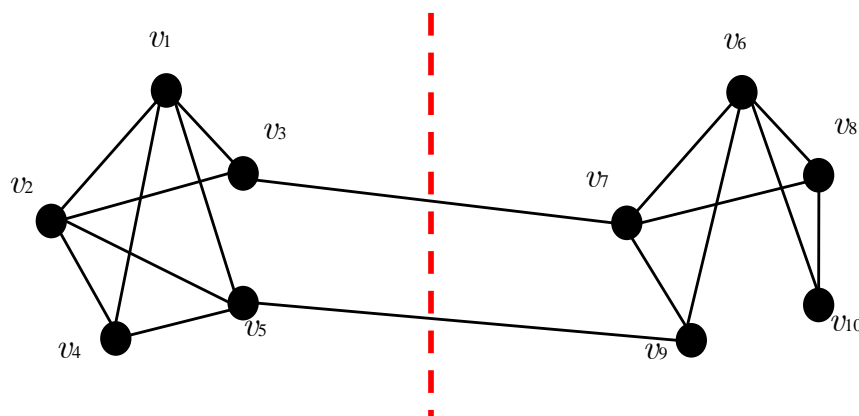


图 1-2: 顶点为 $V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}\}$ , 费德勒向量为 $X = (-0.3520, -0.3341, -0.1890, -0.3750, -0.2704, 0.3235, 0.2041, 0.3862, 0.1008, 0.4576)$ 的图。

了费德勒向量的正数分量(或负数分量)对应的点是连通的。基于此很容易将图分割成两部分。图1-2是一个具有10个顶点, 且费德勒向量为 $X = (-0.3520, -0.3341, -0.1890, -0.3750, -0.2704, 0.3235, 0.2041, 0.3862, 0.1008, 0.4576)$ 的图, 如图所示, 费德勒向量的正负值各自对应的顶点, 刚好完美的组成了图的两个部分。其中顶点 $V_1, V_2, V_3, V_4, V_5$ 对应的费德勒向量分量值都是大于零, 位于图中左半部分。其他顶点对应的值小于零, 对应图中右侧部分。当然这是一个简单的实例, 当图的结构比较复杂的时候, 分割就不会如此完美。但人们在此基础上进行的相关的推广研究, 并进行了广泛的应用。



F. Perazzi等人<sup>[52]</sup>提出了一种有效的检测潜在突出前景区域的方法。该方法假设在图像边界上的大部分像素都显示出非突出背景的性质，通过对根据图像超像素颜色相似性定义的拉普拉斯图的特征值分析来识别突出区域。此外，该方法利用费德勒向量的识别特性，设计了一个基于支持向量机的分类器，能够确定一个图像是否包含任何显著的对象，这在以前的研究中是被忽视的。P. Y. Chen等人<sup>[53]</sup>针对社交网络中的深层社区发现（Deep Community Detection）问题，提出了将深层社区作为多阶段节点删除来检测的问题，即在每个阶段最大化一种新的中心性度量，即局部费德勒向量中心性（LFVC）。顶点或边的LFVC由费德勒向量分量定义，当从图中删除顶点或边的子集时，LFVC与代数连通度的敏感性相关。此外，经典的基于费德勒向量的谱聚类也被广泛的应用，例如。A. Ozdemir等人<sup>[54]</sup>提出了一种层次共识谱聚类方法来解决大脑功能网络的研究过程中所有受试者所代表的群体的问题。另外，随着互联网的迅速发展，信息可以通过社会网络广泛传播。在当今日益增长的网络环境下，对信息源的定位至关重要。可以应用费德勒向量将网络划分为几个节点集群，在其中选择观察者。然后利用最大似然源估计量，根据观察者提供的信息，对该源进行定位<sup>[55]</sup>。

综合来看，费德勒向量的应用主要有两个特点，第一是所有的具体问题一定要可以建模成一个图的划分（或分割）问题。这一点从费德勒向量理论的提出就可以得到。上文中提到的费德勒是从代数学的角度研究图的结构，本质上来讲，就是利用费德勒向量的分量，将图划分成两个或多个自身连通的子图，如图1-2所示。第二个特点是，多数费德勒向量的应用，需先求解它。就是要求解图的拉普拉斯矩阵的特征向量。然后，根据向量分量来分析图的结构。这其实是一个复杂度较高（求矩阵特征向量）且单向（先求值，在分析结构）的应用过程。而本文的研究出发点是深入研究费德勒向量与图结构的对应关系，旨在刻画出图结构与费德勒向量之间的关系，进而在实际应用过程中，实现能够通过结构与费德勒向量的关系，在不求解费德勒向量的基础上，达到通过图结构的特点直接分析费德勒的应用的目的。

### 1.3 复杂网络加边算法研究概述

在本节，我们将对复杂网络领域中的加边算法进行简要概述。目前，已有的加边算法，多数是以优化一个或多个网络指标为目标。因此，我们将重点论述与网络指标相关的加边算法。

#### 1.3.1 问题描述

网络加边问题对应于图论学科中的图的增广问题（Graph Augmentation Problem）<sup>[56]</sup>，首先我们提炼一下问题的模型。本文重点论述简单图的加边问题，即网络的拓扑图是无向无权无重边的简单图。那么加边问题就可以总结为：

**问题1:** 令 $G = (V, E)$ 表示连通无向简单图, 其中 $E$ 表示图的边的集合,  $V$ 表示图的顶点的集合, 若 $d(i, j)$ 表示图中顶点 $i$ 与顶点 $j$ 的距离, 令 $CS = \{(i, j) | d(i, j) \geq 2\}$ 表示可以添加的边的集合, 若 $S \subseteq CS$ , 则 $G + S$ 表示图 $G$ 中添加了 $S$ 中的所有边的新图。令 $\mathcal{M}(G)$ 表示图 $G$ 的某个指标参数。则图的增广问题就是: 给定整数 $k$ , 求子集 $S \subseteq CS$ 且 $|S| = k$ , 使得 $\mathcal{M}(G + S)$ 最大 (或最小)。实际应用中对问题1的求解一般有两种方法, 一种是直接求解含有 $k$ 个元素的集合, 也就是一起找到 $k$ 条可以添加的边来达到最大化或最小化 $\mathcal{M}(G + S)$ , 但因为有的指标 $\mathcal{M}(G)$ 比较复杂, 这种方法一般比较难于实现。因为人们一般都会用一种近似的方法, 即先找打一条边 $e$ , 使得 $\mathcal{M}(G + e)$ 最大或最小, 然后重复 $k$ 次。

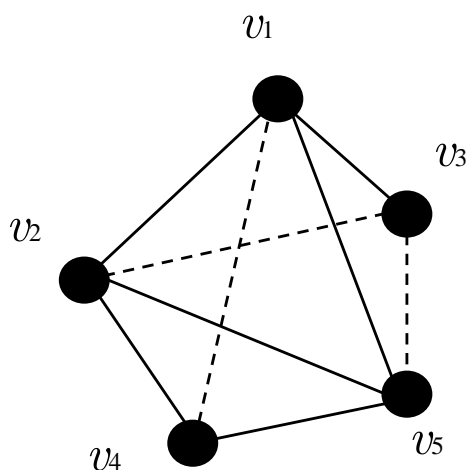


图 1-3: 5个顶点的图

如图1-3所示, 5个顶点的图, 其中可以作为候选边的有 $v_1v_4$ ,  $v_2v_3$ ,  $v_3v_5$ , 我们考虑简单情况, 当假设 $k = 1$ 时, 求添加哪条边可以使得代数连通度增加最大? 我们可以通过简单的穷举地方法计算, 添加哪一条边。令 $\alpha(G)$ 表示图 $G$ 的代数连通度,  $G + v_1v_4$ 等表示图 $G$ 加上边 $v_1v_4$ 。则通过MATLAB计算可得,  $\alpha(G + v_1v_4) = 1$ ,  $\alpha(G + v_2v_3) = 1.5858$ ,  $\alpha(G + v_3v_5) = 1.5858$ 。因此, 可以得到, 添加边 $v_2v_3$ 或 $v_3v_5$ 可以使得代数连通度增加值最大。可以看出, 当候选边比较少的时候, 可以通过穷举法, 求出所有候选边添加到网络中之后代数连通度值, 再判断选哪一条边。但是当网络中顶点数目较多, 特别是稀疏网络, 候选边的数目巨大的时候, 要反复求解矩阵的特征值, 计算代价过大, 因此穷举法是不可行的, 需要设计有效的近似算法。

以下,我们将针对不同的网络指标,结合不同的应用背景简述一下相应问题1 的算法分类。

### 1.3.2 最大化代数连通度的加边算法

代数连通度 (Algebraic Connectivity) <sup>[57]</sup> 是指图拉普拉斯矩阵的第二小特征值,它是代数图论中衡量图的连通性和稳定性的重要指标 <sup>[58]</sup>。一般认为,较大的代数连通度的网络具有较好的连通性和系统鲁棒性。而任意添加一条边到网络中,都使得网络的代数连通度增加,因此,如问题1 所述,人们考虑的是在网络加边的过程中,如何取得更好的性能,得到最大的代数连通度。针对不同的网络背景,学者们提出了多种最大化代数连通度的加边策略。Kim 研究了如何在有限条件下利用最大化代数连通度来增加多智能体系统的稳定性 <sup>[59]</sup>, Ibrahim 等基于度量代数连通度的变化设计了网络维持算法 (Network Maintenance Algorithm) 来优化无线传感器网络 <sup>[60]</sup>, Alenazi 等在主干网络中,应用代数连通度设计优化系统,提高网络连通性,增加系统抗攻击能力 <sup>[61]</sup>, Wei 等利用代数连通度优化航线设计 <sup>[35]</sup>。总体来看,目前有效的最大化代数连通度的加边算法,大部分都要计算原图的特征值和特征向量。例如, Ghosh 和 Boyd <sup>[33]</sup> 考虑了增加多个边以使代数连通度最大化的类似问题。他们提出了一种基于菲德勒向量的贪心扰动启发式算法,通过一条一条地增加边来最大化代数连通度。Kim <sup>[59]</sup> 提出了一种基于特征方程的二分算法。因此,如何降低算法的计算复杂度,以应对现在动辄几万个顶点的的大规模网络,是一个值得深入研究的问题。

### 1.3.3 最大化网络传输容量的加边算法

传输容量 (Traffic Capacity) 是由 Arenas 等人 <sup>[62]</sup> 提出的网络传输模型中,被用于刻画网络数据传输能力的一个指标。随着人们对其研究的深入,不断有新的研究成果出现 <sup>[34, 63-68]</sup>。假设网络中的每个顶点既能产生数据包,也能传输数据包,某一时刻,每个顶点都产生或传输数据包,这样,网络中某一时刻整个网络所产生的数据包数目将会有有一个边界值。这个值就定义为网络传输容量。即当网络中某一时刻产生的数据包大于网络的传输容量时,网络将出现拥堵。反之,则网络通畅。网络在动态变化中,人们希望得到更好的传输能力,即传输容量越大,网络传输数据的能力越强。但是,与人们的直观感觉相反,当某些边添加到网络中时,传输容量反而会减小。也就是说,对于交通网络而言,添加不恰当的边,反而会引起交通堵塞。因此人们研究的重点之一就是如何提高网络的传输容量,通常有两种方法,一种称为“软方法”,是以改善数据在传输过程中的路由策略为主要研究方向。另一种称为“硬方法”,是以调整网络的拓扑结构为研究方向,调整网络拓扑结构常见的方式是网络中顶点或边的增减。于是对于传输容量这个网络指标,也会有问题1这样的研究,即网络中增加哪些边可以使得

得到的新网络的传输容量最大。对此问题的研究具有非常重要的意义，人们可以根据研究结果合理的进行现有的资源再分配，从宏观上提高网络的传输能力，缓解网络拥堵，提高网络的承载能力，特别是在计算机通信、交通等领域尤为重要。

### 1.3.4 最小化平均最短路径的加边算法

小世界 (Small World) 现象是社会网络的一个理想属性，因为它保证了社会网络顶点之间的路径都比较短，从而信息能够有效地在网络上传播信息。随着小世界网络模型<sup>[10]</sup>的引入，许多研究致力于分析网络特性。如，两点之间的最短路径，它是指网络连接这两顶点之间最小长度和的边的集合。其在实际网络中有广泛的应用<sup>[69-73]</sup>。进一步而言，网络中的平均路径长度 (Average Shortest Path Length) 是测量消息传递的端到端延迟的重要指标，也是衡量网络小世界现象的一个指标。因此网络中的平均路径长度也引起了学者们的关注。特别是，通过在顶点对之间添加一条或多条附加边来减少网络的平均路径长度进而有效地提高网络的传输效率和同步能力<sup>[14, 37, 38, 74-79]</sup>。同时，学者们还考虑了在多核处理器设计中增加射频链路的问题<sup>[80-82]</sup>。显然，更小的网络平均路径长度，意味着更高的网络通讯性能，添加任意一条边到网络中，都会影响到网络中的某些最短路径，而且网络中的任意点对之间的最短路径长度不会增加。然而，要计算网络的平均最短路径长度，需要计算出网络中所有顶点对之间的最短路径长度，然后求平均值。整个网络的所有点对的最短路径的计算本身就是一个算法复杂度极高的问题，另外还需要计算添加不同边的影响，这使得最小化网络平均最短路径长度问题的计算代价非常高。因此，目前学者主要研究一些高效的近似算法。

### 1.3.5 基于顶点中介中心性的加边算法

除去以上网络指标外，还有基于网络中介中心性 (Betweenness Centrality)<sup>[83]</sup>的加边算法，网络中顶点的中介中心性度量了该顶点参与其他任意顶点对之间的最短路径程度。一个节点的中心性或在网络中的排名之间的计算可以在多项式时间内完成，但在无向无权的图上需要 $O(mn)$ 时间<sup>[84]</sup>，其中 $m$ 表示网络中边的数目， $n$ 表示网络中顶点的个数，这对于大型网络显然是不可行的。因此，当前的主要研究也是集中的近似算法的方向。从直观上讲，在假设流量遵循最短路径的前提下，顶点的中介中心性的值测量顶点对流经网络的流量的影响。而具有较大的中介中心性的顶点，在网络中往往具有更大的影响力，具有高中间性的节点有大概率接收到消息，因此称为中心节点。例如：在交通网络分析领域，中间中心性似乎与机场的效率正相关<sup>[85]</sup>。而在社交网络中，中间中心性值高的人对经纪人非常有益<sup>[86]</sup>。因此，研究添加边对网络中顶点的中介中心性影响具有应用意义。关于中介中心性的加边算法研究主要是给定网络中的一个顶点，考虑如何增加一定数目的边，使得该点的中介中心性增加最大。

如, E.Bergamini 等人<sup>[87]</sup>提出了Maximum Betweenness Improvement (MBI) 算法, G. D. Angelo 等人<sup>[88]</sup>提出的TopK算法等。事实上, 网络中顶点的中介中心性并是越大网络性能越好, 中介中心性大的顶点, 意味着有更多的最短路径通过该点, 在考虑网络传输问题的时候, 认为这样的节点往往更容易产生拥堵, 所以在研究基于网络传输容量的加边算法的时候, 常以最小化网络中顶点的最大中介中心性为目标<sup>[67, 89]</sup>。

### 1.3.6 其他加边算法

网络的同步性 (Synchronization) 有许多定义, 其中之一, 可以用网络的拉普拉斯矩阵的最大特征值与第二小特征值的比值来定义, 比值越小意味着网络的同步性越好。另外, 网络的代数连通度也可以作为同步性的一个度量。同步在生物学、化学、医学和工程学等许多领域都很有用。例如, 在传感器网络中, 高度同步的网络连接结构, 有助于顶点在较短的时间内达成一致性 (Consensus), 从而提高网络的持久寿命<sup>[13]</sup>。P. Schultz等人<sup>[90]</sup>以树型网络为出发点, 研究了添加边是如何改变网络的同步性的。结果显示, 即使增加一条边, 也可以实质性地改变系统的同步行为, 并且连接不同层次树的短环对拉普拉斯特征值的影响最大, 特别是当它们是在靠近根节点的地方创建的。L.X. Yang等人<sup>[15]</sup>则研究了候选边的添加和删除对基本振荡网络同步的影响, 数值模拟显示, 边的特性对振荡网络的同步有重要影响, 具体来讲, 添加网络社团之间的边有助于增强网络同步性, 而添加社团内部的边对网络同步性影响不大。

结构可控性 (Structural controllability) 将经典的可控性概念扩展到具有不确定边的网络情况。不严格地说, 如果网络对几乎所有边权重的实现都是可控的, 那么它在结构上是可控的。针对有向动力网络拓扑结构的优化修改问题, X. M. Chen等人<sup>[19]</sup>提出了一种结构可控性的优化方法。更准确地说, 考虑到有向动态网络的结构, 提出了一个框架, 以找出需要添加到网络拓扑中的最小有向边数, 从而实现结构可控系统。另外, 加边的思路和方法也可以用于经典的社交网络社团检测问题, S. Ghiasifard 等人<sup>[20]</sup>使用主题模型为具有共同兴趣的用户生成虚拟边, 来度量用户之间的相似性, 即为那些相似性超过某个阈值的用户添加虚拟边。此外, 为了减少用户之间无用边的影响, 通过测量用户主题的相似性来加权网络, 从而生成一致的社区, 其中只包含一个主题或一组一致的主题。同样对于加边策略对不同网络鲁棒性的影响也被学者们深入研究<sup>[16-18]</sup>。

综上所述, 网络加边算法因为其重要的应用价值被广泛的研究, 本质上来讲, 加边是网络结构的变化, 而结构的细微变化都会对各方面的网络性能产生影响。因此加边算法的选择也面临着无法统一的问题, 往往一个加边算法对某一个指标参数而言, 对网络性能的提高是有效的, 但对另外的指标参数却是无效的。同时, 加边算法的研

究也面临着网络规模过大的情况下的效率问题，因此，对于大规模（一万个顶点）或超大规模（百万个顶点以上）的网络的近似加边算法研究是目前的研究热点。另外，也需要引入更多的角度和理论研究加边算法，例如，本文引入代数角度研究网络传输容量，引入社会学的结构洞理论研究平均网络最短路径长度问题。本文首先研究的费德勒向量与网络结构的关系，深入了解了网络的结构之后，针对网络的三个常用指标指标，代数连通度，传输容量，平均最短路径长度，提出了相应的加边策略。旨在为深入理解网络结构，研究网络结构的变化所产生的影响提供思路。

## 1.4 本文主要工作

费德勒向量的重要意义在于其把图的代数特征，及拉普拉斯矩阵的特征值和特征向量，与图的拓扑结构特征联系起来，使得人们可以通过特征值的分量的大小，正负等来研究图的结构。但是在应用过程中人们也面临一些问题，比如，要应用费德勒向量必须要计算它，就是要计算矩阵的特征向量，所以当图的规模较大的时候，就要付出很大的计算代价。这一点就严重制约了费德勒向量在实际问题中的应用进程。针对这个现象，本文研究了图的结构特征如何与费德勒向量相对应并取得了一定的研究成果，这些成果是费德勒向量理论发展的有力补充，也为费德勒向量的应用提供了部分理论支撑。具体结果如下：

首先，我们定义了 $\alpha$ -匹配图（ $\alpha$ -matching）。这类图的最小度点恰恰对应了费德勒向量的分量的最大值。并且我们证明了树图是 $\alpha$ -匹配图。进一步抽样统计结果显示：大部分图都是 $\alpha$ -匹配图。

其次，我们定义了Fiedler-Breadth-First图，这类图的特点是，如果以费德勒向量分量的最大值对应的顶点为根生成广度优先图，则费德勒向量分量的最小值一定在改图的最高层。接着我们证明了一类这样的图。同样抽样统计实验表明大部分图都是Fiedler-Breadth-First图。

在深入研究费德勒向量与图结构之间的关系的基础上，我们将以上得到的结论用到网络结构变化过程中的加边算法设计中，针对几种应用问题设计了相应的算法。

1. 我们设计了最大化代数连通度的算法。该算法主要依据我们研究费德勒向量极值的结论，即最小度对应费德勒向量分量的最大值，以及以最大值为根的广度优先图中，最小值在最高层的图的结构与费德勒向量分量的对应关系。该算法将在费德勒向量分量的最大值和最小值对应点之间加边问题转化为在最小度点和离最小度点最远的点之间的加边问题。有效的解决了利用费德勒向量时候计算费德勒向量的的高计算量问题。在较大规模是随机网络上和大规模真实网络上测试，我们的算法在有效性和效率上都优于目前已有的算法。具有较高的应用价值。

2. 接着我们分析, 在网络加边的过程中, 最大化代数连通度与网络传输容量 (Traffic Capacity) 之间呈现正相关性, 因此, 我们根据两者之间的关系设计了最大化传输容量的加边算法。该算法的重要意义在于从代数的角度研究了传输容量问题, 将代数图论与网络数据传输容量联系起来, 为深入研究网络数据的传输问题提供了新的思路。

3. 网络中顶点的结构洞 (Structural Holes)<sup>[91]</sup>刻画了该顶点在网络中的重要性。我们基于顶点的结构洞的值提出了一种减小网络平均最短路径长度的算法。结构洞是社会学中的一个概念, 主要是用来研究个体在网络结构中的影响力。将这个概念以及相关的社会学中的研究结论引入一些网络问题的研究可以进一步拓宽研究视角, 取得更好的研究成果。

## 1.5 论文组织结构

本论文共有7章, 第一章为绪论, 主要介绍本项目研究的背景和意义, 以及简述费德勒向量和常见加边算法。第2章主要介绍本论文主要研究问题所涉及到的数学理论和基础模型。主要是图论相关的理论和网络相关的定义和模型。第3章主要研究费德勒向量的理论, 其中, 重点研究了费德勒向量分量的最大值和最小值与图中顶点的对应关系。第4章在上一章得到的关于费德勒向量极值的理论基础上设计了最大化代数连通度的算法。第5章研究了改善网络数据传输能力问题, 重点研究了最大化网络传输容量的加边算法, 基于代数连通度和传输容量之间的正相关性提出了新的算法。第6章研究了网络的平均最短路径问题, 基于结构洞理论, 提出了最小化平均最短路径的加边算法。第7章是本文的总结和进一步研究计划。本文的研究框架如图1-4所示。

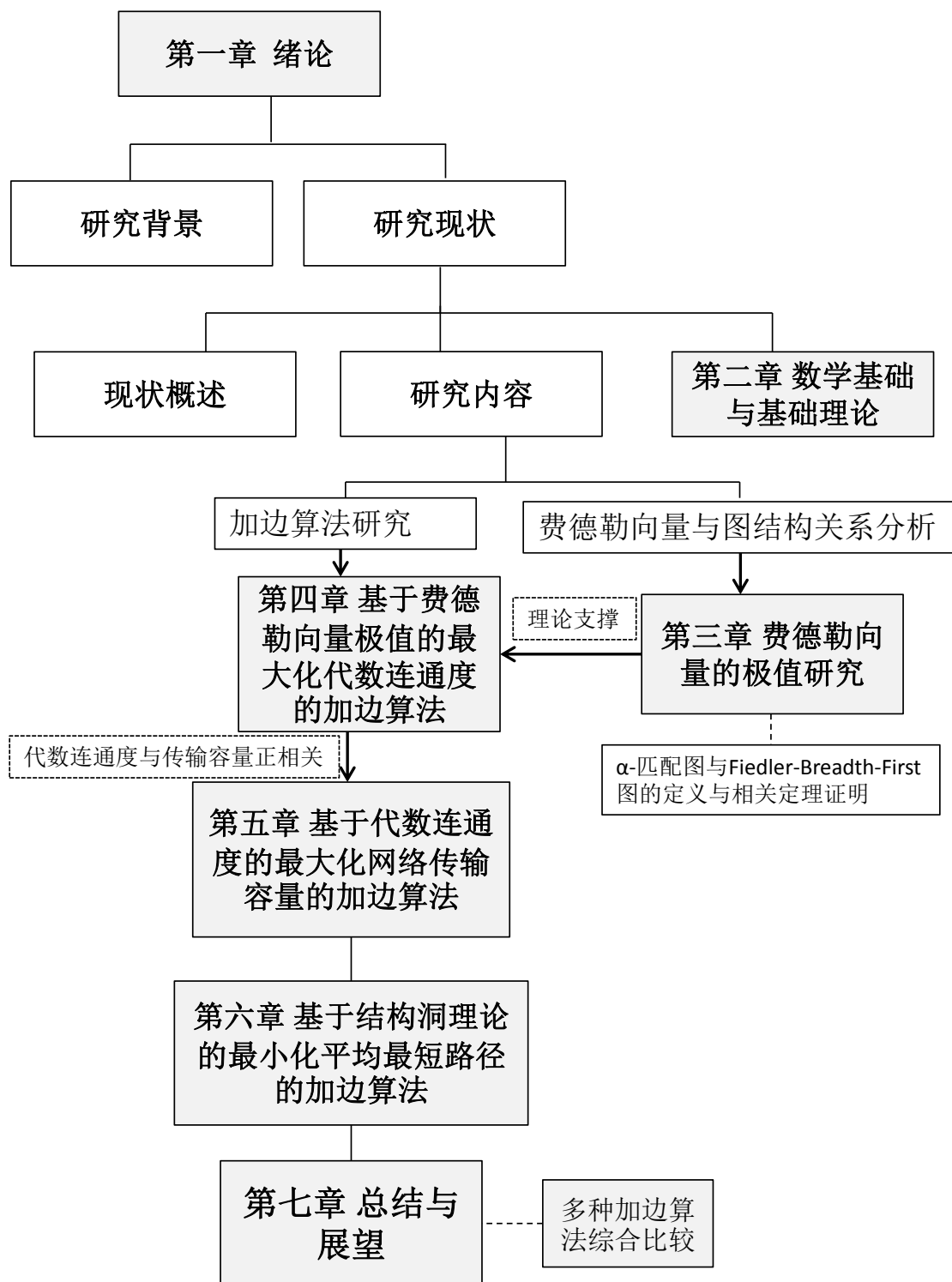


图 1-4: 论文的研究框架图。



## 第二章 数学基础与基础理论

### 2.1 图论基础

设  $G = (V, E)$  表示图，其中， $V = \{v_1, v_2, \dots, v_n\}$  是标号顶点或顶点的集合，且  $|V| = n$ 。  $E$  是边的集合。即，若顶点  $v_i$  和顶点  $v_j$  之间有一条边相连，则记为  $v_i v_j \in E$ ，也记为  $e = v_i v_j$ ，此时，我们称顶点  $v_i$  与  $v_j$  相邻(adjacent)，顶点  $v_i$  与  $v_j$  是边  $v_i v_j$  的端点。  $E^c$  表示图  $G$  的边集的补集。即  $v_i v_j \in E^c$  当且仅当  $v_i v_j \notin E$ ，这时我们称图  $G^c = (V, E^c)$  是图  $G$  的补图。若图  $G = (V, E)$  的每条边都有一个权重值(weight)，则称为赋权边，该图称赋权图(weighted graph)或赋权网络(weighted network)。用三元组  $G(V, E, W)$  表示赋权图。其中  $W$  表示权重集，它的每个元素是与边集  $E$  中的边一一对应。

如图2-1，2-2所示，5个顶点的图  $G$  和其补图  $G^c$ ，对于图  $G$  而言，顶点集合  $V = \{v_1, v_2, v_3, v_4, v_5\}$ ，边的集合  $E = \{v_1 v_2, v_1 v_3, v_1 v_5, v_2 v_4, v_1 v_5, v_4 v_5\}$ 。且图  $G$  的每条边权重均为1，也称之为无权图。若一条边的两个端点是同一个顶点，则称这条边为环。若两个顶点之间有两边（或两条以上）边相连，则称这两个顶点之间有重边。若图  $G$  无环，无重边，则称为简单图。图2-1即为简单图。本文中，在不加特殊说明的时候所提到的图均为无向，无权重的简单图。图  $G$  的邻接矩阵  $A = (a_{ij})$  是一个对称  $(0,1)$  矩阵，其中  $a_{ij} = 1$  当且仅当图中顶点  $v_i$  与  $v_j$  相邻，如矩阵 (2-1) 所示。与顶点  $v_i$  相邻的顶点的个数，称为顶点  $v_i$  的度 (degree)，记为  $d(v_i)$ <sup>1</sup>。如图2-1所示，顶点  $v_1$  的度为  $d(v_1) = 3$ 。设  $D$  是对角矩阵，且  $D(i, i) = d(v_i)$ ，其中  $i = 1 \dots n$ 。则  $D$  称为图  $G$  的度矩阵。图  $G$  中一个点边交替出现的序列  $p = v_{i_0} e_{i_1} v_{i_1} e_{i_2} \dots e_{i_k} v_{i_k}$ ，且  $v_{i_j} \in V$ ， $e_{i_j} \in E$ ， $j = 1, \dots, k$ ，

<sup>1</sup>注：本文中，按常规用法，用  $d(v_i, v_j)$  表示顶点  $v_i$  与顶点  $v_j$  之间的最短路径长度

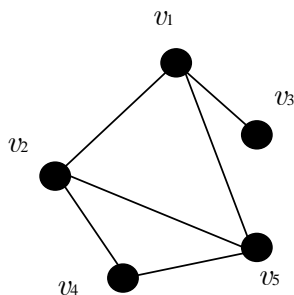


图 2-1: 5个顶点的图  $G$ 。

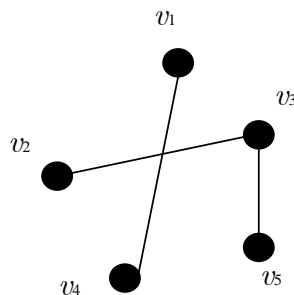


图 2-2: 5个顶点的图  $G$  的补图  $G^c$ 。

若序列 $p$ 中,没有重复的顶点,则称序列 $p$ 为连接顶点 $v_{i_0}$ 与顶点 $v_{i_k}$ 的路径(path),常常省略边,只用顶点表示,即 $p = v_{i_0}v_{i_1}\cdots v_{i_k}$ ,且 $k$ 为路径 $P$ 的长度。如图 $G$ 中,顶点 $v_3$ 到顶点 $v_4$ 的路径, $p = v_3v_1v_5v_4$ ,且路径长度为3。由 $G$ 可以看出,两个顶点之间的路径常常不止一条,顶点 $v_3$ 到顶点 $v_4$ 的路径还有一条 $p' = v_3v_1v_2v_4$ 。连通(connected):在图 $G$ 中,两个顶点间,至少存在一条路径,称这两个顶点连通的(connected);反之,称作非连通(unconnected)。显然,图(2-1)是连通图。对于连通图 $G$ ,连接两个顶点 $v_i$ 和 $v_j$ 之间的所有路径中,最短的一条,定义为顶点 $v_i$ 和 $v_j$ 之间的最短路径。其长度称为顶点 $v_i$ 与 $v_j$ 的距离。如果图 $G$ 中,任意两个顶点之间存在一条路径,则称图 $G$ 是连通图。若图 $G$ 中,任意两个顶点之间有且仅有一条路径,则 $G$ 又称为树,常用 $T$ 表示。满足 $|E| < |V|\log|V|$ 的图,称为稀疏(sparse)图;反之,称为稠密(dense)图。

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \end{pmatrix}. \quad (2-1)$$

图的拉普拉斯矩阵(Laplacian Matrix)定义为 $L = D - A$ ,其中 $D = \text{diag}(d_1, \dots, d_n)$ 是图的度矩阵,而 $d_i$ 是顶点 $v_i$ 的度。令 $\mu_1(G) \geq \mu_2(G) \geq \cdots \geq \mu_{n-1}(G) \geq \mu_n(G) = 0$ 表示拉普拉斯矩阵 $L$ 的特征值。其中,第二小的特征值 $\mu_{n-1}(G)$ 定义为代数连通度(Algebraic Connectivity)<sup>[57]</sup>,表示为 $\alpha(G)$ 。与代数连通度对应的特征向量叫做费德勒向量(Fiedler Vector),用 $X$ 表示,即 $LX = \alpha X$ 。

部分符号定义为说明,将在后文具体应用过程中给出相应的定义和说明。

## 2.2 主要模型及参数

### 2.2.1 ER随机网络模型

本文中我们将用ER随机图对实际网络进行模拟建模。用MATLAB生成ER图,验证算法有效性。1951年R. Solomonoff等人<sup>[92]</sup>和1959年Erdős-Rényi<sup>[93]</sup>各自独立提出了ER随机图模型,以下我们将简称ER图,或在不混淆的情况下,我们也简称为随机图。ER图是最古老的图模型,用 $G_n(p)$ 表示ER图。其中, $n$ 表示网络中顶点的个数, $p$ 表示任意两个顶点之间有边相连的概率。ER随机图的生成过程如下:在 $n$ 个顶点的网络中,任意两个顶点之间以概率 $p$ 连接一条边。ER随机图网络模型是经典的网络模型,在各种研究和学习的过程中被广泛的应用。图的预期边是数目是 $pn(n-1)/2$ 条,预期

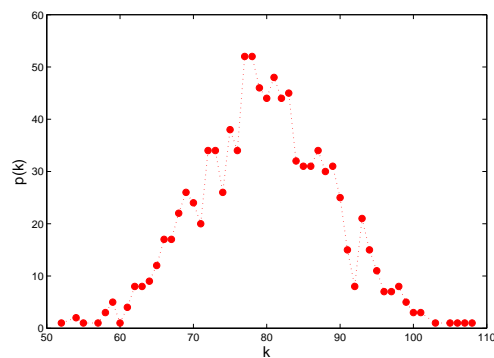
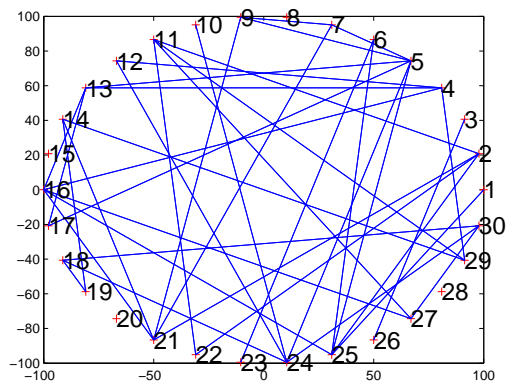


图 2-3: MATLAB生成的 $ER$ 图,  $G_{30}(0.1)$ 。图 2-4: MATLAB生成的 $ER$ 图 $G_{1000}(0.08)$ 的度分布。

的平均度为 $\langle k \rangle = p(n-1)$ 。其都分布 $P(k)$ 服从Poisson分布。因此, 顶点具有几乎相同的度数, 都接近平均度 $\langle k \rangle$ , 如图2-3和图2-4所示。

### 2.2.2 BA无标度网络模型

本文中使用的另一个网络模型是 $BA$ 网络模型。同样, 我们用MATLAB生成 $BA$ 网络模型, 验证算法有效性。Barabási 和Albert<sup>[94]</sup>于1999年首次提出无标度网络模型 (Scale-free Network)。以下我们简称为 $BA$ 无标度模型, 也称为 $BA$ 图。本论文中将用 $G(n, n_0, m_0)$ 表示 $BA$ 无标度网络, 其中 $n$ 表示最终网络的规模, 即网络形成后的顶点的个数,  $n_0$ 表示网络开始时顶点的个数,  $m_0$ 表示每次新加入一个网络顶点的时候, 与之相连的顶点的个数。 $BA$ 网络模型模拟了一个网络的动态生成过程, 它的生成主要体现了实际网络的两个主要特性, 即增长性和优先连接性:

(1) 增长性 (growth): 网络由 $n_0$ 个相互连接的顶点开始, 每次新加入一个顶点, 直到网络中总共有 $n$ 个顶点为止。新加入的顶点与原有网络中 $m_0$ 个顶点相连接。显然要满足 $m_0 \leq n_0$ 。

(2) 优先连接性 (preferential attachment): 新增加的顶点与已有的任意顶点 $i$ , 以一定的概率 $p_i$ 相连接,  $p_i$ 表示为以下关系式:

$$p_i = \frac{d_i}{\sum_j d_j}, \quad (2-2)$$

式中,  $d_i$ ,  $d_j$ 表示顶点 $i$ 和顶点 $j$ 的度, 变量 $j$ 遍历所有网络中已存在的顶点。

增长性体现了网络的形成过程, 在实际网络中有诸多体现, 很多现实中的网络都是通过不断扩大和不断增长而形成的, 例如互联网中不断生成的新网页, 人际网络中不断增加的新朋友, 新论文对已有论文的不断引用, 航空网络中新机场的建造等等。

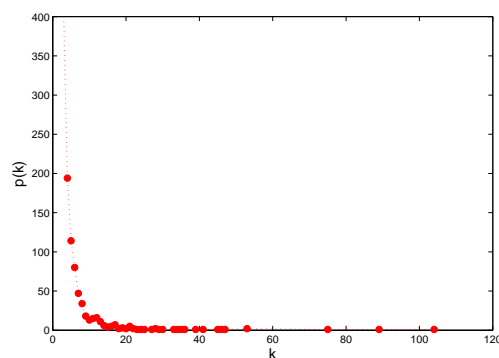
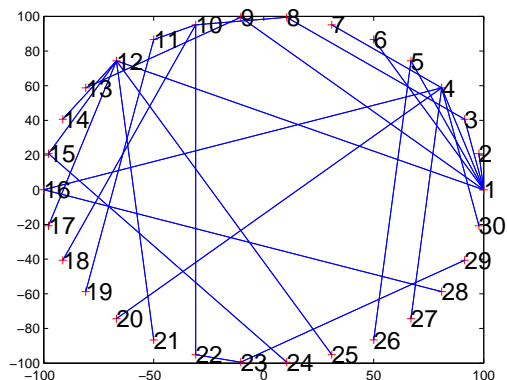


图 2-5: MATLAB生成的BA图 $G(30, 2, 1)$ 。图 2-6: MATLAB生成的BA图 $G(1000, 4, 3)$ 的度分布。

而优先连接性则体现了“富者更富”的特性，新的顶点在加入时会倾向于与有更多连接的顶点相连，例如新网页一般会链接到已有的知名的网络站点，希望引起更多的关注，新加入社群的人往往首先会想与社群中的知名人士（具有更多人际关系的人）结识，新机场或火车站的会优先考虑建立与航线更多的大机场之间的航线等等。而BA网络模型的度分布可以用幂律函数近似描述。如图2-5和图2-6所示。

### 2.2.3 流量模型

在本文中，我们采用广泛使用的流量模型来测量数据包的动态运行进度。在该模型中，网络的每个顶点都是路由器和主机，不仅可以生成数据包，而且可以转发数据包。在网络运行期间，每个时刻，网络中随机生成 $R$ 个数据包，并且从随机选择的源到目的地转发数据包。每个时刻，新接收的数据包将被推送到源顶点的信息队列的末尾，并将先进先出（FIFO）规则应用于每个队列。所有顶点都具有基于某路由策略的转发 $C$ （ $C = 1$ ）个数据包到邻居顶点的能力。一旦数据包被传递到目的地，它们就会从网络中被删除。

我们使用由Arenas 等人<sup>[62]</sup>提出的有序参数对 $(\eta, R)$ 来刻画网络中数据包的传输过程中的状态

$$\eta(R) = \lim_{t \rightarrow \infty} \frac{\langle \Delta W \rangle}{R \Delta t}, \quad (2-3)$$

其中 $\langle \Delta W \rangle = W(t + \Delta t) - W(t)$ ， $\langle \dots \rangle$ 表示 $\Delta t$ 时间内的平均数。 $W(t)$ 表示 $t$ 时刻网络中的数据包总量。为了测量传输容量，在我们的研究中采用了广泛使用的数据包生成速率关键值 $R_c$ 。当 $R < R_c$ 时，网络处于自由流阶段，否则，当 $R > R_c$ 时，网络发生拥塞，这意味着网络不能处理由网络生成的所有数据包。

### 2.2.4 中介中心性（Betweenness Centrality）

中介中心性（Betweenness Centrality）刻画了网路中顶点的中心性，被用来理论

上估计网络中顶点的相对负荷情况。某个顶点的中介中心性定义为通过该顶点的最短路径条数与网路中最短路径总条数的比值，公式如下：

$$BC(v_j) = \sum_{v_i, v_k} \frac{\delta_{(v_i, v_k)}(v_j)}{\delta_{(v_i, v_k)}}, \quad (2-4)$$

其中 $BC(v_j)$ 表示顶点 $v_j$ 的中介中心性， $v_i, v_j, v_k \in V$ ， $\delta_{(v_i, v_k)}(v_j)$ 表示顶点 $v_i$ 到顶点 $v_k$ 的最短路径中通过顶点 $v_j$ 的总条数， $\delta_{(v_i, v_k)}$ 表示顶点 $v_i$ 到顶点 $v_k$ 的最短路径总条数。本文中计算网络的中介中心性直接应用MATLAB的算法工具包，因此这里不再介绍具体方法。

从定义可以看出，具有较大中介中心性的顶点，意味着有较多的最短路径经过，因此，网络中的最大中介中心性的顶点往往是人们主要研究的对象。例如，文章<sup>[95]</sup>和<sup>[66]</sup>中，研究人员认为，网络中拥堵，常常发生在最大中介中心性顶点出，因此，其可以被用于理论估计传输容量。下面我们根据文献<sup>[66]</sup>中的内容，给出简要说明。给定了一个特定的网络结构。由于中介中心性值最大的顶点容易发生拥塞，并且拥塞会迅速扩散到整个网络，因此我们只需考虑该顶点的数据流量平衡即可。由于数据包是沿着从源到目的地的最短路径传输的，因此网络中产生的一个数据包通过最大中介中心性顶点 $i$ 的概率为 $B_i / \sum_{j=1}^n (B_j)$ （注意 $B_i$ 是顶点 $i$ 的中介中心性的值， $n$ 是网络中顶点个数）。假设，每个时刻网络中有 $R$ 个数据包生成。因此，每个时刻，最大中介中心性的顶点接收到的数据包平均值 $Q_{in}$ 如下等式：

$$Q_{in} = RD \frac{B_{L_{max}}}{\sum_{j=1}^n (B_j)}, \quad (2-5)$$

其中 $D$ 是网络平均最短路径长度， $L_{max}$ 是最大中介中心性的顶点的下标。另一方面，每一个时刻，最大的中介中心性的顶点能够传递的数据包个数为 $Q_{out} = C_{L_{max}}$ （注意，这里 $C_i$ 是每个时刻顶点 $i$ 向终点传输的数据包的数目。当拥堵出现的时候，意味着具有最大中介中心性的顶点接受到的数据包的数目，大于向外传递的数据包的数目。即 $Q_{in} \geq Q_{out}$ 。那么

$$R_c D \frac{B_{L_{max}}}{\sum_{j=1}^n (B_j)} = C_{L_{max}}. \quad (2-6)$$

因为 $\sum_{j=1}^n (B_j) = n(n-1)D$ ，等式(2-6)可以简化为：

$$R_c = \frac{n(n-1)C_{L_{max}}}{B_{L_{max}}}. \quad (2-7)$$

我们注意到，在我们的模型中， $C_{L_{max}} = 1$ 。那么下列表达式成立。

$$R_C \approx \frac{n(n-1)}{BC_{max}}, \quad (2-8)$$

其中， $n$ 是网络中节点的个数， $BC_{max} = \max(BC(v)), v \in V$ 。

### 2.2.5 结构洞理论 (Structural Holes Theory)

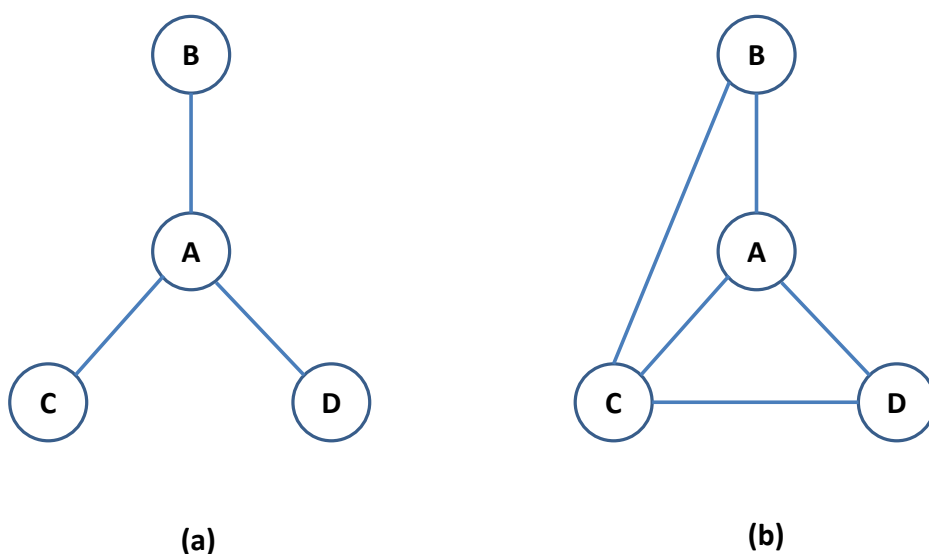


图 2-7: 结构洞说明图

社会网络是由一组社会行动者（如个人或组织）、一组二元关系以及行动者之间的其他社会互动组成的社会结构。社会网络中一个重要的概念就是结构洞(Structural Holes)，最初由Ronald Burt<sup>[91]</sup>在《结构洞：竞争的社会结构》一书中提出来用于社会网络研究的一个概念。结构洞的研究涉及到社会学、经济学和计算机科学等多个学科。结构洞概念的引入，最初是为了解释社会资本差异的起源。以社会网络中的人际关系图为例，如图2-7，其中，结构洞多的人掌握的信息、资源与权力比网路中其他人具有更大的位置优势。图(a)中，A 具有3个结构洞，即A的3个邻点之间互不相连，因此，A比其他3人拥有信息优势和控制优势。而图(b)中，A与C结构洞数都是1，因此这两个人的优势相同，都优于另外两人。

### 2.2.6 相关系数

#### 1、皮尔森相关系数 (Pearson Correlation Coefficient)

皮尔森相关系数主要度量两个事物或两个数据变量之间线性相关性。皮尔森相关系数也称为积差相关，相关系数定义为两个变量之间的协方差和标准差的比值。该系数反映了两个变量之间的线性相关性的强弱程度，相关系数的绝对值越大说明两个变量之间的相关性越强。当相关系数的值大于零时，说明两个变量呈正相关，即一个变量的值越大则另一个变量值也越大；当相关系数值小于零时，表明两个变量呈负相关，即一个变量的值越大则另一个变量值反而越小；当相关系数值等于零时，表明两个变量不是线性相关的。显然，当相关系数的值等于1 或-1 是时，表明两个变量可以由直线方程来描述，即所有的样本点都落在同一条直线上。

#### 2、斯皮尔曼等级相关系数 (Spearman's Rank Correlation Coefficient)

在统计学中,斯皮尔曼等级相关系数<sup>[96]</sup>是衡量两个变量的依赖性的非参数指标。他的适用范围广泛，只需要两个变量的观测值是成对出现的等级评定数据，或者是由连续变量观测数据转化得到的等级数据，不考虑两个变量的总体分布形态、样本容量的大小，都可以应用斯皮尔曼等级相关性来研究。斯皮尔曼等级相关系数利用单调方程来评价两个统计变量的相关性。如果观测数据中没有重复值，并且两个变量完全单调相关时，斯皮尔曼相关系数的值为+1或-1。假设两个随机变量分别为 $X$ 、 $Y$ （也可以看做两个集合），斯皮尔曼相关系数表明 $X$ 和 $Y$ 的相关方向。如果当 $X$ 增加时， $Y$ 趋向于增加，斯皮尔曼相关系数则为正。如果当 $X$ 增加时， $Y$ 趋向于减少，斯皮尔曼相关系数则为负。若 $X$ 增加时 $Y$ 没有任何趋向性则斯皮尔曼相关系数为零。

### 2.2.7 平均路径长度 (Average path length)

平均路径长度是网络拓扑中的一个概念，定义为所有可能的网络节点对沿最短路径的平均步数。它是衡量网络上信息或大众运输效率的一种手段。平均路径长度 (APL) 定义如下：

$$APL = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)} \quad (2-9)$$

其中 $d_{ij}$ 表示从顶点 $i$ 到顶点 $j$ 的最短路径长度， $n$ 表示网络中节点的个数。如算法2-1所示，本文中计算网络平均最短路径长度的方法是，先用Floyd-Warshall算法计算出网络中所有点对之间的最短路径长度，然后再计算其平均值。整个计算平均路径长度的时间复杂度为 $O(V^3)$ 。Floyd-Warshall（弗洛伊德算法）算法，简称Floyd 算法，是计算图中任意两点间的最短路径的一种算法，可以正确处理任何图中的最短路径问题。该算法的基本思路就对图中每一对顶点，进行 $n$ 次比较( $n$ 为图顶点的个数)，每一次比较(假设当前比较的顶点标号为 $i, j$ )原来顶点 $i$ 与顶点 $j$ 之间的路径长度是否大于加入顶点 $k$ 后

的路径长度，如果大于，就更新顶点 $i$ 与顶点 $j$ 之间的路径及长度值，最终计算得到更新后的每对顶点之间的最短路径和长度值。本文主要用于求解无向，无权的图的任意两点之间的最短路径长度。该算法的时间复杂度为 $O(V^3)$ 。

---

#### 算法 2-1 求平均最短路径长度算法

---

输入: 图 $G$  的邻接矩阵 $A$ 。

步骤1:初始化图的距离矩阵 $D$ ，将邻接矩阵变为邻接距离矩阵，两顶点之间不相邻时赋值为无穷大 $\text{inf}$ ，对角元素为0。平均路径长度 $APL = 0$ ，图中顶点个数为 $n$ 。

步骤2:Floyd算法求解任意两点的最短距离

for  $i \leftarrow 1$  to  $n$  do

for  $j \leftarrow 1$  to  $n$  do

for  $k \leftarrow 1$  to  $n$  do

$D[j, k] \leftarrow \min\{D[j, k], D[k, j] + D[i, j]\}$

返回 $D$

步骤3: $APL \leftarrow \text{sum}(\text{sum}(D)) / (n * (n - 1))$

输出: 平均路径长度 $APL$ 。

---

## 2.3 本章小结

本章对论文中即将用到的基本概念和模型进行了描述。其中，包含，图论相关的概念和模型，特别是本文主要工作相关的内容，如代数连通度，费德勒向量，传输模型，网络平均最短路径长度等等。另外还有一些基础定义和理论读者可以参考图论相关书籍。为了读者能更容易的理解，部分概念将在后文中具体研究内容部分给予说明。



## 第三章 费德勒向量的极值研究

### 3.1 相关符号和定义

首先我们介绍一下与本章内容相关的符号与定义。设  $G = (V, E)$  表示一个无向, 无权重的简单图, 其中,  $V = \{v_1, v_2, \dots, v_n\}$ 。设  $A$  和  $B$  点集  $V$  的子集, 则我们用  $d(A, B)$  表示集合  $A$  和  $B$  之间的距离, 且  $d(A, B) = \min\{d(v_i, v_j) | v_i \in A, v_j \in B\}$ , 其中  $d(v_i, v_j)$  表示顶点  $v_i$  与顶点  $v_j$  之间的距离。前文已给出图的拉普拉斯矩阵定义, 即  $L = D - A$ 。事实上, 定义图的  $(0, 1)$  关联矩阵 (Incidence matrix)  $D$ , 其中,  $D(i, j) = 1$  当且仅当顶点  $i$  是边  $j$  的一个端点。即, 关联矩阵的行表示图的顶点, 列表示图的边。则图的拉普拉斯矩阵也可以表示为  $L = DD^T$ 。从以上定义可以得到, 拉普拉斯矩阵是半正定的对称矩阵, 因此, 拉普拉斯矩阵的特征值是非负的。于是有  $LI = 0$ , 其中  $I$  是单位矩阵, 所以可用  $\mu_1(G) \geq \mu_2(G) \geq \dots \geq \mu_{n-1}(G) \geq \mu_n(G) = 0$  表示拉普拉斯矩阵  $L$  的特征值。其中, 第二小的特征值  $\mu_{n-1}(G)$  定义为代数连通度 (algebraic connectivity)<sup>[57]</sup>, 表示为  $\alpha(G)$ 。与代数连通度对应的特征向量叫做费德勒向量 (Fiedler vector), 用  $X$  表示, 即  $LX = \alpha X$ 。由矩阵分析知识可知, 对于给定的一个图, 费德勒向量是一组值, 不失一般性, 本文主要研究归一化之后的一个费德勒向量, 即令  $X = (x_1, x_2, \dots, x_n)$ , 其中,  $\sum_{i=1}^n (x_i^2) = 1$ 。图3-1, 度序列为  $(d_1, d_2, \dots, d_{10}) = (2, 2, 2, 3, 5, 3, 3, 2, 2, 2)$ 。则图3-1 的度矩阵定义为  $D = \text{diag}(2, 2, 2, 3, 5, 3, 3, 2, 2, 2)$ 。邻接矩阵  $A$ , 度矩阵  $D$ , 拉普拉斯矩阵  $L$ , 如下3-1, 3-2所示:

$$A = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \end{pmatrix}, D = \begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 5 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix} \quad (3-1)$$

$$L = D - A = \begin{pmatrix} 2 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & 0 & 3 & -1 & -1 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & -1 & 5 & -1 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & -1 & -1 & 3 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 3 & 0 & -1 & -1 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 2 & 0 & -1 \\ 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & -1 & 0 & 2 \end{pmatrix} \quad (3-2)$$

结合3-2, 由 $LX = \alpha X$ , 利用MATLAB程序, 可以求得图3-1的费德勒向量为:  
 $X = (0.5138, 0.4383, 0.3145, 0.1283, -0.0531, -0.0689, -0.2901, -0.3209, -0.2450, -0.4170)$ 。同时可得代数连通度为 $\alpha = 0.5543$ 。

从费德勒向量的定义还可以得到如下重要性质:

1) 任意向量 $Y = (y_1, y_2, \dots, y_n)$ ,  $YLY = \frac{\sum_{v_i, v_j \in E} (y_i - y_j)^2}{\sum_{k=1}^n y_k^2}$ 。

由拉普拉斯矩阵的定义, 结合图结构可以得到此结论。

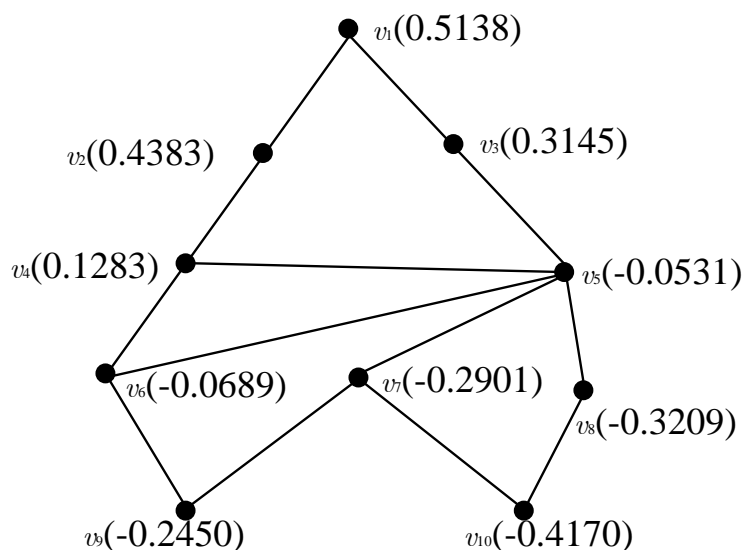
2) 令费德勒向量为 $X = (x_1, x_2, \dots, x_n)$ , 则 $\sum_{i=1}^n x_i = 0$ 。

因为 $LI = 0$ , 并且 $L$ 是半正定的矩阵, 零元是 $L$ 最小的特征值。而 $G$ 是连通图, 我们容易有上述性质1) 得到单位矩阵 $I$ 是 $YLY = 0$ 的唯一线性独立解。这就意味着零元是简单的特征值, 其他所有的特征值都是正数, 且与之对应的特征向量都是与 $I$ 正交的。即,  $XI = \sum_{i=1}^n x_i = 0$ 。

我们定义下标集 $\{1, 2, \dots, n\}$ 的子集如下:

$$\begin{aligned} i_0(X) &= \{j | 1 \leq j \leq n, x_j = 0\}, \\ i_-(X) &= \{j | 1 \leq j \leq n, x_j < 0\}, \\ i_+(X) &= \{j | 1 \leq j \leq n, x_j > 0\}. \end{aligned} \quad (3-3)$$

上述定义是将费德勒向量的分量的下标进行了分割, 其中 $i_0(X)$ 是指费德勒向量分量等于零的下标的集合, 同样 $i_-(X)$  ( $i_+(X)$ ) 指费德勒向量分量为负数 (正数) 的下标的集合。事实上, 我们在研究费德勒向量的时候, 提到向量的分量 $x_1, x_2, \dots$  往往是与标号的顶点 $v_1, v_2, \dots$  相对应的, 也就是说在不引起混淆的情况下下标集合 $\{1, 2, \dots, n\}$ 可以表示费德勒向量的分量, 也可以表示图的顶点集。因此公式3-3也可以解释为 $i_0(X)$ 是



### 3.2 费德勒向量的研究现状与成果

将图用矩阵表示，如邻接矩阵和关联矩阵等，从矩阵的角度对图进行研究属于代数图论的范畴。更早的图相关的矩阵的研究基于这种连接或不连接的0和1的表示，随后人们逐渐发现，这种二值化表示的矩阵所能承载的图本身结构的信息是十分有限的，研究人员尝试以更丰富的矩阵来表示图的结构。因此图的拉普拉斯矩阵被提出来，相对与邻接矩阵，拉普拉斯矩阵中加入了图的顶点度数。如此其表达的信息量增大了。在此基础上，M. Fiedler 在1975 年的文章<sup>[22]</sup>中首次将图的拉普拉斯矩阵特征向量即费德勒向量与图的拓扑结构之间建立起了联系。提出了多个影响深远的定理，这些定理显示了费德勒向量分量的值的大小决定了图中对应顶点与其他顶点的位置关系。如，大于某一负数的分量对应的点集是相互连通的。另外这些定理为网络数据挖掘提供了聚类方法的理论基础。该方法在数据挖掘中有广泛的应用。

**定理1.** ([22], 3.3) 设 $G$ 是简单连通图，且费德勒向量为 $X = (x_1, x_2, \dots, x_n)$ 。顶点集为 $V = \{v_1, v_2, \dots, v_n\}$ 。下标集 $N = \{1, 2, \dots, n\}$ 对于任意 $r \geq 0$ ，令

$$M(r) = \{i \in N \mid x_i + r \geq 0\}, \quad (3-5)$$

则基于集合 $M(r)$ 的图 $G$ 的导出子图，记为 $G(r)$ ，是连通的。

该定理陈述的是可以用一个数值，将图划分出一个连通分支。参见图3-1，顶点 $v_8$ 的费德勒向量分量对应为 $x_8 = -0.3209$ ，依据以上定理的假设，我们令 $r = 0.3209$ ，则 $M(r) = \{v_1, \dots, v_9\}$ ，由图3-1可以看出， $M(r)$ 导出子图是连通的。

矩阵的性质与图的结构性质相关的关键理论是矩阵理论中的不可约性质。 $n$ 阶方阵 $A$ 是可约的，如果存在一个排列阵 $P$ 使得 $P'AP$ 为一个分块的上三角阵，否则称该矩阵是不可约的。具体如下定义可知：

**定义4.** <sup>[97]</sup> 设 $A \in R^{n \times n} (n \geq 2)$ ，若存在置换矩阵 $P$ ，使得公式3-6成立，其中， $B, C, D$ 是阶数不小于1的方阵， $O$ 是零矩阵，则称 $A$ 是可约的，否则称矩阵 $A$ 是不可约的。

$$PAP^T = \begin{pmatrix} B & C \\ O & D \end{pmatrix} \quad (3-6)$$

**定理2.** 设无向，无权的简单图 $G$ 的邻接矩阵为 $A$ ，则图 $G$ 是连通的当且仅当 $A$ 是不可约的。

定理2的证明是显然的, 假设 $A$ 是可约的, 则 $\exists P$ , 使得公式3-6成立, 则从图论角度来看, 子矩阵 $B$ 对应的图中的顶点, 与子矩阵 $C$ 和 $D$ 对应的图中的顶点之间是不存在相连的边的。因此子矩阵 $B$ 对应的子图, 与子矩阵 $C$ 和 $D$ 对应的子图是不连通的。

定理2为图的代数角度研究和结构角度研究搭建了桥梁, 在研究图或子图是否连通时, 就可以研究其对用的矩阵是否可约。因此定理1的证明思路就是往证集合 $M(r) = \{i \in N \mid x_i + r \geq 0\}$ 对应的矩阵式不可约的。具体过程参考 [22], 这里不再赘述。由矩阵理论可以知道若 $LX = \alpha X$ , 则 $L(-X) = \alpha(-X)$ , 也就是说, 如果 $X$ 是图的费德勒向量, 那么 $-X$ 也是图的费德勒向量, 因此, 定理2很容易得到其对应的推论, 如下:

**推论1.** ([22], 3.4) 设 $G$ 是简单连通图, 且费德勒向量为 $X = (x_1, x_2, \dots, x_n)$ 。顶点集为 $V = \{v_1, v_2, \dots, v_n\}$ 。下标集 $N = \{1, 2, \dots, n\}$ 对于任意 $r \leq 0$ , 令

$$M(r) = \{i \in N \mid x_i + r \leq 0\}, \quad (3-7)$$

则基于集合 $M(r)$ 的图 $G$ 的导出子图, 记为 $G(r)$ , 是连通的。

定理1是图的分割问题的重要理论基础, 也是数据挖掘中谱聚类方法的理论依据。该定理是在图中分割出一部分连通子图, 同时我们自然会思考其余的顶点构成的图结构有什么特点。M. Fiedler在定理1基础上, 继续深入研究, 试图能够在图中找到更多的连通分支, 期望能够找到更好的图的多分支的划分方式, 以下定理就是一个重要的成果。

**定理3.** ([22], 3.11) 设 $G$ 是连通图, 且费德勒向量为 $X = (x_1, x_2, \dots, x_n)$ 。顶点集 $V = \{v_1, v_2, \dots, v_n\}$ , 令 $v_k$ 是图 $G$ 的割点, 设 $G_0, G_1, \dots, G_r$ 表示图 $G$ 中删除割点 $v_k$ 及与之相连的所有边之后得到的所有连通分支。那么:

(i) 如果 $x_k > 0$ , 那么只存在一个包含负费德勒向量分量的连通分支, 其他所有连通分支中的点对应的费德勒向量分量的值都大于 $x_k$ 。

(ii) 如果 $x_k = 0$ 并且存在一个连通分支同时包含正值和负值的分量, 那么这样的连通分支有且仅有一个, 并且其他连通分支对应的分量值全部为零。

(iii) 如果 $x_k = 0$ 并且没有连通分支既包含正值分量, 有包含负值分量, 那么, 每个连通分支中的分量值只能是大于零, 小于零, 等于零其中的一种。

该定理给出了特殊的顶点(割点)对应的费德勒向量分量的值对图的结构或连通性的影响。割点对应费德勒向量分量的值决定了删掉该点之后各个联通分支的费德勒

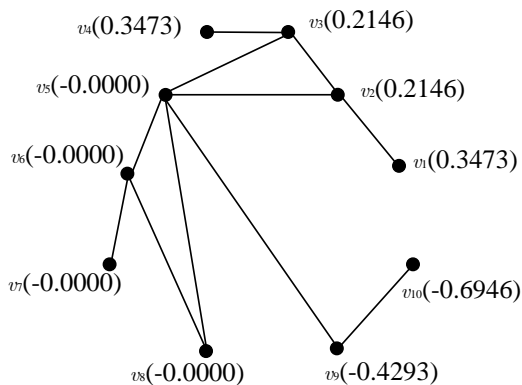


图 3-2: 费德勒向量为  $X=(0.3473, 0.2146, 0.2146, 0.3473, 0.0000, 0.0000, 0.0000, 0.0000, -0.4293, -0.6946)$  的图。

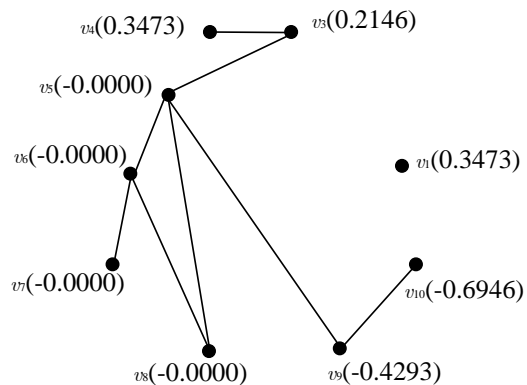


图 3-3: 删除割点  $v_2$  后的连通分支

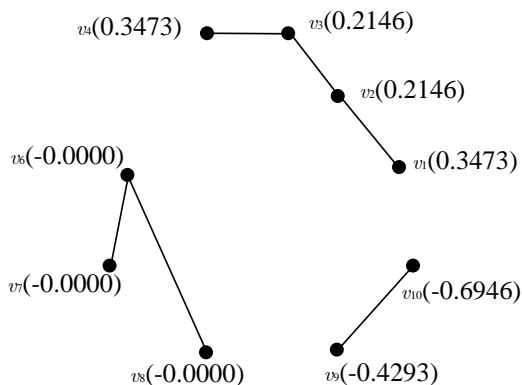


图 3-4: 删除割点  $v_5$  后的连通分支

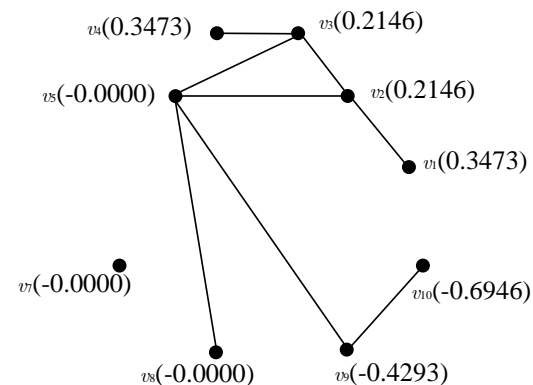


图 3-5: 删除割点  $v_6$  后的连通分支

向量分量的值。比较遗憾的是该定理是从代数角度刻画图结构的单向的理论。也就是说给定了割点的费德勒向量的值,研究图的结构(连通分支)。而并不能从图的结构(连通分支)方向研究点的费德勒向量分量。该定理可以作为定理1的一个推广,比如其中第一个结论,很容易由定理1得到。由上文提到的性质2)可以得到 $\sum_{i=1}^n x_i = 0$ ,而 $x_k > 0$ ,所以一定存在 $x_l < 0$ ,不失一般性,假设删除割点 $v_k$ 之后有 $r$ 个连通分支, $G_0, G_1, \dots, G_r$ ,且 $x_l$ 位于连通分支 $G_0$ 中,因此,只要考虑其他连通分支,并证明所有其他连通分支中的点的费德勒向量分量的值都大于 $x_k$ 即可。不妨假设存在一个顶点 $v_p$ 且不在连通分支 $G_0$ 中,其对应的值 $x_p < x_k$ ,则 $\exists \delta$ ,使得 $x_k - \delta \geq x_p$ ,考虑由集合 $M(s) = \{s \in N \mid x_s \leq x_k - \delta\}$ 生成的导出子图 $G'$ ,由定理1可知, $G'$ 是连通子图。同时可知所有小于 $x_k$ 的顶点都是连通的。即其他连通分支中的顶点的费德勒向量值都大于 $x_k$ 。另外两个结论也是类似的思路,具体参见[22]。

为了更好的理解定理3,我们给出几个实例图,图3-2是费德勒向量为 $X=(0.3473, 0.2146, 0.2146, 0.3473, 0.0000, 0.0000, 0.0000, 0.0000, -0.4293, -0.6946)$ 的图。图3-3显示原图删除割点 $v_2$ (大于零)后,剩下两个连通分支,一个只有单点 $v_1$ ,且该点对应的值大于 $v_2$ 点对应的值。另一个连通分支由顶点 $v_3, v_4, \dots, v_{10}$ 组成,显然该连通分支包含负值的费德勒向量分量。该图验证了定理中的第一个结论,图中删除一个费德勒向量大于零的割点的情况。图3-4显示了原图删除割点 $v_5$ (等于零)后,共产生三个连通分支。恰好,每个连通分支只有大于零,小于零,等于零的值。该图对应了定理中的第二个结论。最后,图3-5显示了原图删除割点 $v_6$ (等于零)后(对应了定理中的第三个结论),产生了两个连通分支,一个只有零值,一个既包含负值也包含正值。这个定理进一步显示了费德勒向量(分量值的正负)与图的结构(子图连通性)之间的关系。

费德勒向量的提出,为人们从代数角度研究图的结构奠定了理论基础,相当长时间里,研究人员不断的发掘费德勒向量的应用,就其本质来说,都是用费德勒向量值完成图的划分。例如,谱聚类方法,图像分割,图数据聚类等等。但另一方面,由于代数工具的缺乏,使得费德勒向量刻画图的结构的研究进展缓慢。直至2014年,J. Urschel<sup>[50]</sup>等对费德勒向量理论进一步研究,得出了如下定理:

**定理4.** ([50], Theorem 4.2) 设 $G$ 是简单连通图,则存在图 $G$ 的费德勒向量 $X$ 使得由 $i_+(X) \cup i_0(X)$ 与 $i_-(X)$ 得到的导出子图都是连通的。

该定理指出任意无向无权重的简单图都存在一个费德勒向量,使得费德勒向量分量的非负值对应的点和负值对应的点分别连通。这个结论改进了M. Fiedler原有的结论,进一步确定了费德勒向量的临界值0对于图结构中连通性的影响。

2017 年, M. Begué<sup>[51]</sup>得出了另外一个结论: 具有正费德勒向量分量的点与具有负费德勒向量的点之间的距离不大于2。具体定理如下:

**定理5.** (*[51], Lemma 3.1*) 设 $G$ 是简单连通图, 图 $G$ 的费德勒向量 $X$ , 则 $d(i_+(X), i_-(X)) \leq 2$ 。

其中,  $d(i_+(X), i_-(X))$  表示集合 $i_+(X)$ 对应的导出子图的顶点集与集合 $i_-(X)$ 对应的导出子图的顶点集之间的距离。

### 3.3 费德勒向量极值的研究现状与成果

费德勒向量极值研究主要关注费德勒向量分量中最大值最小值对应的图的顶点的结构特点。该内容的研究最早源于2011年, Moo K. Chung等人受热方程解的极值的启发, 研究了极值直径图 (为了便于说明, 这里我们采用J.Lefèvre 等人的定义), 即若图 $G$ 的费德勒向量分量的最大值和最小值对应顶点之间的距离等于图的直径 (图中任意两顶点之间的距离最大值), 则称该图具有极值直径, 是极值直径图。Moo K. Chung等人提出: 树是极值直径图。若图 $G$ 中, 任意两个顶点之间有且仅有一条路径, 则 $G$  又称为树。换句话说, 树的费德勒向量分量的最大值与最小值对应的点是该树所有点中距离最远的两个点。但是, 很快关于该猜想的反例被提出了<sup>[47, 48]</sup>。Evans 等人<sup>[47]</sup> 提出了一种费德勒玫瑰 (Fiedler rose) 图, 不具有上述的特点。如图3-6 显示, 这幅图看起来像一朵有着弯曲的“叶子”、长长的“茎”和许多“花瓣”的玫瑰, 因此称它为“费德勒玫瑰”。通过MATLAB计算可以得, 费德勒向量的极值不是最远的茎的两端, 而是在较长的路径端点和花瓣之间。

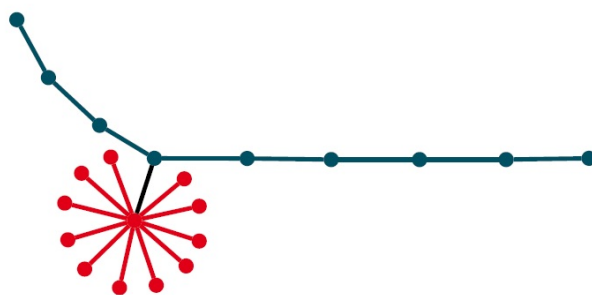


图 3-6: 费德勒玫瑰, 极值点之间的距离不等于图中点对最远距离的反例, 来源于参考文献<sup>[47]</sup>。

费德勒玫瑰图显示: 树中存在不具有极值直径的图, 受此启发J.Lefèvre 等人<sup>[48]</sup>继续深入研究了费德勒玫瑰, 并定义了满足极值直径性质的图, 重点解决了哪些费德勒玫瑰图具有极值直径。



**定义5.** 设 $G = (V, E)$ 表示一个简单图,  $X$ 是费德勒向量, 如果下列条件满足, 则称图 $G$ 具有极值直径性质。

1) 向量 $X$ 只有一个最大值, 一个最小值;

2) 令 $u$ 是向量 $X$ 最大值对应的顶点,  $v$ 是向量 $X$ 最小值对应的顶点, 顶点 $u, v$ 之间的距离等于图 $G$ 的直径 (图中任意两顶点之间的距离最大值)。

**定义6.** 给定三个整数,  $s, t, p$ , 则玫瑰树 $R(s, t, p)$ , 如图3-7所示, 按以下方式构造: 形状类似费德勒玫瑰图, 路径长度为 $s + t + 1$ , 具有 $p$ 个分支的星图, 其中心点连接在路径的第 $s + 1$ 个顶点上。

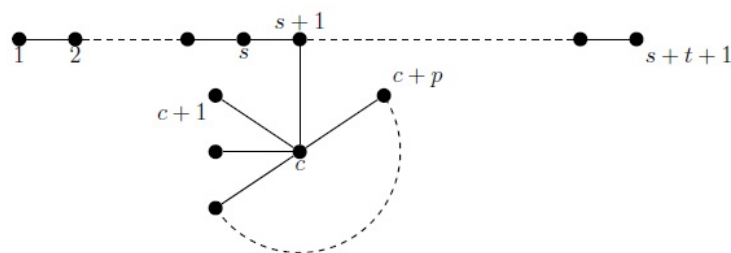


图 3-7: 玫瑰树。来源于参考文献<sup>[48]</sup>。

**定理6.** <sup>[48]</sup> 令 $R(s, t, p)$ 是上述定义的玫瑰树。则存在两个函数 $f_m, f_M : (N - \{0, 1, 2\})^2 \rightarrow R$ 使得:

- 1) 对于 $p \leq f_m(s, t)$ ,  $R(s, t, p)$ 是极值直径图;
- 2) 对于 $p > f_M(s, t)$ ,  $R(s, t, p)$ 是极值直径图;

应用定理6, J.Lefèvre等人进一步将图划分, 提出了极值直径图的一个充分条件。通过基本的图论分析, 得到树的直径对应的顶点在叶子节点上, 其中叶子顶点是图的1度点, 又称悬挂点。因此J.Lefèvre等人的另一个重要结论是, 树的费德勒向量极值对应叶子顶点。这个结论与本论文的一个结论基本一致, 但本论文从另一个角度进行了解读和证明。近期, H. Gernandt等人<sup>[49]</sup>进一步研究了J.Lefèvre等人的研究成果, 证明了更特殊的关于树的费德勒向量极值的结论如下:

**定理7.** <sup>[49]</sup> 令 $T$ 表示如图3-8所示的树, 且直径 $d(T) \geq 2$ , 费德勒向量 $X = (x_1, \dots, x_n)$ , 则下列性质成立。

- 1) 费德勒向量的极值点位于悬挂点;
- 2) 以下断言中有一个是成立的:
  - (i) 路径 $v_1, \dots, v_k$ 上的顶点对应费德勒向量分量的值是单调递增或递减的;
  - (ii) 若 $x_1, x_k \geq 0$ , 则存在 $x_{k'} \geq 0$ , 满足 $x_1 \geq x_2 \geq \dots \geq x_{k'}$ 和 $x_{k'} \leq x_{k'} + 1 \leq \dots \leq x_k$ ;
  - (iii) 若 $x_1, x_k \leq 0$ , 则存在 $x_{k'} \leq 0$ , 满足 $x_1 \leq x_2 \leq \dots \leq x_{k'}$ 和 $x_{k'} \geq x_{k'} + 1 \geq \dots \geq x_k$ 。

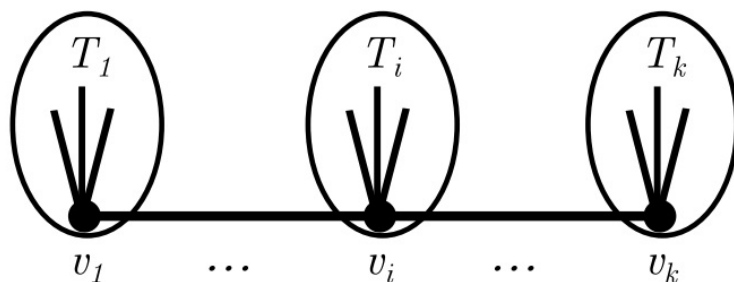


图 3-8: 树图, 路径 $v_1, \dots, v_k$ 上的每个顶点上都连接一颗子树 $T_i$ 。来源于参考文献<sup>[49]</sup>。

以上结果表明, 目前对于费德勒向量极值的研究主要集中在特殊图类-树图。而且是特殊的树图, 例如费德勒玫瑰图, 玫瑰树图等, 并且取得了一定成果, 但对于一般图的结论很少, 接下来, 本文将对一般图的费德勒向量分量的极值进行研究。

### 3.4 本论文关于费德勒向量极值的主要理论成果

这一节我们主要研究费德勒向量分量的最大值和最小值对应的顶点的结构性质。为了便于理解, 我们重新看前文提到的图3-9, 结合上一节叙述的内容, 对于树图而言, 费德勒向量最大值位于叶子节点(最小度点)。依据这个结论我们考虑其一般性, 通过对图3-9分析可以看到, 其费德勒向量分量的最大值(0.3473)也对应了图的最小度顶点( $v_1$ )。由此, 我们首次提出这种费德勒向量分量最大值和图的最小度点相对应的图类的概念,  $\alpha$ -匹配图如下:

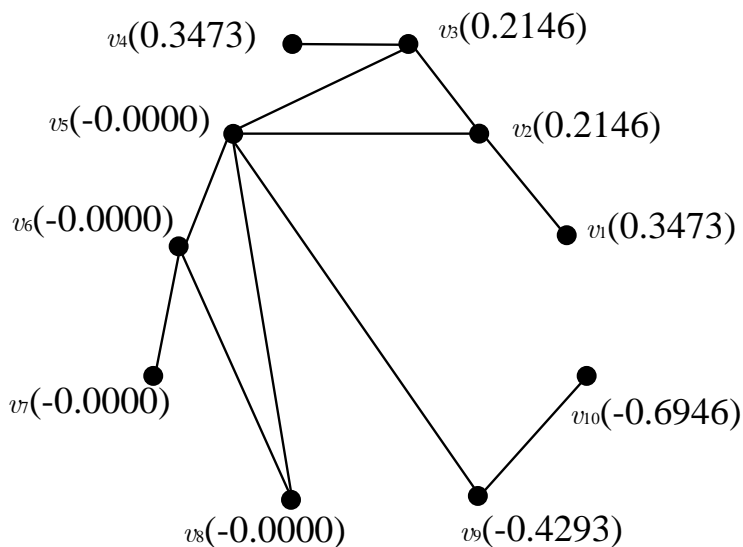


图 3-9: 费德勒向量为 $X=(0.3473, 0.2146, 0.2146, 0.3473, 0.0000, 0.0000, 0.0000, 0.0000, -0.4293, -0.6946)$ 的图。

表 3-1:  $\alpha$ -匹配图的百分比 (%), 随机图 $G_n(p)$ 

p	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75
n=100	98.1	98	98.8	98	98.9	97.7	98.7	99.1	98.3	99
n=300	99.1	99.4	99.3	99	99.2	99.3	99.5	99.1	99.5	99.3
n=500	99.1	99.5	99.6	99.8	99.2	99.6	99.5	99.5	99.8	99.2

### 3.4.1 $\alpha$ -匹配图

设 $G$ 是简单连通图, 且费德勒向量为 $X = (x_1, x_2, \dots, x_n)$ 。设 $X$ 的最大分量为 $x_{max}$ 。顶点集为 $V = \{v_1, v_2, \dots, v_n\}$ 。

**定义7.** 设 $G$ 是简单连通图, 费德勒向量 $X$ 的最大分量 $x_{max}$ 对应的顶点为 $v_{i_0}$ , 若顶点 $v_{i_0}$ 度是图 $G$ 中最小的。则我们称图 $G$ 为 $\alpha$ -匹配图。

显然图3-9是 $\alpha$ -匹配图, 接下来我们用随机图验证一下 $\alpha$ -匹配图的存在性。我们用 $ER$ 图 $G_n(p)$ 进行仿真实验, 对于不同的参数 $n$ 与 $p$ 组合, 生1,000个随机图, 然后检验 $\alpha$ -匹配图的百分比, 结果如表格3-1结果所示,  $\alpha$ -匹配图比例非常高, 这是个很好的结果, 意味着 $\alpha$ -匹配图类覆盖了大部分图, 而且从表中可以看出: 当图 $G_n(p)$ 中, 任意两点之间存在边的概率 $n$ 不变时候, 趋势上呈现出, 顶点数目 $n$ 越大,  $\alpha$ -匹配图的比例越高。就整体而言,  $n$ 和 $p$ 值越大,  $\alpha$ -匹配图的比例越高。

为了进一步证明 $\alpha$ -匹配图的存在性, 以下我们将证明树图是 $\alpha$ -匹配图。

**定理8.** 树是 $\alpha$ -匹配图。

**证明.** 设 $X = (x_1, x_2, \dots, x_n)$ 是树 $T$ 的费德勒向量。不失一般性, 设所有 $X$ 的最大分量值中 (假设有多个),  $x_{i_0}$ 对应的顶点 $v_{i_0}$ 的度最小 (不一定是整个图的最小度)。由费德勒向量的定义可知,  $x_{i_0} > 0$ 。以下, 我们用反证法来证明: 假设顶点 $v_{i_0}$ 的度大于1 (树的最小度是1), 即顶点 $v_{i_0}$ 不是树 $T$ 的最小度点。则 $v_{i_0}$ 点是割点, 树 $T$ 删除 $v_{i_0}$ 后有两个或两个以上的连通分支。设 $T_0, \dots, T_r$ 是树删除 $v_{i_0}$ 点以及与之相连的边后的所有连通分支。又因为 $x_{i_0}$ 是最大值, 所以 $x_{i_0} > 0$ 。由定理3(i), 可得, 有一个连通分支, 设为 $T_j$ , 包含费德勒向量的负值, 其他所有连通分支中的点对应的费德勒向量的分量值, 设为 $x_s$ , 均有 $x_s > x_{i_0}$ 。显然, 这与 $x_{i_0}$ 是最大值矛盾, 因此顶点 $x_{i_0}$ 的度只能是1度。证毕。□

图3-10显示了一颗 $\alpha$ -匹配树, 树的结构与其他一般图比较而言, 相对简单。树中没有回路, 任意两点之间有且只有一条路径。图中显示, 顶点 $v_4, v_6, v_8$ 都是最小度点, 且他们对应的费德勒向量分量值都是最大值0.4。事实上, 正如表格3-1显示的, 本论文中提到的所有图, 都是 $\alpha$ -匹配图, 而定理8证明了 $\alpha$ -匹配图的存在性, 这些将作为下一

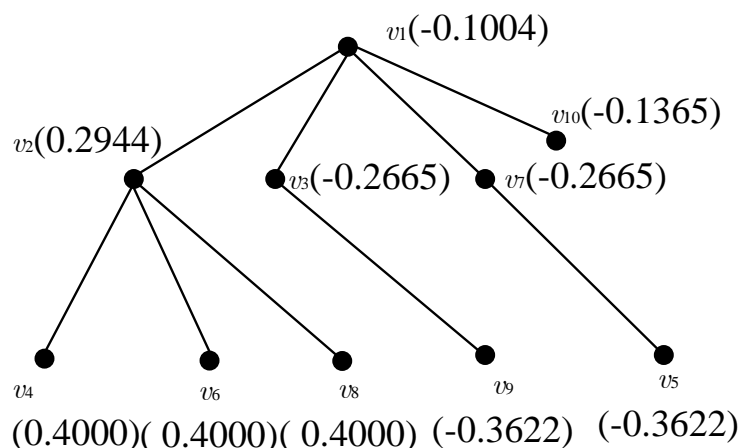


图 3-10: 费德勒向量为 $X=(-0.1004, 0.2944, -0.2665, 0.4000, -0.3622, 0.4000, -0.2665, 0.4000, -0.3622, -0.1365)$ 的树。

章中，基于费德勒极值的加边算法的部分理论依据。而更多的关于 $\alpha$ -匹配图的理论和充分条件将是我们后期的主要研究课题。

$\alpha$ -配匹图的定义主要是为了研究费德勒向量分量的最大值与图的结构，换句话说，是为了确定其在图中的位置。通过上述的验证和证明，可以得到结论，大部分图的费德勒向量分量的最大值对应的顶点是图中最小度点。这样，就可以初步确定费德勒向量分量的最大值的位置，而接下来，我们进一步考虑，费德勒向量分量的最小值对应的顶点有什么结构特点。受到前文提到的费德勒玫瑰的启发，我们思考，费德勒向量分量最小值的顶点，应该距离最大值的顶点距离比较远。因此，我们在下一节将对距离最大值最远的顶点进行研究。

### 3.4.2 Fiedler-Breadth-First图

首先我们回顾一下，对于有根图的顶点的层概念：有根图中，一个顶点的层定义为该点与根点的距离加1。图3-10是一棵以图中顶点 $v_1$ 为根的广度优先搜索树。且第一层只有根底点， $v_1$ ，第二层有 $v_2, v_3, v_7, v_{10}$ 四个顶点。其余顶点位于最高层，第三层。接下来，我们仍然从图类角度，定义Fiedler-Breadth-First图。

**定义8.** 设图 $G$ 是以顶点 $v_{i_0}$ 为根进行广度优先搜索得到的广度优先图，且顶点 $v_{i_0}$ 对应的是费德勒向量的最大分量 $x_{i_0}$ 。如果费德勒向量的最小值在图 $G$ 的最高层，我们就称图 $G$ 为Fiedler-Breadth-First图。

图3-11显示了一个5层的Fiedler-Breadth-First图。如图所示，以最大值（0.7622）对应的顶点 $v_1$ 为根，得到的广度优先搜索图，最小值（-0.2569）对应的顶点恰好

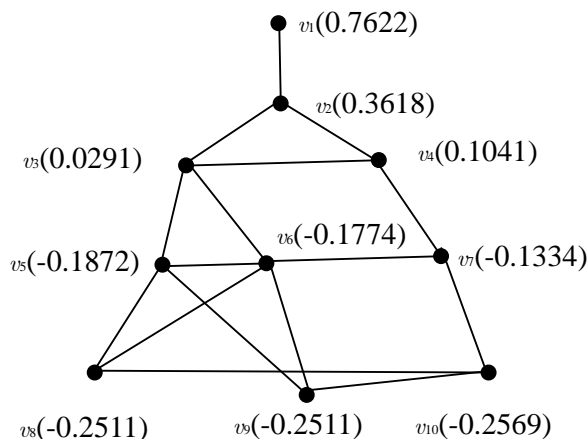


图 3-11: 一个以 $v_1$ 为根的5层的breadth-first graph。费德勒向量为 $X=(0.7622, 0.3618, 0.0291, 0.1041, -0.1872, -0.1774, -0.1334, -0.2511, -0.2511, -0.2569)$ 。顶点 $v_1$ 是根，其对应费德勒向量的最大分量值(0.7622)。且最小值(-0.2569)位于图的最高层。

表 3-2: *Fiedler-Breadth-First*图的百分比 (%), 随机图 $G_n(p)$

p	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75
n=100	94	99.9	99.8	99.8	99.8	99.3	99.3	98.5	98.6	97.8
n=300	100	100	99.9	100	99.7	99.4	99.1	99	98.3	97
n=500	100	99.7	100	99.9	100	99.6	100	98.8	99.8	98.6

在最高层第5层。同样，我们用 $ER$ 图 $G_n(p)$ 进行仿真实验，对于不同的参数 $n$ 与 $p$ 组合，我们生1,000个随机图，检验*Fiedler-Breadth-First*图的百分比。如表格3-2所示，*Fiedler-Breadth-First*图的百分比非常高，因此我们可以说大部分图都是*Fiedler-Breadth-First*图。而且从表格可以看出，顶点的数目越大，百分比数值越大，也就是说顶点数多的大规模图，*Fiedler-Breadth-First*图的比例更高。

接下来，为了证明*Fiedler-Breadth-First*图的存在性，我们将给出一类如图3-11所示的*Fiedler-Breadth-First*图。并给出严格的理论证明。设 $V_+ = \{v_i : x_i > 0\}$ ， $V_- = \{v_i : x_i < 0\}$ ，且 $V_0 = \{v_i : x_i = 0\}$ ，其中 $x_i$ 表示 $X$ 中对应顶点 $v_i$ 的特征分量。

**定理9.** 设图 $G$ 是以顶点 $v_{i_0}$ 为根进行广度优先搜索得到的广度优先图，其中顶点 $v_{i_0}$ 对应的费德勒向量分量的最大值 $x_{i_0}$ 。图 $G$ 的最高层数是 $l$ 。设 $V(k)$ 表示 $k$ 层上的顶点的集合，令 $X_{\min}(k) = \min x_i$ ，其中 $v_i \in V(k)$ 。如果代数连通度 $0 < \alpha < 1$ ， $l \geq 4$ ， $X_{\min}(l-2) > 0$ ， $X_{\min}(l-1) < 0$ ，且 $X_{\min}(l) < 0$ ，那么 $X_{\min}(k) \leq X_{\min}(k-1)$ ，其中 $2 \leq k \leq l$ 。

**证明.** 首先，由定理1可得， $V_- \cup V_0$ 是连通的。因为 $X_{\min}(l-2) > 0$ 且 $X_{\min}(l-1) < 0$ ，则可得，对于任意 $1 \leq k \leq l-2$ ，有 $X_{\min}(k) > 0$ 。如若不然，假设存在 $k^* < l-2$ ，使

得  $X_{\min}(k^*) \leq 0$ 。这与  $V_- \cup V_0$  连通矛盾。又因为  $LX = \alpha X$ ，我们可得，

$$x_i(d_i - \alpha) = \sum_{v_i v_j \in E} x_j, \quad (3-8)$$

其中  $d_i$  是顶点  $v_i$  的度。且

$$\alpha x_i = \sum_{v_i v_j \in E} (x_i - x_j). \quad (3-9)$$

等式3-9显示，对于任意  $x_{k^*} > 0$  的顶点  $v_{k^*}$ ，存在一个它的邻接顶点  $v_{k^{**}}$  使得  $x_{k^{**}} < x_{k^*}$ 。由于  $X_{\min}(1) = x_{i_0}$  是费德勒向量的最大值，我们可以得到  $X_{\min}(1) > X_{\min}(2)$ 。因此，对  $k$  做归纳法，可以得到，对于任意  $2 \leq k \leq l-2$ ， $X_{\min}(k) < X_{\min}(k-1)$ 。而当  $k = l-1$ ，由题设可得  $X_{\min}(l-1) < X_{\min}(l-2)$ 。设  $v_{i^*}$  对应  $X_{\min}(l-1) = x_{i^*}$ 。因为  $X_{\min}(l-1) < 0$  且  $X_{\min}(l-2) > 0$ ，所以顶点  $v_{i^*}$  存在一个位于  $l-2$  层的邻接点  $v_{j^*}$ ，并且  $x_{j^*} > 0 > x_{i^*}$ 。然而，由于  $0 < \alpha < 1$  和等式3-9，可知顶点  $v_{i^*}$  存在一个邻接顶点  $v_{j^{**}}$  对应  $x_{j^{**}} < 0$ 。因此顶点  $v_{i^*}$  的度  $d_{i^*}$  必然大于1，这样，我们修改等式3-8如下：

$$x_{i^*}(1 - \alpha) + x_{i^*}(d_{i^*} - 1) = x_{j^*} + \sum_{j \neq j^*, v_{i^*} v_j \in E} x_j. \quad (3-10)$$

因为  $0 < \alpha < 1$ ，我们可得  $x_{i^*}(1 - \alpha) < x_{j^*}$ 。因此可得， $x_{i^*}(d_{i^*} - 1) > \sum_{j \neq j^*, v_{i^*} v_j \in E} x_j$ 。所以，根据  $d_{i^*} \geq 2$ ，顶点  $v_{i^*}$  存在一个邻点  $v_{j^{***}}$  对应  $x_{j^{***}} < x_{i^*}$ 。显然，顶点  $v_{j^{***}}$  位于  $l$  层。因此， $X_{\min}(l) < X_{\min}(l-1)$ 。证毕。□

由定理9，我们可以很容易得到以下推论。

**推论2.** 设图  $G$  是以顶点  $v_{i_0}$  为根进行广度优先搜索得到的广度优先图，其中顶点  $v_{i_0}$  对应的费德勒向量分量的最大值  $x_{i_0}$ 。图  $G$  的最高层数是  $l$ 。设  $V(k)$  表示  $k$  层上的顶点的集合，令  $X_{\min}(k) = \min x_i$ ，其中  $v_i \in V(k)$ 。如果代数连通度  $0 < \alpha < 1$ ， $l \geq 4$ ， $X_{\min}(l-2) > 0$ ， $X_{\min}(l-1) < 0$ ，且  $X_{\min}(l) < 0$ ，则图  $G$  是 *Fiedler-Breadth-First* 图。

引理2给出形如图3-11的 *Fiedler-Breadth-First* 图类的证明。本质上来看，*Fiedler-Breadth-First* 图是给出了费德勒向量分量极值之间的拓扑结构关系，可以根据费德勒向量最大值确定最小值的位置。事实上， $\alpha$ -匹配图和 *Fiedler-Breadth-First* 图的研究都是为了确定费德勒向量极值在图中的位置。旨在能够通过图的拓扑结构指标确定极值对应的图的顶点。而通过实验验证，我们可以得到确定费德勒极值的方法，即首先根据图的最小度顶点确定最大值，然后以此顶点为根广度优先遍历图，在最远处确定费德勒向量的最小值。该方法将在下一章基于费德勒向量极值的最大化代数连通度的加边算中得到应用。

表 3-3: 既是 *Fiedler-Breadth-First* 图又是  $\alpha$ -匹配图的百分比 (%), 随机图  $G_n(p)$

p	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75
n=100	94	98.9	98.8	98.8	98.8	98.3	98.3	98.5	97.6	97.4
n=300	98.1	97.9	97.9	98.1	97.7	98.4	97.3	99	97.2	97.2
n=500	98.2	98.1	98.4	97.4	97.6	96.6	97.5	97.8	97.8	98.2

### 3.5 讨论

上一节中, 主要研究了如何通过图的结构指标确定费德勒向量的极值的位置, 其中,  $\alpha$ -匹配图的研究目的是确定费德勒向量的最大分量, *Fiedler-Breadth-First* 图的研究是为了通过最大值和最小值的关系确定最小值的位置。从整个研究过程和结果来看, 还有一些问题值得深入研究。

1、两种图类的充分条件需要进一步研究,  $\alpha$ -匹配图和 *Fiedler-Breadth-First* 图具有很好的费德勒向量极值和结构指标的匹配性。但实验验证, 并不是所有的图都是这两类图。因此, 两类图的充分条件甚至充要条件值得我们进一步研究。

2、本节中, 主要对两种图类分别分析讨论, 事实上, 同时具有  $\alpha$ -匹配性和 *Fiedler-Breadth-First* 性的图更具有实际应用意义 (下一章主要的一个理论应用), 这里我们给出实验验证, 理论证明将在我们以后的工作中开展。同样, 我们采用随机图验证, 每组  $n$  和  $p$  组合, 生成 1000 个  $G_n(p)$  图, 计算既是 *Fiedler-Breadth-First* 图又是  $\alpha$ -匹配图的百分比, 如表 3-3 所示, 实验生成的所有图中, 同时具有两种图类性质的图的比例仍然很高, 但比单独具有一种图类性质的图的比例略低。且比例高低与参数  $n$  和  $p$  的值的关联不明显。

### 3.6 本章小结

费德勒向量是图的伴随矩阵拉普拉斯矩阵的一个特征向量, 属于代数图论的研究范畴, 它的重要意义在于是图的结构性质研究和代数研究的桥梁。本章主要介绍费德勒向量的理论进展。M. Fiedler 最早提出费德勒向量, 旨在用代数的角度研究图以及图的导出子图的连通性。其重要的一个结论是根据费德勒向量分量的某个值可以将图划分成不同的连通分支。这个理论研究成果被广泛的应用到数据挖掘领域。另外, 继 M. Fiedler 之后, 费德勒向量的研究主要聚焦在其极值的方向。尤其对于树图, 取得了一些研究成果。本章在已有结果的基础上, 继续深入研究费德勒向量的极值与图结构的关系。为此我们提出了两个定义, 定义了新的图类, 即  $\alpha$ -匹配图和 *Fiedler-Breadth-First* 图, 这两类图的费德勒极值具有显著特点。大量的统计实验显示, 大部分图都是这两类图, 为此, 我们找到并证明了两种这样的图。进一步充实了费德勒向量的理论研究, 并为下一步研究奠定了基础。而通过费德勒向量极值与图结构指标的关系, 我们可以

在不求解费德勒向量的情况下，找到其最大值和最小值对应的顶点，这也为下一章基于费德勒向量极值的最大化代数连通度加边算法提供了部分理论依据。



## 第四章 基于费德勒向量极值的最大化代数连通度的加边算法

代数连通性是图拉普拉斯矩阵的第二个最小特征值，是各种网络系统（如多代理网络系统）的一个基本性能度量。这一章，我们重点讨论如何在网络中添加一条边，通过最大化代数连通性来提高网络连通性和鲁棒性。大多数有效的代数连通性最大化算法都需要直接计算，这会导致时间复杂度很高，特别是对于大型网络。本章中，在上一章分析费德勒向量极值与结构指标关系的基础上，我们提出了一种启发式算法，即最小度和最大距离算法，不需要计算代数连通性。该算法在大型随机网络和自治系统网络中进行了测试。结果表明，与其它算法相比，该算法是有效的，运行时间短。因此，它可以应用于非常大的网络，特别是大型稀疏网络。

### 4.1 研究背景和意义

代数连通度是指图拉普拉斯矩阵的第二小特征值，它是代数图论中衡量图的连通性和稳定性的重要指标<sup>[58] [57]</sup>，自1973年Fiedler提出概念以来，人们对它的理论研究取得了丰硕的成果<sup>[98]</sup>。近年来代数连通度在以图为模型的系统中的应用逐渐引起人们的关注。多智能体系统（multi-agent systems）的许多应用引起了人们巨大的研究兴趣，在多智能体网络中如何达到一致性是多智能体中的一个基本问题。Olfati-Saber等<sup>[99]</sup>指出：代数连通度是衡量多智能体网络系统中一致性算法收敛速度的一个指标，并且，多智能体网络系统的代数连通度越大，基于此系统的一致算法的收敛速度越快。因此，代数连通度在系统和控制领域得到了广泛的研究<sup>[100-104]</sup>。

同时，代数连通度作为网络稳定性和鲁棒性的一个指标也具有应用意义，如，在有限条件下可以利用最大化代数连通度来增加多智能体系统的稳定性。还可以度量代数连通度的变化设计网络维持算法（network-maintenance algorithm），进而来优化无线传感器网络。此外在主干网络中，应用代数连通度设计优化系统，可以提高网络连通性，增加系统抗攻击能力。代数连通度还可以用来优化航线设计等。

现在的研究和应用往往要面向规模庞大，结构复杂，动态多变的网络系统，而传统的代数连通度相关理论多数是在静态，小规模图（graph）的基础上研究发展的，并且代数连通度是矩阵的特征值，在大规模网络中计算代价过大。因此我们需要更深入的研究代数连通度理论，以适应新形势的应用需求，同时探索更多的应用，这对代数连通度的研究提出了新的挑战：能否用结构指标（如：直径，围长，团数，度，独

立数等)近似计算代数连通度,降低计算代数连通度的代价,使其更好的应用在大规模的复杂的网络系统中。

随着硬件技术的发展和人们需求的增加,网络变得越来越大,即网络中的节点数和边数变得越来越多,而网络所承载的数据量也越来越大,这样对网络的要求也越来越高,如何更好的设计网络,以满足人们日益增加的需求催生了各种网络优化设计方法。考虑到经济成本和客观条件约束,在原有的网络中添加一些关联的边,变得很有应用意义,如:增加城市之间航空网络的新航线,主干网络中增加边进而提高网络的连通性,增加系统抵抗攻击的能力等等。

综上所述,本章将讨论在网络中如何增加边进而能最大程度的增加代数连通度。我们称之为**最大化代数连通度问题**(the maximizing algebraic connectivity problem (MACP))。显然,最大化代数连通度问题是我们前文提到的加边问题的特例之一。目前,最大化代数连通性的最有效算法需要直接计算它,这导致很高的时间复杂度,特别是对于大型网络。我们提出了一种启发式算法,最小度和最大距离(MDMD)算法,基于费德勒向量的分析,不需要计算代数连通性。所提出的算法在大随机网络和自治系统的网络(AS)对等信息中进行测试。结果表明,该算法是有效的,并且比其他算法具有更短的运行时间。因此,它可以应用于非常大的网络,特别是对于大型稀疏网络。

## 4.2 国内外研究现状及发展动态分析

### 4.2.1 问题的研究现状与成果

最大化代数连通度问题的研究有很大的应用价值。如:在<sup>[61]</sup>中,Alenazi等人提出了一种启发式算法,通过增加链路来优化骨干网络,从而通过最大化代数连通性来实现更高的网络恢复能力。在<sup>[60]</sup>中,Ibrahim等人提出了一种通过增加一组中继来提高无线传感器网络的代数连通性的网络维护算法。在航空网络中增加新航线的过程中,Wei等<sup>[35]</sup>通过最大化代数连通性,设计了鲁棒的空中交通网络。

围绕着MACP问题,研究人员设计了多种算法,各种算法的总结如下,Ghosh和Boyd<sup>[33]</sup>考虑了增加多个边以使代数连通性最大化的类似问题。他们提出了一种基于费德勒向量的贪心扰动启发式算法,通过一个一个地增加边缘来最大化代数连通性。Wang和Mieghem<sup>[105]</sup>提出了关于MACP的两种策略:一种是在最小度顶点和任意随机顶点之间加一条边;另一种与Ghosh和Boyd所提算法类似,在代数连通度距离最大的两点之间添加一条边。Martinez<sup>[106]</sup>提出了多种加边策略用以增加代数连通度。这些策略包括随机增加法、最小度节点、特征向量(与Wang和Mieghem相同)、相似度节点和最大度节点等。Kim<sup>[59]</sup>提出了一种基于特征方程的计算有效的二分算法。与以上研

究不同, Sydney 等<sup>[107]</sup>研究了高代数连通度的单边重新布线问题。Ibrahim 等<sup>[60]</sup>将链加法问题转化为半定规划优化问题的算法。

另外, 关于最大化代数连通度的文献还有, 例如: Kim 和 Mesbahi<sup>[108]</sup>提出了加权图上状态依赖图拉普拉斯代数连通度最大化的迭代算法。Ogiwara 等<sup>[109]</sup>研究了最大化或局部最大化代数连通度的图。

### 4.2.2 最大化代数连通度问题研究存在的问题

如前所述, 目前已有大量的文献对最大化代数连通度进行研究, 并取得了显著的成果。但是我们注意到, 一部分算法如<sup>[33, 59]</sup>是基于图的拉普拉斯矩阵的特征值或特征向量, 这将导致但网络规模很大时算法的计算复杂度过高而难以实现。特别是当今社会, 网络发展迅猛, 网络规模动辄几万个节点, 甚至更多, 这种情况下算法的时效性将大大降低。而另一部分算法如<sup>[105, 106]</sup>则是基于图的拓扑结构指标(如: 度、聚类系数等)。虽然算法的时间复杂度较低但同样结果的准确度也低。因此迫切需要研究出具有较低时间复杂度且较高准确度的相关算法。事实上, 代数连通度不能被进一步应用的原因是由于在应用的过程中需要准确计算出它的值, 也就是拉普拉斯矩阵的特征值, 众所周知, 计算高维度的矩阵的特征值的时间复杂度很高。也因此最近学者们也在探索代数连通度的低复杂度计算方法如<sup>[110, 111]</sup>。

综合上述分析, 我们考虑了代数连通度与其他图不变量之间的关系, 并提出了一种启发式算法, 最小度和最大距离(MDMD)算法, 不需要计算特征值(或特征向量)。仿真结果表明, 对于大规模网络, 尤其是大型稀疏网络, MDMD是有效的, 可以获得更短的运行时间。

## 4.3 预备知识

为了便于读者阅读, 我们首先将相关的定义和问题介绍一下。

### 4.3.1 问题定义

一个网络可以用简单的无向图表示。因此, 首先提供与图有关的一些符号和定义。我们用 $G = (V, E)$ 表示一个无向, 无权重的连通图, 其中 $V = \{v_1, v_2, \dots, v_n\}$ 表示有 $|V| = n$ 点的顶点集, 并且 $E$ 表示图的边的集合。令 $v_i v_j$ 表示图的一条边, 即顶点 $v_i$ 和顶点 $v_j$ 相连。如果两个顶点有一条边相连, 我们也说这两个顶点相邻。图的邻接矩阵 $A = [a_{ij}]$ 是一个 $(0, 1)$ 对称矩阵,  $a_{ij} = 1$ 当且仅当顶点 $v_i$ 和 $v_j$ 被一条边所连接。顶点 $v_i$ 的度 *degree* 是指图中所有和它相连接的边的条数。图的拉普拉斯矩阵 (*Laplacian matrix*) 定义为 $L = D - A$ , 其中 $D = \text{diag}(d_1, \dots, d_n)$ 是图的度矩阵, 而 $d_i$ 是顶点 $v_i$ 的度。令 $\mu_1(G) \geq \mu_2(G) \geq \dots \geq \mu_{n-1}(G) \geq \mu_n(G) = 0$ 表示拉普拉斯矩阵 $L$ 的特征值。

其中, 第二小的特征值 $\mu_{n-1}(G)$  定义为代数连通度 (algebraic connectivity)<sup>[57]</sup>, 表示为 $\alpha(G)$ 。与代数连通度对应的特征向量叫做费德勒向量 (Fiedler vector), 用 $X$  表示, 即 $LX = \alpha X$ 。M. Fiedler [22]证明如下等式,

$$\alpha(G) = \min_{X \neq 0, XI=0} \frac{\sum_{v_i v_j \in E} (x_i - x_j)^2}{\sum_{k=1}^n x_k^2}. \quad (4-1)$$

已知图 $G$  的补图 (complement) 用 $G^c = (V, E^c)$ 表示, 补图 $G^c$  与原图 $G$  有着相同的顶点集合 $V$ , 但是在补图 $G^c$  中, 两个顶点相连当且仅当这两点在原图 $G$  中不相邻。我们使用符号 $G + e$  表示在原图 $G$ 的基础上添加一条新边 $e$ 。这样, 最大化代数连通度问题可以描述为以下模型:

$$\begin{aligned} & \text{maximize } \alpha(G + e) \\ & \text{subject to } e \in E^c. \end{aligned} \quad (4-2)$$

代数连通度是衡量图的连通性的一个重要指标。例如: 图 $G$  是连通度当且仅当 $\alpha(G) > 0$ <sup>[57]</sup>。以下定理显示了, 给已知图添加边会使得代数连通度增加。

**定理10.** <sup>[112]</sup>  $\mu_1(G + e) \geq \mu_1(G) \geq \mu_2(G + e) \geq \mu_2(G) \geq \cdots \geq \mu_{n-1}(G + e) \geq \mu_{n-1}(G) \geq \mu_n(G + e) \geq \mu_n(G) = 0$ 。

更多的关于代数连通度的内容请参阅<sup>[98]</sup>。

### 4.3.2 与本章相关的工作

在这一节中, 我们首先回顾一下解决MACP的三种方法。它们是穷举法, 贪心扰动启发式算法<sup>[33]</sup>, 和二分算法<sup>[59]</sup>。这三种算法都是基于拉普拉斯特征值或特征向量的算法。

最简单直接的方法就是穷举法 (brute-force algorithm (BFA)), 这种方法将每一条补图 $G^c$  中的边 $e$  作为候选边, 然后逐条添加到原图 $G$  中, 并计算新图 $G + e$  的代数连通度, 最后选出是代数连通度增加最大的边。显然, 这种方法的计算时间复杂度是最高的, 即使使用对称QR算法<sup>[113]</sup> 来计算代数连通度, 该算法的时间复杂度也要 $O(4mn^3/3)$ , 其中 $m = |E^c|$  且 $n = |V|$ 。

Ghosh 等<sup>[33]</sup> 提出一种贪心扰动启发式算法, 记为Ghosh。令 $X = (x_1, x_2, \cdots, x_n)$  为费德勒向量。若我们考虑归一化的费德勒向量时候, 公式4-1可以化简为4-3

$$\alpha(G) = \min_{X \neq 0, XI=0} \sum_{v_i v_j \in E} (x_i - x_j)^2. \quad (4-3)$$

因此, 当 $\alpha(G)$  单特征值时候,  $(x_i - x_j)^2$  给出了代数连通度 $\alpha(G)$  的增长的一阶近似, 其中 $v_i v_j \in E^c$ 。显然, 除去费德勒向量的一次计算外, 该算法的时间复杂度是 $O(2mn)$ 。不难看出, 该算法的时间复杂度依赖于补图的边数, 因此, 对于原图边数较少, 补图边数较多的稀疏图而言, 该算法的时间消耗是巨大的。

Kim<sup>[59]</sup> 提出了一种二分算法, 记为Kim, 该算法利用了久期方程 (secular equation)<sup>[114]</sup>。此算法首先计算拉普拉斯矩阵的每一个特征值 $\mu_k(G)$  和对应的特征向量 $X_k$ , 其中 $k = 1, 2, \dots, n$ 。然后再划分了区间 $[\mu_{n-1}(G), \mu_{n-2}(G)]$  并且根据久期方程反复选择子区间。其主要利用了图的代数连通度的性质,  $\alpha(G) = \mu_2 \leq \alpha(G + e) = \mu'_2 \leq \mu_3$ , 其中,  $\alpha(G + e) = \mu'_2$  表示图中增加了一条边之后的代数连通度。得到修改后的久期方程如下:

$$f(\mu(G')^*) = 1 + \rho \sum_{k=2}^n \frac{(v_k(i) - v_k(j))^2}{\mu_k - \mu(G')^*} \quad (4-4)$$

其中,  $\mu(G')^*$ 是图 $G$ 加一条边之后的假设代数连通度值,  $\mu_k$ 表示图 $G$ 的拉普拉斯矩阵的第 $k$ 个特征值。4-5显示当图 $G$  添加任意一条边图 $e$  时, 任意 $\mu_k$ 与 $\mu_{k+1}$ 之间的值与图 $G + e$  的第 $k$ 个特征值之间的关系。

$$f(\mu(G')^*) = \begin{cases} > 0 & \mu(G')^* > \mu(G') \\ = 0 & \mu(G')^* = \mu(G') \\ < 0 & \mu(G')^* < \mu(G') \end{cases} \quad (4-5)$$

根据等式4-4和4-5, 我们可以选择一条边, 添加到图中, 然后反复猜测得到的特征值, 并根据方程评估猜测值是否过高、过低或猜测值是否正确。Kim利用这个思路, 提出了对分算法, 该算法从原始图 $G$ 的 $\mu_2$ 到 $\mu_2$ 范围开始。然后, 使用等式4-4和4-5, 将每个候选边分类为生成大于、等于或小于 $\mu_2$ 与 $\mu_2$ 中点的 $\mu'_2$ 。选择加入后使得 $\mu'_2$ 较大的边, 搜索空间减少一半, 仅包括这些边所在的范围。然后重复, 直到搜索空间足够小。最终确定使得代数连通度增加最大的候选边。

由于, Kim的方法在 $\mu_{n-1}(G)$ 与 $\mu_{n-2}(G)$ 之间进行插值计算, 所以该方法不适用与 $\mu_{n-1}(G) = \mu_{n-2}(G)$ 的图。最终通了理论推导和实验验证, Kim 算法的实际时间复杂度约为 $O(5.7mn)$  (详见<sup>[59]</sup>)。

综上所述, 本文选择穷举法BFA作为我们算法有效性的比较, 选择两种已有的启发式算法验证本文算法的效率。其中, BFA算法的估计精度最高但计算代价最大, Ghosh算法的时间复杂度最低。

本章节工作的目的是设计一个有效的算法, 不需要计算特征值 (或特征向量), 以最大限度地提高代数连通度。具体细节将在下一节中描述。

## 4.4 最小度-最大距离 (MDMD) 算法

### 4.4.1 MDMD 算法

在这一节中,我们将介绍了MACP的MDMD算法。不同于<sup>[33]</sup>和<sup>[59]</sup>那样,我们的算法不用计算拉普拉斯矩阵的特征值(或特征向量),而是寻找不相邻的具有最小度的顶点和距离该点最远的顶点,这两个顶点分别记为 $\text{min-degree vertex}$ 和 $\text{max-distance vertex}$ 。

我们策略的第一步是通过选择图 $G$ 中的最小度数顶点来确定 $\text{min-degree vertex}$ 。但是,往往图的最小度顶点不止一个(第二步中需要选择的最高层的顶点也不止一个)。为了进一步缩小选择范围,提高算法有效性,本文中,引入了顶点的可扩展性中心( $EC$ )指标,用来确定 $\text{min-degree vertex}$ 。顶点 $v_i$ 的 $EC$ 定义为与其相邻的顶点的度的总和, $EC$ 可以通过公式(4-6)求得。顶点 $v$ 的 $EC$ 值的意义在于:我们可以通过顶点 $v$ 的邻点来评价它的影响力。令 $M$ 为图 $G$ 的最小度集合,则 $\text{min-degree vertex } v_{i_0}$ 可以通过等式(4-7)找到。如果有多个 $\text{min-degree vertex}$ ,我们将随机选取一个。

$$EC(v_i) = \sum_{v_i v_j \in E} d(v_j). \quad (4-6)$$

$$v_{i_0} = \arg \min EC(v_i), v_i \in M. \quad (4-7)$$

第二步,我们将选取 $\text{max-distance vertex } v_{j_0}$ 。顶点 $v_i$ 和 $v_j$ 之间的距离 $\text{distance}$ 定义为连接顶点 $v_i$ 和 $v_j$ 的最短路径的长度。我们令 $N$ 为距离顶点 $v_{i_0}$ 距离最远的点的集合。显而易见, $N$ 可以由以顶点 $v_{i_0}$ 为根的广度优先搜索确定。因此 $\text{min-degree vertex } v_{j_0}$ 可以通过公式(4-8)找到。同样,如果有多个 $\text{min-degree vertex}$ ,我们将随机选取。

$$v_{j_0} = \arg \min EC(v_j), v_j \in N. \quad (4-8)$$

基于以上分析,提出了MDMD算法如4-1所示:

值得注意的是,MDMD不需要计算拉普拉斯矩阵及其特征值(或特征向量)。时间复杂度很大程度上取决于利用广度优先搜索的步骤4。因此,时间复杂度为 $O(|V|+|E|)$ ,而 $O(|E|)$ 在 $O(1)$ 和 $O(|V|^2)$ 之间变化,其值由图的稀疏程度有关,也就是说当图中边比较少的时候, $O(|E|)$ 也比较小。这意味着,MDMD算法特别适合于大型稀疏网络。

### 4.4.2 MDMD算法的理论分析

首先,我们简单回顾一下本节将用到的相关定义和符号。令 $G = (V, E)$ 为具有 $n$ 个顶点 $v_1, \dots, v_n$ 的图,并且令 $X = (x_1, x_2, \dots, x_n)$ 为图 $G$ 的费德勒向量,其中费德勒

**算法 4-1** 最小度最大距离加边算法 (MDMD)

输入: 图 $G$  的邻接矩阵 $A$ 。

步骤1: 计算顶点 $v_i$  的度 $d_i$ , 其中 $i = 1, \dots, n$ 。

步骤2: 添加具有最小度的顶点到 $M$ 集合中。

步骤3: 根据等式(4-7)在 $M$ 集合中选择 $\min\text{-degree}$ 点 $v_{i_0}$ , 如果有多个 $\min\text{-degree}$ 点, 我们将随机选取一个。

步骤4: 通过广度优先搜索找到距离顶点 $v_{i_0}$ 最远的顶点, 并且将他们添加到集合 $N$ 中。

步骤5: 根据等式(4-8)在 $N$ 集合中选择 $\max\text{-distance}$ 点 $v_{j_0}$ , 如果有多个 $\max\text{-distance}$ 点, 我们将随机选取一个。

输出:  $e = \{v_{i_0}, v_{j_0}\}$ 。

向量的分量 $x_i$  与顶点 $v_i$  相对应, 因此他们的下标相同。令 $x_{i_0}$  为费德勒向量分量中的最大值,  $x_{j_0}$  为最小值。正如4.3.2 中所描述 $(x_i - x_j)^2$  给出了代数连通度 $\alpha(G)$  的给出了增长的一阶近似。换句话说, 等式(4-9) 是模型(5-3)的近似解:

$$e = v_i v_j = \arg \max (x_i - x_j)^2, v_i v_j \in E^c. \quad (4-9)$$

因此, 我们可以考虑图中哪些点对用了费德勒的最大值和最小值。找到最大值和最小值对应的顶点, 两点之间的边, 就是改变代数连通度最大的边。由上一章关于费德勒向量价值的研究结果可以得到, 大部分图是 $\alpha$ -匹配图, 即费德勒向量最大值对应了最小度点。而Fiedler-Depth-First图的研究显示, 具有最小值的顶点距离最大值顶点最远。所以, 我们根据这些性质设计算法是合理的。

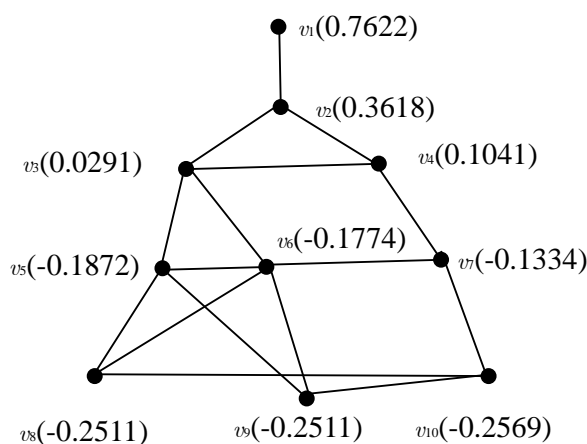


图 4-1: 一个以 $v_1$ 为根的5层的breadth-first graph。费德勒向量为 $X=(0.7622, 0.3618, 0.0291, 0.1041, -0.1872, -0.1774, -0.1334, -0.2511, -0.2511, -0.2569)$ 的图, 且其最高层有多个顶点。

另外, 我们重新看图4-1, 可以发现图中的最小度点不止一个。它们分别为:  $v_8$ ,  $v_9$ 和 $v_{10}$ , 并且通过公式(4-6)求得它们对应的 $EC$ 分别为: 12, 12和9。显然, 最小 $EC$ 值

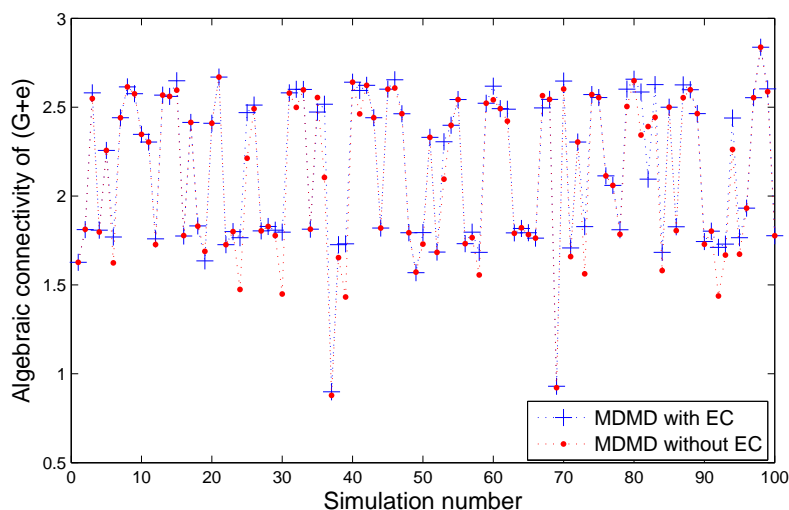


图 4-2: 带有 $EC$ 和不带 $EC$ 的MDMD算法比较, 其中 $p = 0.004$ 且 $|V| = 3,000$ .

对应的顶点 $v_{10}$ 才是我们要选择的具有最大费德勒向量分量的顶点。因此我们采用计算顶点的 $EC$ 值, 并选择最小 $EC$ 值对应的顶点的策略。进一步验证计算 $EC$ 取最小值的策略的有效性, 我们生成100个 $ER$ 图, 比较应用 $EC$ 和不应用 $EC$ , 算法MDMD的差别。如图4-2所示, 结果表明, 大多数带有 $EC$ 的算法结果具有更大的代数连通度值。两个数据集的平均值分别为2.0988和2.0748。这个结果表明, 与不含 $EC$ 的MDMD相比, 带有 $EC$ 的MDMD算法代数连通度提高了约1.157%。同样地, 两个数据集(包括或不包括 $EC$ )的方差分别为0.1747和0.1774。这一结果进一步证实了带 $EC$ 的MDMD算法更稳定。

## 4.5 数值实验

在这一部分中, 我们将MDMD算法与4.3.2中提到的三种算法进行了比较。模拟使用 $ER$ 图和九个自治系统的对等信息图(来自从2001年3月31日至2001年5月26日的俄勒冈路由视图)。所有算法都在MATLAB中实现和测试。

首先, 与<sup>[59]</sup>一样, 我们生成100个具有100个顶点且 $p = 0.7$   $ER$ 图。图4-3显示了四种算法的仿真实验结果。可以看到, 四种算法的计算结果很接近。图4-4显示了四种算法的计算时间比较结果, 结果显示MDMD算法和Ghosh算法时间很接近, 且用时较短, 而另外两种算法耗时较长。

接下来, 我们生成100个具有3000个顶点并且 $p = 0.004$ 的随机图。由于BFA算法和Kim算法需要的计算时间过高, 这里我们仅比较MDMD算法和Ghosh算法。如图4-5和图4-6所示, MDMD计算的结果与Ghosh算法计算的结果相近, 但所用时间要小很多。



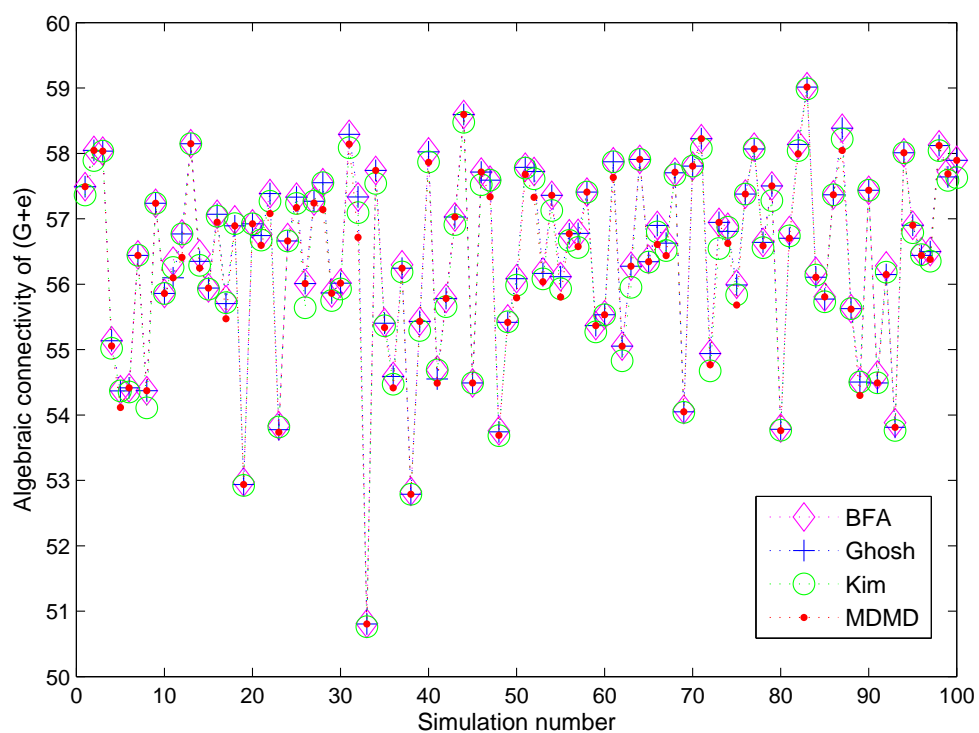


图 4-3: 100个 $p = 0.7$  且 $|V| = 100$ 的ER图的 $\alpha(G + e)$ 的比较结果。

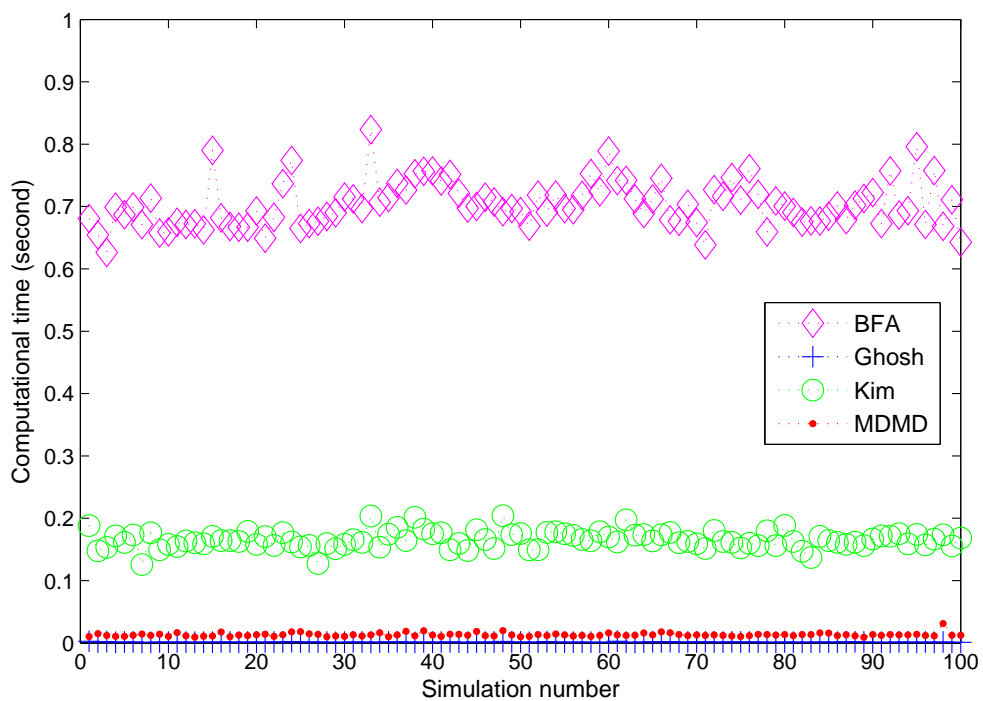


图 4-4: 100个 $p = 0.7$  且 $|V| = 100$ 的ER图的计算时间比较结果。

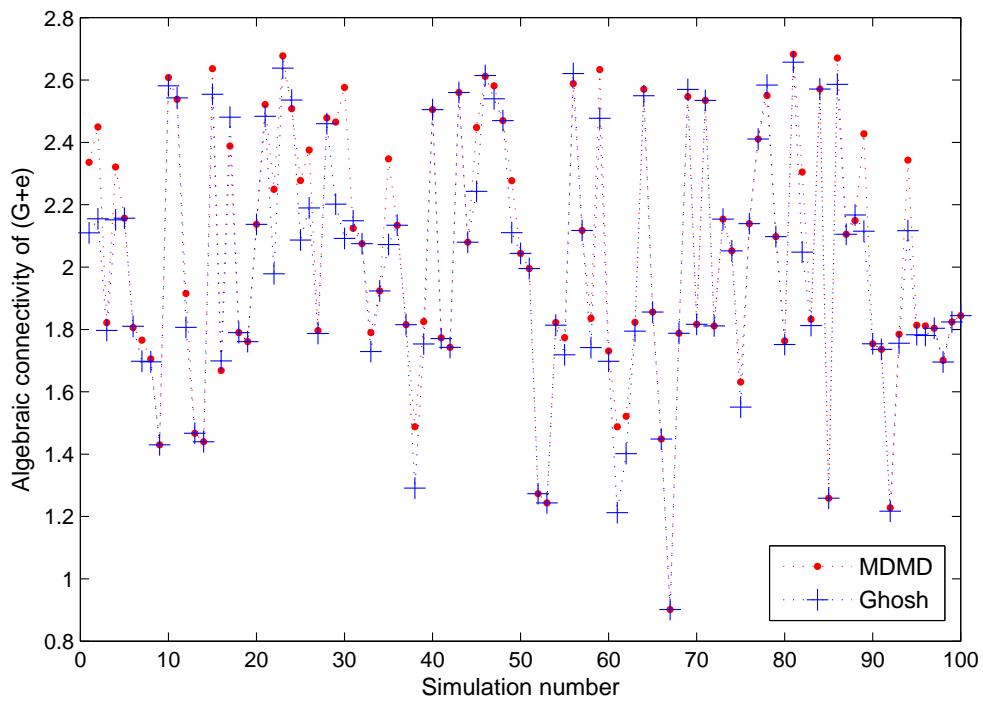


图 4-5: 100个 $p = 0.004$  且 $|V| = 3000$ 的ER图的 $\alpha(G + e)$ 的比较结果。

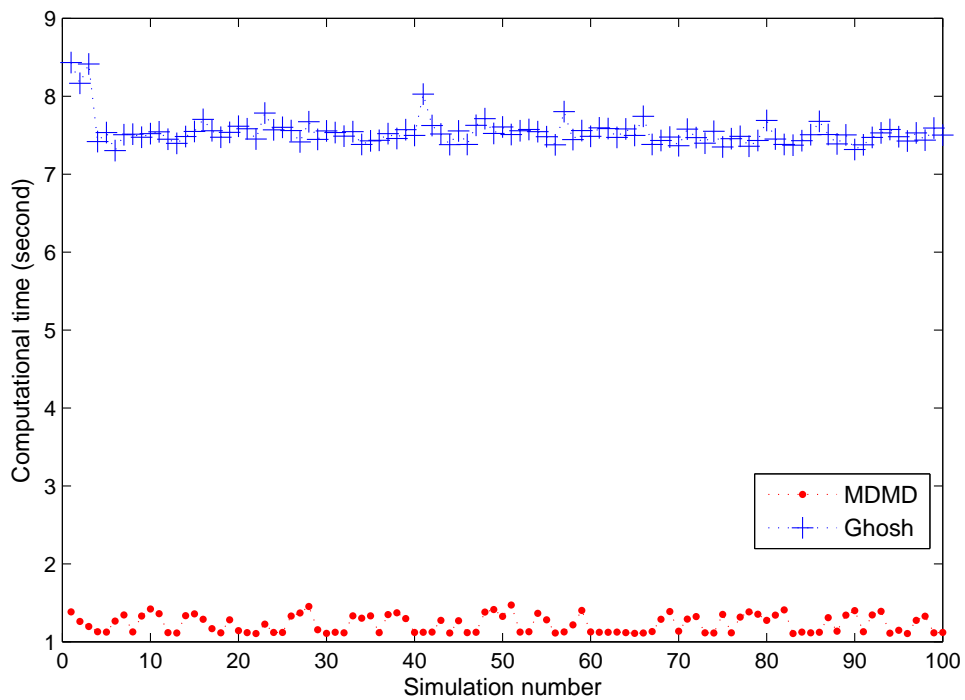


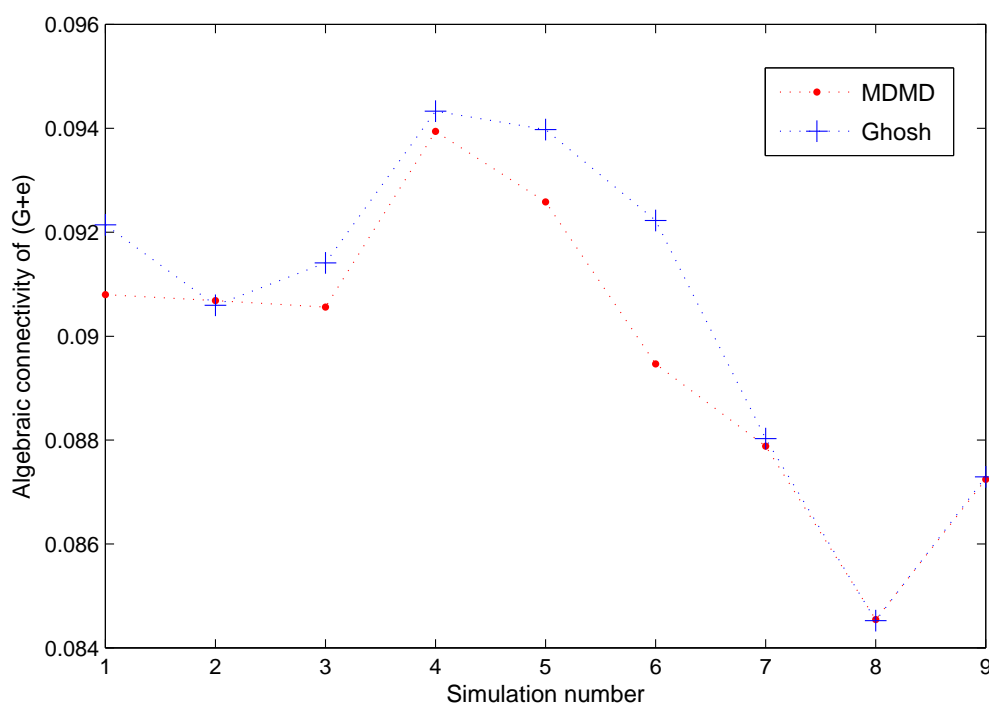
图 4-6: 100个 $p = 0.004$  且 $|V| = 3000$ 的ER图的计算时间比较结果。

最后，我们使用九个自治系统的对等信息图（来自从2001年3月31日至2001年5月26日的俄勒冈路由视图）来进行仿真实验。这些图来自参考文献<sup>[115]</sup>，表格6-4 显示了这些

表 4-1: 既是 *Fiedler-Breadth-First* 图又是  $\alpha$ -匹配图的百分比 (%), 随机图  $G_n(p)$ 

no.	1	2	3	4	5	6	7	8	9
vertices	10,670	10,729	10,790	10,859	10,886	10,943	11,011	11,051	11,174
edges	22,002	21,999	22,469	22,747	22,493	22,607	22,677	22,724	23,409

图的顶点数和边数。如图4-7 和图4-8所示, MDMD 算法准确度上略低于Ghosh 算法, 但是其运行时间相比于Ghosh 算法却少很多。

图 4-7: 9个实例图的  $\alpha(G+e)$  比较结果。

上述实验, 在上一节算法理论分析的基础上, 进一步验证了算法的效率和有效性。

## 4.6 本章小结

在本章中, 我们研究了如何增加链接(边)到网络, 进而最大限度地提高代数连通度, 这是和网络的一致性和鲁棒性密切相关的。代数连通度被定义为图(或网络)拉普拉斯矩阵的第二最小特征值。大多数现有的算法必须直接计算它, 这占用了太多的运行时间。本章的第一个贡献是提出了一种基于最小度和最大距离的启发式算法(MDMD), 它不需要计算拉普拉斯矩阵及其特征值(或特征向量)。仿真结果表明, 在大规模稀疏网络中, MDMD可以获得比其他算法更好的性能。此外, 可以将MDMD应用到添加多条边的问题中, 即通过逐条地增加边来最大化代数连通度。第二个贡献是,

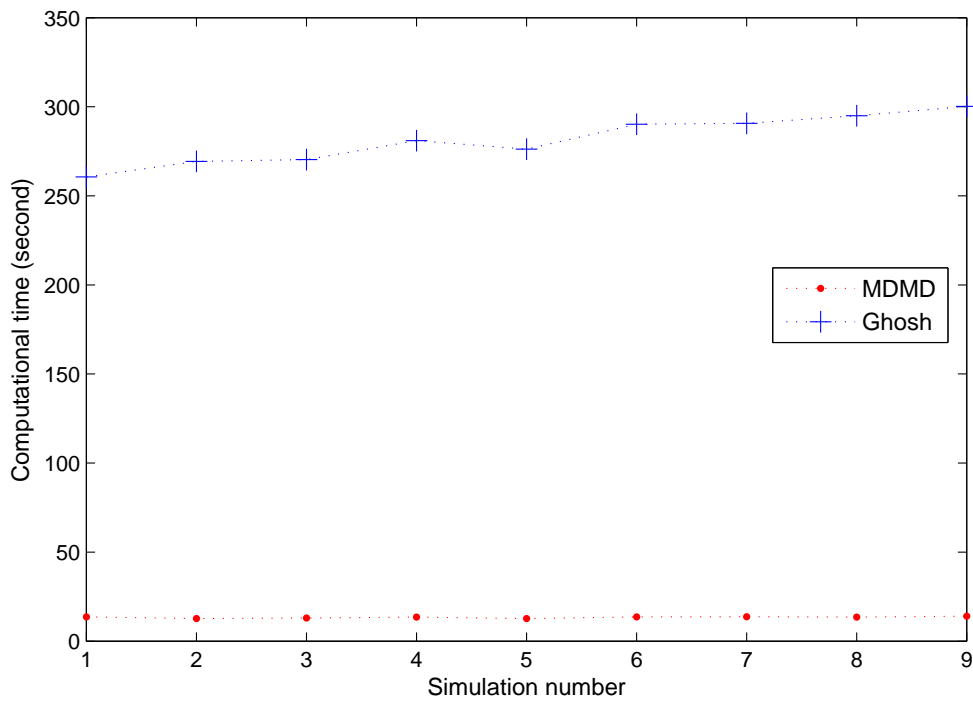


图 4-8: 9个实例图的计算时间比较。

MDMD的有效性表明，根据代数连通度和其他图不变量之间的关系来估计代数连通性的变化是可行的和有效的。因此，对于将来的工作，我们希望通过图不变量（如度、距离等）来估计代数连通性。我们还将继续修改算法在加权图和有向图中运行。

## 第五章 基于代数连通度的最大化网络传输容量的加边算法

本章我们将主要研究如何通过网络加边提高无标度网络的传输效率。在分析图拉普拉斯矩阵第二最小特征值代数连通性与传输容量之间的关系的基础上,提出了一种有效的加边法策略,即最大代数连通性增量边(MACIE)。现有的方法是基于拓扑结构指标,如网络的路径和度,需要昂贵的计算时间,不同于现有的网络加边策略,MACIE通过最大化代数连通性来提高传输效率,从而缩短运行时间。仿真结果表明,与以往的结构孔缩小策略相比,MACIE具有更高的效率和性能。

### 5.1 研究背景和意义

无标度拓扑性质<sup>[94]</sup>在现实的网络中有广泛的应用,例如在全球运输网络<sup>[116]</sup>、无线传感器网络<sup>[36]</sup>、移动网络中<sup>[117]</sup>。计算工具的最新进展和日益增长的传输大数据的需求开辟了许多无标度网络相关的挑战。如何提高网络传输能力,减轻网络拥塞是其中之一。本章讨论了如何提高无标度通信网络的传输效率。

随着通信和交通的快速发展,公共交通网络和计算机网络等实际网络的规模正在急剧增加,如图5-1, 5-2。因此,不可避免地要添加到现有网络的链接。近年来,通过链路增加来优化网络性能已经引起了极大的兴趣。Wang 等人<sup>[67]</sup>提出了一种基于结构空穴理论的链路附加策略,以提高交通容量。Jalili 等人<sup>[118]</sup>提出了一种通过添加新的跨社区链路来增强网络同步性的框架。Chakraborty 等人<sup>[119]</sup>为一个字符串(线性)图提供了锚节点的位置的解析理由,其中所有添加的链接都发生了事件。Zhang 等人<sup>[68]</sup>针对无标度无线传感器网络提出了一种链路增强策略,以提高网络的鲁棒性和业务容量。在本章中,我们专注于如何添加边到无标度通信网络,以提高网络传输效率。一般来说,传输性能取决于路由策略和网络的结构特征。因此,有两种策略可以提高网



图 5-1: 航空网络, 寰宇一家航空网络地图



图 5-2: Facebook网络全球友谊图, Facebook实习工程师保罗·巴勒(Paul Butler)。

络对拥塞的恢复能力。一个是“软”的策略，通过设计有效的路由策略。另一个是改变底层网络结构的“硬”策略。在本文中，我们研究了“硬”策略，当采用最短路径路由策略时，采用链路添加策略。

## 5.2 国内外研究现状及发展动态分析

有许多将边缘添加到网络中的策略。捷联链路策略 (SL)<sup>[34]</sup> 是在现有网络中增加最短路径长度的节点间的快捷链路。低中介中心性优先 (LBF) 策略<sup>[64]</sup> 涉及节点间中心性。另一种策略 (IE)<sup>[65]</sup> 根据最短的最短路径长度和低的中间性来选择添加链路。最近, Bai 等人<sup>[63]</sup> 提出了一种基于节点间距离、链路间距离和最短路径长度的混合链路添加策略。所有这些策略都有效地增加了交通容量,但是由于计算每两个节点之间的最短路径长度而计算成本很高(时间复杂度为 $O(|V|^3)$ )<sup>[34]</sup>)。与上述策略不同, Wang等人<sup>[67]</sup> 提出了一种基于结构空穴理论的减小结构空洞 (RSH) 策略,以提高交通容量。理论上, RSH策略运行在 $O(L^2(L + |E|))$ , 其中 $L$  表示所加的边的条数。以往的方法都是基于拓扑结构指标,如网络的路径和程度等。这促使我们考虑,我们是否可以开发一个有效的方法,基于代数指标,如网络的代数连通性,以提高交通容量。

代数连通度是图拉普拉斯矩阵的第二最小特征值,与网络的性质密切相关,如:鲁棒性<sup>[120]</sup>, 分布式优化算法的收敛速度<sup>[121]</sup>, 以及一致性问题<sup>[101]</sup>。本章的研究思路如5-3所示,我们首先研究交通容量和代数连通性之间的关系。实验证明,代数连通度与传输容量(或点中介中心性)有显著的相关性。然后,我们设计了一个链路增加策略,称为**最大代数连通度增量边** (MACIE),通过最大化代数连通性来增强流量容量。这是从代数图论的角度研究交通容量的新尝试。MACIE策略运行在 $O(L(|V| + |E|))$ ,这意味着它比RSH策略更快。进一步的实验表明,MACIE在交通容量最大化方面也优于RSH。

本章的主要贡献如下。

- 1) 我们研究了代数连通度与传输容量(或中间中心性)之间的相关性,发现它们具有高度相关性。这为研究网络传输效率提供了一个新的视角。
- 2) 提出了一种基于代数连通度的提高传输效率的有效策略。从代数图论的角度研究网络传输效率。

## 5.3 预备知识

在先前的研究中,由Arenas 等人<sup>[62]</sup>提出的有序实数对 $(\eta, R)$ 被广泛的应用于衡量传输容量。

$$\eta(R) = \lim_{t \rightarrow \infty} \frac{\langle \Delta W \rangle}{R \Delta t} \quad (5-1)$$

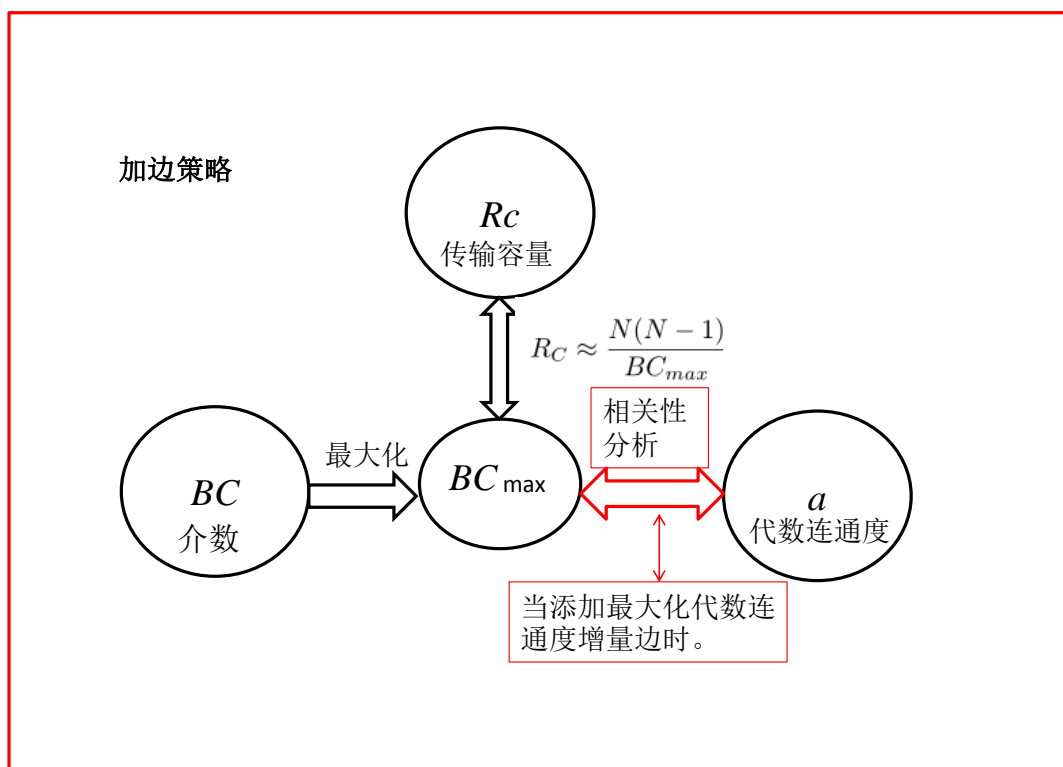


图 5-3: 研究思路。

其中 $\langle \Delta W \rangle = W(t + \Delta t) - W(t)$ ,  $\langle \dots \rangle$  表示 $\Delta t$ 时间内的平均数。 $W(t)$  表示 $t$  时刻网络中的数据包总量。 $R$  是数据包的生成率。表示每一个时刻系统中生成的数据包的个数, 传输容量 $R_c$  是与生成率相关的一个关键值, 用来刻画网络从自由流动状态过度到拥堵状态。当 $R < R_c$ , 可以得到 $\Delta W = 0$  则 $\eta = 0$ 。那么显然网路处于自由流动状态。然而, 当 $R > R_c$ 时, 可以看到,  $\Delta W > 0$  且 $\eta > 0$ 。那么, 单位时间内, 网络中生成的数据包的数量将会大于从网路中删除的数据包的数量, 这就意味着网络中的数据包的数量越来越多, 拥堵就会出现。

我们将在仿真试验中使用Barabási-Albert<sup>[94]</sup> (BA) 无标度网络模型, 用 $G(n, n_0, m_0)$  表示, 其中 $n$ 表示最终的网络规模, 即网络中节点的个数。 $n_0$  是起始网络的节点个数, 并且, 当一个新的节点添加到网络中时, 它将于已有网络中的 $m_0$  个旧节点相连。

接下来, 我们简单回顾一下与本章相关的定义和符号。详细内容参考第二章。设 $G = (V, E)$ 表示一个无向, 无权重的连通图, 其中 $V = \{v_1, v_2, \dots, v_n\}$  是具有 $|V| = n$ 个点的顶点集, 而,  $E$  表示具有 $|E| = m$ 条边 (或连接) 的边集。设 $A$  是图 $G$ 的邻接矩阵。图的拉普拉斯矩阵定义为:  $L = D - A$ 其中 $D$  是图 $G$  的度矩阵。另 $\mu_1(G) \geq \mu_2(G) \geq \dots \geq \mu_{n-1}(G) \geq \mu_n(G) = 0$  表示拉普拉斯矩阵 $L$ 的特征值。则其第二小的特征值 $\mu_{n-1}(G)$  称作代数连通度 (algebraic connectivity)<sup>[57]</sup>, 用 $\alpha(G)$ 来表示。图 $G$  的补

图 $G^c = (V, E^c)$ 是与原图 $G$ 具有相同顶点的图,但补图中两个顶点相邻当且仅当这两个顶点在原图中不相邻。我们使用 $G + e$ 表示,在原图 $G$ 的基础上添加一条新边。

代数连通度是衡量图连通度的一个重要指标。例如,图 $G$ 是连通的当且仅当 $\alpha(G) > 0$  [57]。以下定理刻画了:图中添加一条边将使得代数连通度增加。

**定理11.** [112]  $\mu_1(G + e) \geq \mu_1(G) \geq \mu_2(G + e) \geq \mu_2(G) \geq \cdots \geq \mu_{n-1}(G + e) \geq \mu_{n-1}(G) \geq \mu_n(G + e) \geq \mu_n(G) = 0$ .

中介中心性 (Betweenness Centrality (BC)) [83]是用来衡量网络中一个点的中心性的指标,其别用与各种网络系统中。顶点的中介中心性计数了通过它的最短路径的分数。它被定义为:

$$BC(v_j) = \sum_{v_i, v_k} \frac{\delta_{(v_i, v_k)}(v_j)}{\delta_{(v_i, v_k)}}, \quad (5-2)$$

其中 $BC(v_j)$ 表示节点 $v_j$ 的中介中心性,  $v_i, v_j, v_k \in V$ ,  $\delta_{(v_i, v_k)}(v_j)$ 表示节点 $v_i$ 到节点 $v_k$ 的最短路径中通过节点 $v_j$ 的总条数,  $\delta_{(v_i, v_k)}$ 表示节点 $v_i$ 到节点 $v_k$ 的最短路径总条数。

## 5.4 与本章相关的工作

### 5.4.1 软方法

软方法是针对网络数据包的流动,设计不同的路由策略,在每个顶点上引导数据包向目的地转发。硬方法主要是关注网络结构的变化,引起传输容量的变化。本章中,主要研究在一定的路由策略下,如何添加边,来增加网络的传输容量。我们将基于有效路由策略和最短路径路由策略下进行对比试验。

**最短路径路由策略:**提高传输容量的经典软方法,在该路由策略下,数据包沿着起点到终点之间的最短路径移动。该方法是以数据包尽快到达目的地为目标。每个数据包生成之后,选择离目的地最短的路径方向移动。但也因此,网络中通过最短路径多的顶点更容易发生阻塞。所以学者们常常用网络中顶点的最大中介中心性来估计传输容量。

**有效路由策略**该策略是一种经典的提高传输容量的软方法,由uG. Yan等人 [122]提出的一种提高复杂网络运输效率的路由策略。该策略不使用最短路径的路由策略,而是给出一种广义的路由算法来寻找有效路径。考虑到网络中实际路径上,最大度的节点非常容易引起交通拥挤,因此,改善交通拥挤和控制交通拥挤的一种有效方法是将中心节点的交通负荷重新分配给其他非中心节点。直观上来讲,数据在传输时,寻找有效的路径不是以到达目的地的最短路径为标准,而是以到达目的地的过程中,经过的所有节点的度和最小为标准。



### 5.4.2 减小结构洞策略 (RSH)

Burt<sup>[91]</sup>引入了结构洞的概念来分析社会问题。在社会网络中,结构洞较多的个体控制和处理信息较多,具有一定的位置优势。在图(或网络)中,顶点的结构洞是相邻顶点之间断开的顶点对。Wang等人<sup>[67]</sup>指出,网络中结构洞数目较大的顶点更倾向于有较多的最短路径通过,因此,减小结构洞数目,可以最大可能的影响网络中的最短路径长度。因此,RSH策略是找到网络中具有最大结构洞数目的顶点,然后在其邻点中加边。理论上,RSH计算时间复杂度为 $O(L^2 * (L + |E|))$ ,其中 $L$ 是添加边的数目, $|E|$ 是原网络中边的数目。

## 5.5 MACIE算法

如我们所知,向图中添加边可以增加代数连通度。它启发我们可以通过增加代数连通度来提高交通容量。为了尽可能地增加连通性,我们首先定义最大代数连通性问题:

$$\begin{aligned} & \text{maximize } \alpha(G + e) \\ & \text{subject to } e \in E^c. \end{aligned} \quad (5-3)$$

那么,我们可以定义**最大化代数连通度增量**(*maximum algebraic connectivity increment* (MACI))边。

**定义9.** 设 $G = (V, E)$ 是无向,无权重的连通图,其补图为 $G^c = (V, E^c)$ . 如果 $e \in E^c$ 是问题(5-3)的解那么边 $e$ 称作图 $G$ 的**最大代数连通度增量边**。

本章中,我们应用最小度最大距离算法(MDMD)<sup>[123]</sup>来寻找最大代数连通度增量(MACI)边。MDMD是一个时间复杂度为 $O(|V| + |E|)$ 的近似算法。

另一方面,当数据包从起点向目的地以最短路径为路由策略传输时,下面的表达式<sup>[95] [66]</sup>可以被用于理论估计传输容量。

$$R_C \approx \frac{n(n-1)}{BC_{max}}, \quad (5-4)$$

其中, $n$ 是网络中节点的个数, $BC_{max} = \max(BC(v)), v \in V$ 。

因此,我们考虑当MACI边被加入时,代数连通度与顶点的最大点中介中心性之间的关系。我们首先利用BA网络模型进行实验,并生成一个 $G(600, 5, 3)$ 图。令 $L$ 表示向网络中添加的边的总条数, $m$ 表示原始网络中的边的条数。我们定义分式 $f_r = L/m$ 用以表示添加的边的量。这里设 $f_r = 0.5$ 。在实验中,被选择MACI边逐一增加。然后,

计算了代数连通度和最大点中介中心性。如图5-4所示，随着MACI边的增加（代数连通度的增加），点中介中心性明显减小。

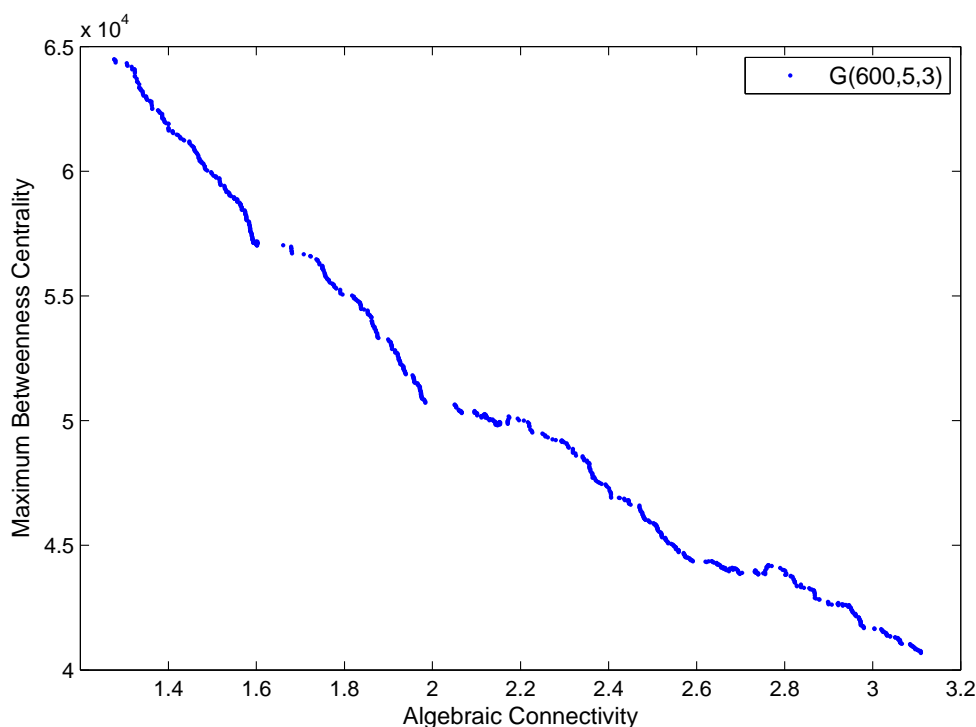


图 5-4: 当最大代数连通度增量边逐条添加的时候，代数连通度与最大点中介中心性的关系。

进一步，我们生成三种图 $G(600, 5, 2)$ ， $G(800, 5, 3)$ 和 $G(1000, 5, 4)$ ，来研究代数连通度与最大点中介中心性之间的斯皮尔曼等级相关系数（Spearman's rank correlation coefficient）和皮尔森相关系数(Pearson correlation coefficient)，并且为了更具有代表性，每种图我们生成100个个例。在实验中MACI边被逐条添加，且设 $f_r = 0.5$ 。如图5-5和图5-6所示，最大点中介中心性与代数连通度高度相关。换句话说，增加MACI边可以有效地降低最大点中介中心性，因此，根据最大点结数与传输容量的关系5-4，也可以用来调高网络的传输容量。

基于以上分析，我们将最大代数连通度增量边算法描述如下：算法5-1。

#### 算法 5-1 MACIE

输入:  $(A, L)$ ，其中 $A$ 是图 $G$ 的邻接矩阵， $L$ 是需要添加的边的条数。

步骤1: 初始化 $i = 1$ 。

步骤2: 查找最大代数连通度增量边 $e_i$ ， $1 \leq i \leq L$ 。

步骤3: 更新 $G = G + e_i$ 和 $i = i + 1$ 。

步骤4: 如果 $i \leq L$  转到步骤2。

输出: 添加 $L$ 条边后的图 $G'$ 的邻接矩阵 $A'$ 。

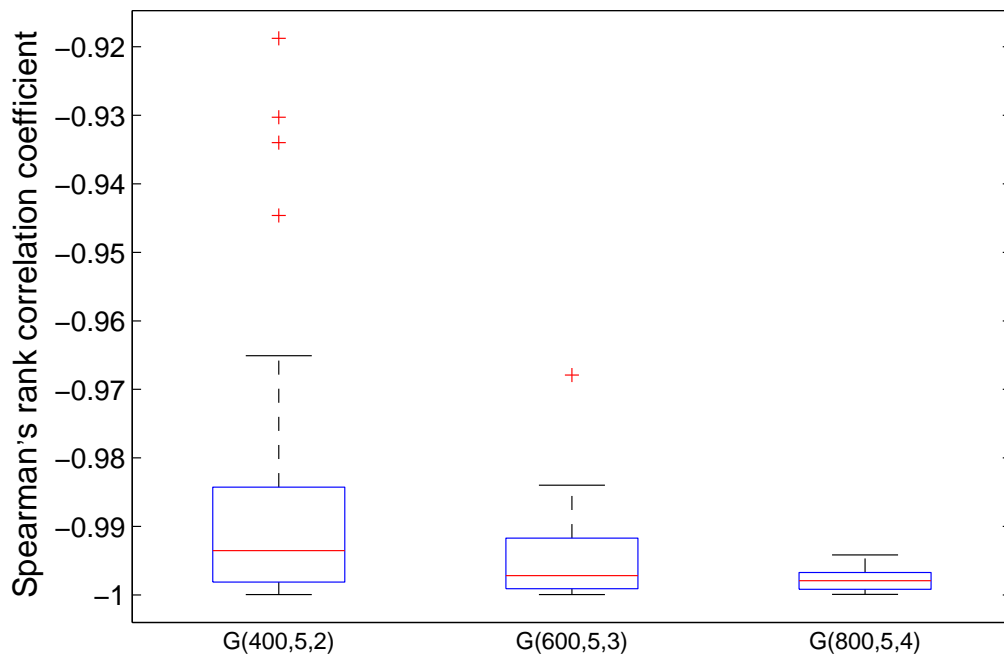


图 5-5: 代数连通度与最大点中介中心性斯皮尔曼等级相关系数的盒形图。有100个重复值。

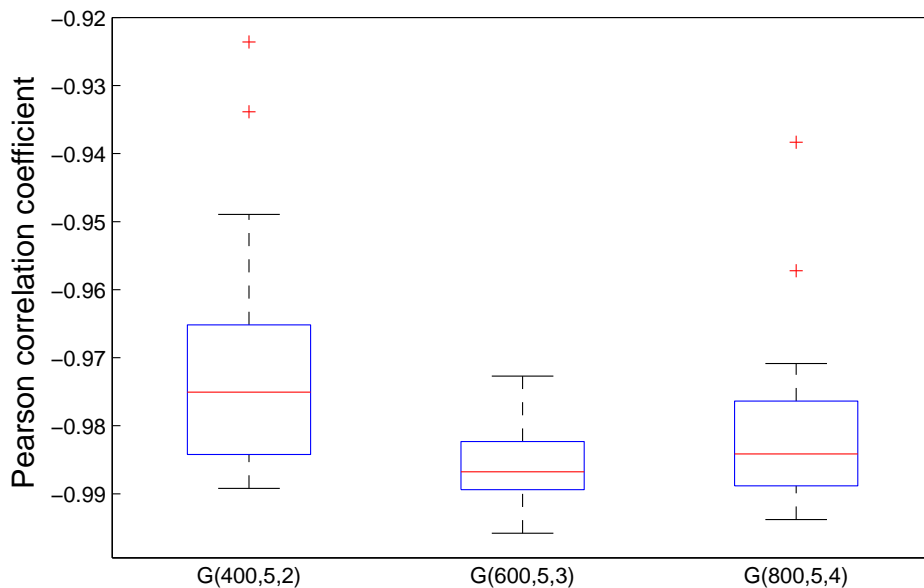


图 5-6: 代数连通度与最大点中介中心性皮尔森相关系数的盒形图。有100个重复值。

注意到MACIE算法的时间复杂度主要取决于在第4步中，MDMD算法被用于求出最大代数连通度增量边。而MDMD算法的时间复杂度是 $O(|V| + |E|)$ 。因此整

个MACIE算法的时间复杂度为 $O(L(|V| + |E|))$ ，其中 $L$  是添加的边的条数。

## 5.6 数值实验

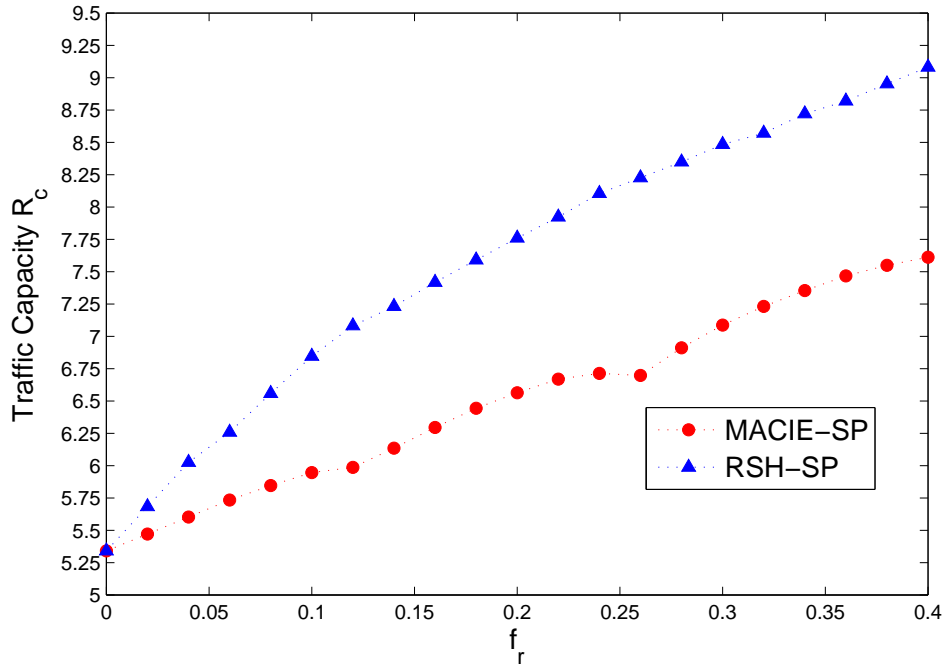


图 5-7: 网络传输容量 $R_c$  v.s.  $f_r$ , 其中, 应用最短路径路由策略。BA网络参数为 $N = 400$ ,  $n_0 = 5$ , and  $m_0 = 3$ .

在这一部分中, 我们给出在无标度网络上的MACIE策略和RSH策略 [67]的仿真结果。我们使用最短路径路由 (shortest path (SP)) 策略和有效路由 (efficient routing (ER)) 策略。BA 网络参数设为 $G(400, 5, 3)$ 。公平起见, 每种图生成100个, 最终取平均值。图5-7和图5-8 刻画了不同 $f_r$  的值下MACIE 策略和RSH 策略得出的传输容量 $R_c$ 。对比两种策略, 例如当 $f_r = 0.4$ , 我们可以看到, 对于最短路径路由策略, MACIE的 $R_c$ 小于RSH策略, 分别为7.612 和9.815。但是, 应用有效路由策略时, MACIE策略的 $R_c$ 是69.089, 比RSH策略的24.209增长了168.863%。实际上, 在最短路径路由策略下, 数据包总是沿着源和目标之间的最短路径移动, 并且总是在有效的路由策略下沿着具有最小的顶点度数之和的路径进行路由。如上所述, RSH旨在减少最大结构洞。显然, 减少顶点的结构洞也可以减少通过顶点的最短路径的数量。因此, RSH对最短路径路由策略更为敏感。MACIE尽可能增加网络连通性, 显然, 它可以提高传输容量。此外, 我们还研究了MDMD算法, 即最大代数连通性的网络加边算法, 发现它在最小度顶点和最远顶点之间添加边。MDMD 算法的边添加策略使更多路径能够通过最小度或次最小度点。而最小度的点直观上来讲, 会有较少的最短路径通过, 也意味着较少发生拥堵。因此, MACIE比RSH 更有效。

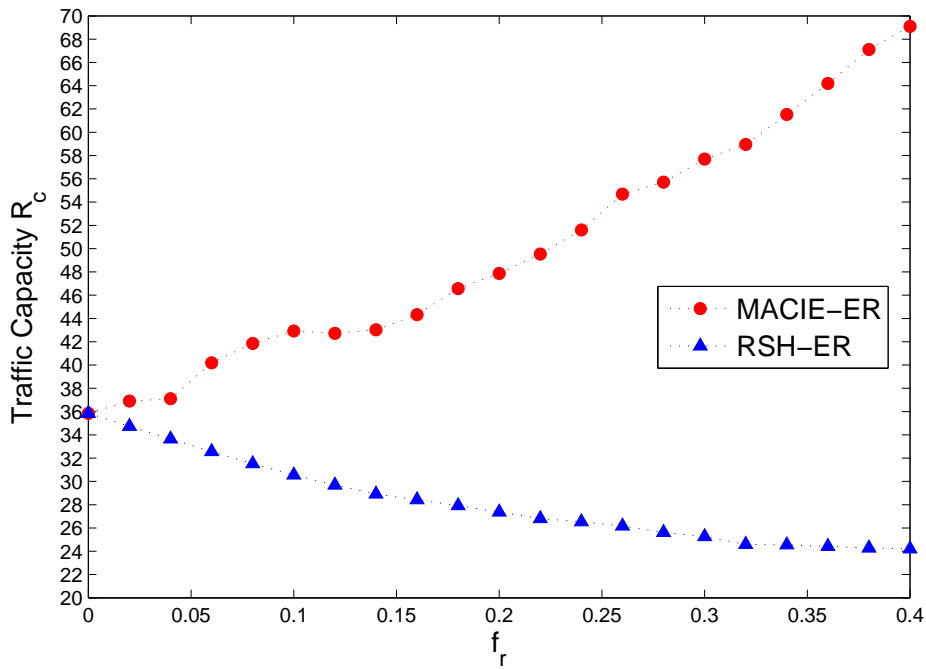


图 5-8: 网络传输容量 $R_c$  v.s.  $f_r$ , 其中, 应用有效路由策略。BA网络参数为 $N = 400$ ,  $n_0 = 5$ , and  $m_0 = 3$ 。

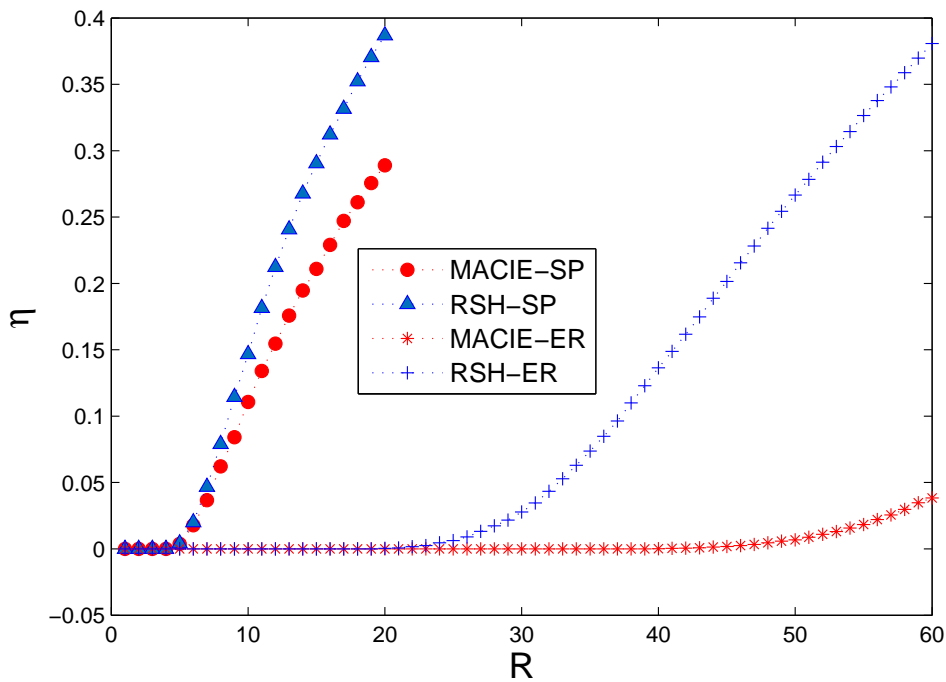


图 5-9: 有序参数 $\eta$ 与数据包生成率 $R$ , 最短路径策略和有效策略下的RSH 与MACIE比较。BA网络参数为 $N = 400$ ,  $n_0 = 5$ ,  $m_0 = 3$ , and  $f_r = 0.2$ 。

接下来, 我们设 $f_r = 0.2$ 并且生成一个BA网络 $G(400, 5, 3)$ 来评估不同 $R$ 时的 $\eta(R)$ 值。如图5-9所示, 当应用最短路由策略的时候, 两种策略的 $R_c$ 值都是5。但是对于有效路由

策略，MACIE策略增长了96%，从25增长到49。这个结果显示，与RSH策略相比，我们提出的MACIE策略加边后的网络传输容量要更大。

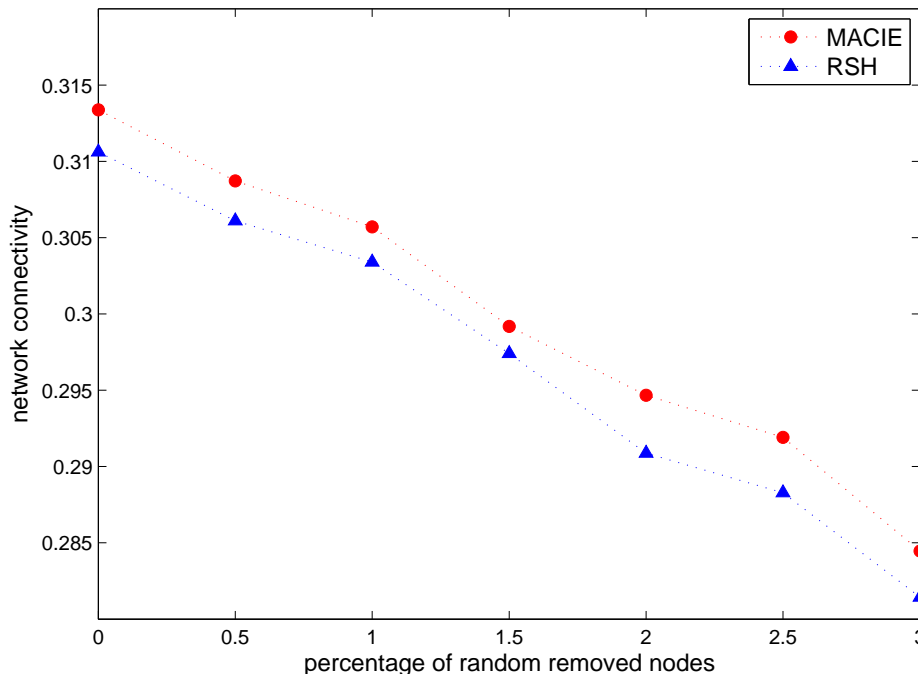


图 5-10:  $BA$ 无标度网络的误差和攻击容忍度,  $BA$ 网络参数设为 $N = 1000$ ,  $n_0 = 5$ , 且 $m_0 = 3$ 。

最后，我们比较MACIE策略和RSH策略对网络鲁棒性的影响。正如文章<sup>[124]</sup>所述，指标**网络连通性**（network connectivity ( $NC$ ））可以用于评价 $BA$ 网络在随机攻击下节点失效的容错能力。并且较大的 $NC$ 值意味着网络有较高的节点失效的容错能力。 $NC$ 定义如下：

$$NC = \frac{1}{n(n-1)} \sum_{i \neq j} \frac{1}{d_{ij}} \quad (5-5)$$

其中 $d_{ij}$ 表示从顶点 $i$ 到顶点 $j$ 的最短路径长度， $n$ 表示网络中节点的个数。本实验中，我们比较在随机节点攻击下， $BA$ 网络 $G(1000, 5, 3)$ 的 $NC$ 的值。我们设置 $f_r = 0.2$ 。如图5-10所示，在MACIE策略和RSH策略下， $NC$ 值都随着节点的随机移除而减小。然而，MACIE策略得到较大的 $NC$ 值。因此，相比于RSH策略，MACIE策略可以得到更高鲁棒性的网络。

综上所述，上述仿真结果表明，MACIE在提高网络传输效率方面优于RSH，在随机攻击下保持较高的网络鲁棒性。

## 5.7 本章小结

随着网络数据的爆炸性增长,提高网络传输能力和减少拥塞已成为一个热点问题。与此同时,网络的规模也在不断扩大。其中一个表现是网络边的生长。因此,本章中重点讨论了如何在网络中添加边以增强其数据的传输容量。代数连通度是图拉普拉斯矩阵的第二小特征值,它与网络性能密切相关。本章分析了传输容量与代数连通性之间的相关性,得出当最大代数连通性增量边逐个加入网络时,代数连通度与传输容量具有显著的正相关关系。然后提出了一种边添加策略,称为最大代数连通度增量边(MACIE),通过最大化代数连通度来提高传输容量。这是从代数图论的角度来研究传输能力的一种尝试。仿真结果表明,MACIE策略是一种有效的策略,其性能优于RSH策略。

## 第六章 基于结构洞理论的最小化平均最短路径的加边算法

### 6.1 研究背景和意义

在过去的十年中,网络数据呈现了爆炸式增长。这种网络化数据的一个重要特性是小世界 (Small world) 现象:即使在大尺寸的网络中,两个节点 (vertex or node) 之间的平均最短路径长度 (Average shortest path length) 也很小。其中,路径长度定义为连接两点之间的所有路径中,最短的边 (edge or link) 的数目。两点之间的最短路径是一个非常理想的属性,因为它能够推进和放大网络中的动态过程,例如信息传播或病毒式营销,信息可以通过两点之间的最短路径,快速的从起点,传播到终点。因此,一个连接良好的网络,对网络数据的用户和所有者都是有价值的。

同样,在对于现实中的网络数据传输问题,从起点发送到目的地的任何消息都通过中间顶点传播。如果必须经过许多中间节点,则数据包会导致端点到端点的延迟。由于任何一对节点都可以充当起点-终点的对,所以希望网络中的所有路径尽可能少地进入中间节点。换言之,如果网络中的平均最短路径长度较低,网络中的任何业务将引起更少的平均延迟。通常,这些图的边被认为是加权的。权重可以根据网络的上下文指示建立边、延迟或其他因素的成本。因此平均路径长度在衡量网络工程的具体性能方面起着重要的作用,优化它们有助于提高网络性能。因此优化网络的平均最短路径长度在结构优化领域也引起了广泛的关注。

这一章,主要关注的是对于一个网络,如何向其添加多条边,才能够使得网络的平均最短路径长度最小,即最小化平均路径长度问题 (Minimizing Average Shortest Path Length)。正如N. Parotsidis等人在文章<sup>[79]</sup>中提到的:最小化平均路径长度有广泛的应用意义,比如,社会网络 (Social network) 中,不断有新的朋友关系产生;或者罪犯关系网中,新的关联关系不断形成,新的链接边产生了,人们更关注哪些边添加之后对整个网络的影响最大,会使网络中的点的关系更加紧密。

另外,通过调整节点或边来降低平均最短路径长度,可以有效地提高传输效率和同步能力<sup>[77, 78, 125]</sup>。Chang 等人<sup>[81, 126]</sup>考虑了多核处理器设计中增加射频链路的问题。他们的目的是减少多核体系结构中的通信延迟,通过在诸如网状网络之类的常规互连方案之上添加额外的射频链路。然而,射频互连需要更多的面积,不能取代传统的互连,添加单个RF互连可以显著提高延迟。通常,较低的平均最短路径长度降低了网络中的整体通信成本,因此,最小化网络的平均最短路径长度,使网络呈现小世界的特性,



具有显著的应用价值。

## 6.2 国内外研究现状及发展动态分析

### 6.2.1 问题的研究现状与成果

Myelson和Taigiu<sup>[80]</sup>已经证明,在所有顶点对之间增加 $k$ 条边以最小化加权平均最短路径长度的一般问题是NP-hard,并且设计多种添边策略,以最小化每对顶点之间的平均最短路径长度。Ogras和Marculescu<sup>[127]</sup>考虑了在常规网络上设计用于VLSI设计的高效片上网络(networks-on-chip)的长距离边。在这种情况下,长边的长度是链接上的权重。Pickavet和Demeester<sup>[128]</sup>在同步数字体系(SDH)网络中设计了用于边恢复的启发式算法。他们的四阶段算法的关键部分之一是局部优化,他们试图通过增加额外的边来改善备用容量的拓扑结构。在这种情况下,边上的权重是备用容量。Gaur等人<sup>[38]</sup>研究了将任意未加权和无向网络转换为小世界网络的几种确定性边附加策略。他们发现最好的策略是在任意网络中增加一个额外的(长)边,以最小化网络的平均路径长度。

N. Parotsidis等人在文章<sup>[79]</sup>研究的问题是从一组候选边的 $k$ 边子集的问题,其添加到网络导致其平均最短路径长度的最大减少。他们将问题作为组合优化问题,并证明它是NP-hard的,并且已知的近似技术是不适用的。他们描述了一种有效的方法来计算单个边缘插入对平均最短路径长度的精确效果,以及用于有效估计这种效应的几种启发式算法。Haifeng Du等人<sup>[75]</sup>分析了网络直径与平均路径长度的关系,将平均路径长度优化问题转化为直径减小到2的问题,提出了一种基于模因算法的数学模型。Andrew Gozzard等人<sup>[129]</sup>提出了两个有效的算法引入一个新的边最小化平均路径长度的问题。在最坏的情况下,第一个算法在 $O(V^4)$ 中运行,并且在最坏的情况下,第二算法运行在 $O(V^3 \log V)$ 时间内。

综上所述,最小化平均最短路径的加边算法在社交网络分析,网络数据挖掘等方向都有广泛的应用,也取得了一系列的研究成果,但仍然有一些问题值得深入研究。例如,部分准确度较高的算法<sup>[129]</sup>,计算时间复杂度过高,无法适应实际大规模网络的应用。部分近似算法<sup>[38] [128] [79]</sup>,也因为需要计算网络中的所有路径,导致计算时间复杂度居高不下。因此,面对当今社会的大规模动辄十几万规模的网络数据来说,急需低复杂度高准确率的算法。另一方面,目前已有的大部分算法都需要整个网络的全局信息,这个在实际应用的时候是很难获得的。因此基于局部信息的算法是有待解决的问题。本章将利用社会学中结构洞理论,结合图论算法,提出一种基于结构洞的强强联合的最小化网络平均最短路径的算法。社会学理论可知,网络中具有较大的结构洞的顶点意味着掌握着更多的资源,在网络中具有更重要的地位,我们提出的算法在网

路中具有第一大和第二大的结构洞的两个非邻顶点之间添加边。因此是一种强强联合的加边策略。该算法只需网络中的局部信息，即顶点的结构洞数，因而，算法的计算时间复杂度较低，而且在实际应用中极具价值。

### 6.2.2 预备知识

我们简单回顾一下与本章相关的定义和符号。详细内容参考第二章。设  $G = (V, E)$  表示一个无向，无权重的连通图，其中  $V = \{v_1, v_2, \dots, v_n\}$  是具有  $|V| = n$  个点的顶点集，而  $E$  表示具有  $|E| = m$  条边的集合。平均最短路径长度（APL）定义如下：

$$APL = L(G) = \frac{\sum_{i \neq j} d_{ij}}{n(n-1)} \quad (6-1)$$

其中  $d_{ij}$  表示从顶点  $i$  到顶点  $j$  的最短路径长度， $n$  表示网络中节点的个数。我们先介绍一下常用的网络最短路径算法，这样读者可以更好的理解网络中的最短路径的计算过程和难点，具体有如下几种：

单源最短路径（single-source shortest path（SSSP））算法：作为前置条件，这种算法需要给定一个有权有方向的图  $G$ ，以及一个称为源点的顶点  $v$ ，该算法用于计算给定一个顶点和其他所有顶点之间的最短路径长度，这里的长度指路径上所有边的权重之和。例如著名的迪杰斯特拉（Dijkstra）算法、Bellman-Ford算法、SPFA算法等，以迪杰斯特拉（Dijkstra）算法为例，该算法的基本思路是以给定点为核心点，逐渐向外层扩展，搜索，直到不能扩展为止，在逐层的扩展过程中，每个点到给定点的最短路径被记录下来。扩展完成，则所有到给定点的最短路径即全部找到。限于本章讨论内容只是无权，无向的简单图，可以将单源最短路径算法等同于以给定点为根点的广度优先搜索，显然，此时，算法的时间复杂度为  $O(m)$ ，其中， $m$  表示图中边的数目。

所有节点之间的最短路径（all-pairs shortest path（APSP））算法：该算法用于计算图中所有点对之间的最短路径。可以通过以每个点作为源点使用Dijkstra等单源最短路径算法来求得图中所有点对之间的最短路径。也可以用经典的Floyd-Warshall算法，该算法简洁，清晰，使用了中继点的概念。算法循环  $n$ （图的顶点个数）次，每次都以一个顶点  $k$  为中继点，然后对于图中每一对顶点  $u$  和  $v$ ，计算他们到顶点  $k$  的距离之和，如果小于顶点对  $u$  和  $v$  之间的距离，则更新顶点对  $u$  和  $v$  之间的最短路径。更多的最短路径的算法读者可以参考<sup>[130]</sup>。

接下来，我们介绍一下本章讨论的核心问题：最小化平均路径长度的加边问题。图中有新的边添加，相关的指标也会变化。假设向已知图  $G$  添加一条边  $e$  且  $e \notin E$ ，构成

新图  $G + e$ ，从公式6-2 可以得到：

$$APL' = L(G + e) = \frac{\sum_{i \neq j} d'_{ij}}{n(n-1)} \quad (6-2)$$

其中， $d'_{ij}$ 表示新图  $G + e$  中，顶点  $i$  和顶点  $j$  之间的最短距离。当顶点个数不发生改变的时候， $L(G + e)$  的值只与  $d'_{ij}$  有关，而容易证明所有通过不相邻的两个顶点  $i$  和顶点  $j$  的最短路径上的点对之间的距离，都会在连接顶点  $i$  和顶点  $j$  之后减小（详细参考本章中的定理1）。因此，得到结论：当图中有一条新边加入的时候，整个图的平均最短路径长度将减小。

令  $C$  表示图  $G$  的候选边集，即  $C = \{e : e \notin E\}$ 。则最小化平均最短路径长度问题就是找到  $k$  条边的  $C$  的子集  $S \subseteq C$ ，且  $|S| = k$ ，使得  $L(G + S)$  最小，其中  $G + S$  表示图  $G$  中添加  $S$  中的边之后的新图。模型如下所示：

$$\begin{aligned} & \text{maximize } L(G + S) \\ & \text{subject to } S \in C, |S| = k. \end{aligned} \quad (6-3)$$

Myelson和Taigiu<sup>[80]</sup>已经证明最小化平均最短路径长度问题是NP-hard。因此人们设计了多种近似算法解决这个问题，以下我们将列举多种最小化平均路径长度策略。

## 6.3 相关工作

### （1）最大接近中心性差异（MaxCCD）算法

顶点的接近中心性（Closeness Centrality）测量节点与网络中其余节点的接近程度。它的值可以用通过该节点所有最短路径长度之和的倒数来表示，如公式6-4所示。

$$CC(v_i) = \frac{N-1}{\sum_1^N d(v_i, v_j)} \quad (6-4)$$

其中， $N$ 表示网络节点的个数。 $d(v_i, v_j)$ 表示顶点  $v_i$  和  $v_j$  之间的最短路径长度。顶点之间的接近中心性差异（Closeness Centrality Disparity）定义为两个顶点的接近中心性之差。Gaur等人<sup>[38]</sup>研究了将任意未加权和无向网络转换为小世界网络的几种确定性链路附加策略。他们发现最好的策略是在任意网络中增加一个额外的（长）链路，以最小化网络的平均路径长度。他们提出了一种叫做最大接近中心性差异（MaxCCD）的算法，有效的减小平均最短路径。在MaxCCD加边策略中，新的边添加到具有最大中心性差异的两个点之间。作者认为，网络中具有最大接近中心性的顶点意味着有更多的最短路径经过，而最小的接近中心性的顶点有最少的最短路径经过，这样，连

接这样的两个点，可以有效的减小网络的平均最短路径。与其他算法相比较取得了良好的性能。用前文我们提到的单源最短路径算法计算图中所有顶点的 $CC$ 的时间复杂度为 $O(N^2 \times \log N)^{[131]}$ 。而计算出所有顶点的 $CC$ 值之后，计算最大或最小的 $CC$ 值的顶点的时间复杂度只需要 $O(N)$ 。因此最大接近中心差异算法的计算时间复杂度为 $O(N^2 \times \log N)$ 。

该算法完全依赖于经典的单源最短路径算法，简单理解就是将网络中的所有点为源点的最短路径全部计算出来，然后选择最大接近中心性的顶点和最小中心性的顶点，添加边。因此，需要网络的全局信息，来计算源点的所有最短路径。所以，对于大规模实际网络，部分信息不明确的时候，该算法是受限的。

### (2) 边影响力 (EdgeEffect) 算法

对于图 $G = (V, E)$ ，令 $e = (x, y)$ 为可添加的候选边， $d_{old}(u, v)$ 表示图 $G$ 中，添加边 $e$ 之前顶点 $u$ 与顶点 $v$ 之间的路径长度。定义 $A_x = \{u \in V : d_{old}(u, y) + 1 < d_{old}(u, x)\}$ ， $A_y = \{v \in V : d_{old}(v, x) + 1 < d_{old}(v, y)\}$ ，显然，任意顶点 $v \in A_y$ ，其到顶点 $x$ 的距离要比到顶点 $y$ 的距离小于2以上。这意味着，当边 $e$ 添加之后，顶点 $v$ 到顶点 $y$ 的距离将会减少。也就是说集合 $A_x$ 与集合 $A_y$ 表示，新的一条边 $e$ 添加之后，路径长度受到影响的顶点的集合。集合 $A_x$  ( $A_y$ ) 可以通过以顶点 $x$  ( $y$ ) 为根的广度优先搜索求得。基于集合 $A_x$ 与集合 $A_y$ ，可以计算出，当一条新的候选边 $e = (x, y)$ 添加之后，图中所有点之间引起的最短路径的减小值 $R_G(e)$ 。显然，可以根据候选边的 $R_G(e)$ 值计算出改变添加后图中最短路径的减少值，当所有的候选边的减小值都计算出来时，最佳的候选边也就确定了。基于此，N.Parotsidis等人<sup>[79]</sup>提出了边影响力 (EdgeEffect) 算法。从算法设计过程中可以看到，该算法的计算复杂度主要体现在要以图中每个顶点为根进行一次广度优先搜索。因此，该算法的时间复杂度为 $O(mn)$ ，其中， $m$ 是图的边的数目， $n$ 是顶点的数目。显然当候选边集 $|C| = c$ 时，向图中添加 $k$ 条边，使得网络的平均最短路径最短的边影响力 (EdgeEffect) 算法的时间复杂度为 $O(cmn)$ 。

### (3) 边影响力近似 (EffectEstimation) 算法

在边影响力 (EdgeEffect) 算法基础上，N.Parotsidis等人<sup>[79]</sup>又提出了边影响力近似 (EffectEstimation) 算法。令 $A_x[i]$  ( $A_y[i]$ ) 表示 $A_x$  ( $A_y$ ) 中，添加候选边 $e = (x, y)$ 之后，距离顶点 $x$  ( $y$ ) 最短路径恰好减少 $i$ 的顶点的集合。N.Parotsidis等人认为，当 $i$ 或 $j$ 较小时，即对于到顶点 $x$ 或顶点 $y$ 的距离较小的顶点，候选边 $e = (x, y)$ 添加之后，对整个图的平均最短路径的减小的影响不大。因此，在考虑添加候选边的时候，只需考虑计算 $i + j > d_{old}(x, y)$ 的顶点。因而整个算法的时间复杂度会降低。而集合 $A_x[i]$  ( $A_y[i]$ )，可以在计算 $A_x$  ( $A_y$ ) 时求得。因此，简化后的近似算法时间复杂度为 $O(cm)$ 。

对于边影响力 (EdgeEffect) 算法和边影响力近似 (EffectEstimation) 算法，

N.Parotsidis等人应用三种数据，验证了算法的有效性。即Facebook数据<sup>[132]</sup>，DBLP数据<sup>[133]</sup>，互联网图数据<sup>[134]</sup>。

#### (4) 路径筛选法 (Path Screening Method) 算法

当一条候选边 $e = xy$ 添加到图中的时候，并不是所有原有的最短路径都会受到影响，因此，人们开始考虑哪些最短路径会受到影响，显然，在加边之后的新图 $G + e$ 中，通过边 $e$ 的所有路径，最有可能在受到影响而减小了。所以，新图中，过新边 $e$ 的路径被重点考虑。令 $e = xy \in C$ ， $C$ 是候选边集。 $s, t \in V$ 。令 $D_{st}(x, y)$ 为候选边 $xy$ 添加后，最短路径受到影响的点对 $st$ 的集合。则下列等式成立：

$$\begin{aligned} st \in D_{st}(x, y) & \text{ if } d'(s, t) < d(s, t) \\ st \notin D_{st}(x, y) & \text{ if } d'(s, t) = d(s, t). \end{aligned} \quad (6-5)$$

其中， $d'(s, t)$ 表示添加候选边之后，新图中顶点 $s$ 与顶点 $t$ 之间的最短路径长度。则，当边 $e = xy$ 添加后，最短路径的减小值 $u_{xy}$ 即可用如下公式求得：

$$u_{xy} = \sum_{(s, t) \in D_{st}(x, y)} (d(s, t) - d'(s, t)). \quad (6-6)$$

事实上，进一步分析，我们可以知道，当边 $e = xy$ 添加后，真正受到影响的是这样的点对 $(s, t)$ ，在新图 $G + e$ 中，顶点 $s$ 和顶点 $t$ 之间的最短路径经过 $xy$ 边。因此我们只需考虑在图 $G + e$ 中，最短路径经过 $e = xy$ 的点对即可。令 $D_{st}^{SP}(x, y)$ 表示新图 $G + e$ 中，最短路径经过边 $e = xy$ 的点对 $(s, t)$ 的集合。则减小值 $u_{xy}$ 可以进一步整理如下：

$$\begin{aligned} u_{xy} &= \sum_{(s, t) \in D_{st}^{SP}(x, y)} (d(s, x) + d(x, y) + d(y, t)) \\ &\quad - (d(s, x) + d'(x, y) + d(y, t)) \\ &= \sum_{(s, t) \in D_{st}^{SP}(x, y)} (d(x, y) + d'(x, y)). \end{aligned} \quad (6-7)$$

因此，若令 $\delta = d(x, y)$ ，则

$$u_{xy} = \sum_{(s, t) \in D_{st}^{SP}(x, y)} (\delta - 1). \quad (6-8)$$

所以，研究的角度就改变了。我们只需遍历所有的图中的最短路径，并在每条最短路径上调整 $\delta$ 的值。基于以上思路，M. Papagelis<sup>[14]</sup>提出了路径筛选法 (Path Screening Method) 加边算法，令 $\bar{L}(G)$ 表示图 $G$ 中的平局路径长度。则路径筛选法分两个步骤，第一步先计算出图 $G$ 中所有的路径，时间复杂度为 $O(n(n \log n + m))$ （相当于运行

了Johnson's算法一次)。然后对每条路径 $O(n^2)$ 对点, 需要计算 $\delta O(L(G)L(G))$ 次。因此路径筛选法算法的计算时间复杂度为 $O(n(n\log n + m)) + O(n^2 L(G)L(G))$ 。

### (5) FirstMinAPL算法

Andrew Gozzard等人<sup>[129]</sup>针对向网络添加一条边, 使得网络的平均最短路径长度最小的问题, 提出了两个精确的算法, 下面我们简介一下第一个性能较好的称为FirstMinAPL的算法。

**引理1.** 设 $G = (V, E)$ 是一个简单无向图, 令 $p = v_1, v_2, \dots, v_k$ 表示从顶点 $v_1$ 到顶点 $v_k$ 的最短路径。对于任意整数 $1 \leq i \leq j \leq k$ , 令 $p_{ij} = v_i, v_2, \dots, v_j$ 表示路径 $p$ 的从顶点 $v_i$ 到顶点 $v_j$ 的子路径。则路径 $p_{ij}$ 是顶点 $v_i$ 与顶点 $v_j$ 之间的最短路径。

事实上, 假设路径 $p_{ij}$ 不是顶点 $v_i$ 与顶点 $v_j$ 之间的最短路径, 即意味着存在另外一条不同于路径 $p_{ij}$ 的路径 $p_{ij}^*$ , 使得路径 $p_{ij}^*$ 的长度小于路径 $p_{ij}$ 的长度。则我们考虑顶点 $v_1$ 到顶点 $v_k$ 的最短路径 $p$ , 令 $p^* = v_1, \dots, v_{i-1} + v_i, \dots, v_j + v_{j+1}, \dots, v_k$ , 其中 $p_{ij}^* = v_i, \dots, v_j$ , 根据假设路径 $p^*$ 的长度小于路径 $p$ 的长度。这与引理的前提是矛盾的。引理1显示, 在一段较长的最短路径上, 任意两点之间的子路径也是此两点之间的最短路径。这意味着最短路径之间存在一定的包含关系, 也就是说我们在实际计算平均最短路径的时候, 有一部分最短路径会出现重复计算的情况, 这也是计算最短路径的算法普遍计算复杂度高的原因。基于引理1, 人们就可以考虑如何减少重复计算, 降低计算复杂度的问题。

**引理2.** 设 $G = (V, E)$ 是一个简单无向图, 令 $d_{ij}$ 表示顶点 $v_i$ 与顶点 $v_j$ 之间的最短路径长度,  $e_{kl} = v_k v_l$ 表示顶点 $v_k$ 与顶点 $v_l$ 之间的边。 $G' = (V, E + e_{kl})$ , 令 $d'_{ij}$ 表示顶点 $v_i$ 与顶点 $v_j$ 在图 $G'$ 中的最短路径长度。则 $d'_{ij} = \min(d_{ij}, d_{ik} + d_{lj} + 1)$ 。

对于引理2我们可以从三种情况来考虑。

(1) 当顶点 $v_k$ 或顶点 $v_l$ 中的一个顶点在路径 $p_{ij} = v_i, v_2, \dots, v_j$ 上时, 不失一般性, 设顶点 $v_k$ 位于路径 $p_{ij}$ 上, 在新图 $G' = (V, E + e_{kl})$ 中考虑 $d'_{ij}$ , 显然我们可以将路径 $p'_{ij}$ 分成两种可能, 一种是 $p'_{ij} = p_{ij}$ , 令一种是 $p'_{ij} = p_{ik} + p_{lj} + e_{kl}$ 。则结论成立。

(2) 若顶点 $v_k$ 和顶点 $v_l$ 都在路径 $p_{ij}$ 上, 有引理1可知, 结论成立。

(3) 若顶点 $v_k$ 和顶点 $v_l$ 都不在路径 $p_{ij}$ 上, 情况同(1)。

引理1告诉我们, 当一条边添加到图中的时候, 在计算点对之间的最短路径的时候, 有一些路径是重复计算了, 那么引理2就告诉我们添加一条边到图中的时候, 哪些边受到影响以及计算最短路径长度的方法。在次引理的基础上, Andrew Gozzard等人提出了FirstMinAPL算法。该算法首先运行Floyd-Warshall算法, 求出所有点对之间的最短

路径，并将结果存储在 $ASP$ 矩阵中。引理2可以看出，添加一条边之后的任意点对的最短路径长度的计算完全基于原图中最短路径的计算。所以将候选边与矩阵 $ASP$ 相结合，就可以计算出新图中平均最短路径长度的变化。因此，FirstMinAPL算法是在 $ASP$ 矩阵基础上，利用穷举的方法，将图的平均最短路径长度的减小值求出来。然后得到最小值对应的边。由算法的过程可以得到，该算法的时间复杂度为 $O(V^2E + V^3)$ ，其中，Floyd-Warshall算法运行时间复杂度为 $O(V^3)$ ，穷举每条边的时间复杂度为 $O(V^2E)$ 。显然，最坏的情况时间复杂度为 $O(V^4)$ 。

## 6.4 基于结构洞的平均最短路径优化算法

本节中，我们将提出基于结构洞理论的最小化平均最短路径长度的算法。社会学中，一个人的结构洞的数量，代表了这个人的社会影响力，结构洞数越大的人，在社会中掌握的资源越多，对整个社交网络的掌控力度越大。在实际生活中的地位也越高。因此人们研究社会网络的过程中，往往重点关注结构洞数目较多的网络节点。而在应用过程中，人们往往用网络约束指数（Network Constraint Index）来计算和衡量结构洞数量。网络约束指数（NCI）描述了个人直接或间接地与其他个人保持联系，并且，在图中计算网络约束指数只需要顶点的局部信息。

$$C_{ij} = (P_{ij} + \sum_{q \neq i, j} P_{iq}P_{qj})^2, \quad (6-9)$$

其中，顶点 $q$ 是顶点 $i$ 和顶点 $j$ 之间的任意的第三个点。对于一般图， $P_{ij} = 1/d_i$ ，其中， $d_i$ 表示顶点 $i$ 的度数。显然，当顶点 $i$ 是度数为1的悬挂点时， $P_{ij} = 1/d_i = 1$ ，此时，顶点 $i$ 的 $C_{ij}$ 值取得最大，即 $C_{ij} = 1$ 。

$$C(i) = \sum C_{ij}. \quad (6-10)$$

如公式6-10所示，则顶点 $i$ 的网络约束指数 $C(i) = \sum C_{ij}$ ，其中，顶点 $j$ 是图中顶点 $i$ 的邻接点，即 $ij \in E$ 。由网络约束指标的定义可知，一个顶点的网络约束指标越大，则该顶点的结构洞越小。所以，在社交网络中，与结构洞相反，网络约束指标越小的顶点对于网络的控制能力越强。因此，在应用中，可以直接计算网络中个体的网络约束指标来衡量网络中个体的结构洞数。

直观上来讲，图中顶点的结构洞数目，是指该顶点的所有邻点中，互不相连的顶点对的数目。如图6-1，顶点 $v_1$ 的结构洞数目大于顶点 $v_3$ ，其临点 $v_2$ 与 $v_3$ 不相临，也就是说 $v_2$ 与 $v_3$ 之间的路径很大可能要经过顶点 $v_1$ 。因此直观上，结构洞数目较大的顶点，经过他的最短路径也较多。我们的算法思路就是将图中最大和第二大结构洞数目的

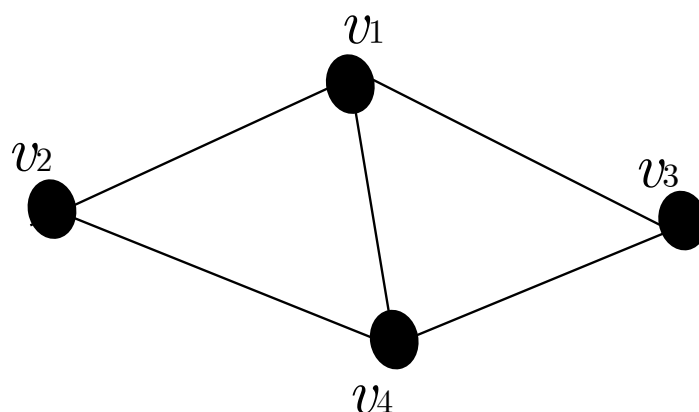


图 6-1: 结构洞示意图

顶点连接，尽可能的将最短路径经过较多的顶点相连，这样图中的最短路径能够受到影响最大。基于此思路，我们设计了最小化平均最短路径长度算法。首先我们根据公式  $C(i) = \sum C_{ij}$  与公式 6-9 求出图中所有顶点的网络约束指数，然后选择其中最小值对应的顶点，以及与最小值顶点不相连的第二小的值对应的顶点。最后得到需要添加边的两个顶点。接下来，只需更新以上两个顶点的网络约束指数，然后重复找最大和第二大值的过程。直到得到需要的添加的边的数目。具体算法如 6-1 所示：

---

**算法 6-1 基于结构洞的加边算法**


---

**输入：**图  $G$  的邻接矩阵  $A$ ，需要添加的边的数目  $k$ ，以及选边集  $C$ 。

**步骤1：**根据公式 6-9 与 6-10，计算所有候选边集  $C$  中，顶点  $v$  网络约束指标  $C(v)$ 。

**步骤2：**令  $e = v_i v_j \in S$ ，其中， $v_i$  是网络约束指标最小的顶点，顶点  $v_j$  是与顶点  $v_i$  不相邻的且网络约束除顶点  $v_i$  之外最小的顶点。

**步骤3：**根据公式 6-9 与 6-10，更新顶点  $v_i$  是与顶点  $v_j$  的网络约束指标。

**步骤4：**重复步骤 2，直到  $|S| = k$ 。

**输出：** $S$ 。

---

接下来，分析一下算法的时间复杂度，注意到，我们的算法时间复杂度主要取决于步骤一，网络约束指数的计算。设图  $G$  有  $n$  个顶点， $m$  条边，候选边集  $|C| = c$ ，要求添加的边的集合  $|S| = k$ ，图中顶点的平均度数为  $d = 2 \times (m)/n$ ，则计算一个顶点的网络约束指标的时间复杂度为  $O(d^2)$ ，那么，我们取最坏情况，设  $C = E^c$ ，且只考虑添加



表 6-1: 计算时间复杂度比较

算法	计算时间复杂度
MaxCCD	$O(n^2 \log n)$
EdgeEffect	$O(n^5)$
EffectEstimation	$O(n^4)$
Path Screening	$O(n^3 + n^2 \times L(G)^2)$
FirstMinAPL	$O(n^4 + n^3)$
我们的方法	$O(n^3)$

一条边的情况，则整个算法的时间复杂度为：

$$\begin{aligned}
 O(\sum(d^2 \times n)) &= O(\sum((2 \times m/n)^2 \times n)) \\
 &= O(\sum((4 \times m^2/n))) \\
 &= O(n^3).
 \end{aligned} \tag{6-11}$$

最后，我们综合比较一下以上6种算法的计算时间复杂度。公平起见，我们主要考虑向图添加一条边的情况。此时，候选边集就是原图边集的补集，即 $C = E^c$ ，所以 $O(C) = O(E) = n^2$ ，且 $k = 1$ 。比较结果如表6-1所示。可以看出，就添加一条边的情况而言，我们的算法具有明显优势。

同时，从算法的设计过程我们可以看到，对于给定候选边的情况，我们的算法只需要基于候选边中的点，即可以求出需要添加的边的集合，而不需要整个图的所有信息，当然，这也是我么算法复杂度较低的原因，更重要的是在实际网络应用过程中，整个网路的信息，包含点的度数，全部路径等信息，是不容易得到的。因此，我们的算法更具有应用意义。

## 6.5 数值实验

这一节中，我们将提出的基于结构洞的加边算法，与上一节提到的5种算法进行比较。首先，我们将问题限定为：向图添加哪一条边，使得图的平均最短路径长度能够减小最大。我们的实验过程如下：首先生成一个ER随机图 $G(100, 0.08)$ ，分别计算6种算法选择的边，加入到图中得到的图的平均最短路径长度和算法时间，然后重复生成100个图，最后求出100个图对应的平均值，比较6种算法的结果和运行时间，注意，从仿真结果来看，无论是平均最短路径长度（APL），还是计算用时，对算法的评价来讲都是越小越好。

如图6-2、图6-3和表格6-2所示，从算法计算时间的角度来看，我们的算法和MaxCCD算法一样，在计算时间上几乎是最快的，而EdgeEffect算法和FirstMinAPL算法计算时间相对较长，这一点也和我们之前分析的计算时间复杂度相吻合。从算法有效性角度

来看, EdgeEffect算法和FirstMinAPL 算法效果最好, 我们的算法求得的平均最短路径长度相对较大, 但表格6-2 显示, 我们的计算结果比结果最好的算法 (EdgeEffect算法和FirstMinAPL 算法), 仅仅相差0.237%。而我们的计算时间仅仅是EdgeEffect 算法的5/100000。当然, 与MaxCCD 算法相比, 无论是计算时间还是有效性, 我们的算法整体性能上还处于劣势。

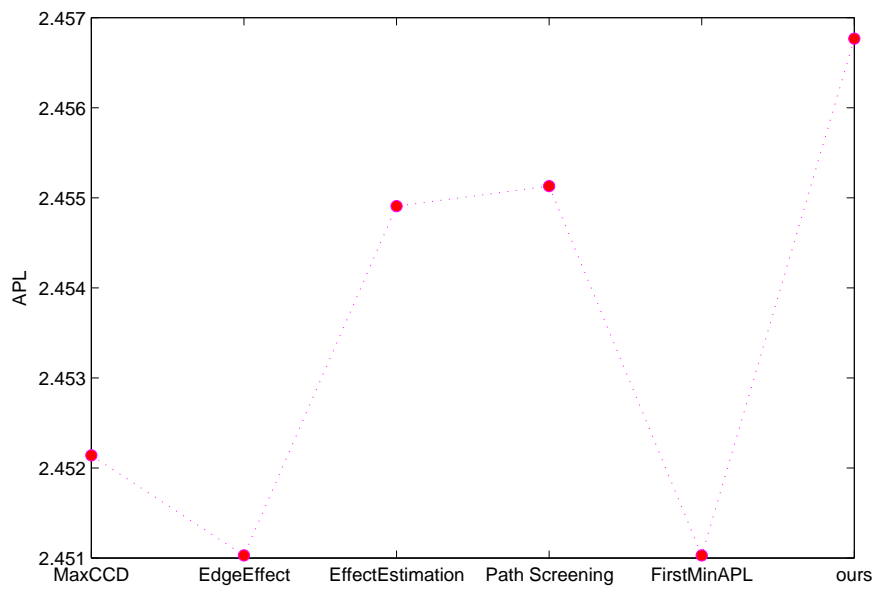


图 6-2: 6种算法平均最短路径长度比较,  $G(100, 0.08)$ , 100个图的平均值。

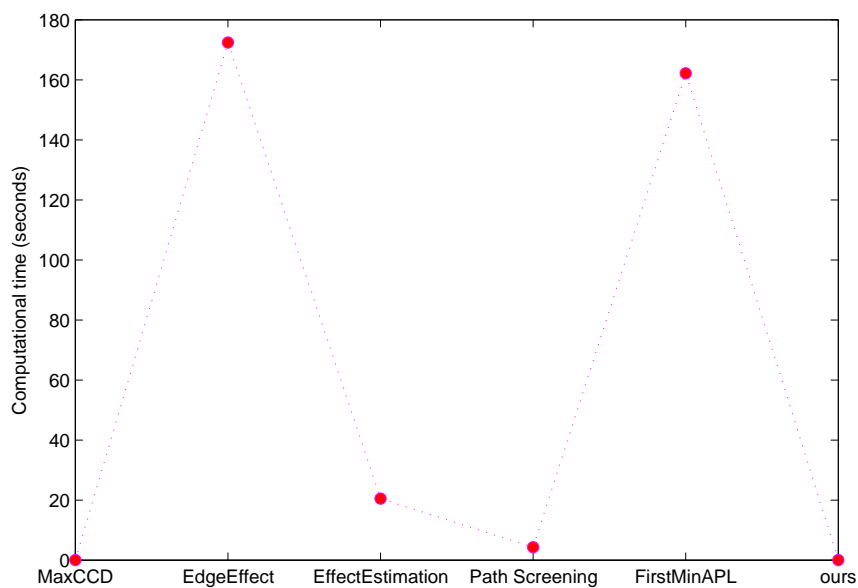


图 6-3: 6种算法计算时间比较,  $G(100, 0.08)$ , 100个图的平均值。

表 6-2: 算法对比

算法	平均最短路径长度	计算时间
MaxCCD	2.4521	0.0099
EdgeEffect	2.4510	172.4195
EffectEstimation	2.4549	20.5034
Path Screening	2.4551	4.3074
FirstMinAPL	2.4510	162.1580
我们的方法	2.4568	0.0010

进一步，我们再生成一个BA图 $BA(100, 5, 2)$ ，比较平均最短路径长度和计算时间。从图6-4、图6-5和表格6-3可以看出，我们的算法在计算时间上是最快的。与MaxCCD算法相比，我们的计算结果相差0.54%，而我们的算法计算用时不足他的1/10。因此我们的算法依然体现了优异的性能。

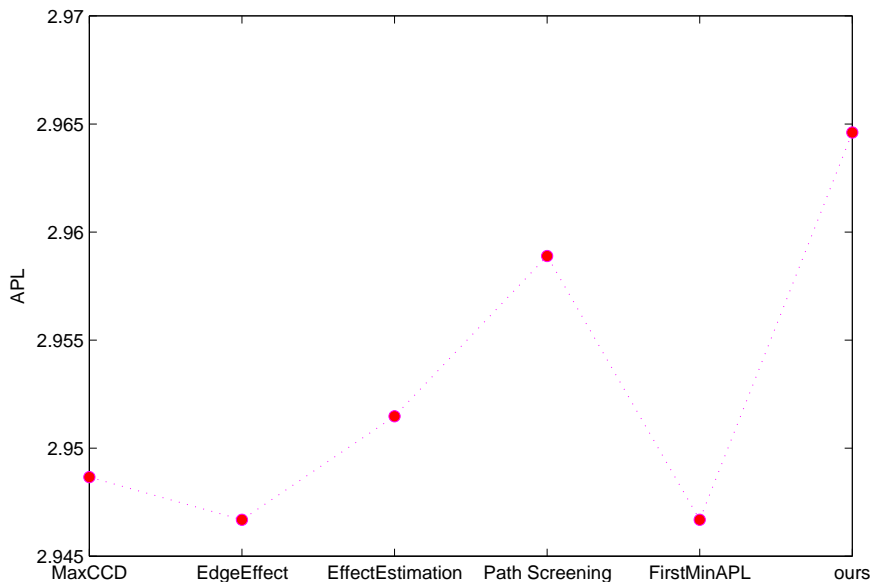


图 6-4: 6种算法平均最短路径长度比较， $BA(100, 5, 2)$ ，100个图的平均值。

综合考虑以上的实验结果，MaxCCD 算法和我们的算法效率和有效性差异不明显，且与其他算法相比，有效性差别不大，而效率方面的优势显著。因此，为了进一步验证我们算法的性能，我们将在较大规模的图上继续测试，由于EdgeEffect 算法等其他算法计算时间复杂度相对较高，因此，我们仅仅验证MaxCCD 算法和我们的算法在较大规模图上的有效性和效率。

如图6-6、图6-7和表格6-4所示，在较大规模的BA图上测试，我们算法明显优于MaxCCD 算法，特别是在效率上，我们的算法用时仅仅是它的0.35%。从MaxCCD 算法设计过程可以看到，该算法基于全局信息，需要求解图中所有的最短路径，因此

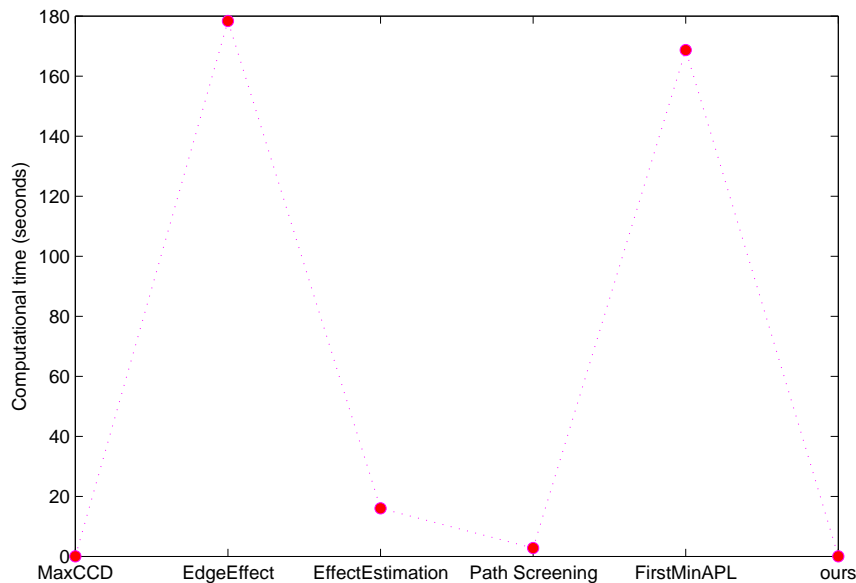
图 6-5: 6种算法计算时间比较,  $BA(100, 5, 2)$ , 100个图的平均值。

表 6-3: 算法对比

算法	平均最短路径长度	计算时间
MaxCCD	2.9487	0.0096
EdgeEffect	2.9467	178.3398
EffectEstimation	2.9515	16.0247
Path Screening	2.9589	2.7867
FirstMinAPL	2.9467	168.6492
我们的方法	2.9646	0.0008

当图的规模增大时, 其计算时间将会显著增加。而我们的算法基于结构洞理论, 仅仅需要与顶点的邻接点相关的信息。因此, 当图的规模增加是, 计算时间的增加不显著。而我们的方法中, 将两个不相连的结构洞大的顶点连接, 相当于将社交网络中资源控制能力最大的两个人联系起来, 强强联手, 确实可以提高网络中消息的传播速度。

综合本章所有实验来看, 当图的规模较小(例如100个顶点的图)的时候, 针对向图添加一条边, 使得图的平均最短路径长度最小化的问题, FirstMinAPL 算法和EdgeEffect 算法最有效, 但他们的计算时间相对最长。MaxCCD算法的效率最高, 我们提出的算法性能居中, 优势不显著。但当图的规模较大时(例如800个顶点的图),

表 6-4: 算法对比

算法	平均最短路径长度	计算时间
MaxCCD	3.9042	9.4730
我们的方法	3.9039	0.0334

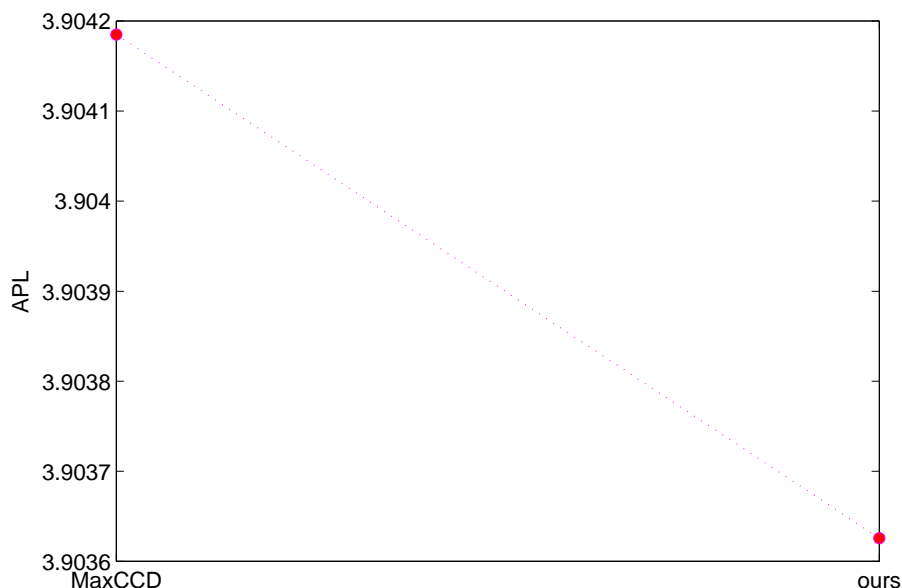


图 6-6: 2种算法平均最短路径长度比较,  $BA(800, 5, 2)$ , 100个图的平均值。

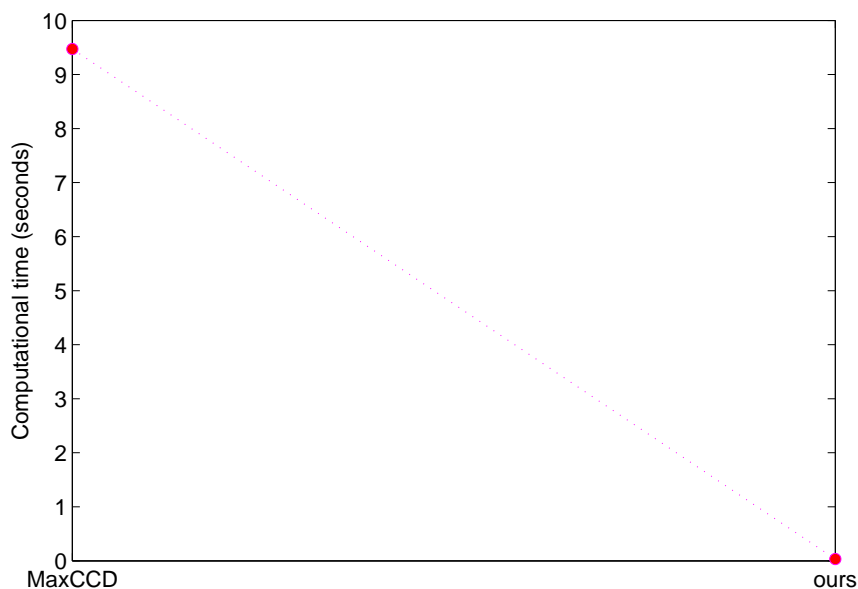


图 6-7: 2种算法计算时间比较,  $BA(800, 5, 2)$ , 100个图的平均值。

FirstMinAPL算法和EdgeEffect算法的计算代价非常昂贵, 而我们提出是算法是效率最高的。并且有效性也优于MaxCCD算法。因此, 我们提出的算法性能明显优于其他已有算法, 而且非常适合大规模网络应用。

## 6.6 本章小结

本章主要讨论最小化平均最短路径长度的加边问题。网络的小世界属性是指实际

网络中任意两个顶点之间的路径都很短。实际网络大都呈现小世界属性，许多网络在动态演化过程中也是向着小世界的方向变化。平均最短路径长度是刻画网络是否具有小世界属性的一个指标。因此，当网络中需要添加新的链接（边）时，如何才能最小化平均最短路径长度，近可能的加速网络的小世界进程是人们很关心的问题。另外，平均最短路径长度也代表了网络中信息传输的延迟情况，平均最短路径长度越小，网络的信息传播延迟越小。因此，人们在构建通信网络的时候，往往需要考虑最小化平均最短路径长度的问题。而图中点对之间的最短路径的计算时间复杂度较高，需要研究降低计算复杂度的算法。另一方面，最小化平均路径长度的问题已经被证明是NP-hard问题。在此背景之下，本章节对最小化最短路径长度问题进行研究。针对网络中添加一条边最小化平均最短路径长度问题提出了一种基于结构洞理论的算法，该算法主要特点是不需要计算网络中的点对之间的最短路径，解决了大规模网络中最短路径计算代价昂贵的问题，同时，该算法只与网络中点和邻点的信息有关，不需要全局信息。因此具有实际应用意义。同时，本章从多个角度验证所提出的算法的有效性和效率。我们综述了目前已有的5种最小化最短路径长度的加边算法，分别用顶点规模较小的BA图模型和ER随机图来验证比较所有的算法的性能，实验结果显示，我们提出的算法与其他算法对比，是有效的。同时，效率也优于其他算法。最后，我们用较大规模（800个顶点）的BA模型验证算法。此时我们提出的算法在有效性和效率上都具有明显优势。

## 第七章 总结与展望

### 7.1 工作小结

费德勒向量的重要意义在于其把图的代数特征, 及拉普拉斯矩阵的特征值和特征向量, 与图的拓扑结构特征联系起来, 使得人们可以通过特征值的分量的大小, 正负等来研究图的结构, 比如, 费德勒向量分量的非负的值对应的图的顶点时连通的。但是在应用过程中人们也面临一些问题, 比如, 要应用费德勒向量必须要计算它, 就是要计算矩阵的特征向量, 所以当图的规模较大的时候, 就要付出很大的计算代价。这一点就严重制约了费德勒向量在实际问题中的应用进程。针对这个现象, 本文反其道而行, 重点研究了图的结构特征如何与费德勒向量相对应并得出了如下结论: 首先我们找到了一类图, 这类图的最小度点恰恰对应了费德勒向量的分量的最大值。我们定义这种图为 $\alpha$ -匹配图。并且我们证明了树图是 $\alpha$ -匹配图。进一步我们进行了抽样统计, 得出结论是大部分图都是 $\alpha$ -匹配图。然后, 我们又找到另一类图, 这类图的特点是, 如果以费德勒向量分量的最大值对应的顶点为根生成广度优先图, 则费德勒向量分量的最小值一定在改图的最高层。我们定义这类图为Fiedler-Breadth-First图, 接着我们也证明了一类这样的图。同样抽样统计实验表明大部分图都是Fiedler-Breadth-First图。接着, 我们将以上得到的结论用到实际应用中, 针对几种问题设计了相应的算法。

1、我们设计了最大化代数连通度的算法。该算法主要根据最小度对应费德勒向量分量的最大值, 以及以最大值为根的广度优先图中, 最小值在最高层的图的结构与费德勒向量分量的对应关系。模拟实验证明我们的算法在准确度和时间消耗上都优于目前已有的算法。

2、接着我们根据代数连通度与网络传输容量之间的关系, 又设计了一种通过添加网络链接来调高传输容量的算法。仿真实验显示, 我们的算法能够明显提高网络中的传输容量, 减少网络拥堵, 其性能优于已有的算法。

3、针对网络的最小化平均最短路径长度问题, 我们提出了一种新的加边算法。仿真实验显示, 我们的近似算法在准确度上略低于穷举法, 但时间复杂度上明显低于穷举法, 与已有的近似算法比较, 我们的算准确度和时间复杂度也都有明显优势。

这三个加边算法是针对不同的网络指标提出的, 文章最后, 我们比较一下三个算法针对不同的网络指标(代数连通度, 传输容量, 平均最短路径长度, 最大中介中心性)性能。针对向网络中添加一条边, 如何使得网络的性能指标更优的问题, 我们生成100个图随机图 $G(100, 0.08)$ , 对于三种不同的方法计算出不同的指标取平均值进行比

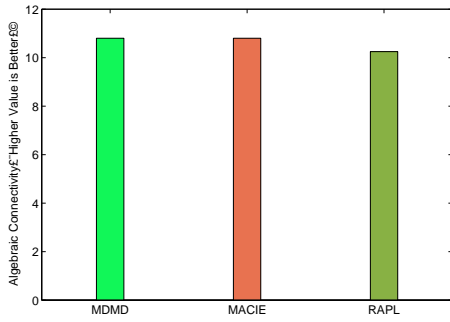


图 7-1: 向网络中添加一条边使得代数连通度增加最大, 三种算法比较, 100个图 $G(100, 0.08)$ 平均值。

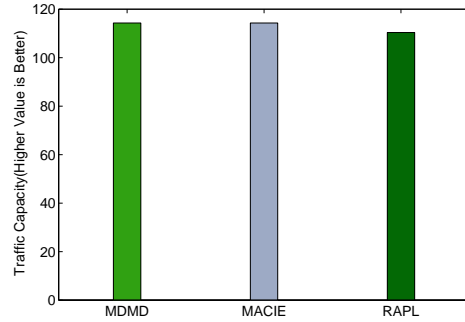


图 7-2: 向网络中添加一条边使得传输容量增加最大, 三种算法比较, 100个图 $G(100, 0.08)$ 平均值。

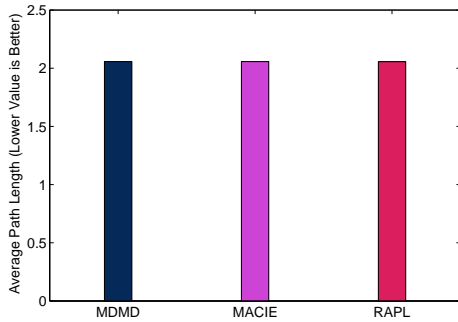


图 7-3: 向网络中添加一条边使得平均路径长度减小最大, 三种算法比较, 100个图 $G(100, 0.08)$ 平均值。

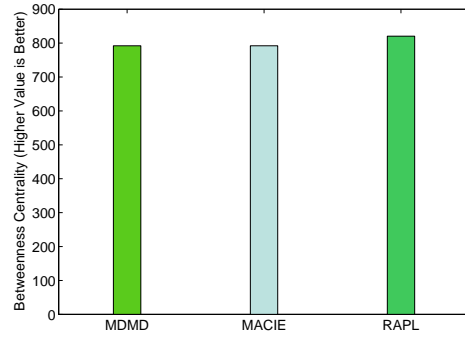


图 7-4: 向网络中添加一条边使得最大中介中心性增加最大, 三种算法比较, 100个图 $G(100, 0.08)$ 平均值。

较。另外，本文第五章中提出的基于代数连通度的最大化传输容量的加边算法，本质上应用了第四章提出的最大化代数连通度加边算法，因此，正两种算法选择的候选边是一样的。为了便于显示，以下我们将第六章提出的最小化网络平均最短路径长度的加边算法简称为RAPL 算法。另外，值得注意的是，我们提出MACIE算法，本质上应用了MDMD选择候选边，因此，两种方法选出来的候选边相同，也可以通过以下的实验验证。

如图7-1 所示，对于代数连通度而言，MDMD算法最有效，加入该方法选择出来的候选边之后图的代数连通度的值比RAPL算法得到的结果高5.37%,分别为10.8001 与10.2497。如图7-2 所示，对于传输容量指标，MACIE算法最有效，其得到的结果比RAPL算法高3.54%。分别为114.2765与110.3683。如图7-3 所示，对于平均最短路径而言，RAPL算法的结果与MDMD或MACIE算法的结果相差不大，分别为2.0567与2.0569。实验结果说明MDMD算法和和MACIE算法也能很好的减小平均路径长度。最后，如图7-4 所示，RAPL算法得到的最大中介中心性的值最大是820.3169，



而其他两种算法得到的结果是792.0265。说明RAPL算法更具有优势，回顾RAPL算法，不难发现，网络中具有做大结构洞的顶点被添加了一条边，而结构洞大的顶点，意味着有更多的最短路径经过，意味着有更大的中介中心性。给这样的顶点添加一条边，无疑增加了该点的中介中心性。至此，本论文中的各种加边策略相比较，逻辑上是合理的。

## 7.2 未来工作展望

回顾本文的研究内容和过程，还有很多问题值得进一步的研究，这些也将成为我们以后的研究方向之一。

1、 $\alpha$ -匹配图和Fiedler-Breadth-First图的充要条件值得深入研究，本文提出这两个概念的目的是要确定图中费德勒向量极值的位置，希望能够从结构上找到费德勒向量的极值，而不是通过计算得到。

2、既是 $\alpha$ -匹配图又是Fiedler-Breadth-First图的研究。对于这两类图的研究也可以进一步了解图的代数指标与结构指标的关系。下一步工作，我们可以首先重点考虑树图是Fiedler-Breadth-First图的证明。然后再深入研究一般图。

3、费德勒向量与图的结构指标的关系在数据聚类方法中的应用。费德勒向量是数据的谱聚类方法的理论基础之一，在谱聚类的过程中，可以考虑如何与结构指标相结合，提高聚类的有效性。

## 参考文献

- [1] Boccaletti S., Latora V., Moreno Y., et al., Complex networks: Structure and dynamics[J]. Physics Reports, 2006, 424 (4): 175–308
- [2] 汪小帆, 李翔, 陈关荣, [M]. 清华大学出版社, 2006
- [3] Lu J., Chen G., Ogorzalek M.J., et al., Theory and applications of complex networks: Advances and challenges[C]. In IEEE International Symposium on Circuits and Systems, 2013, 2291–2294
- [4] Callahan D., Shakarian P., Nielsen J., et al., Shaping operations to attack robust terror networks[C]. In International Conference on Social Informatics, 2012, 13–18
- [5] West B.J., Geneston E.L., Grigolini P., Maximizing information exchange between complex networks[J]. Physics Reports, 2008, 468 (1): 1–99
- [6] Chen Z., Wu J., Xia Y., et al., Robustness of interdependent power grids and communication networks: A complex network perspective[J]. IEEE Transactions on Circuits and Systems II Express Briefs, 2017, (99): 1–1
- [7] Wu J., Zeng J., Chen Z., et al., Effects of traffic generation patterns on the robustness of complex networks[J]. Physica A Statistical Mechanics and Its Applications, 2018, 492: 871–877
- [8] Ma J., Wang H., Zhang Z., et al., An efficient routing strategy for traffic dynamics on two-layer complex networks[J]. International Journal of Modern Physics B, 2018, 32: 1850155
- [9] Lin Y., Lui J.C.S., Jung K., et al., Modelling multi-state diffusion process in complex networks: theory and applications[J]. Journal of Complex Networks, 2018, 2 (4): 431–459
- [10] Watts D.J., Strogatz S.H., Collective dynamics of 'small-world' networks[J]. Nature, 1998
- [11] Barabasi A.L., Albert R., Emergence of scaling in random networks[J]. Science (New York, N.Y.), 1999, 286: 509–12

- [12] Manzano M., Calle E., Ripoll J., et al., Epidemic and cascading survivability of complex networks[C]. In International Workshop on Reliable Networks Design and Modeling, 2015
- [13] Jalili M., Yu X., Enhancement of synchronizability in networks with community structure through adding efficient inter-community links[J]. IEEE Transactions on Network Science and Engineering, 2017, 3 (2): 106–116
- [14] Papagelis M., Refining social graph connectivity via shortcut edge addition[J]. Acm Transactions on Knowledge Discovery from Data, 2015, 10 (2): 1–35
- [15] Yang L.X., Jiang J., Impacts of link addition and removal on synchronization of an elementary power network[J]. Physica A: Statistical Mechanics and its Applications, 2017, 479: 99–107
- [16] Cui P., Zhu P., Wang K., et al., Enhancing robustness of interdependent network by adding connectivity and dependence links[J]. Physica A Statistical Mechanics and Its Applications, 2018, 497: 185–197
- [17] Kazawa Y., Tsugawa S., On the effectiveness of link addition for improving robustness of multiplex networks against layer node-based attack[C]. In Computer Software and Applications Conference, 2017
- [18] Wang X., Cao J., Rui L., et al., A preferential attachment strategy for connectivity link addition strategy in improving the robustness of interdependent networks[J]. Physica A Statistical Mechanics and Its Applications, 2017, 483: 412–422
- [19] Chen X., Pequito S., Pappas G.J., et al., Minimal edge addition for network controllability[J]. IEEE Transactions on Control of Network Systems, 2018, (99): 1–1
- [20] Ghiasifard S., Khadivi S., Asadpour M., et al., Improving the quality of overlapping community detection through link addition based on topic similarity[C]. In International Symposium on Artificial Intelligence and Signal Processing, 2015
- [21] Hua J., Deng G., Ma Z., et al., A new community detection algorithm based on adding and deleting links[C]. In IEEE 2nd International Conference on Big Data Analysis, 2017, 412–422

- 
- [22] Fiedler M., A property of eigenvectors of non-negative symmetric matrices and its application to graph theory[J]. Czechoslovak Mathematical Journal, 1975, 25 (4): 619–633
- [23] Pothen A., Simon H.D., Liou K.P., Partitioning sparse matrices with eigenvectors of graphs. *siam j matrix anal appl*[J]. *Siam J.matrix Anal.appl*, 1990, 11 (3): 430–452
- [24] Donetti L., Muñoz M.A., Detecting network communities: a new systematic and efficient algorithm[J]. *Journal of Statistical Mechanics Theory and Experiment*, 2004, (10): 10012
- [25] Chen P.Y., Hero A.O., Deep community detection[J]. *IEEE Transactions on Signal Processing*, 2014, 63 (21): 5706–5719
- [26] Ulrik B., Bobo N., Asymmetric relations in longitudinal social networks[J]. *IEEE Transactions on Visualization and Computer Graphics*, 2011, 17 (12): 2283
- [27] Chen P.Y., Hero A.O., Local fiedler vector centrality for detection of deep and overlapping communities in networks[C]. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, 2014, 1120–1124
- [28] Shi J., Malik J., Normalized cuts and image segmentation[C]. In *Conference on Computer Vision and Pattern Recognition*, 1997
- [29] Akaydin B., Manguoğlu M., Parallel image segmentation using fiedler vector[C]. In *Proceedings of the IASTED International Conference on Computer Graphics and Imaging, CGIM 2013*
- [30] Spielman D.A., Teng S., Spectral Partitioning Works: Planar Graphs and Finite Element Meshes[M]. *University of California at Berkeley*, 1996
- [31] Orponen P., Schaeffer S.E., Local clustering of large graphs by approximate fiedler vectors[C]. In *International Conference on Experimental and Efficient Algorithms*, 2005, 524–533
- [32] Shao H., Mesbahi M., On the fiedler vector of graphs and its application in consensus networks[C]. In *American Control Conference*, 2015, 1734–1739

- [33] Ghosh A., Boyd S., Growing well-connected graphs[C]. In Decision and Control, 2006 IEEE Conference on, 6605–6611
- [34] Huang W., Chow T.W., Effective strategy of adding nodes and links for maximizing the traffic capacity of scale-free network[J]. Chaos, 2010, 20 (3): 033123
- [35] Wei P., Chen L., Sun D., Algebraic connectivity maximization of an air transportation network: The flight routes’ addition/deletion problem[J]. IEEE Transactions on Intelligent Transportation Systems, 2014, 61 (2): 13–27
- [36] Qiu T., Zhao A., Xia F., et al., Rose: Robustness strategy for scale-free wireless sensor networks[J]. IEEE/ACM Transactions on Networking, 2017, 25: 2944–2959
- [37] Gozzard A., Ward M., Datta A., Converting a network into a small-world network: Fast algorithms for minimizing average path length through link addition[J]. Information Sciences, 2018, 422
- [38] Gaur N., Chakraborty A., Manoj B.S., Delay optimized small-world networks[J]. IEEE Communications Letters, 2014, 18 (11): 1939–1942
- [39] Merris R., Laplacian matrices of graphs: a survey[J]. Linear Algebra and Its Applications, 1994, 197-198 (2): 143–176
- [40] Pardalos P.M., Handbook of optimization in complex networks[J]. Springer Optimization and Its Applications, 2012, 58: 0111222–2001
- [41] Das K.C., The laplacian spectrum of a graph[J]. Computers and Mathematics with Applications, 2004, 48 (5): 715–724
- [42] Cvetković D.M., Doob M., Sachs H., Spectra of graphs[J]. 1980, 6 (2): 73–92
- [43] Li J.S., Zhang X.D., On the laplacian eigenvalues of a graph[J]. Linear Algebra and Its Applications, 1998, 285 (1 – 3): 305–307
- [44] Biyikoğlu T., Leydold J., Stadler P.F., Laplacian eigenvectors of graphs[J]. Lecture Notes in Mathematics, 2007, 1915
- [45] Banuelos R., Burdzy K., On the “hot spots” conjecture of j. rauch[J]. Journal of Functional Analysis, 1997, 164 (1): 1 – 33

- 
- [46] Burdzy K., Werner W., A counterexample to the "hot spots" conjecture[J]. *Annals of Mathematics*, 1999, 149 (1): 309–317
- [47] Evans L.C., The fiedler rose: On the extreme points of the fiedler vector[J]. *Mathematics*, 2011
- [48] Lefèvre J., Fiedler vectors and elongation of graphs: A threshold phenomenon on a particular class of trees[J]. *arXiv:1302.1266*, 2012
- [49] Gernandt H., Philipp Pade J., Schur reduction of trees and extremal entries of the fiedler vector[J]. *arXiv:1807.01084v1*, 2018
- [50] Urschel J.C., Zikatanov L.T., Spectral bisection of graphs and connectedness[J]. *Linear Algebra and Its Applications*, 2014, 449 (5): 1–16
- [51] Begué M., Okoudjou K.A., Invertibility of graph translation and support of laplacian fiedler vectors[J]. 2017: 1–21
- [52] Perazzi F., Sorkine-Hornung A., Sorkine-Hornung A., Efficient salient foreground detection for images and video using fiedler vectors[C]. In *Eurographics Workshop on Intelligent Cinematography and Editing*, 2015
- [53] Chen P.Y., Hero A.O., Deep community detection[J]. *IEEE Transactions on Signal Processing*, 2015, 63 (21): 5706–5719
- [54] Ozdemir A., Bolaños M., Bernat E., et al., Hierarchical spectral consensus clustering for group analysis of functional brain networks[J]. *IEEE Transactions on Biomedical Engineering*, 2015, 62 (9): 2158–2169
- [55] Zhong J., Kai N., He Z., Detecting information source in diffusion networks[C]. In *IEEE International Conference on Computer Communication and the Internet*, 2016
- [56] Filippidou I., Kotidis Y., Effective and efficient graph augmentation in large graphs[C]. In *IEEE International Conference on Big Data*, 2017, 875–880
- [57] Fiedler M., Algebraic connectivity of graphs[J]. *Czechoslovak Mathematical Journal*, 1973, 23 (23): 298–305

- [58] Feddema J.T., Byrne R.H., Abdallah C.T., Algebraic connectivity and graph robustness[J]. Prod.sandia.gov, 2009
- [59] Kim Y., Bisection algorithm of increasing algebraic connectivity by adding an edge[J]. IEEE Transactions on Automatic Control, 2009, 55 (1): 170–174
- [60] Ibrahim A.S., Seddik K.G., Liu K.J.R., Improving connectivity via relays deployment in wireless sensor networks[C]. In Global Telecommunications Conference, 2007. GLOBECOM 07. IEEE, 1159–1163
- [61] Alenazi M.J.F., Cetinkaya E.K., Sterbenz J.P.G., Cost-efficient algebraic connectivity optimisation of backbone networks[J]. Optical Switching and Networking, 2014, 14 (4): 107–116
- [62] Arenas A., Díazguilera A., Guimerà R., Communication in networks with hierarchical branching.[J]. Physical Review Letters, 2001, 86 (14): 3196–3199
- [63] Bai Y., Liu S., Zhang Z., Effective hybrid link-adding strategy to enhance network transport efficiency for scale-free networks[J]. International Journal of Modern Physics C, 28 (08)
- [64] Jiang Z.Y., Liang M.G., Guo D.H., Enhancing network performance by edge addition[J]. International Journal of Modern Physics C, 2011, 22 (11): 1211–1226
- [65] Zhang S., Liang M.G., Jiang Z.Y., et al., Effective strategy of adding links for improving network transport efficiency on complex networks[J]. International Journal of Modern Physics C, 2014, 25 (06): 1450014
- [66] Zhao L., Lai Y.C., Park K., et al., Onset of traffic congestion in complex networks[J]. Phys Rev E Stat Nonlin Soft Matter Phys, 2005, 71 (2): 026125
- [67] Wang D., Liu E., Liu D., et al., Rsh: A link-addition strategy for capacity enhancement in scale-free networks[J]. IEEE Communications Letters, 2015, 19 (12): 2110–2113
- [68] Zhang Z., Liu S., Yang Y., et al., A link-adding strategy for improving robustness and traffic capacity in large-scale wireless sensor networks[J]. Cluster Computing, 2018, (4): 1–8

- 
- [69] Liu H., Jin C., Yang B., et al., Finding top-k shortest paths with diversity[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30 (3): 488–502
- [70] Making a small world smaller: Path optimization in networks[J]. IEEE Transactions on Knowledge and Data Engineering, 2018, 30 (8): 1533–1546
- [71] Faizian P., Mollah M.A., Yuan X., et al., Random regular graph and generalized de bruijn graph with  $k$  -shortest path routing[J]. IEEE Transactions on Parallel and Distributed Systems, 2018, 29 (1): 144–155
- [72] Basciani F., Demidio M., Ruscio D.D., et al., Automated selection of optimal model transformation chains via shortest-path algorithms[J]. IEEE Transactions on Software Engineering, 2018, (99): 1–1
- [73] Zhang Y., Song S., Shen Z.J.M., et al., Robust shortest path problem with distributional uncertainty[J]. IEEE Transactions on Intelligent Transportation Systems, 2017, (99): 1–11
- [74] Shimizu N., Mori R., Average shortest path length of graphs of diameter 3[C]. In 2016 Tenth IEEE/ACM International Symposium on Networks-on-Chip (NOCS), 1–6
- [75] Du H., He X., Du W., et al., Optimization of the critical diameter and average path length of social networks[J]. Complexity, 2017, 2017 (5): 1–11
- [76] Mauricio B., Louis M P., Synchronization in small-world systems[J]. Physical Review Letters, 2001, 89 (5): 054101
- [77] Jost J., Joy M.P., Spectral properties and synchronization in coupled map lattices.[J]. Physical Review E Statistical Nonlinear and Soft Matter Physics, 2002, 65 (1-2): 016201
- [78] Ashton D.J., Effect of congestion costs on shortest paths through complex networks.[J]. Physical Review Letters, 2005, 94 (5): 058701
- [79] Parotsidis N., Pitoura E., Tsaparas P., Selecting shortcuts for a smaller world[C]. In Proceedings of the 2015 SIAM International Conference on Data Mining, 28–36



- [80] Meyerson A., Tagiku B., Minimizing average shortest path distances via shortcut edge addition[C]. In Proceedings of the 12th International Workshop and 13th International Workshop on Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques, 2009
- [81] Chang M.C.F., Socher E., Tam S.W., et al., Rf interconnects for communications on-chip[C]. In International Symposium on Physical Design, ISPD 2008, Portland, Oregon, Usa, April, 2008, 78–83
- [82] Chang M.F., Cong J., Kaplan A., et al., Cmp network-on-chip overlaid with multi-band rf-interconnect[J]. 2007: 191–202
- [83] Freeman L.C., A set of measures of centrality based on betweenness[J]. Sociometry, 1977, 40 (1): 35–41
- [84] Brandes U., A faster algorithm for betweenness centrality\*[J]. Journal of Mathematical Sociology, 2001, 25 (2): 163–177
- [85] Malighetti P., Martini G., Paleari S., et al., The impacts of airport centrality in the eu network and inter-airport competition on airport efficiency[J]. Mpra Paper, 2009
- [86] Everett M.G., Valente T.W., Bridging, brokerage and betweenness[J]. Soc Networks, 2016, 44: 202–208
- [87] Bergamini E., Crescenzi P., D’Angelo G., et al., Improving the betweenness centrality of a node by adding links[J]. Journal of Experimental Algorithmics, 2017, 23 (2)
- [88] D’Angelo G., Severini L., Velaj Y., On the maximum betweenness improvement problem[J]. Electronic Notes in Theoretical Computer Science, 2016, 322: 153–168
- [89] Li G., Hao Z.F., Huang H., et al., A maximum algebraic connectivity increment edge-based strategy for capacity enhancement in scale-free networks[J]. Physics Letters A, 2019, 383 (17): 2046 – 2050
- [90] Schultz P., Peron T., Eroglu D., et al., Tweaking synchronisation by link addition[J]. arXiv preprint: arXiv:1603.07157, 2016

- 
- [91] Burt R.S., Structural holes: The social structure of competition.[M]. Harvard University Press, 1992
- [92] Solomonoff R., Rapoport A., Connectivity of random nets[J]. Bulletin of Mathematical Biophysics, 1951, 13 (2): 107–117
- [93] Erdos P., Renyi A., On random graphs[J]. Publicationes Mathematicae, 1959, 6 (4): 290–297
- [94] Barabasi A.L., Albert R., Emergence of scaling in random networks.[J]. Science (New York, N.Y.), 1999, 286 (5439): 509–512
- [95] Guimerà R., DíazGuilera A., Vegaredondo F., et al., Optimal network topologies for local search with congestion[J]. Physical Review Letters, 2002, 89 (24): 248701
- [96] L. Myers J., Well A., Lorch R., Research Design and Statistical Analysis, 2nd ed.[M]. Japanese Journal of National Medical Services, 2013
- [97] 冯果忱., 数值分析[M]. 高等教育出版社, 2007
- [98] Abreu N.M.M.D., Old and new results on algebraic connectivity of graphs[J]. Linear Algebra and Its Applications, 2007, 423 (1): 53–73
- [99] Olfati-Saber R., Fax J.A., Murray R.M., Consensus and cooperation in networked multi-agent systems[J]. Proceedings of the IEEE, 2007, 95 (1): 215–233
- [100] Li X., Chen M.Z., Su H., et al., Distributed bounds on the algebraic connectivity of graphs with application to agent networks[J]. IEEE Transactions on Cybernetics, 2017, 47 (8): 2121–2131
- [101] Georgiou O., Algebraic connectivity of keyhole random geometric graphs[J]. IEEE Communications Letters, 2016, 20 (10): 2079–2082
- [102] Liu S., Chen P.Y., Hero A.O., Accelerated distributed dual averaging over evolving networks of growing connectivity[J]. IEEE Transactions on Signal Processing, 2017, (99): 1–1
- [103] Manaffam S., Behal A., Bounds on the smallest eigenvalue of a pinned laplacian matrix[J]. IEEE Transactions on Automatic Control, 2017, (99): 1–1

- [104] Yu W., Chen G., Cao M., Consensus in directed networks of agents with nonlinear dynamics[J]. IEEE Transactions on Automatic Control, 2011, 56 (6): 1436–1441
- [105] Wang H., Miegheem P.V., Algebraic connectivity optimization via link addition[C]. In International Conference on Bio-Inspired MODELS of Network, Information and Computing Sytems, 2008, 22
- [106] Martínez M.C.S., Robustness optimization via link additions[J]. Universitat Politècnica De Catalunya, 2009
- [107] Sydney A., Scoglio C., Gruenbacher D., Optimizing algebraic connectivity by edge rewiring[J]. Applied Mathematics and Computation, 2013, 219 (10): 5465–5479
- [108] Kim Y., Mesbahi M., On maximizing the second smallest eigenvalue of a state-dependent graph laplacian[J]. IEEE Transactions on Automatic Control, 2006, 51 (1): 116–120
- [109] Ogiwara K., Fukami T., Takahashi N., Maximizing algebraic connectivity in the space of graphs with a fixed number of vertices and edges[J]. IEEE Transactions on Control of Network Systems, 2017, 4 (2): 359–368
- [110] Lorenzo P.D., Barbarossa S., Distributed estimation and control of algebraic connectivity over random graphs[J]. IEEE Transactions on Signal Processing, 2014, 62 (21): 5615–5628
- [111] Aragues R., Shi G., Dimarogonas D.V., et al., Distributed algebraic connectivity estimation for undirected graphs with upper and lower bounds[J]. Automatica, 2014, 50 (12): 3253–3259
- [112] Cvetković D.M., Doob M., Sachs H., Spectra of graphs : theory and application[J]. Math Hungar, 1995, 6 (2): 191–195
- [113] Golub G.H., Van Loan C.F., Matrix computations[J]. Mathematical Gazette, 1996, 47 (5 Series II): 392–396
- [114] Bunch J.R., Nielsen C.P., Sorensen D.C., Rank-one modification of the symmetric eigenproblem[J]. Numerische Mathematik, 1978, 31 (1): 31–48

- 
- [115] Leskovec J., Krevl A., Snap datasets: Stanford large network dataset collection. <http://snap.stanford.edu/data/>, 2014
- [116] Andor H., Fi I., London A., et al., Complex network analysis of public transportation networks: A comprehensive study[C]. In International Conference on MODELS and Technologies for Intelligent Transportation Systems, 2015, 371–378
- [117] Xia X., Wang X., Wei W., et al., Measuring the mobile app market: A complex network approach[J]. Information Technology Journal, 2013, 12 (11): 2090–2100
- [118] Jalili M., Yu X., Enhancement of synchronizability in networks with community structure through adding efficient inter-community links[J]. IEEE Transactions on Network Science and Engineering, 2016, 3 (2): 106–116
- [119] Chakraborty A., Vineeth B.S., Manoj B.S., Analytical identification of anchor nodes in a small-world network[J]. IEEE Communications Letters, 2016, 20 (6): 1215–1218
- [120] Bilal K., Manzano M., Khan S.U., et al., On the characterization of the structural robustness of data center networks[J]. IEEE Transactions on Cloud Computing, 2013, 1 (1): 1–1
- [121] Liu S., Chen P.Y., Hero A.O., Accelerated distributed dual averaging over evolving networks of growing connectivity[J]. IEEE Transactions on Signal Processing, 2018, (99): 1–1
- [122] Gang Y., Tao Z., Bo H., et al., Efficient routing on complex networks[J]. Physical Review E Statistical Nonlinear and Soft Matter Physics, 2005, 73 (2): 046108
- [123] Li G., Hao Z.F., Huang H., et al., Maximizing algebraic connectivity via minimum degree and maximum distance[J]. IEEE Access, 2018, 6: 41249–41255
- [124] Crucitti P., Latora V., Marchiori M., et al., Error and attack tolerance of complex networks[J]. Physica A: Statistical Mechanics and its Applications, 2004, 340 (1-3): 388–394
- [125] Barahona M., Pecora L.M., Synchronization in small-world systems[J]. Physical Review Letters, 2002, 89 (5): 054101

- [126] Chang M.F., Cong J., Kaplan A., et al., Cmp network-on-chip overlaid with multi-band rf-interconnect[C]. In IEEE International Symposium on High PERFORMANCE Computer Architecture, 2008, 191–202
- [127] Ogras U.Y., Marculescu R., "it's a small world after all": Noc performance optimization via long-range link insertion[J]. IEEE Transactions on Very Large Scale Integration Systems, 2006, 14 (7): 693–706
- [128] Pickavet M., Demeester P., A zoom-in approach to design sdh mesh restorable networks[J]. Journal of Heuristics, 2000, 6 (1): 107–130
- [129] Gozzard A., Ward M., Datta A., Converting a network into a small-world network: Fast algorithms for minimizing average path length through link addition[J]. Information Sciences, 2017, 422: 282–289
- [130] Cormen T.H., Leiserson C.E., Rivest R.L., et al., Introduction to Algorithms, Third Edition[M]. The MIT Press, Cambridge, 2009
- [131] Okamoto K., Chen W., Li X.Y., Ranking of closeness centrality for large-scale social networks[C]. In International Workshop on Frontiers in Algorithmics, 2008
- [132] Viswanath B., Mislove A., Cha M., et al., On the evolution of user interaction in facebook[C]. In Acm Workshop on Online Social Networks, 2009
- [133] Ley M., The dblp computer science bibliography: Evolution, research issues, perspectives[C]. In String Processing and Information Retrieval, International Symposium, Spire, Lisbon, Portugal, September, 2002
- [134] Mislove A., Druschel P., Bhattacharjee B., et al., Growth of the flickr social network[C]. In Proc. of WOSN 2008

# 攻读博士学位期间取得的研究成果

一、已发表（包括已接受待发表）的论文，以及已投稿、或已成文打算投稿、或拟成文投稿的论文情况（只填写与学位论文内容相关的部分）：

序号	作者（全体作者，按顺序排列）	题 目	发表或投稿刊物名称、级别	发表的卷期、年月、页码	相当于学位论文的哪一部分（章、节）	被索引收录情况
1	<b>Gang Li,</b> Zhi-Feng Hao, Han Huang, Hang Wei	Maximizing Algebraic Connectivity via Minimum Degree and Maximum Distance	IEEE ACCESS, 影响因子: 3.557, JCR 一区刊物	vol. 6, pp. 41 249 - 41 255, Aug. 2018.	第三章和第四章	SCI
2	<b>Gang Li,</b> Zhi Feng Hao, Han Huang, Hang Wei, Xue Ming Yan	A Maximum Algebraic Connectivity Increment Edge-based Strategy for Capacity Enhancement in Scale-free Networks	PHYSICS LETTERS A , 影响因子: 1.863, JCR 二区刊物	vol. 383, no. 17, p. 2046-2050, Jun. 2019.	第五章	SCI
3	<b>Gang Li,</b> Zhi Feng Hao, Han Huang	A structural hole theory based link addition algorithm for a smaller world	INFORMATION SCIENCES, 影响因子: 4.305, JCR 一区刊物	拟成文投稿	第六章	SCI

二、与学位内容相关的其它成果（包括专利、著作、获奖项目等）

博士在读期间发表的其他学术论文：

署名排序	题目	发表或投稿刊物名称、级别	发表的卷期、年月、页码
第一作者	A Construction Graph-Based Evolutionary Algorithm for Traveling Salesman Problem	Simulated Evolution and Learning. SEAL 2017. Lecture Notes in Computer Science, Springer (EI)	2017,vol 10593
第四作者	Human-computer cooperative brain storm optimization algorithm for the two-echelon vehicle routing problem	2016 IEEE Congress on Evolutionary Computation (CEC), 2016. (EI)	2016:2676-2681.
第四作者	A human-machine cooperative approach for combinatorial optimization problem	International Conference on Machine Learning and Cybernetics. IEEE (EI)	2017:838-843
第四作者	A Real Adjacency Matrix-Coded Differential Evolution Algorithm for Traveling Salesman Problems	Bio-Inspired Computing - Theories and Applications. (EI)	BIC-TA 2016, Part II, CCIS 682, pp. 135 - 140, 2016,
第六作者	A real adjacency matrix-coded evolution algorithm for highly linkage-based routing problems	International Journal of Bio-Inspired Computation. (SCI)	二审

## 致谢

首先，我衷心感谢我的导师郝志峰教授，是郝志峰教授多年来一直的鼓励和支持，才使我能够在读博的道路上坚持走下来。郝志峰教授待人和善，心胸开阔，知识渊博，思维敏捷，不时的奇思妙想和悉心的教导常常使我茅塞顿开。此外，特别感谢他给我提供了广泛的学术交流条件和机会。他谦和的为人和严谨的学术态度，为我树立了一辈子的学习典范，也将激励和鞭策着我在研究和人生的道路上励精图治，开拓创新。

其次，感谢我们研究团队的师兄师姐师妹师弟们，特别感谢黄瀚师兄，多年来一直对我的鼓励和指导。祝愿大家在学术上能够百尺竿头，更进一步。

最后，感谢我挚爱父母亲，夫人和女儿，他们在背后的默默支持一直是我前进的动力。为了论文和实验，我丢掉了与她们一起的陪伴时光，心中甚感愧疚。感谢她们的理解和支持，使我在负重前行的道路上更加坚定有力。祝愿他们身体健康，幸福快乐。



3. 答辩委员会对论文的评语

(主要包括: 1.对论文的综合评价; 2.对论文主要工作和创造性成果的简要介绍; 3.对作者掌握基础理论、专业知识程度、独立从事科研工作能力以及在答辩中表现的评价; 4.存在的不足之处和建议; 5.答辩委员会结论意见等)

费德勒向量是研究图结构与图性能指标的重要工具, 其有效刻画了图的结构特点与图的代数指标的关系, 而图的加边问题是图动态变化引起相关性能变化的研究领域的热点问题之一。

本文分析了费德勒向量与图结构的关系, 从代数的角度研究了网络的结构性质, 并提出了快速有效的最大化代数连通度和传输容量的加边算法, 另外, 还结合社会科学中的结构洞理论提出了最小化网络平均最短路径的加边算法, 本文为大规模网络扩容提供了高效可行的加边算法, 同时为多角度研究网络结构提供思路实验结果表明, 作者提出的方法具有较强的创新性。

论文选题具有一定的前沿性和重要的实用价值。作者查阅了大量国内外相关文献与资料, 对本领域的研究现状和最新进展了解全面, 评述得当。

论文工作反映出作者具有较好的理论基础和深入扎实的专业知识以及独立从事科学研究工作的能力。论文结构完整, 写作规范, 分析严谨, 工作量充足。

作者在答辩过程中叙述清晰, 回答问题正确。

答辩委员会经无记名投票表决, 一致同意李刚通过博士学位论文答辩, 建议授予工学博士学位。

论文答辩日期: 2019 年 5 月 29 日                      答辩委员会委员 5 人  
表决票数: 同意毕业及授予学位 (5) 票;  
              同意毕业, 但不同意授予学位 (0) 票;  
              不同意毕业 (0) 票  
表决结果 (打“√”): 通过 (√); 不通过 ( )  
决议: 同意授予博士学位 (√)      不同意授予博士学位 ( )

答辩委员会成员签名	吴尔宽 (主席)			
	杨晓伟		徐家如	王甲海

答辩秘书签名	何毅
--------	----