



单位代码 10635

学 号 112013321001534

西南大學

硕士学位论文

基于 k-shell 分解的复杂网络节点重要度研究

论文作者：宋起超

指导教师：邓 勇 教授

学科专业：计算机应用技术

研究方向：复杂网络

提交论文日期：2016 年 4 月 22 日

论文答辩日期：2016 年 5 月 29 日

学位授予单位：西南大学

中 国 • 重 庆

2016 年 5 月

独创性申明

学位论文题目：基于k-shell分解的复杂网络节点重要度研究

本人提交的学位论文是在导师指导下进行的研究工作及取得的研究成果。论文中引用他人已经发表或出版过的研究成果，文中已加了特别标注。对本研究及学位论文撰写曾做出贡献的老师、朋友、同仁在文中作了明确说明并表示衷心感谢。

学位论文作者：梁超超

签字日期：2016年6月2日

学位论文版权使用授权书

本学位论文作者完全了解西南大学有关保留、使用学位论文的规定，有权保留并向国家有关部门或机构送交论文的复印件和磁盘，允许论文被查阅和借阅。本人授权西南大学研究生院(筹) 可以将学位论文的全部或部分内容编入有关数据库进行检索，可以采用影印、缩印或扫描等复制手段保存、汇编学位论文。

(保密的学位论文在解密后适用本授权书，本论文：☐ 不保密，☐ 保密期限至 年 月止)。

学位论文作者签名：梁超超

导师签名：邓马

签字日期：2016年6月2日

签字日期：2016年6月2日

目 录

摘要	I
ABSTRACT	III
第一章 绪论	1
1.1 引言	1
1.2 复杂网络概述	2
1.2.1 复杂系统	2
1.2.2 复杂网络发展过程	2
1.2.3 复杂网络统计特性	7
1.3 复杂网络节点重要度	12
1.4 本文组织结构	14
第二章 复杂网络节点重要度	17
2.1 基于节点度中心性的评估方法	17
2.1.1 度中心性	17
2.1.2 半局部中心性	17
2.2 基于路径的评估方法	18
2.2.1 介数中心性	19
2.2.2 接近中心性	20
2.3 基于特征向量的评估方法	22
2.3.1 特征向量中心性	22
2.3.2 α 中心性	23
2.3.3 累计提名法	23
2.4 基于随机游走的评估方法	23
2.4.1 PageRank算法	24
2.4.2 LeaderRank算法	25
2.4.3 HITS算法	26
2.4.4 自动信息汇集算法	28

2.4.5	SALSA算法	28
2.5	基于节点位置的评估方法	29
2.5.1	经典k-shell分解算法	29
2.5.2	MDD算法	29
2.5.3	加权k-shell分解算法	32
2.5.4	最小k-shell算法	32
2.5.5	KCK算法	32
2.6	本章小结	33
第三章	改进的加权k-shell分解算法	35
3.1	k-shell分解算法	35
3.2	加权k-shell分解	35
3.3	改进的加权k-shell分解	36
3.4	实验与分析	42
3.4.1	实验数据	42
3.4.2	k-shell分解实验	43
3.4.3	传播性能实验	44
3.4.4	脆弱性实验	46
3.5	本章小结	50
第四章	基于有效距离的k-shell分解算法	51
4.1	有效距离理论	51
4.2	基于有效距离的k-shell分解方法	52
4.3	仿真实验与分析	54
4.3.1	实验数据	54
4.3.2	Kendall秩相关性实验	55
4.3.3	Susceptible-Infected疾病传播实验	56
4.4	本章小结	57
第五章	总结与展望	59
5.1	本文总结	59
5.2	未来工作展望	60
参考文献		63

发表文章目录	71
简历	73
致谢	75

<http://www.ixueshu.com>

基于k-shell分解的复杂网络节点重要度研究

计算机应用技术 专业 硕士研究生 宋起超

指导教师 邓勇 教授

摘要

随着科技的进步,人们对复杂网络的研究也日渐深入。复杂网络作为一个新兴学科已经融入到我们的日常生活,从具体的交通网、河流网到抽象的社交网、互联网等都属于复杂网络的研究范畴。复杂网络是一门交叉学科,涉及的领域包括物理学、生物学、社会学、管理学等。复杂网络的研究包括复杂网络结构、复杂网络特性、复杂网络节点重要度等。其中,复杂网络节点重要度的研究是这篇文章主要研究方向。复杂网络节点重要度的研究可以为人们生活带来很大的便利,比如人们对交通网络的研究可以帮助人们预防交通拥堵、交通事故等情况,人们对于社交网络的研究可以帮助人们切断重要传染源、预防流行疾病的爆发等。

目前,复杂网络节点重要度的研究已有很多成果,比如度中心性、介数中心性、特征向量中心性、k-shell分解等多种评估节点重要度的算法。虽然人们提出了很多评估节点重要度的方法,但是这些方法在某些情况下都有很大的局限性,人们一直试图寻找更有效、更快捷、更精确的评估重要度的方法。本文首先介绍了复杂网络发展史和复杂网络的相关统计特性。然后介绍了多种节点重要度评估方法,主要包括度中心性评估方法、基于路径的评估方法、基于特征向量的评估方法、基于随机游走的评估方法和基于节点位置的评估方法。随后,提出了两种节点重要度评估方法:第一种,改进的加权k-shell分解方法;第二种,基于有效距离的k-shell分解方法。

本文的工作内容主要包括以下几个方面:

(1) 提出改进的加权k-shell分解方法

原有的加权k-shell分解方法在计算加权度方面还存在不足。于是,本文通过改进节点加权度的计算提出了一种改进的加权k-shell分解方法,且在六个真实的网络Blogs、Email、Net-science、Roget、USAir、Yeast分别进行了k-shell分解实验、SIR传染病实验和脆弱性实验来验证改进方法的性能。在SIR传染病实验中,模型达到平衡状态后,改进的k-shell分解算法比原有的方法所得到的传染节点要多。由此可以看出,改进的k-shell分解算法排序得到的前30个节点的传播能力比原有的方法要强。在脆弱性实验中,大多数网络的实验中改进方法排序得到的前30个节点的网络脆弱性之和比原有的方法排序得到的前30个节点网络脆弱性之和要大,这说明改进方法的排序得到的前30个节点对网络的影响力从总体上来说比原有的方法排序得到的前30个节点对网络的影响力要大,这进一步证实了使用改进算法排序所得到的节点的重要度要比原有的方法有一定的优越性。

(2) 提出基于有效距离k-shell分解方法

传统的k-shell分解算法考虑的都是无权、加权的情况,但是没有考虑有效距离的因素。因此本文提出了基于有效距离的k-shell分解方法。为了验证提出方法的有效性和可行性,

又在四个真实网络C.elegans、etscience、polblogs、USairport 上分别进行了kendall相关性实验和SI疾病传播实验。celegans网络中, 基于有效距离的k-shell分解实验结果明显优于其它两种方法。在netscience网络中经典k-shell分解比基于有效距离k-shell分解结果要好一些。在polblogs网络中基于有效距离的k-shell分解实验结果比其它两种方法优秀。在USAirport网络中, 基于有效距离k-shell分解明显优于其它两种方法。综上所述, 本文提出的基于有效距离的k-shell分解方法在大多数情况性比原有的加权k-shell分解方法和经典k-shell分解方法要优越。

关键词: 复杂网络, 节点重要度, SIR, 脆弱性, 有效距离

Study of nodes influence in complex network based on k-shell decomposition

SongQichao

Supervised by Prof. DengYong

ABSTRACT

With the advent of technology, we have in-depth studies in complex network. As a new-rising subject, it has already come into our daily life now. From the specific aspect to the abstract aspect, such as traffic network, river network, social network, and internet network and so on, all of them are within the scope of the study. Complex network is an interdisciplinary subject. It contains physics, biology, sociology, management and so on. The research contents of complex network contains the structure of complex network, the character of complex network and the nodes influence of complex network and so on. By the way, the research of nodes influence in complex network is the main work of this paper. The study of nodes influence is significant to people's life. The study of traffic network will help us to avoid traffic congestion and traffic accident et al. The study of social network will help human to isolate infection sources and break the chain of transmission.

At present, the research of nodes influence has fruitful achievements, such as degree centrality, betweenness centrality, eigenvector centrality, k-shell decomposition and so on. Although there are so many method to evaluate nodes influence, these method still has its own defect. And researchers are always trying to look for a more effective, more precise and easier method to evaluate nodes influence in complex network. In this paper, firstly, we introduced the history of complex network and related statistical property of complex network. Secondly, we introduced some common methods of evaluating nodes influence in complex network. These common methods include degree centrality method, path-based method, feature vector method, random walker method and nodes location method. And then, we proposed two methods to evaluating nodes influence. One of these method is an improved k-shell decomposition based on potential weights. Another method is k-shell decomposition based on effective distance.

The main work of this paper could be summarized as follows:

- a) Proposed an improved weighted k-shell decomposition

The existing weighted k-shell decomposition is defective in calculating weighted degree. So we improved the existing weighted k-shell decomposition via improving the method of calculating weighted degree in this paper. To verify the performance of improved method, we did some experiments in six real networks, Blogs, Email, Net-science, Roget, USAir and

Yeast. In the SIR spreading experiment, the improved method could get more infected nodes after the model getting balance. The result of SIR spreading method shows that improved method is better than existing method for top-30 nodes. In the vulnerability experiment, the top-30 nodes of improved method could get bigger value than the top-30 nodes of existing method. The result of vulnerability experiment shows that the top-30 nodes of improved method is more influential than the top-30 nodes of existing method. And it further proves that the improved method is superior to existing method.

b) Proposed a k-shell decomposition based on effective distance

Common method based on k-shell decomposition consider factors often including weight and unweight et al. But never take the effective distance into consideration. Now, we proposed a k-shell decomposition based on effective distance in this paper. To verify the performance of our method, we did two experiments in four real networks, C.elegans, netscience, polblogs and USAirport. The experiment results is showed as follows: In C.elegans network, the k-shell decomposition based on effective distance is better than the two other methods obviously. In the netscience network, the classical k-shell decomposition is better than k-shell decomposition based on effective distance. In the polblogs network, k-shell decomposition is better than the two other methods inconspicuously. In the USAirport network, the better than the two other methods obviously. To sum up, the k-shell decomposition method based on effective distance is superior to the exiting k-shell decomposition in most case.

Keywords: complex network, nodes influence, SIR model, vulnerability, effective distance.

第一章 绪论

1.1 引言

本文中一个重要的概念的是复杂网络。然而鄙人才疏学浅，对于复杂网络以的理解与解释可能有失偏颇。希望各位对我的错误进行指正和纠正，本人将虚心接受。众所周知，任何事物的变化都需要一个过程，从量变而达到质变。也许随着科学的发展，复杂网络在未来会变得不复杂，但是现阶段仍然需要研究学者们继续探究。所谓复杂网络，直观上可以理解为复杂的网络。那么什么是网络？什么是复杂的网络？

“网”字来源于象形文字，虽经过千年变化，依然保留了起象形含义，看起来像很多点、线交织在一起，形成了“网”。“络”古汉语字典中解释为网状物，由“网”引申而来。在现实生活中，也随处可以见到网络的身影，小到蜘蛛网、渔网、纱窗等。大到河流网、交通网等,还有抽象意义上的互联网、信息网、神经网络、通信网等。随着现代信息技术和互联网的发展，网络与人们之间产生了密不可分的关联。随着科学的发展，研究学者开始使用网络来探究自然科学。到目前为止，网络大体经历了3个发展阶段。第一阶段，规则网络[1]。网络的第一阶段也是网络刚刚产生萌芽的阶段，该阶段的产生很大程度上要归功于伟大的数学家欧拉。他在18世纪研究了一个著名问题——格尼斯堡七桥问题。格尼斯堡是欧洲的一个小镇，有一次欧拉到达该小镇，发现该小镇上有一个有趣的问题，当地人们无法解决。于是，欧拉采用数学方法将问题进行抽象建模，将格尼斯堡的四块相互陆地抽象成四个点，将架在陆地之间的桥抽象成线，最后得到一个拥有四点、七线的图。格尼斯堡七桥问题用数学语言描述则变为：将图中任意一点设为起点，是否存在一条经过每边一次而又返回起点的回路。而后欧拉通过研究发现，这个问题的答案是否定的。通过对七桥问题的研究，欧拉开创了一个新的数学分支——图论。目前为止，图论对于网络研究有着无可取代的作用。第二阶段，随机网络。事物的年发展需要一个过程。欧拉的七桥问题过去将近两个世纪后，图论发展有了新的突破。20 世纪60 年代，随机图论(random graph theory) 产生。该理论由Erdos 和Renyi 提出,标志着复杂网络系统性理论研究的开端。在随机图理论中，两个节点之间连边与否不再是确定的事情，而是根据一个概率决定[2]。由此生成的网络叫做随机网络，随机网络在接下来的近半个世纪的时间里，被很多研究学者用来研究和分析真实网络，为很多网络研究提供理论基础。第三阶段，复杂网络。随着计算机数据处理和运算能力的飞速发展，随机网络的不足也逐渐显现。研究学者们逐渐发现，现实生活中的大量网络既不是规则的也不是随

机的，这些真实网络所显现的特性也处于一种中间状态。怎样描述、研究这些既不规则，也不随机的网络是一个具有开创性、挑战性的课题。这也标志着复杂网络（complex network）时代的到来[2]。

科学的进步，促进了网络的发展；网络的发展也促进了科学的进步。随着科学的进步与人们对于网络认识的加深，复杂网络成为一个新的研究方向[3]。说到复杂网络，不得不提及复杂系统这一概念。复杂系统的复杂性主要有以下几点：第一点，复杂系统中的个体成员数量非常庞大；第二点，复杂系统中每个个体的行为特征没有统一定论，相对复杂；第三点，复杂系统中个体与个体之间的相互关系非常复杂；第四点，所有个体组成的系统结构十分复杂。面对诸多难题，需要科学的理论方法对复杂系统进行研究探讨。复杂网络作为研究复杂系统的有效工具也应运而生。复杂网络通过研究微观上个体行为、个体相互关系来揭示宏观上复杂系统的总体特征[4, 5]。

1.2 复杂网络概述

1.2.1 复杂系统

复杂系统是一门新兴学科，是科技发展的产物。复杂系统相对于简单系统就好比牛顿经典力学相对于广义相对论，其发展过程是一个从量变达到质变的过程。复杂系统比简单系统规模庞大、结构复杂、关系交错，并且比简单系统有智能性、自组织性、自适应性等。由于复杂系统包含的因素很多，目前对于复杂系统的定义没有统一定论。复杂系统是一个多学科交叉的系统，其研究方法也不同于以往的研究方式，传统的使用数学推理的方法已经失效。但是随着计算机计算性能的提高，人们开始采用计算机对复杂系统进行模拟研究。复杂系统的发展主要经历了三个阶段：简单系统、随机系统和复杂系统[6]。简单系统的特点是包含的元素数目少，元素间的关系简单，具有可控性、可预见性，可以通过经典的图论方法进行研究。随机系统比简单系统包含的元素多、元素间的相对关系也不再单一，通常采用随机图论方法对其进行研究。复杂系统比随机系统更复杂，其包含的元素众多、信息量极大、元素间的相互关系非常复杂，具有智能性、自组织性、自适应性、不稳定性、非线性、不确定性和不可预测性等特性，已经不能采用单一的、传统的方法进行了[7]。复杂系统涉及众多学科，比如生物学、物理学、社会学、经济学、计算机学等等。随着科技的发展，人们将陆续解开复杂系统的神秘面纱，让复杂系统为人们的生活提供更加便利的理论指导[8]。

1.2.2 复杂网络发展过程

网络的发展经历了二百多年的时间，大体分为三个阶段，其体系也日渐成熟。（1）第一阶段：图论的产生。1736年，伟大的数学家、图论之父欧拉到访普

鲁士东部的小镇Königsberg。该小镇有一条河流穿过，该河流将小镇分割成四部分，有七座桥将小镇的所有陆地链接在一起，小镇的构型与桥的分布如图1.1所示。当地人们中流传着一个有趣的问题：怎样能将七座桥走遍，每座桥只经过一次，最后回到起点。欧拉对这个问题产生了极大的兴趣。他将这个问题用数学方法进行抽象并建立模型，将四块被分割的大陆抽象成四个点A、B、C、D，七座桥分别抽象成连接这四个点的线段，如图2所示。如此，则问题用数学语言描述为，在图2中是否存在一条从任意点出发、经过每边一次又返回起点的回路？欧拉认为倘若存在一条这样的回路，则图1.2中任意一点都有偶数条边。然而图1.2中所有点的边都是奇数，所以不存在这样的回路。欧拉对这一问题的解决为数学这一学科开创了新的分支——图论。现如今，人们对复杂网络的研究也是建立在图论这一基础上的，图论的发展也不断推动着人们对网络的认识和研究。

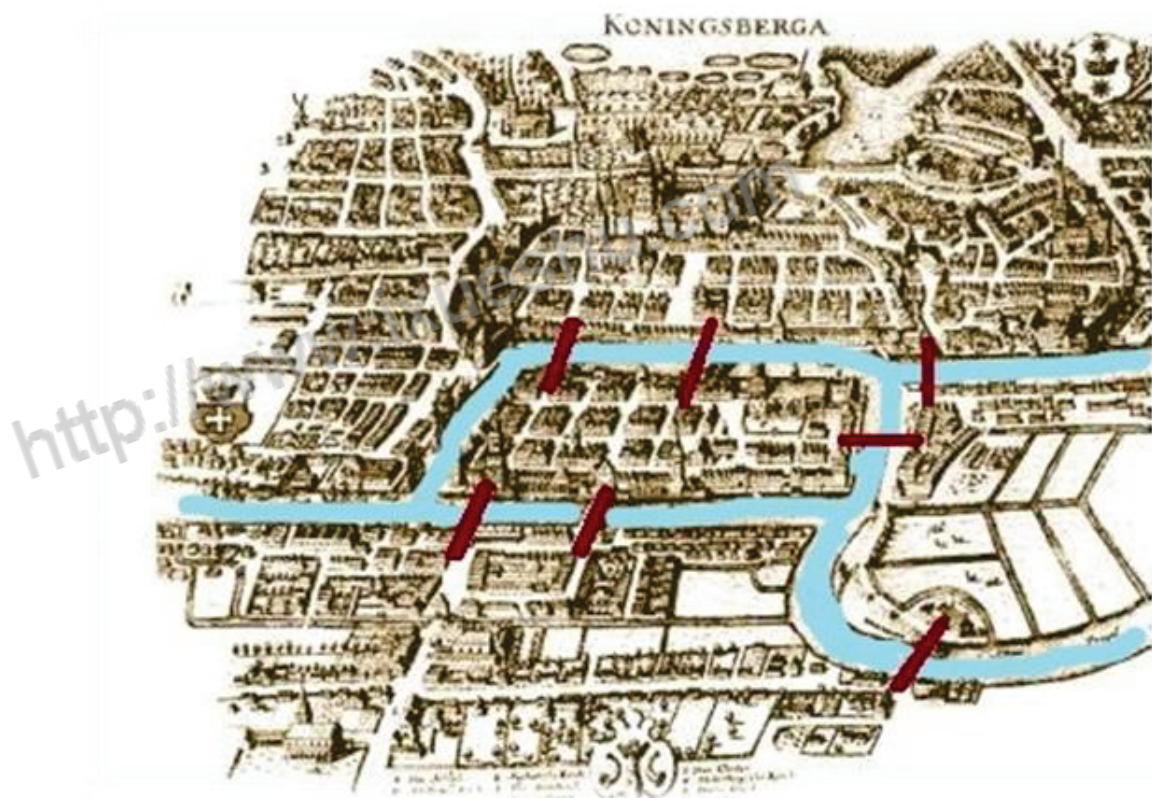


图 1.1: 格尼斯堡小镇

欧拉时代研究的主要是规则网络。所谓规则网络是指点与点之间是按照一定的规则进行确定的链接。如图1.3为规则网络一个示例。（2）第二阶段：随机图论的产生与发展。在欧拉创立图论后近两百年的时间里，图论并没有太大的研究进展。直到1936年，也就是欧拉访问格尼斯堡200年后，有一本名为《ON THE

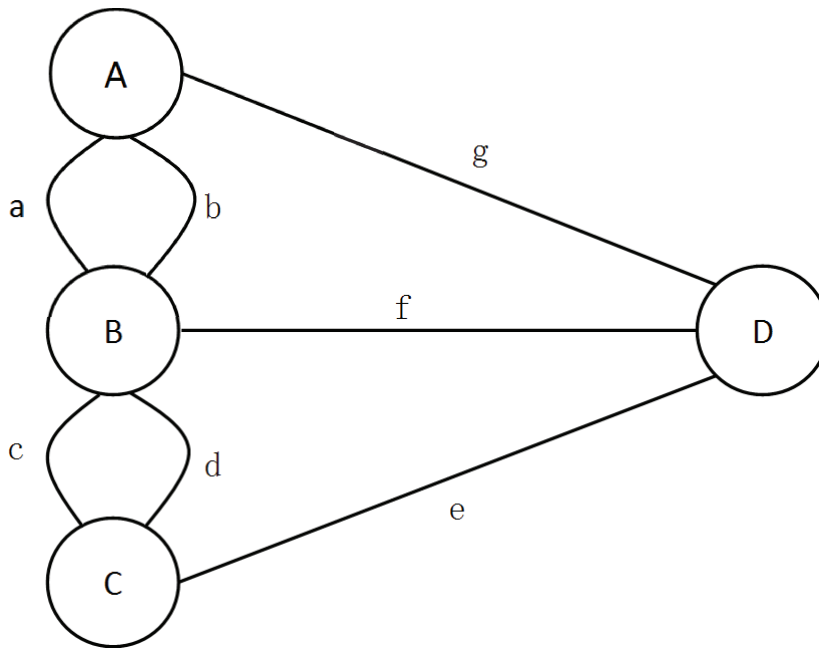


图 1.2: 格尼斯堡小镇抽象图

EVOLUTION OF RANDOM GRAPHS》的书籍问世了，该书由Erdos和Renyi撰写而成。此书的出版标志着随机图论的产生，对图论的研究发展有着重大意义[9]。随机图论与经典图论的不同之处在于随机图论将概率论与图论相结合，认为任意两点之间链接的建立并不是固定的，而是有一定的概率 p 。设一个网络有 N 个节点，每隔时间 t ，任选两点以概率 p 进行链接。图1.4为随机图的一个图例。在随机图模型中，节点度的分布服从Poisson分布。在此后相当长一段时间内，随机图论都是研究网络的主要理论之一[10]。

(3) 第三阶段：复杂网络的发展。所谓的复杂是相对简单而言的，复杂网络也属于网络的一种。复杂网络的诞生与信息技术和计算技术的发展是脱离不开的。在信息技术没有飞速发展之前，人们发现复杂性问题的技术达不到要求。在信息科学得到长足发展之后，人们有了优良的技术工具，看待问题、发现问题的角度得到了质的飞跃，以前没有发现的经典网络的不足之处也一一显露出来。1998年，Watts和Strogatz在Nature上发表了一篇重要文章——Collective Dynamics of ‘Small World’ Networks[8]。Watts和Strogatz在这篇文章中构建了一个小世界网络，数据来源与哈佛大学Stanley Milgram所做的小世界实验，同时这篇文章中还揭示了小世界网络具有较高的聚合系数何较短的平均路径长度[11]。图1.5是WS小世界网络模型。图1.6是NW小世界网络模型。

Stanley Milgram是美国一名社会心理学家，通过大量社会调查发现世界上任意两人之间的平均距离不超过6。按照他的理论，我们最多只需要再认识5个人，就能认识美国总统奥巴马。这个理论也被称为六度分离（Six Degrees of

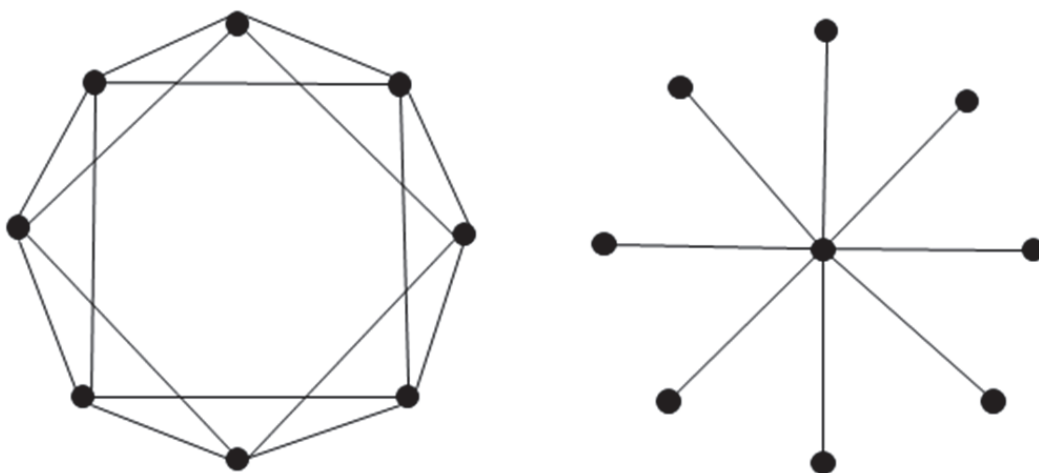


图 1.3: 规则图示例

Separation)。在Milgram的第一次社会实验中，他招募了约160个志愿者，通过志愿者的亲人、朋友等为他把信送达马萨诸塞州的一位证券经纪人。事实上，参与此次实验的志愿者只有50个左右，而最后被送到这位证券经纪人手里的信也只有三封。第一次实验由于多重因素导致成功率很低，但是Milgram经过多番摸索和实验，最终送出了300多封信并将成功率提高到了97%。他从自己所做过的这些实验中，得出了一个结论：某个特定世界很小[12]。

Milgram的实验过程和结果有一定可信度，然而并不完善，不足以使人信服。为了验证“六度分离”的正确性，研究学者们又做了大量实验。Kevin Bacon游戏是众多实验中比较出名的一个。Kevin Bacon 是20世纪60年代美国的著名演员。Kevin Bacon游戏的主要思想是想把其他的任意的演员与Bacon 建立联系，不管是认识Bacon还是不认识Bacon的。实验结果发现，所有演员的与Bacon的距离的平均数不超过6。Bacon游戏再一次验证了六度分离假设的正确性。

演艺圈里有Bacon游戏，科研界里也有自己的小世界游戏——Erdos 数。Erdos对数学的贡献是数学史上辉煌的一笔，他也因此而名流青史。Erdos一生著作繁多，是一个高产高质的科研工作者，与他合作过的人也不计其数[13]。Erdos数的定义是这样的：与Erdos本人直接合作发表过文章的人的Erdos数为1，而与Erdos数为1的人直接合作发表过文章的人的Erdos 数为2，以此类推。Erdos数的实验结果同样也证明了“小世界”理论的正确性。比如与Erdos 完全不认识的比尔盖茨，其Erdos为4。科学巨匠爱因斯坦的Erdos 数为2。

上述三种小世界实验虽然在一定程度上验证了小世界特性，然而其规模相对较小、网络构造也相对简单，不具有推广性。为解决这一不足之处，Watts 等人于2001年利用Internet 网络建立了一个世界性的大规模实验。他们建立了一个专

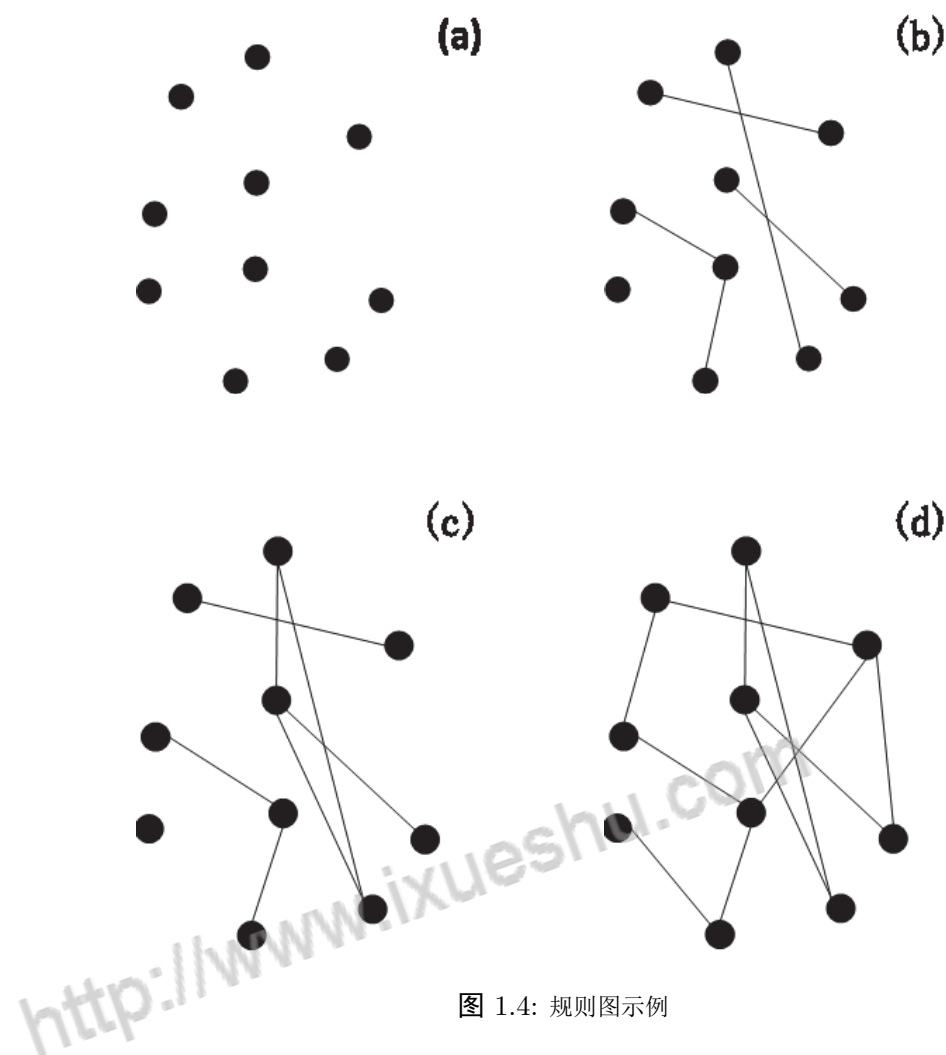


图 1.4: 规则图示例

门用于做小世界实验的网站。该网站的注册人群结构合理，数量庞大。志愿者们注册登录网站后，会被要求向特定的目标用户发送邮件。如果志愿者认识目标对象，则直接发送邮件。如果志愿者不认识目标对象，则向网站推荐合适的朋友，网站再联系他的朋友，向目标用户发送邮件。此次实验耗时较长，规模也比以往大很多，包含了13个国家，6万多名志愿者。据统计，志愿者平均转发57次，便可将邮件送达目标用户。

继Watts发表的小世界文章之后，Albert和Barabasi于1999年发表了一篇名为《Emergence of Scaling in Random Networks》的文章，文中揭示了网络的无标度特性[14]。同时也掀起了复杂网络研究的新一轮热潮。如图1.7所示，是BA网络的演化过程示例。

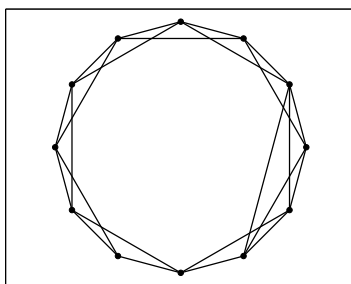
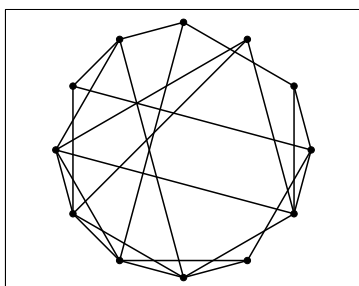
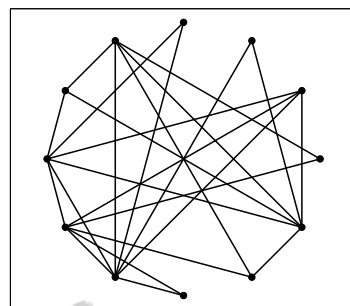
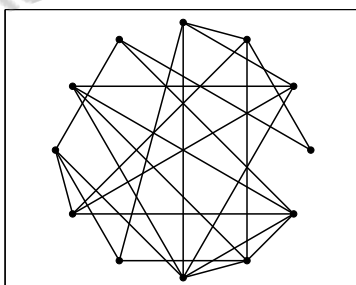
(a) $p=0.1$ (b) $p=0.3$ (c) $p=0.7$ (d) $p=1.0$

图 1.5: WS小世界网络

1.2.3 复杂网络统计特性

1.2.3.1 网络的表示

通过对复杂网络的研究，人们发现了复杂网络的多种统计特性，比如平均路径长度、聚类系数、度分布等[15–17]。为表述方便，设网络 $G = (V, E)$ ，其中 V 表示网络所有节点集合， E 表示网络所有边的集合， n 为所有点的总数和， m 为所有边的总数和。图实际上是节点与节点之间二元关系的集合。网络 G 中任意两

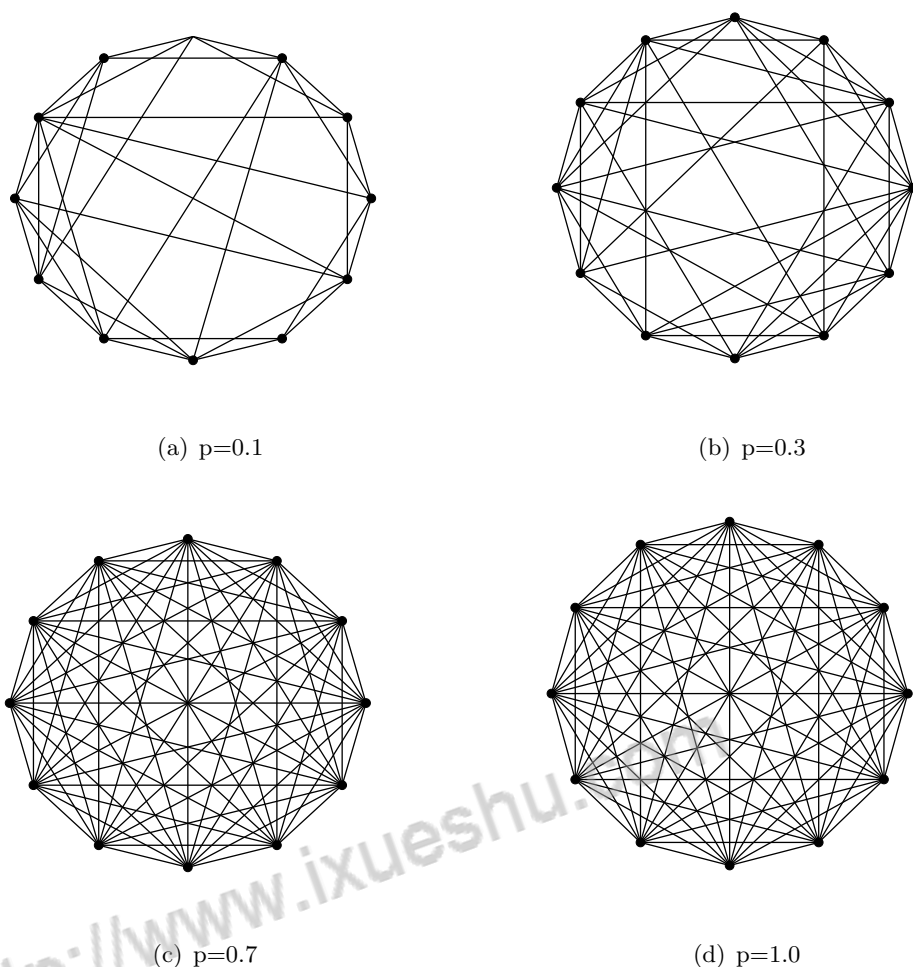


图 1.6: NW小世界网络

点 V_i 、 V_j 所对应的边 e_{ij} 与 e_{ji} 相同, 则该网络为无向网络 (undirected network)。否则, 若网络 G 中存在两点 V_i 、 V_j 所对应的边 e_{ij} 与 e_{ji} 不相同, 则该网络为有向网络 (directed network)。图为有向网络和无向网络示意图。若网络 G 中每条边都赋有权重值, 则该网络为有权网络 (weighted network)。若网络 G 中的边都没有权重值, 则该网络为无权网络 (unweighted network)。权重值在不同的网络中由不同的含义。在交通网中, 权重值可以表示每条街道的车流量或人流量。在社交网络中, 权重值可以表示每个人之间关系的影响力。图1.8为各种类型网络示意图。

为了研究方便, 人们使用矩阵对图进行表示。常用的两种矩阵有邻接矩阵和关联矩阵, 其中邻接矩阵的应用最为广泛。邻接矩阵用来描述网络中节点与节点之间的连通关系, 能反应网络整体的拓扑结构。对于一个含有 n 个节点网络, 需

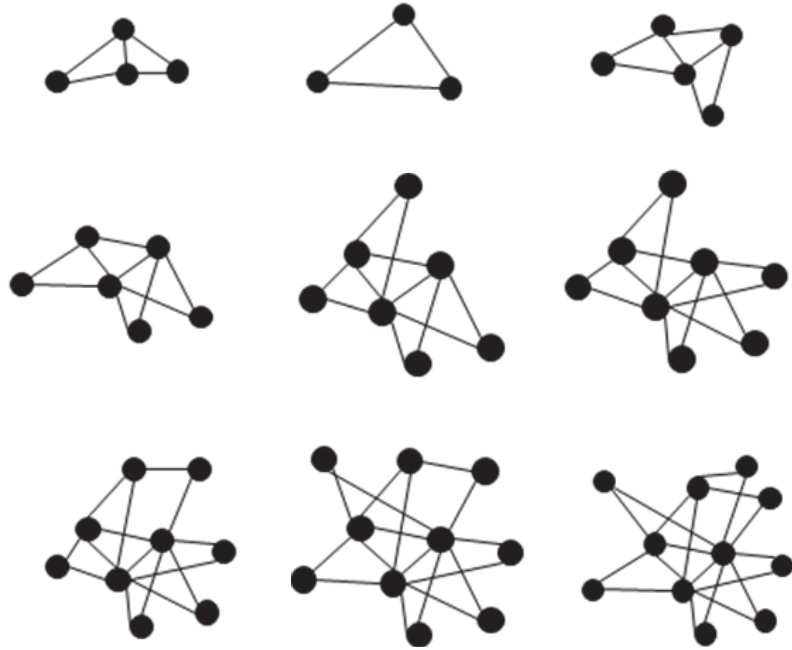


图 1.7: BA网络演化示意图

要构造一个 $n \times n$ 的邻接矩阵:

$$Mat = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix}$$

对于无向无权网络, 矩阵Mat中第i行、第j列的元素的取值规则如下:

$$a_{ij} = \begin{cases} 1 & \text{若第i节点与第j节点相互连通} \\ 0 & \text{若第i节点与第j节点不连通} \end{cases}$$

对于无向加权网络, 矩阵Mat中第i行、第j列的元素的取值规则如下:

$$a_{ij} = \begin{cases} w_{ij} & \text{若第i节点与第j节点相互连通} \\ 0 & \text{若第i节点与第j节点不连通} \end{cases}$$

除了邻接矩阵外, 还有一种常用的表示网络的方式, 那就是关联矩阵。关联矩阵包含的是网络节点和边的邻接关系, 比邻接矩阵包含更多的网络结构信息。设G是一个含有n个节点, m条边的网络, 则其关联矩阵Matrix是一个 $n \times m$ 的矩

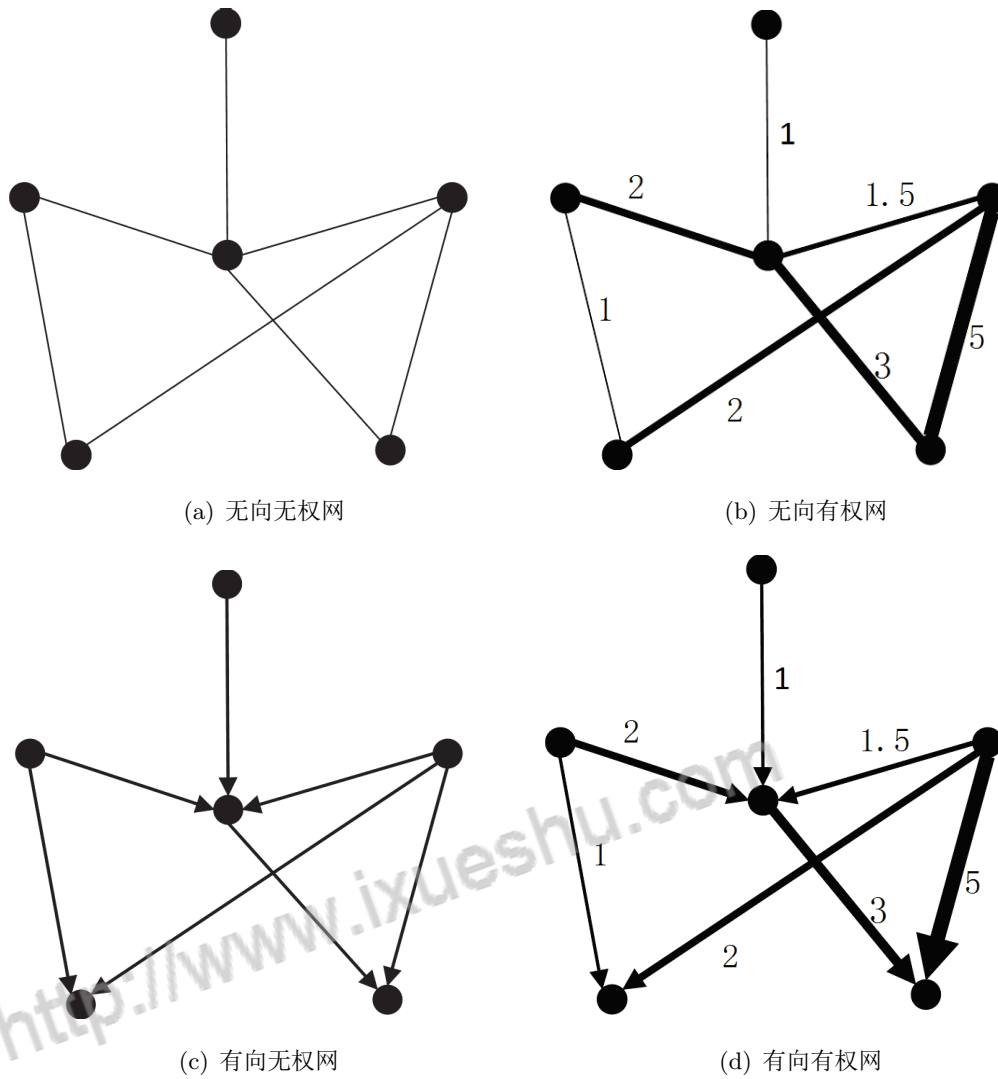


图 1.8: 各类型网络示意图

阵，其中每个元素的定义如下：

$$a_{ij} = \begin{cases} 1 & \text{若第} i \text{节点与第} j \text{条边相互联结} \\ 0 & \text{若第} i \text{节点与第} j \text{条边不联结} \end{cases}$$

图1.9为邻接矩阵与关联矩阵表示的例子。其邻接矩阵表示为：

$$Mat1 = \begin{bmatrix} 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \end{bmatrix}$$

因图1.9中含有5个节点、6条边，其关联矩阵是一个 5×6 的矩阵，如下所示：

$$Mat2 = \begin{bmatrix} 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

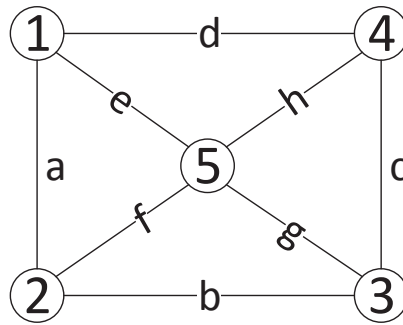


图 1.9: 邻接矩阵与关联矩阵示例

1.2.3.2 测地线和网络直径

测地线这一概念来源于大地测量学。对于网络中任意两点 V_i 和 V_j ，点 V_i 与 V_j 之间的测地线是两点之间的最短距离。对于无权网络，两点之间的测地线是两点之间边数最少的一条通道。对于加权网络，两点之间测地线是两点之间最小的加权通道。设 L 表示从点 V_i 到点 V_j 之间所有通道的集合。则测地线表示为

$$d_{ij} = \text{Min}(L)$$

网络直径，是指网络中任意两点 V_i 、 V_j 之间距离的最大值。网络直径 D 表示为

$$D = \text{Max}(L)$$

1.2.3.3 平均路径长度和网络效率

平均路径长度定义为：网络中所有节点之间测地线的平均值，表示为

$$\langle l \rangle = \frac{1}{\frac{1}{2}N(N-1)} \sum_{i \neq j} d_{ij}$$

网络效率可以反映网络的平均连通度的好坏。其定义为：

$$E = \frac{1}{N(N-1)} \sum_{i \neq j} d_{ij}$$

1.2.3.4 真实网络统计特性

随着复杂网络学科的发展，人们对各种真实网络的统计特性进行了总结分析[16]。如表3.8所示，其中 N 表示各网络节点的个数， M 表示网络中的边数， $\langle k \rangle$ 表示网络的平均度， L 表示网络的平均路径长度， C 表示网络的平均聚类系数。各种网络详细信息见文献[18]。

表 1.1: 真实网络统计特性

网络	类型	N	M	$\langle k \rangle$	L	C
Adjnoun	有向	112	425	3.795	2.599	0.086
AS-22july06	无向	22963	48436	4.219	3.905	0.350
Celegans	有向	297	2345	7.896	3.992	0.169
Cond-mat	无向	16726	47594	5.691	6.628	0.737
Dolphins	无向	62	159	5.129	3.357	0.303
Football	无向	115	613	10.661	2.508	0.403
Karate	有向	34	78	2.294	1.274	0.285
Lesmis	有向	77	254	3.299	2.400	0.287
Netscience	无向	1589	2742	3.451	5.823	0.878
Polblogs	有向	1490	19025	12.768	3.390	0.172
Power	无向	4941	6594	2.669	18.989	0.107

1.3 复杂网络节点重要度

复杂网络节点众多、结构复杂、个体行为复杂，如此复杂的网络，应该采用什么方法了解网络结构、衡量网络复杂度呢？应该采用什么方法评估个体对整体网络的影响呢？研究学者们为了了解网络整体结构、网络内部联系和网络性能等，提出了很多方法，比如级联性、传播性和同步性等。复杂网络节点重要度评估在复杂网络研究领域是一个有深远影响和重要意义的领域。正如上述提到复杂网络是一个交叉学科，复杂网络节点重要度的研究也具有多学科交叉性[19]。研究复杂网络节点重要度在很多不同的领域都具有重要意义，比如互联网、社交网、交通网、电力网等领域。这对于防控网络攻击、控制谣言传播、控制交通拥堵、控制传染病的扩散将产生重要影响[20–23] 倘若我们可以快速、有效地找到某个重要节点或某些重要节点群，即使对这些节点采取相应措施，就可以避免或减少损失。所谓重要节点指的是那些对网络整体结构有重大影响的节点[24, 25, 19]。如何有效、高标地识别重要节点对于很多领域都具有重要的实际意义，例如在生物学、信息学、医学、计算机学、物理学、经济学等领域。2014 年，在韩国爆发的中东呼吸综合征EMRS(Middle East Respiratory Syndrome)，对韩国乃至世界人民的生命健康都产生严重的威胁。如何找到有效措施控制传染源和切掉传染途径是一个急需解决的问题[26, 27]。复杂网络节点重要度也称为节点中心性。研究至

今,已有很多研究成果。研究学者们根据不同网络和需求采用不同的评估方法来刻画节点的重要程度。下面对这些评估方法进行简单介绍。

第一类:基于节点度中心性的评估方法。基于节点度中心性的评价方法是所有评价方法中最为简便、直观的评价方法。顾名思义,这种评价方法与节点的度有关系。基于节点度中心性的评价方法是利于网络节点的度来判断节点在网络中的影响力。节点度越大,该节点的重要性越高;节点的度越小,该节点重要性越小[28]。当然,也有学者认为节点的重要性不仅仅由自己的自身的度决定,还受到邻居节点的影响[29]。基于节点度中心性的评价方法,有一定理论意义和实际应用价值。然而,该方法也有一定的局限性,该评价指标只着眼于局部的节点和节点的度,所考虑的因素不够全面[30]。

第二类:基于路径的评估方法。路径是评估节点重要度的一个重要指标,这里的路径既指节点之间的距离,又指节点与节点之间节点的个数[31]。在交通网中,节点间的距离将会影响节点之间信息传递的效率。同样流量的话,距离越长,效率越慢;距离越短,效率越快。在社交网络中,两个节点之间节点的个数也会影响到信息传递的效率和准确性。其它因素相同的条件下,节点数目越多,所需的中转时间越长,效率越慢;节点数目越少,所需的中转时间越短,效率越高。基于路径的评价指标又可以详细的分为离心中心性指标、接近中心性评价指标、介数中心性评价指标、社团中心性评价指标等[32]。

第三类:基于特征向量的评估方法。特征向量这种评估方法不同于基于度中心性评估标准和基于路径的评估标准,它利用网络的邻接矩阵,计算出网络的最大特征值,这样既考虑了节点邻居的数量,又考虑了节点邻居的重要性。特征向量中心性更加强节点所处的周围环境(节点的邻居数量和质量),它的本质是一个节点的分值是它的邻居的分值之和,节点可以通过连接很多其他重要的节点来提升自己的重要性,分值比较高的节点要么和大量一般节点相连,要么和少量其他高分值的节点相连[? 33]。从传播的角度看,特征向量中心性适合于描述节点的长期影响力,如在疾病传播、谣言扩散中,一个节点的特征向量中心性约大说明该节点距离传染源更近的可能性越大,是需要防范的关键节点[34]。

第四类:基于随机游走的评估方法。随机游走主要是用来分析、解决互联网中各网页间的关系。随机游走模型通过分析用户对网页的访问量来判断网页的重要程度。如果将互联网看作一个大型网络,则每个网页可以抽象成网络中的点,某一点被访问的频率越高,则该点的重要度越大。随机游走模型中的常见算法有PageRank、LeaderRank、HITS等算法。PageRank顾名思义是对网页进行排序的算法,该算法中每一个网页的跳转概率都是一样的。但是现实中,每个网页的跳转概率由于种种原因,其跳转概率是不一样的[35]。而且PageRank算法中,参数的获取需要实验获取,取值也随条件改变而改变。为了解决这些问题,

LeaderRank 算法应运而生[36]。LeaderRank算法通过添加节点在网络中的双向边和添加背景节点对PageRank算法进行了改进，这样便解决了PageRank算法中所存在的问题[37]。

第五类：基于节点位置属性的评估方法基于节点位置属性的评估方法，从名字可以看出，该方法与节点在网络中的位置有关。基于节点位置属性的评估方法根据节点在网络中的位置判断节点的重要度。节点位置越处于中心，则其重要性越强。2010年，Kitsak等人提出了k-shell 分解方法，该方法通过判断节点在网络中的位置判断节点的重要度[38]。节点的k-shell值越大，则节点越重要；反之，则节点越不重要[39]。Kitsak 通过实验证明该方法比度中心性和介数中心性的方法在评估节点重要度方面更优秀。在随后的几年，很多研究学者对该方法进行了改进，陆续提出了MDD方法、加权k-shell 方法等方法[40–42]。

1.4 本文组织结构

本文共有五章内容详细介绍如下：

第一章，绪论。主要介绍了本文的研究背景和复杂网络相关基础理论。

第二章，复杂网络节点重要度。主要介绍了五大类复杂网络节点重要度评估方法，包括基于节点度中心性的评估方法、基于路径的评估方法、基于特征向量的评估方法、基于随机游走的评估方法、基于节点位置的评估方法。

第三章，改进的加权k-shell分解方法。本章对原有的基于边的加权k-shell 分解方法进行了改进。为了验证改进方法的性能，我们分别在六个真实网络上进行了k-shell分解实验、SIR 疾病传播实验和脆弱性实验。实验结果验证了改进方法的有效性和实用性。

第四章，基于有效距离的k-shell分解方法。本章对k-shell分解方法引进了有效距离理论。为了验证改进方法的性能，我们分别在四个真实网络上进行了kendall相关性实验和SI 疾病传播实验。实验结果验证了基于有效距离k-shell 分解方法的有效性和实用性。

第五章，总结与展望。本章对已完成的工作进行了总结与分析，展望了下一步研究的主要方向和思路。

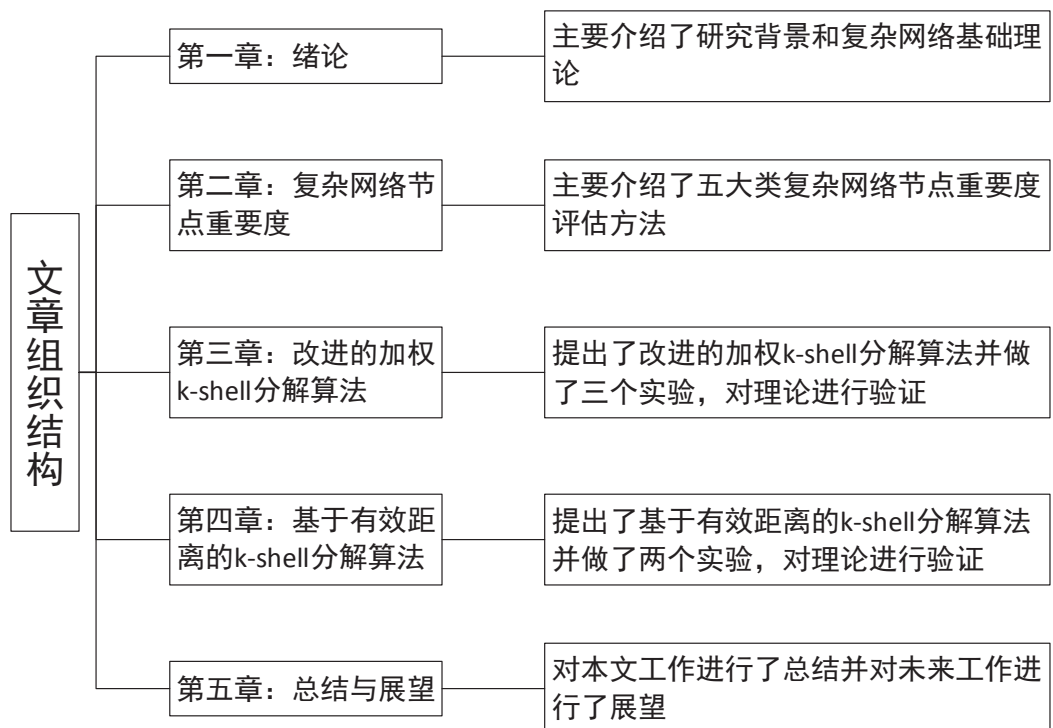


图 1.10: 本文组织结构示意图

第二章 复杂网络节点重要度

复杂网络作为多领域交叉学科，对于帮助人们解决很多问题有重大的理论和实际意义。复杂网络节点重要度的研究对预防和控制疾病传染、控制谣言传播、预测金融危机、解决交通堵塞、揭示生命体内在机制等将产生重要影响[43, 44]。本章主要对一些常见的复杂网络节点重要度评估方法进行相对详细的阐述。

2.1 基于节点度中心性的评估方法

2.1.1 度中心性

节点的度是指节点的邻居节点的总和。某一节点的度越大说明该节点邻居节点越多，则该节点的影响力就越大[19]。所以，节点的度越大，该节点的重要性就越大。一直以来，节点的度中心性都是评估节点重要性的最重要的指标之一。该评估方法计算简单、直观明了、算法复杂度低，很容易运用[45]。在一个拥有 N 个节点的网络中，任意几点 i 的度中心性 C_i 定义如下：

$$C_i = \frac{\sum_j^N x_{ij}}{N-1}$$

其中 N 表示网络中节点的个数， j 表示除节点 i 之外的所有节点。 x_{ij} 表示节点 i 与节点 j 的连接状态，若节点 i 与节点 j 相连，则 $x_{ij} = 1$ ；反之， $x_{ij} = 0$ 。

在有向网络中有两种不同的情况，分为入度和出度。图2.1是入度、出度示例，其中节点A的出度为4、节点B的出度为0、节点A的入度为0、节点B的入度为4。度中心性指标在评价节点重要度方面简洁易懂、操作方便，但是该方法也存在一些问题。度中心性指标知识着眼于某个节点的局部影响力，而未考虑全局因素，比如节点在整体网络中的位置，节点的邻居节点的度等因素。于是，后续很多研究学者对该方法进行了改进，比如王建伟等人提出了基于局部特征的网络节点重要度评价方法，chen等人提出了半局部中心性[46]。

2.1.2 半局部中心性

2010年，王建伟等人提出了一种基于局部特征的网络节点重要度评价方法。该方法不仅考虑了节点本身的度，而且考虑了其邻居节点的度。2011年，Chen等人提出了一种半局部中心性方法（semi-local centrality）[46]。半局部中心性方法增加了一个多阶邻居信息指标来评价节点的影响力，半局部中心性的定义如下：

$$SLC_i = \sum_{j \in \Psi(i)} \sum_{m \in \Gamma(j)} N(m)$$

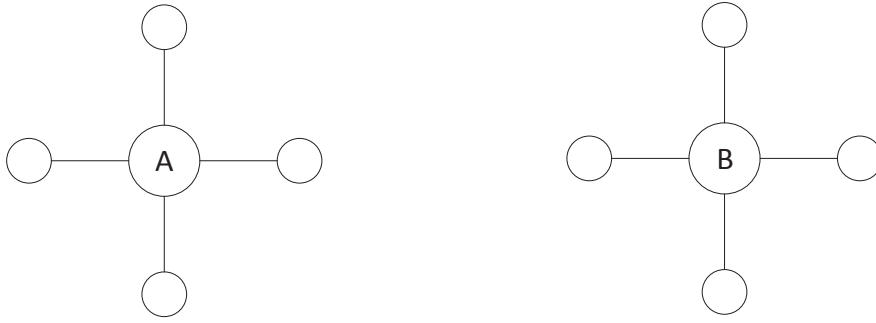


图 2.1: 入度、出度示例

其中 $\Psi(i)$ 为节点 i 的邻居节点集合， $\Gamma(j)$ 为节点 j 的邻居节点集合， $N(m)$ 为节点 m 最近邻居节点数目与次近邻居节点数目之和。半局部中心性评价方法虽然能更好的评估节点重要度，但是该方法需要计算多阶邻居导致算法计算复杂度和算法时间复杂度都大大提高，在实际应用中会受到限制。

2.2 基于路径的评估方法

复杂网络中存在着一些度较小的不起眼的节点，但在网络中发挥重要作用的节点。举个例子，图2.2中节点5的度在该网络中是最小的，但是节点5在图2.2中是不可或缺的，该节点为作用两边的节点充当桥梁的作用，是左边与右边进行信息交流的毕竟之路。若单纯的用度中心性来衡量节点5的重要性是不科学的，因此基于路径的节点重要度评估方法应运而生。基于路径的评估方法主要是通过评估网络节点对信息流的控制能力来判断节点的重要性[47]。节点对信息流的控制能力越强，则节点重要度越高；节点对信息流的控制能力越弱，则节点重要度越低。基于路径的评估方法主要考虑的是节点与节点之间的最短路径[48]。而能影响最短路径的因素有两个：起点与终点之间线路的长度和起点与终点之间节点数目的多少。举个例子说明，图2.2中，节点1设为起点，节点7设为终点，则起点与终点之间四条线段a、b、c、d的长度会影响节点5对信息流的控制，同样的，起点与终点之间线段的数目也会影响节点5对信息流的控制。

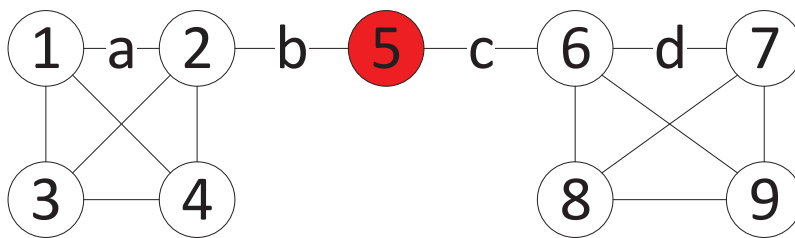


图 2.2: 桥节点示例

在交通网中，节点间的距离将会影响节点之间信息传递的效率。同样流量的话，距离越长，效率越慢；距离越短，效率越快。在社交网络中，两个节点之间节点的个数也会影响到信息传递的效率和准确性。其它因素相同的条件下，节点数目越多，所需的中转时间越长，效率越慢；节点数目越少，所需的中转时间越短，效率越高。基于路径的评估方法又可以详细的分为离心中心性方法、接近中心性评价方法、介数中心性评价方法、社团中心性评价方法等。

2.2.1 介数中心性

2.2.1.1 介数中心性

介数中心性是一个较早的概念，早在1978年，费曼便提出了介数中心性用来评估节点重要度[49]。假设一个网络中最短路径的数目是一定的，若通过某一节点的最短路径数目越多，该节点的介数中心性越大[50]。设网络 $G = (V, E)$ 中有 N 个节点，则节点 i 的介数中心性定义如下：

$$BC(i) = \sum_{i \neq s, i \neq t, t \neq s} \frac{g_{st}^i}{g_{st}}$$

其中， g_{st} 为从节点 s 到节点 t 的所有最短路径的数目， $g_{st}(i)$ 为节点 s 到节点 t 的最短路径中经过节点 i 的最短路径的数目。

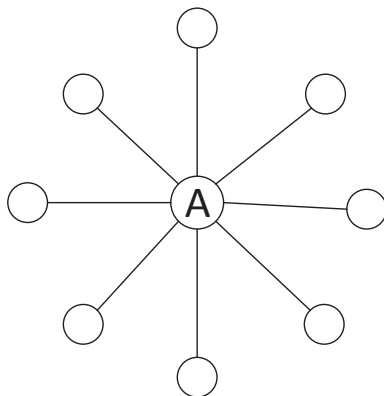


图 2.3: 星型网络

与很多数学问题一样，为了研究方便可以对网络中所有的节点进行归一化。若网络 G 是一星型结构的网络如图2.3所示，中心节点 A 的度为 $N-1$ ，通过节点 A 的最短路径数目为 $(N-1)(N-2)/2$ 。外层节点没有最短路径通过，外层节点的介数都为零。由此可得网络中则归一化后的介数为：

$$C_B(i) = \frac{2}{(N-1)(N-2)} \sum_{i \neq s, i \neq t, t \neq s} \frac{g_{st}^i}{g_{st}}$$

介数中心性是通过计算节点对信息流的控制力来衡量节点重要性的，因此介数中心性在通信网、交通网、电力网等方面有较广泛应用[51, 52]。然而该方法复杂度、计算耗时，在应用时受到一定的限制。介数中心性考虑的是最短路径上信息流的大小，然而在实际情况中只考虑采用最短路径是不切实际的。比如图2.4中，假设图中各个边的权重均为1，节点1到节点7的最短路径为1- > 4- > 7，另外还有两条次短路径分别为1- > 2- > 3- > 7和1- > 5- > 6- > 7。如果我们只采用最短路径，则会是另外两条次短路径处于闲置状态造成资源浪费。而且如果从节点1到节点7的信息流超过最短路径的负荷时，会造成最短路径信息流拥堵，因此在实际应用中，人们通常不会只使用最短路径[53-55]。

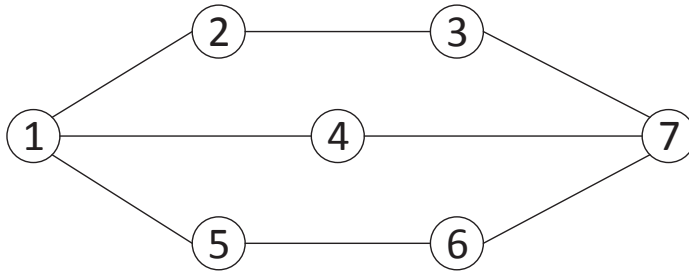


图 2.4: 星型网络

2.2.1.2 流介数中心性

为解决图2.4中的问题，Yan G等人提出了流介数中心性[56]。流介数中心性实际上是对介数中心性的改进，其中中心思想是：在一个网络所有不相同路径中，经过某一点的比例越大，则该点的重要性越大。流介数中心性的定义为：

$$FBC(i) = \sum_{i < j} \frac{g_{ij}^k}{g_{ij}}$$

其中, g_{ij} 为网络中节点 v_i 与 v_j 之间的路径总数, g_{ij}^k 为节点 v_i 与 v_j 之间经过 v_k 的路径数.流介数中心性与介数中心性的不同之处在于：介数中心性只考虑信息流所经过最短路径，而流介数中心性则考虑信息流会经过所有的路径。图2.5与图2.6分别为介数中心性与流介数中心性的计算示例。

2.2.2 接近中心性

接近中心性中的“接近”二字，顾名思义与距离有关。接近中心性的思想是：对于网络中任意一点 i ，计算节点 i 与网络中其它节点之间平均距离，平均距离越小的点，节点的接近中心性越强，平均距离越大的点，节点的接近中心性越弱[57]。节点与其它节点的平均距离越小，则节点与节点之间信息交流的效率越

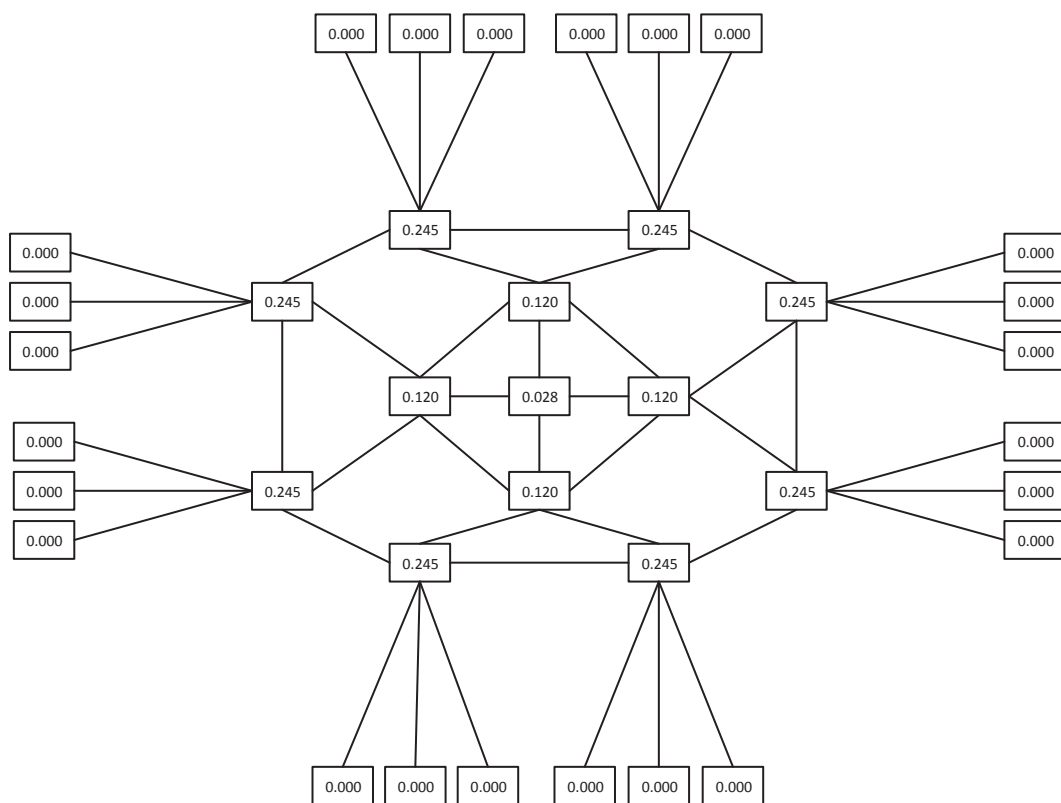


图 2.5: 介数中心性计算示例

高, 从而使得接近中心性越强。节点与其它节点的平均距离越大, 会降低节点与节点之间信息交流的效率, 使得接近中心性越弱[28]。设网络 $G = \langle V, E \rangle$ 中包含 N 个节点, 则对于网络中任意一点 i 的接近中心性定义如下:

$$C(i) = \frac{1}{\sum_j^N d_{ij}}$$

其中, d_{ij} 表示节点 i 与节点 j 的测地线。

网络直径定义为网络 G 中所有节点的离心中心性中的最大值, 网络半径定义为所有节点的离心中心性值中的最小值. 显然, 网络的中心节点就是某个或某些离心中心性值等于网络半径的节点, 一个节点的离心中心性与网络半径越接近就越中心. 要强调的是, 网络直径在复杂网络研究中还有多种不同的定义, 例如Albert 等人研究万维网的时候定义网络直径为网络中所有节点对的最短路径的平均值. 离心中心性的缺点是极易受特殊值的影响, 如果一个节点与大部分节点的距离都很小, 只与极小部分节点的距离很大, 这个节点的离心中心性仍然会取其中的最大值. 接近中心性则采取距离平均值的方式克服了这一缺点.

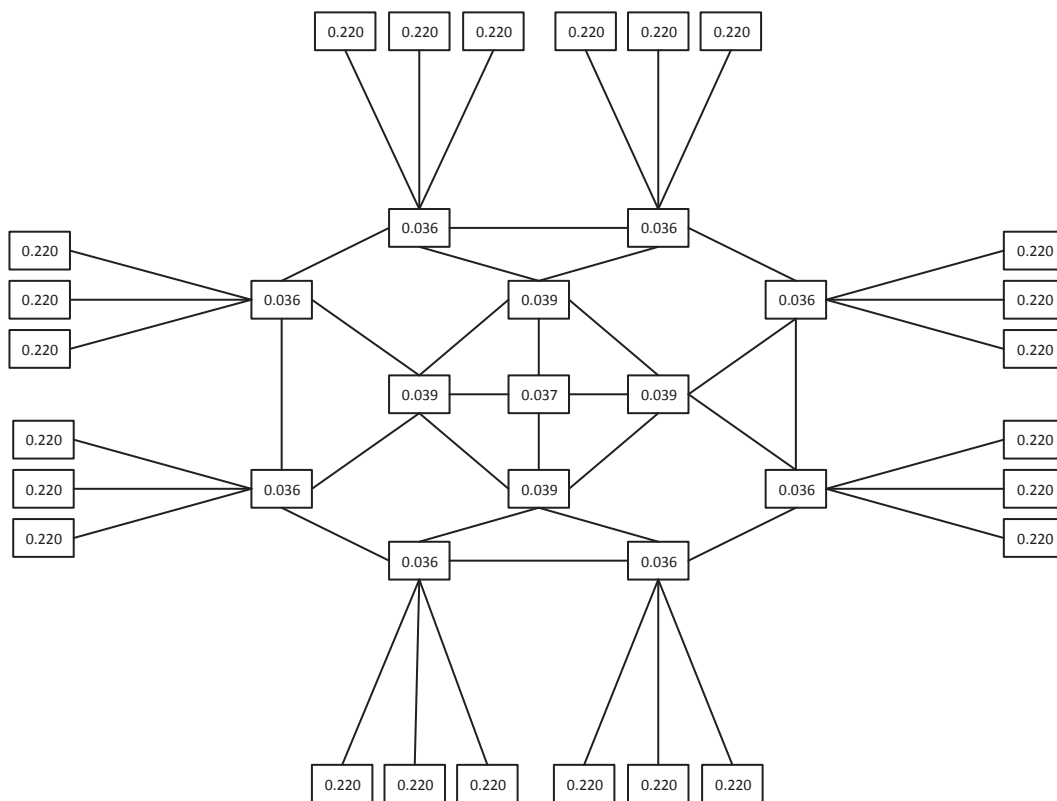


图 2.6: 流介数中心性计算示例

2.3 基于特征向量的评估方法

上述两种评价方法主要是从邻居节点的影响力入手，而未考虑节点本身的影响力。加利福尼亚大学的PHILIP BONACICH 等人于2001 年，提出了一种基于特征向量的评估方法。该方法不仅考虑了邻居节点的影响力，也综合了节点本身的影响力。特征向量评估方法可以分为特征向量中心性、alpha 中心性、累计提名方法。

2.3.1 特征向量中心性

BONACICH等人认为节点影响力不仅由节点的邻居的数量，也与节点本身的影响力有关,于是提出了基于特征向量中心性的评估方法[58]。该方法的中心思想是：某一节点的重要度是由它的邻居节点的重要度之和组成的，邻居节点的重要度之和越大，则该节点越重要。设节点*i*的重要性用 x_i 表示,则该节点的特征向量中心性可以描述为

$$x_i = c \sum_{j=1}^n a_{ij} x_j$$

其中， c 为一调节因子，取值范围为 $(0,1]$ 。上述公式中，某一节点的 x_i 值越大，则该节点的重要度越大；反之，则重要越小。特征向量中心性与以往评估方法的不

同之处在于，它考虑了节点所处的周围环境和节点本身的重要度，考虑的因素较为全面。在特征向量中心中，节点周围邻居的重要度将会影响节点本身，该方法适合用于研究社交网络等。

2.3.2 alpha中心性

除了特征向量中心性外，BONACICH等人还提出另一种解决不对称中心性的方法——alpha中心性（alpha centrality）[59]。alpha中心性可以让每个节点不依赖节点本身或节点的邻居。设 e 为外部影响力向量，则节点的影响力可表示为

$$x = \alpha A^T x + e$$

对上式求解得

$$x = (I - \alpha A^T)^{-1} e$$

特征向量虽然考虑的因素较为全面，然而在处理网络中度非常大的节点时，会出现局部分化现象，使得大多数的EC值都集中于度大的节点，而度相对较小的节点EC值很小。Martin等人为解决这一问题，对BONACICH等人的特征向量中心性进行了改进，基于无回溯矩阵提出了一种选择性中心性方法。该方法在特征向量中心性表现良好时，与特征向量中心性的效果相当。而当特征向量中心性不能有效评估节点重要度时，使用改进的方法能有效避免局部分化现象。

2.3.3 累计提名法

累计提名法由alpha中心性演变而来，alpha中心性中外部影响力 e 为一个固定值，而累计提名法在每次迭代过程中打分值，是一个动态收敛的过程[60]。累计提名法的数学表示形式如下：

$$p_i^t = \frac{\tilde{p}_i^t}{\sum_j \tilde{p}_j^t}$$

R. Poulin等人的实验中，累计提名方法比特征向量中心性的收敛速度更快，对计算机的计算性能要求较低，并且该方法可以比较不同尺度的网络。

2.4 基于随机游走的评估方法

随机游走（Random Walk）顾名思义是随机的、无规则的游走。随机游走的复杂网络模型由Jari Saram'aki等人于2004年提出。Jari Saram'aki等人的实验表明简单的随机游走模型与BA无标度网络模型具有一定的等价关系。随机游走评估方法可以细分为PageRank方法[61]、LeaderRank方法[62, 36]、HITS方法、自动信息汇集算法、SALSA等。

2.4.1 PageRank算法

随着互联网的发展，信息量增大，如何辨别网站中网页的重要性成为一个亟需解决的问题。1998年，谷歌创始人Larry Page 和Sergey Brin提出了PageRank算法，该方法最初用于网页排序。传统的网页排序算法主要根据关键字进行判断，然而该方法对关键字的依赖性过大，容易受到恶意关键字的影响，降低排序结果的有效性。网站有很多链接关系，如网页A包含网页B 的链接，也就是说网页B链向网页A，则网页B 认为网页A是比较重要的网页。一个网页所包含的链接数越多，则该网页的重要性越大。如图2.8 所示，节点 N_1 的链接最多，它的重要度在图2.8是最大的。

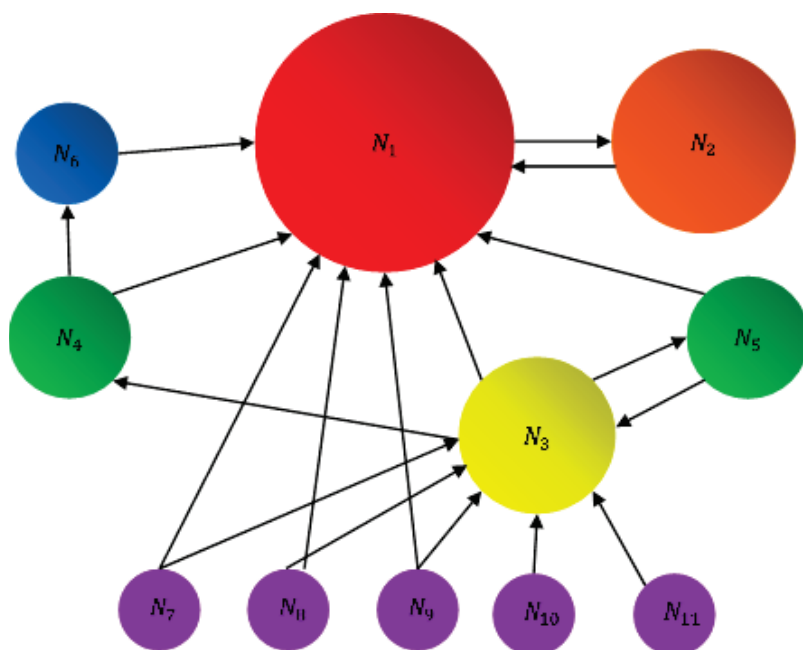


图 2.7: PageRank计算示例

PageRank方法不同于以往的按关键字或标签对网页进行排序方法，该方法的优越之处在于可以避免有些网站进行“恶意关键字”的诱导，从而提高了网页排序结果的可靠性。PageRank方法认为一个网页的重要性与链接该网页的其它邻居网页的数目和邻居网页的重要性有关。如果两个网页拥有同等重要性邻居网页，则拥有邻居网页数目多的那个网页重要性更大。如果两个网页拥有同等数量的邻居网页，则有用较高重要度网页的重要性更大一些。设网页的重要性由PR值决定，PageRank方法在一开始赋予每个网页相同的PR 值，通过迭代，在下一时刻把每个网页的PR值均分给该网页所指向的页面。则每个网页在 t 时刻的重要性 PR_i 计算公式如下所示

$$PR_i(t) = \sum_{j=1}^n a_{ij} \frac{PR_j(t-1)}{k_j^{out}} \quad (2.1)$$

其中 k_j^{out} 为节点 v_j 的出度. 迭代直到每个节点的PR值都达到稳定时为止. PageRank方法是一个优秀的网页排序方法, 但是该方法也存在一定的缺陷. 倘若某个网页的出度为0时, 该网页无法向其它网页传递PR值, 而其它节点还在源源不断的向该节点传递PR值. 为解决这一问题, PageRank 算法在上述过程基础上引入一个随机跳转概率 c . 每一步, 不管一个节点是否为悬挂节点, 其PR值都将以 c 的概率均分给网络中所有节点, 以 $1-c$ 的概率均分给它指向的节点. 该过程实际上是考虑到了现实中网络用户除了通过超链接访问页面之外, 还可以通过直接输入网址的形式对网页进行访问的行为, 从而保证了即使是没有任何入度的网页也有机会被访问到. 其实质是将有向网络变成强连通的, 使邻接矩阵成为不可约矩阵, 保证了特征值1的存在. 由此可得含参数 c 的PageRank 算法:

$$PR_i(t) = (1 - c) \sum_{j=1}^n a_{ij} \frac{PR_j(t-1)}{k_j^{out}} + \frac{c}{n} \quad (2.2)$$

上述公式中的 c 的取值不是固定的, 其取值根据实际情况而定. 当 $c=0$ 时, 公式(2.2)将演变成公式(2.1); 当 c 值趋于0时, 公式(2.2)收敛速度越来越慢; 当 c 取值变大时, 公式(2.2)的收敛速度变快, 但是会导致算法的结果有效性降低; 当 $c=1$ 时, 网络中所有节点的重要度都为 $\frac{1}{n}$. 由拉里佩奇等人提出的PageRank 算法是谷歌搜索引擎的核心算法, 谷歌搜索引擎的成功也充分证明了该方法强大的实用性和有效性. 然而不是所有方法都是完美的, 后续很多研究学者对PageRank方法提出了一系列改进. 由此衍生出了LeaderRank算法、HITS算法、SALAS算法等等.

2.4.2 LeaderRank算法

PageRank算法中, 从一个网页跳转到另一个网页的概率是相等的. 但是实际情况中, 用户对网页的选择会受多种因素影响使得网页的跳转概率不相同. 这些影响因素包括用户的兴趣爱好、网页的吸引力等等. 人们普遍对一些热门的、常用的网站的选择性高于其它网站. PageRank 算法的另一个缺点是公式(2.2)中参数的选取不固定, 需要通过一系列的实验来获取, 参数的不固定性会导致网页排序结果不唯一, 这增加了PageRank方法使用的难度和繁琐性. 为了解决这些问题, lv等人提出了leaderRank算法[36]. 为了解决PageRank算法中网页跳转概率的问题, LeaderRank算法中加入了有一个背景节点(Ground Node) 和该背景节点与网络中所有节点的双向边. LeaderRank算法中, 当用户点击或输入网址访问某个页面时的概率 p_1 与从该页面访问背景节点的概率 p_2 是一样的. 概率 p_2 与网页上的链接成反比, 网页链接越多, 访问背景节点的概率 p_2 越小. 一个原本有 N 个节点的网络 G_1 , 在加入背景节点后变成一个具有 $N + 1$ 个节点强连通网络 G_2 . 设 S_i 表示节点 i 的LeaderRank值, 则LeaderRank算法的具体计算步骤如下:

第1步) 每个节点在初始时刻时 $S_i(0) = 1$, 网络中背景节点取值为 $S_g(0) = 0$ 。

第2步) 在时间 t 时, 节点 V_i 的值为 $S_i(t)$, 如果节点 V_i 指向连接节点 V_j , 则 $a_{ij} = 1$, 否则为 0。假设节点 j 的出度为 k_j^{out} , 则节点 i 在下一时刻 $(t+1)$ 时的 LeaderRank 值为

$$s_i(t+1) = \sum_{j=1}^{N+1} \frac{a_{ij} k_j^{out}}{s_i} (t)$$

第3步) 不断重复第二步, 直到时间步 t_c 时, 节点 i 的分数不变, 收敛于一个定值 $s_i(t_c)$, 则背景节点的值记为 $s_g(t_c)$ 。因此, 经过上面的步骤, 节点 i 的影响力值记为:

$$S_i = S_i(t_c) + \frac{S_g(t_c)}{N}$$

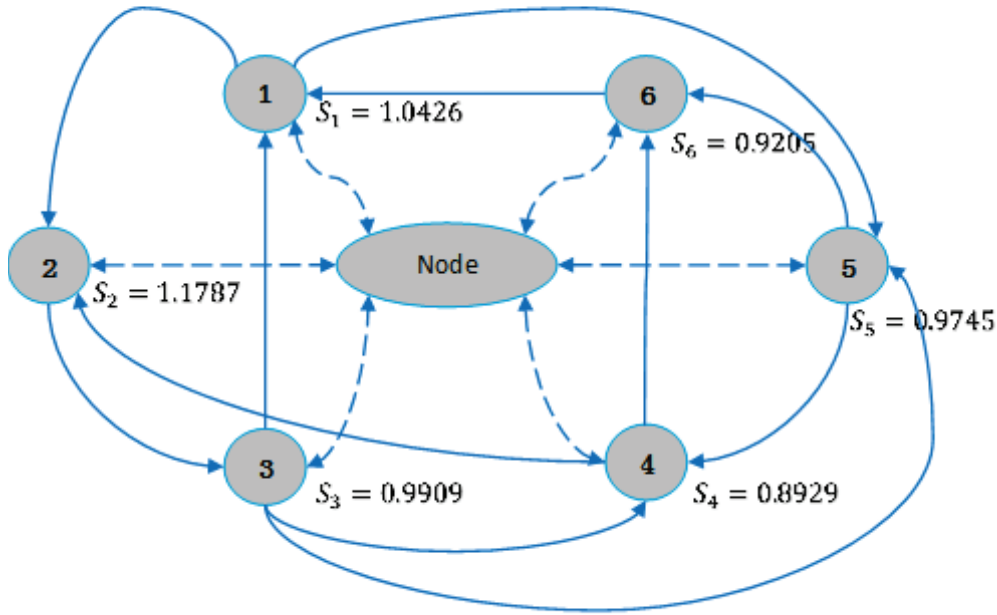


图 2.8: LeaderRank 计算示例

2.4.3 HITS 算法

HITS 算法英文全称为 Hypertext-Induced Topic Search。HITS 算法是对 PageRank 算法的一种改进, 是一种基于链接分析的网页排序算法, 由 Jon Kleinberg 提出。HITS 算法和 PageRank 算法相似, 一般都用来对网页进行排序。HITS 算法的基本

思想是：假设网络中存在权威页面(Authorities) 和枢纽页面(Hubs)，则该算法认为一个好的权威页面会被很多好的枢纽页面链接，而一个好的枢纽页面会链接很多权威页面。权威页面和枢纽页面是一种共生互利的关系，一个好的权威页面会有很多高质量的枢纽页面指向它，而一个好的枢纽页面会指向很多高质量的权威页面。一个网页的Hub值也高，则该网页所指向的权威页面的质量越好，指向好的权威页面的数量越多。一个网页的Authority值越高，则指向该页面的枢纽页面的质量越高，数量也越多。权威页面和枢纽页面通过这种互利关系进行迭代计算，每轮迭代计算后Authority值与hub值都会进行更新，直到这两个值趋于相对稳定时迭代结束。图2.9 为二者之间的关系示例图。

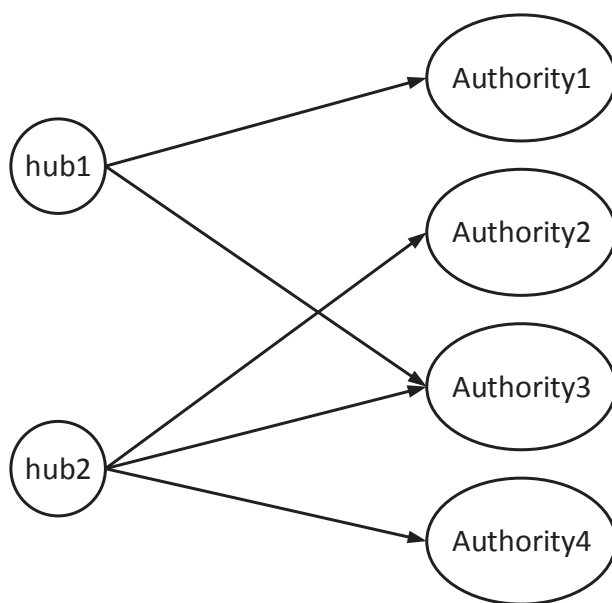


图 2.9: 权威页面与枢纽页面的关系图

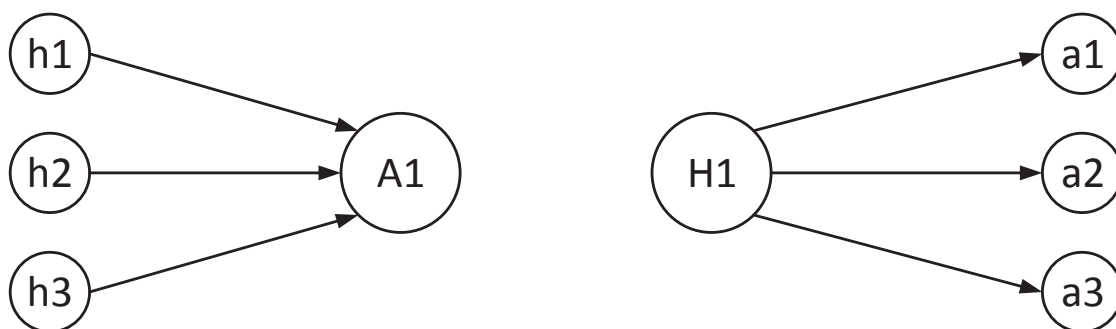


图 2.10: 权威值与枢纽值更新方式示意图

图??为权威值与枢纽值更新方式示意图。其中权威节点A1的权威值计算方式为 $A1 = h1 + h2 + h3$ ，而枢纽节点H1 的枢纽值计算方式为 $H1 = a1 + a2 + a3$ 。

HITS算法应用范围比较广，在搜索引擎、网络分析、自然语言处理等方面都取得了较好的效果。该算法虽然有其独到之处，但也存在一些不足。该算法采用了迭代计算方式，导致计算量大、计算效率不高、资源占用率高等问题，限制了算法的应用范围。HITS算法在查询网页时会查到一些与主题无关的页面，使计算结果发生漂移。HITS的另一个弊端是很容易被恶意操纵，操作者倘若建立一个页面并使该页面指向很多高质量的链接时，会使这个被建立的网页的Hub值和Authority值远大于正常值。

2.4.4 自动信息汇集算法

自动信息汇集(automatic resource compilation,ARC)算法是对HITS算法的一种改进[63]。鉴于HITS算法只考虑网页之间的链接关系而忽略了网页主题信息造成主题漂移现象，Kleinberg等人提出了自动信息汇集算法。ACR算法将不仅考虑了网页链接关系，也考虑了网页主题信息，从而有效避免了主题漂移现象。ACR主要计算方法如下：

- 第1步) 按照被搜索的主题Text对网页进行搜索，并去掉孤立节点，将符合主题的网页汇集到一个集合 Γ 中。
- 第2步) 对集合 Γ 中的网页的权威值和枢纽值进行计算更新。
- 第3步) 对一个含有链接的网页，计算该链接周围包含主题Text的链接数量。如果包含Text的链接数量越多，则该链接的权威值越大；否则，该链接的权威值越小。
- 第4步) 判断权威值与枢纽值是否趋于稳定。如果趋于稳定，则计算结束；如果还未稳定，则回到第二步。

2.4.5 SALSA算法

SALSA算法的英文全称为Stochastic Approach for Link Structure Analysis,即链接结构的随机分析法(stochastic approach for link structure analysis)，该算法将PageRank算法和HITS算法进行了结合，从而融合了这两种算法的优点，是一种优秀的链接分析算法[64]。PageRank算法只考虑用户向前浏览网页的情况，而未考虑用户回退浏览网页的情况。SALSA算法同时考虑了用户前进浏览网页的情况和用户回退浏览网页的情况，而且该算法还采用了HITS 算法中权威网页和枢纽网页的思想。SALSA算法的计算过程如下：

- 第1步) 获取与主题Text相关的网页，去掉孤立节点，得到一个集合 Γ 。

第2步) SALSA算法采用了HITS算法的部分思想,将集合 Γ 划分为二分图生成两个子集A和H,二分图的划分示例如图2.11所示,对集合A和H中的节点的权威值和枢纽值进行计算更新。

第3步) 采用PageRank算法中的随机游走模型。用户随机选中一个子集H,从该子集中随机选中一个节点 V_i 。若 V_i 是一个多边节点,则用户按照相等概率随机选择一条边,从子集H跳到子集A中的与 V_i 相连的某一节点 V_j 。然后节点 V_j 再随机选一条边跳回子集H中的某一邻居节点。如此循环往复,在两个子集A和H中相互跳转。直到最终结果收敛,算法结束。

相比HITS算法而言,SALSA 算法能够更好地避免主题漂移的问题。SALSA算法实际上考虑的是一个基于二部分图的随机游走过程,这一思路也被成功地应用在信息挖掘的另外两个领域中,即基于网络结构的链路预测问题和个性化推荐算法。

2.5 基于节点位置的评估方法

2.5.1 经典k-shell分解算法

Kitsak 等人于2010 年首次提出了节点重要性依赖于其在整个网络中的位置的思想,并且利用K-shell分解获得了节点重要性排序指标(k-shell),该指标时间复杂度低,适用于大型网络,而且比度、介数更能准确识别在疾病传播中最有影响力的节点。近几年不少学者受到这种思想的启发,对k-shell进行了扩展和改进,使其应用范围更广,准确性更好。K-shell分解方法通过递归地移去网络中所有度值小于或等于 k 的节点[65]。K-shell分解是一个层层推进的过程,好像剥洋葱。

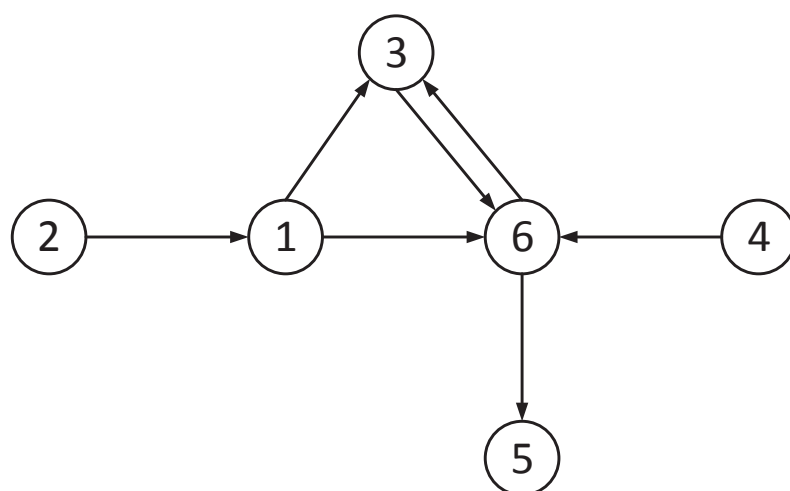
第1步) 去掉度为1 的节点,剩下一个子图,如果该子图中依然有度为1的点则继续删除这些点,直到最后剩下一个不含度为1的节点的子图 G_1 。那些被删除的节点则属于 $k_s = 1$ 的核。

第2步) 跟第一步类似,删除子图 G_1 中度为2的节点,最后得到一个子图 G_2 , G_2 中所有点的度均大于2。

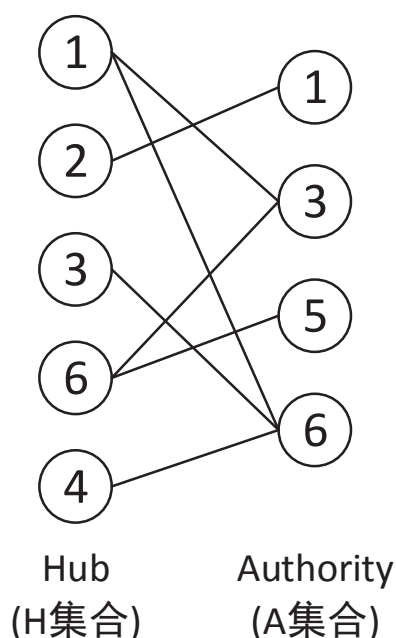
第3步) 以此类推,直到所有的点都被分解到某个核中[29]。Ks分解示意图如图2.12 所示,该网络被划分为3 个不同的层。图2.12(a)为原始网络。

2.5.2 MDD算法

一些学者都认为网络中的Hubs节点或者高介数的节点是传播中最有影响力的节点。这是因为Hubs节点拥有更多的人际关系,而高介数的节点有更多的最短



(a) 原网络图



(b) 二分图

图 2.11: SALSA二分图生成示意图

路径通过,于是疾病控制首先要确定这样的节点。但是,Kitsak 等人通过研究社交网络、邮件网络、病人接触网络、演员合作网络等实证网络并且通过传播动力学的建模分析指出,对于单个传播源情形,Hubs 节点或者高介数的节点不一定是最有影响力的节点,而通过K-shell分解分析确定的网络核心节点(即K-shell值大的节点)才是最有影响力的节点。当然文章也指出,当初始存在多个传播源的时候,传播的规模很大程度依赖于初始传播源之间的距离。在存在多个传播源的情况下,度大的Hub 节点往往比Ks大的节点具有更高传播效率。然而Ks指标赋予大量节点相同的值,例如Barabasi-Albert(BA) 网络模型的所有节点的Ks 值都相等,从而导

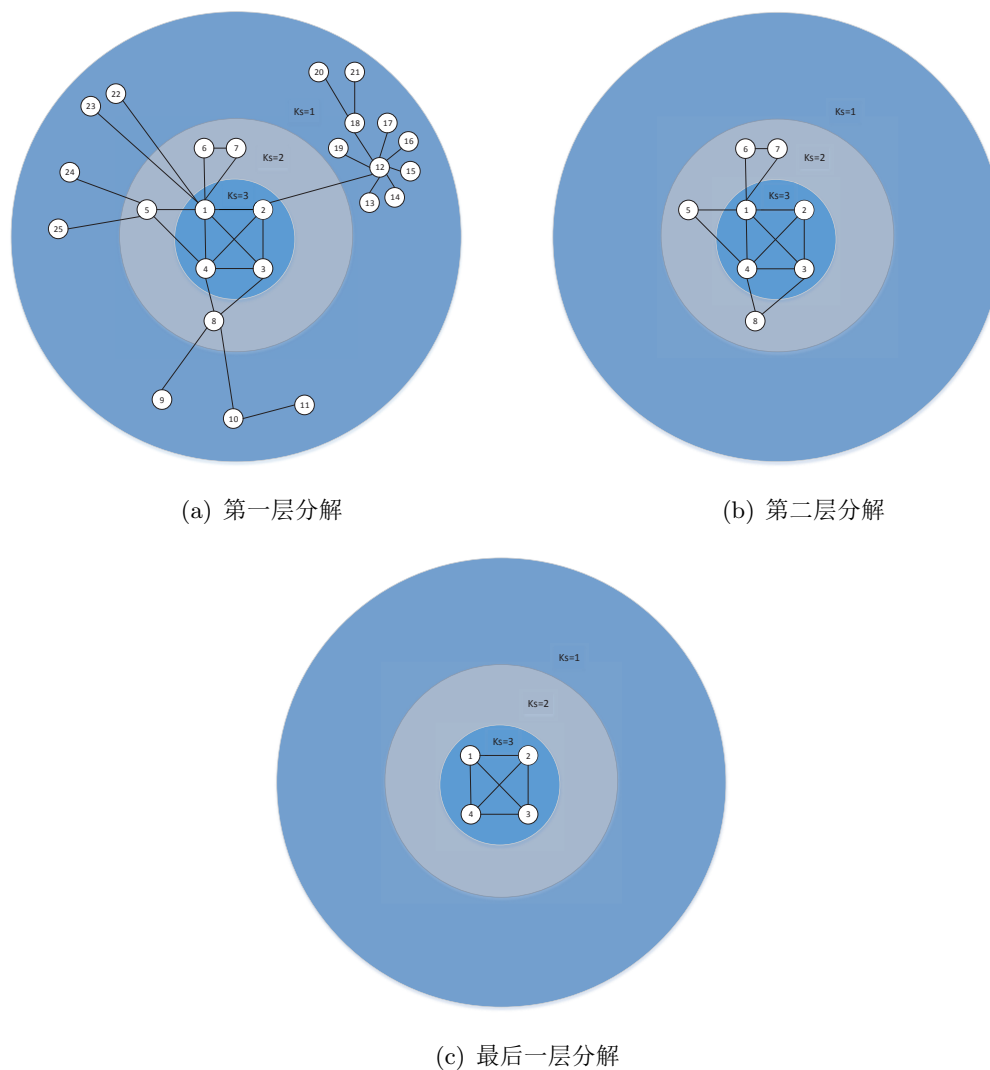


图 2.12: k-shell分解示意图

致Ks指标无法衡量其节点的重要性.

Zeng等人考虑节点的Ks信息和经过Ks分解后被移除节点的信息,提出了混合度分解方法(MDD)[40], 图2.13为MDD算法示意图。MDD的分解流程为:

第1步) 初始状态下,网络中节点的 $k(i)_m$ 值与 $k(i)_r$ 相等。

第2步) 移去网络中 k_m 值最小的节点,这些节点赋值为M-shell。

第3步) 采用公式 $k(i)_m = k(i)_r + \Gamma k(i)_e$ 更新网络剩余节点的 k_m ,并通过递归移除网络中 k_m 值小于或等于M的节点,将这类节点赋值为M-shell, 直至网络中所有剩余节点的 k_m 值大于M.

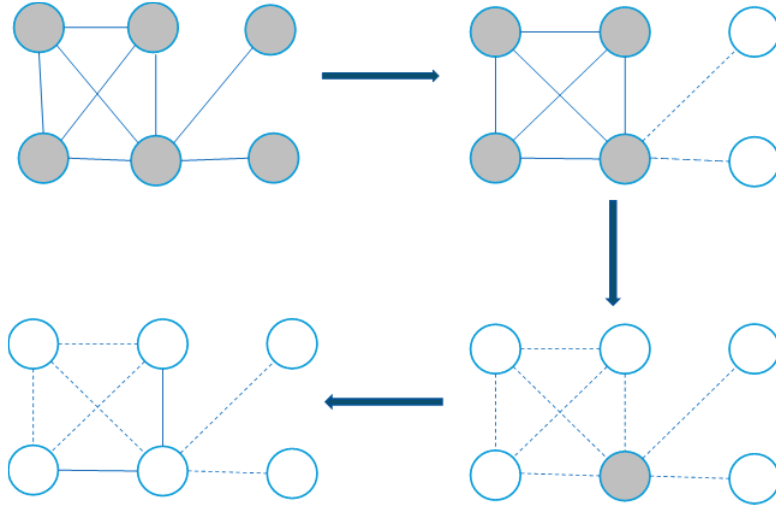


图 2.13: MDD算法示意图

2.5.3 加权k-shell分解算法

Garas等先将加权网络转变成无权网络,再进行经典的Ks分解[41]。魏等人提出了一种基于节点边的加权k-shell分解方法[66],该方法通过节点的连接边的权重,计算节点的加权重,然后再用k-shell分解方法进行分解。在文章[66]中,魏等人通过一系列实验证明了该方法的可行性、实用性和优越性。

2.5.4 最小k-shell算法

任卓明等人发现网络中所有最小k-shell中的节点的k-shell值都是一样的。比如图2.14中,节点5、6、7、8、9会被分到最小k-shell中,但是节点5在网络中的重要性明显比节点6、7、8、9要高,这样就导致k-shell分解无法区分网络中这类节点的重要性,于是任卓明等人提出了基于最小k-shell 节点邻居节点最大k-shell值的评价方法,该指标通过最小k-shell 节点与网络中其它节点的链接关系来区分最小k-shell 中节点的重要性[67],其具体方法如下所示。

$$H(i) = \max\{K_{s_j}\}, j \in \Gamma(i)$$

其中集合 Γ 表示节点 i 所有邻居节点的集合, K_{s_j} 表示节点 i 的第 j 个邻居节点的k-shell值。

2.5.5 KCK算法

Hou等人对度、介数、K-shell三种方法同时进行了结合[68, 69],其结合方法如下公式所示:

$$D(i) = k^2(i) + C_b^2(i) + K_s^2(i)^{\frac{1}{2}}$$

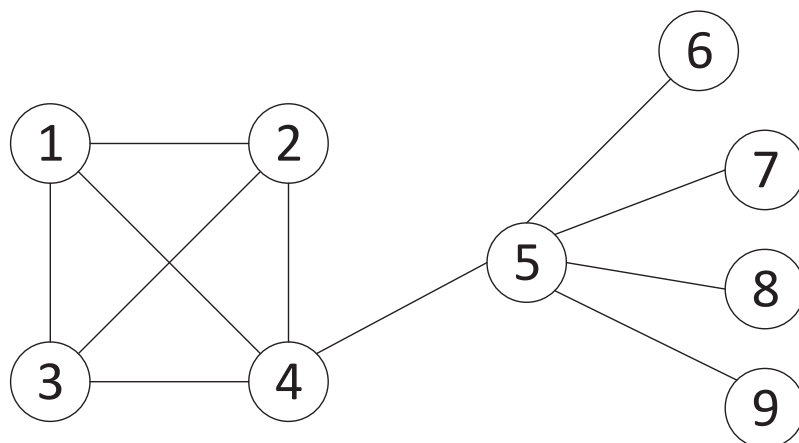


图 2.14: 最小k-shell示例图

其中 $k(i)$ 表示节点 i 的度， $C_b(i)$ 表示节点 i 的介数， $K_s(i)$ 表示节点 i 的ks值。该方法虽然融合了三种方法，但是计算量庞大。

2.6 本章小结

本章主要对复杂网络领域几种重要的节点重要度评估方法进行了讲述。主要讲述了基于节点度中心性的评估方法、基于路径的评估方法、基于特征向量的评估方法、基于随机游走的评估方法、基于节点位置的评估方法。基于节点度中心性的评估方法主要介绍了度中心性和半局部中心性。基于路径的评估方法主要介绍了介数中心性和接近中心性。基于特征向量的评估方法主要介绍了特征向量中心性、alpha中心性和累计提名法。基于随机游走的评估方法主要介绍了PageRank算法、LeaderRank算法、HITS算法、自动信息汇集算法和SALSA算法。基于节点位置的评估方法主要介绍了经典k-shell分解算法、MDD算法、加权k-shell分解算法、最小k-shell算法和KCK算法。

第三章 改进的加权k-shell分解算法

魏等人于2015年提出了一种新的加权k-shell分解。该方法通过改进加权度的计算，从而有效解决了k-shell分解中的单调性等问题。魏等人提出的方法虽然优秀，但是仍有不足之处，本章内容主要对魏等人提出的方法进行了改进，并在六个真实网络上做了k-shell分解实验、SIR传染病实验和脆弱性分析实验来验证改进方法的有效性和实用性。

3.1 k-shell分解算法

k-shell分解算法又叫k-核分解算法，在第二章第五小节有介绍。K-shell分解算法由Kitsak 于2010年提出。K-shell分解是一个层层推进的过程，好像剥洋葱。第一步，去掉度为1的节点，剩下一个子图，如果该子图中依然有度为1的点则继续删除这些点，直到最后剩下一个不含度为1的节点的子图 G_1 。那些被删除的节点则属于 $ks=1$ 的核。第二步，跟第一步类似，删除子图 G_1 中度为2的节点，最后得到一个子图 G_2 ， G_2 中所有点的度均大于2。以此类推，直到所有的点都被分解到某个核中。

3.2 加权k-shell分解

在文章[66]中，魏等人通过改进了k-shell分解中分解度的计算来改进k-shell分解实验。而本文在文章[66]的基础上，对加权度的计算进行了改进，从而改进k-shell分解算法。

一、度的计算

魏等人算法[66]中，度的计算方法如下：

$$DC_i = \sum_j^n e_{ij} \quad (3.1)$$

其中， $e_{ij}=1$ 表示节点i与节点j是连接的；反之，则表示节点i与节点j是断开的。 DC_i 表示节点i的度。

二、加权边的计算

在文献[66]中，对于任意一条边 e_{ij} ，其权重为该边所连接的两个节点 V_i 和 V_j 的节点度之和，计算方法如下：

$$w_{ij} = DC_i + DC_j \quad (3.2)$$

三、节点加权度的计算

对于任意节点 V_i ，其加权度的计算方法如下：

$$DC_i^w = \alpha DC_i + (1 - \alpha) \sum_{j \in \Gamma_i} w_{ij} \quad (3.3)$$

其中， Γ_i 为节点 V_i 所有邻居节点的集合。 α 为调节因子，其取值范围为 $[0,1]$ 。魏的方法在计算每个节点的加权度之后，将加权度进行向上取整，然后按照经典k-shell分解对网络进行分解。文章[66]中的实验表明该方法能有效解决经典k-shell分解中单调性等问题。

3.3 改进的加权k-shell分解

本章主要提出了一种改进的加权k-shell分解算法，该算法主要改进了文章[66]中对加权度的计算。而计算边的权重方法与文献中的一样，具体方法如下：

$$w_{ij} = DC_i + DC_j \quad (3.4)$$

改进的加权度计算方法如下：

$$DC_i^w = DC_i^\alpha \left(\sum_{j \in \Gamma_i} w_{ij} \right)^{1-\alpha} \quad (3.5)$$

其中， Γ_i 为节点 V_i 所有邻居节点的集合。 α 为调节因子，其取值范围为 $[0,1]$ 。当 $\alpha = 1$ 时，则该方法变为经典k-shell分解；当 $\alpha = 0$ 时，则节点完全依赖邻居节点的重要性。计算加权度后，若加权度不为整数则向下取整。然后按照经典k-shell分解对网络进行分解。

为了便于理解改进的k-shell分解方法，在此给出一个计算例子。对于一个无向无权网络，如图3.1所示。

第1步) 计算每条边的权重。图3.1中，节点1的度为1，节点2的度为2，则连接节点1与节点2的边 e_{12} 的权重 $w_{12} = DC_1 + DC_2 = 1 + 2 = 3$ 。同理其余边的权重计算如表格3.1和图3.2所示。

第2步) 按照公式(4)计算各节点的加权度。设调节因子 $\alpha = 0.5$ ，计算节点2的加权度，节点2有两条边 $e_{1,2}$ 、 $e_{2,3}$ 相连，由此可得

$$\Gamma_2 = 1, 3$$

$$\sum_{j \in \Gamma_2} w_{2j} = w_{1,2} + w_{2,3} = 7$$

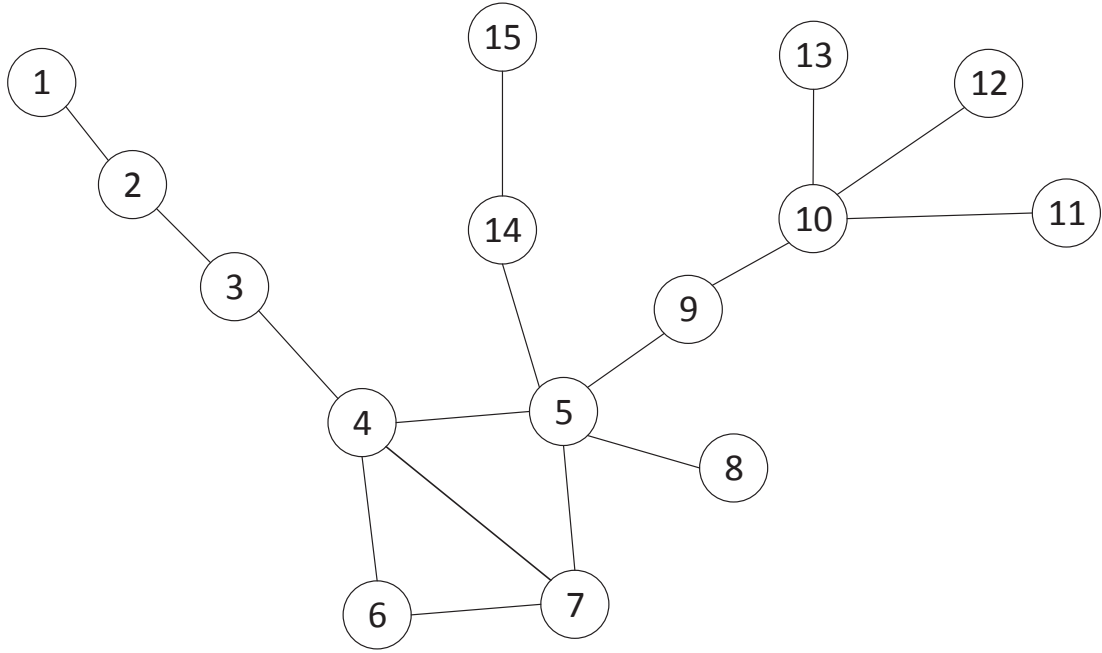


图 3.1: 改进k-shell分解算法示例1

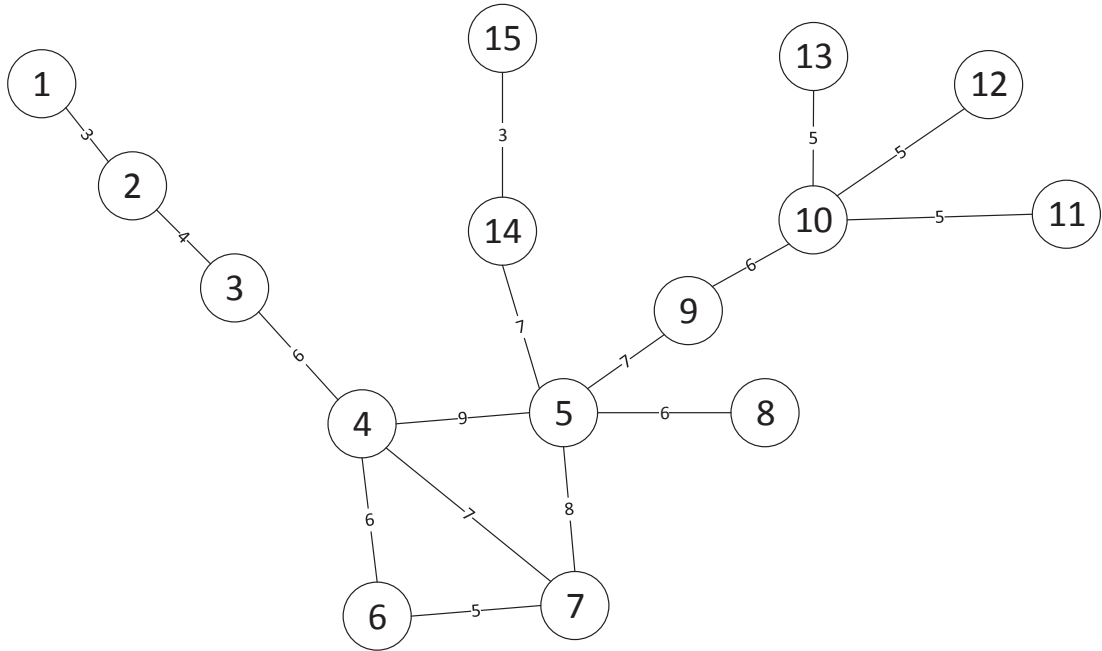


图 3.2: 改进k-shell分解算法:各边的权重计算示意图

$$DC_2^{rw} = DC_2^{0.5} \left(\sum_{j \in \Gamma_i} w_{ij} \right)^{1-0.5} = 2^{0.5} \times 7^{1-0.5} = 3.742$$

各节点加权度的计算如表3.2所示

第3步) 先去将节点1和节点15分配到ks=1的核中，得到子图 G_1 ，如图3.3,重新计

表 3.1: 改进k-shell分解算法权重计算表

e_{ij}	DC_i	DC_j	w_{ij}
$e_{1,2}$	1	2	3
$e_{2,3}$	2	2	4
$e_{3,4}$	2	4	6
$e_{4,5}$	4	5	9
$e_{4,6}$	4	2	6
$e_{4,7}$	4	3	7
$e_{5,7}$	5	3	8
$e_{5,8}$	5	1	6
$e_{5,9}$	5	2	7
$e_{5,14}$	5	2	7
$e_{9,10}$	2	4	6
$e_{10,11}$	4	1	5
$e_{10,12}$	4	1	5
$e_{10,13}$	4	1	5
$e_{14,15}$	2	1	3

表 3.2: 第一轮节点加权度的计算表

节点编号i	i	DC_i	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
1	2	1	3	1.732	1
2	1, 3	2	7	3.741	3
3	2, 4	2	10	4.472	4
4	3, 5, 6, 7	4	27	10.392	10
5	4, 7, 8, 9, 14	5	37	13.601	13
6	4, 7	2	11	4.690	4
7	4, 5, 6	3	20	7.746	7
8	5	1	6	2.449	2
9	5, 10	2	13	5.099	5
10	9, 11, 12, 13	4	21	9.165	9
11	10	1	5	2.236	2
12	10	1	5	2.236	2
13	10	1	5	2.236	2
14	5, 15	2	10	4.472	4
15	14	1	3	1.732	1

算子图 G_1 中各节点的加权度，如表3.3所示。检查是否有加权度不大于1的节点，若有则继续分配到核1中。

第4步) 从表3.3可以观察出节点2的加权度为1，所以应该将节点2分解到 $ks=1$ 的核中。去掉节点2后得到子图 G_2 ，如图3.4 所示。然后，对子图 G_2 中的各节点进行节点加权度的计算，计算结果如表3.4所示。

第5步) 观察表3.4发现表中已经不含有加权度为1的节点了，而表中最小的加权度

表 3.3: 子图 G_1 节点加权度的计算表

节点编号 <i>i</i>	i	DC_i	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
2	3	1	3	1.732	1
3	2, 4	2	10	4.472	4
4	3, 5, 6, 7	4	27	10.392	10
5	4, 7, 8, 9, 14	5	37	13.601	13
6	4, 7	2	11	4.690	4
7	4, 5, 6	3	20	7.746	7
8	5	1	6	2.449	2
9	5, 10	2	13	5.099	5
10	9, 11, 12, 13	4	21	9.165	9
11	10	1	5	2.236	2
12	10	1	5	2.236	2
13	10	1	5	2.236	2
14	5	1	6	2.449	2

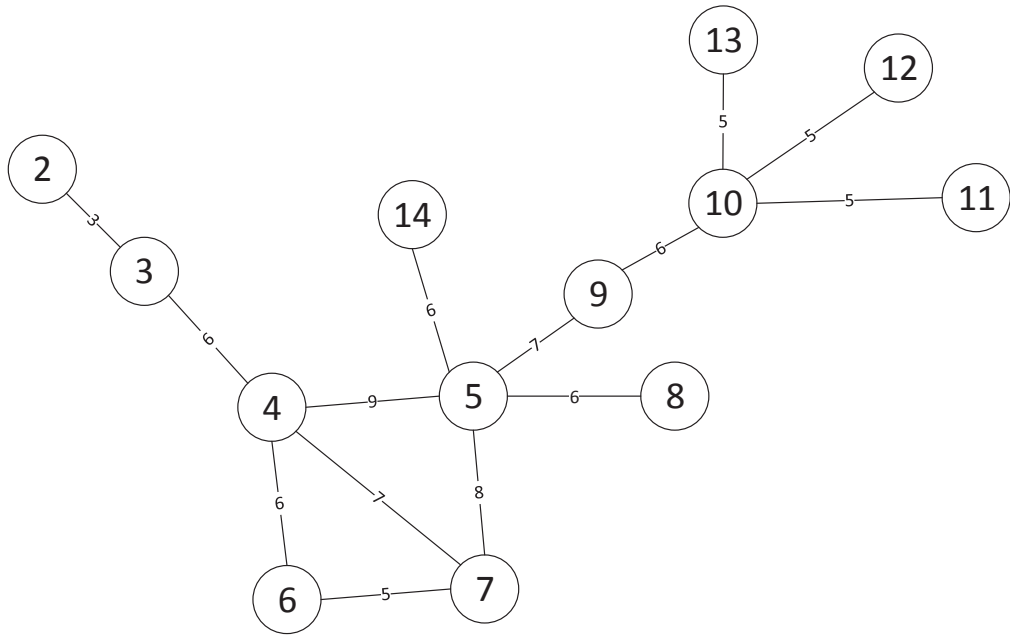


图 3.3: 子图 G_1

为2。于是将加权度为2的节点分配 $ks=2$ 的核中，被分解的节点包括3、8、11、12、13、14。分解后得到子图 G_3 ，如图3.5所示。然后，对子图 G_3 中的各节点进行节点加权度的计算，计算结果如表3.5所示。

第6步) 观察表3.5，可以发现表中只有节点10的度不大于2，于是将节点10分解到 $ks=2$ 的核中。去掉节点10后得到子图 G_4 ，如图3.6所示。计算子图 G_4 中所有节点的加权度，如表格3.6所示。

第7步) 观察表格3.6，可以发现节点9的加权度不大于2，于是将节点9分解到 $ks=2$ 的

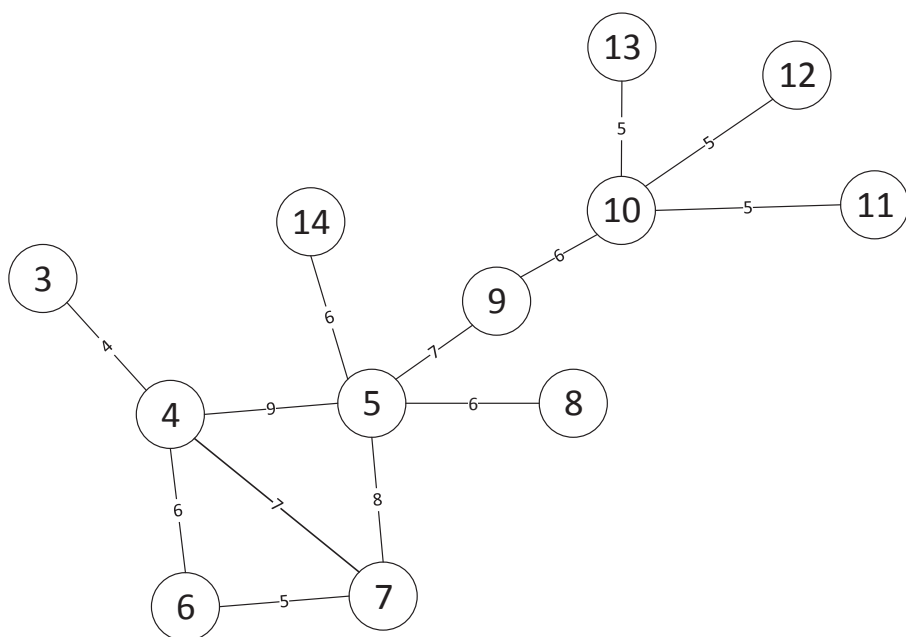


图 3.4: 子图 G_2

表 3.4: 子图 G_2 节点加权度的计算表

节点编号 i	i	DC_i^w	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
3	4	1	4	2.000	2
4	3, 5, 6, 7	4	23	9.592	9
5	4, 7, 8, 9, 14	5	36	13.416	13
6	4, 7	2	11	4.690	4
7	4, 5, 6	3	20	7.746	7
8	5	1	6	2.449	2
9	5, 10	2	13	5.099	5
10	9, 11, 12, 13	4	21	9.165	9
11	10	1	5	2.236	2
12	10	1	5	2.236	2
13	10	1	5	2.236	2
14	5	1	6	2.449	2

表 3.5: 子图 G_3 节点加权度的计算表

节点编号 i	i	DC_i^w	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
4	5, 6, 7	3	17	7.141	7
5	4, 7, 9	3	17	7.141	7
6	4, 7	2	10	4.472	4
7	4, 5, 6	3	17	7.141	7
9	5, 10	2	8	4.000	4
10	9	1	3	1.732	1

核中。去掉节点9后，得到子图 G_5 ，如图3.7所示。计算子图 G_5 中所有几点的加权度，如表3.7所示。

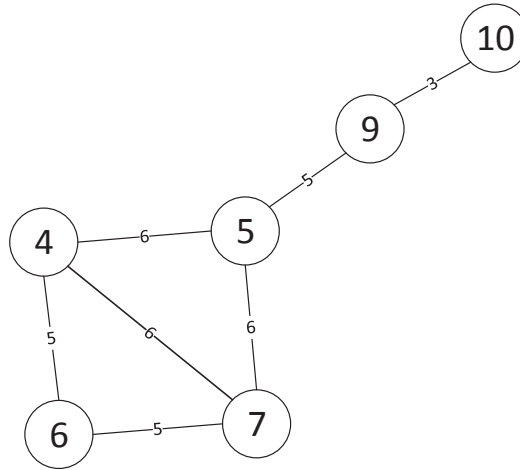


图 3.5: 子图 G_3

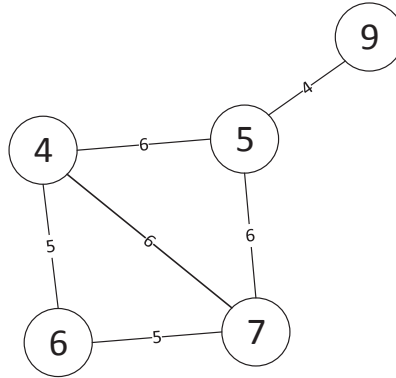


图 3.6: 子图 G_4

表 3.6: 子图 G_4 节点加权度的计算表

节点编号 i	i	DC_i^w	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
4	5, 6, 7	3	17	7.141	7
5	4, 7, 9	3	16	6.928	6
6	4, 7	2	10	4.472	4
7	4, 5, 6	3	17	7.141	7
9	5	1	4	2.000	2

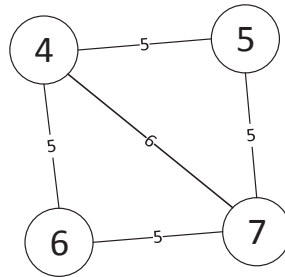


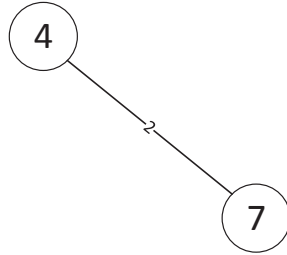
图 3.7: 子图 G_5

表 3.7: 子图 G_5 节点加权度的计算表

节点编号 i	i	DC_i^w	$\sum_{j \in \Gamma_i} w_{ij}$	DC_i^w	$\lfloor DC_i^w \rfloor$
4	5, 6, 7	3	16	6.928	6
5	4, 7, 9	2	10	4.472	4
6	4, 7	2	10	4.472	4
7	4, 5, 6	3	16	6.928	6

第8步) 观察表3.7,可以发现所有子图 G_5 中所有节点的加权度都不小于2, 于是令 $ks=ks+1$, ks 现在为3。

第9步) 观察表3.7, 发现所有节点的加权度都不小于3, 于是令 $ks=ks+1$, ks 现在为4。通过观察可以发现节点5、6的值不大于4, 于是将节点5、6分解到 $ks=4$ 的核中。得到子图 G_6 , 如图3.8所示。

图 3.8: 子图 G_6

第10步) 通过观察可以发现图3.8中所有节点的加权度都不大于4, 于是节点4和节点7分解到 $ks=4$ 的核中。分解结束。

3.4 实验与分析

3.4.1 实验数据

为了验证改进算法的有效性, 我们在六个不同的真实网络上进行了多次试验。这六个真实网络分别是Blogs、Email、Net-science、Roget、USAir和Yeast。六个网络的统计特性如表3.8所示。

- 1) Blogs网络包含1222个节点, 这些节点来源于MSN社交网络, 网络中包含各节点之间的相互关系。该数据集可以从<http://www.cs.bris.ac.uk/~steve/networks/peacockpaper/>进行下载。
- 2) Email网络包含1133个节点, 这些数据来源于是西班牙罗维拉维尔吉利大学的email网络。该数据集可以从<https://www.cs.bris.ac.uk/~steve/networks/peacockpaper/>进行下载。

- 3) Net-science网络包含1589个节点, 该网络是科研合作网络, 该网络中包含都是从事科学研究的人员及其相互关系。该数据集可以从<http://toreopsahl.com/datasets/>进行下载
- 4) Roget网络包含1022个节点, 网络是关于同义词的相互关系网络, 来源于斯坦福大学的图形学项目。该数据集可以从<http://vlado.fmf.uni-lj.si/pub/networks/data/dic/roget/Roget.htm>进行下载。
- 5) USAir网络包含332个节点, 该网络是美国航空网。该数据集可以从<http://vlado.fmf.uni-lj.si/pub/networks/data/default.htm>进行下载。
- 6) Yeast网络包含2361个节点, 该网络是构建酵母菌时蛋白质之间的相互关系网。该数据集可以从<http://vlado.fmf.unilj.si/pub/networks/data/default.htm>进行下载。

表 3.8: 六个网络统计特性:其中 N 表示网络节数, M 表示网络所有边的数目, $\langle k \rangle$ 表示网络总体平均度, k_{max} 表示网络最大节点度。

网络	N	M	$\langle k \rangle$	k_{max}
Blogs	1222	6803	27.3553	351
Email	1133	5451	9.6222	71
Net-science	1580	2741	3.4512	34
Roget	1022	5075	7.1399	28
USAir	332	2126	12.8072	139
Yeast	2361	7182	9.9051	118

3.4.2 k-shell分解实验

为了观察改进方法的性能, 我们利用改进的k-shell分解算法对原有网络进行了分解, 并且与魏等人的方法进行了比较。首先, 我们取调节因子 $\alpha = 0.5$ 进行k-shell分解实验。然后, 由0到1变化 α 的值再进行k-shell分解实验。

一、 α 取值为0.5

当 $\alpha = 0.5$ 时, 对六个网络进行分解的实验结果如表3.9所示 由表3.9, 可以观

表 3.9: 六个网络统计特性:其中 N_w 表示魏等人的方法所得k-shell总数, N_I 表示改进的方法所得k-shell总数.

网络	Blogs	Email	Net-science	Roget	USAir	Yeast
N_w	825	444	94	381	189	590
N_I	502	294	67	219	111	418

察出当 $\alpha = 0.5$ 时, 魏等人的分解方法所得k-shell总数比改进的方法所得k-shell总数要多。

二、 α 取值由0变化到1时

当 α 取值由0变化到1时，对六个真实网络进行分解实验的结果如图3.9所示。由图3.9可以观察到，改进的方法所得k-shell分解总数随着 α 的变化大致呈线性分布，k-shell总数的变化梯度较为均匀。而魏等人的方法所得k-shell总数随着 α 的变化梯度不均匀：当 $\alpha < 0.4$ 时，变化梯度较小；当 $\alpha \geq 0.4$ 时，变化梯度陡增。当然，通过改变 α 的值，可以发现改进的方法和魏等人的方法都可以得到相同的k-shell总数，而且二者k-shell总数的取值范围也相同。

3.4.3 传播性能实验

为了探究改进方法的传播性能，我们通过SIR传染病模型对改进的方法的传播能力与魏等人方法进行了传播实验比较。

一、SIR传染病模型

传染病模型的研究始于20世纪初，Hamer, Ross等人在传染病模型研究中有突出贡献。Kermack与McKendrick于1927年提出了SIR传染病模型。自此以后，SIR传染病模型作为最经典的传染病模型之一沿用至今[70]。

SIR传染病模型中，S(Susceptible)表示易感染疾病者，该类人群有感染疾病的可能性，但不一定感染疾病；I(Infected)表示已感染疾病者，该类人群可能被治愈成为R人群，也可能无法治愈而被淘汰出局；R(Recovered)表示已经从感染状态恢复到健康状态的人权，该类人群因已经获得免疫，将不会被再次感染[71-73]。

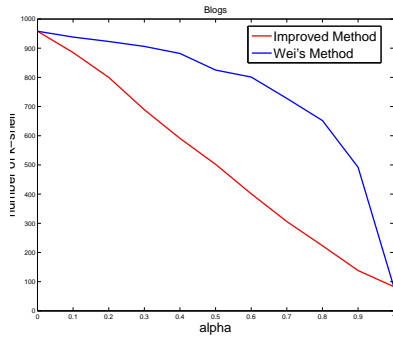
为了方便表述，设网络节点的总个数为 N ，疾病的传染率为 r ，从感染状态I恢复到健康状态R的概率为 g ，易感人群所占的比例为 S ，以感染人群所占的比例为 I ，从感染状态恢复到健康状态的人群比例为 R 。则SIR的演化模型如下列公式所示

$$\begin{cases} N(t) = S + R + I \\ \frac{dS}{dt} = -rNSI \\ \frac{dI}{dt} = rNSI - gI \\ \frac{dR}{dt} = gI \end{cases} \quad (3.6)$$

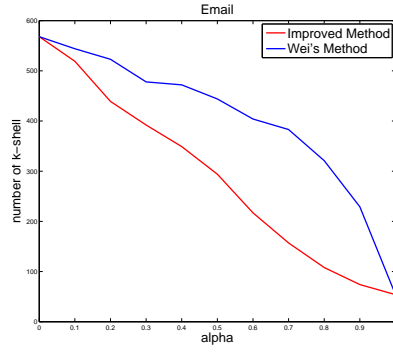
二、SIR传染病实验

为了便于试验研究，我们将改进方法的 α_{iw} 取值和魏等人的方法中 α_{ww} 的取值进行设定。取值原则是让两者能得到近似相等k-shell总数。为此我们定义变化百分比

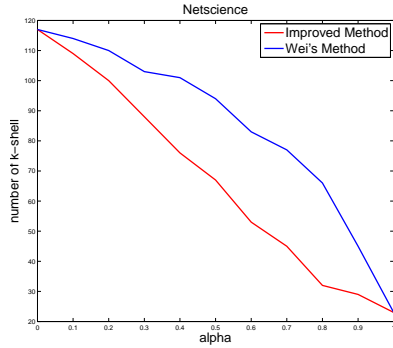
$$\Delta = \frac{|N_w - N_i|}{avg} \quad (3.7)$$



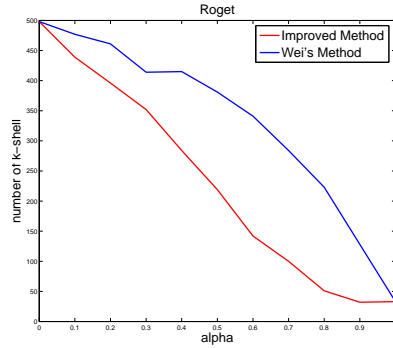
(a) Blogs网络进行k-shell分解



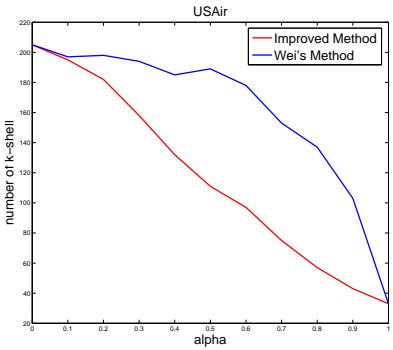
(b) Email网络进行k-shell分解



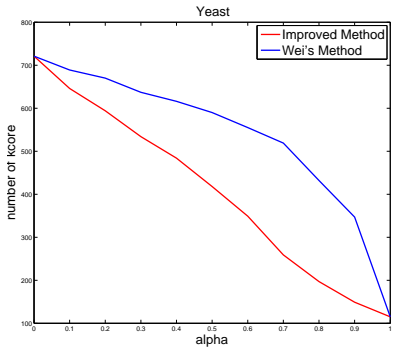
(c) Netscience网络进行k-shell分解



(d) Roget网络进行k-shell分解



(e) USAir网络进行k-shell分解



(f) Yeast网络进行k-shell分解

图 3.9: k-shell分解实验: 其中number of k-shell表示网络分解k-shell的数目, alpha为调节因子。图中红色实线代表改进方法所得k-shell总数。图中蓝色实线代表魏的方法所得k-shell总数。

其中 Δb 表示变化百分比, N_w 表示魏等人方法得到的k-shell总数, N_i 表示改进方法得到k-shell总数, avg 表示 N_w 与 N_i 总和的平均值。如果 $\Delta \leq 1\%$ 我们则认为 N_w 与 N_i 近似相等。

通过k-shell分解实验, α_{iw} 与 α_{ww} 取值如表3.10所示 按照表3.10对每个网络的调节因子进行取值, 然后对各个网络进行k-shell分解实验。将两种方法进行k-

表 3.10: α 取值表: α_{iw} 表示改进方法的调节因子的取值, α_{ww} 表示魏等人方法中调节因子的取值

网络	α_{iw}	N_{iww}	α_{ww}	N_{ww}	Δ
Blogs	0.2	800	0.6	801	0.0006
Email	0.2	439	0.6	444	0.0057
Net-science	0.2	100	0.4	101	0.0100
Roget	0.2	396	0.3	413	0.0050
USAir	0.2	182	0.6	179	0.0083
Yeast	0.2	594	0.5	590	0.0034

shell分解所得的前30个节点进行SIR传播实验。当传染病模型区域稳态时, 改进方法于魏等人方法所得恢复节点数目如表3.11所示 由表3.11可以看出, 当SIR模

表 3.11: 达到稳态所需时间及感染节点数目统计表

网络	Blogs	Email	Net-science	Roget	USAir	Yeast
t_{iw}	9	13	5	13	10	15
N_{riw}	954	749	2.92	914	247	1879
t_{ww}	9	13	5	13	9	15
N_{rww}	942	733	2.82	913	208	1868

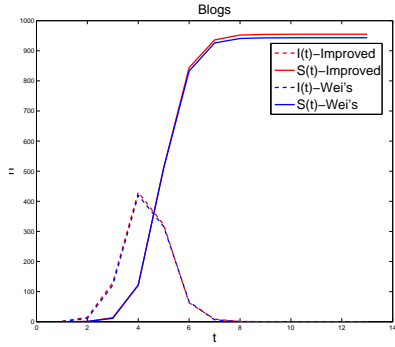
型达到稳态后, 改进的方法所得的被感染节点比魏等人方法所得的被感染节点总数要多。由此说明, 改进方法所得到的前30个节点的传播性比魏等人方法所得到的前30个节点的传播性要大, 也就是说改进方法对网络节点重要度的排序比魏等人方法对网络节点重要度的排序要优越。如图3.10所示是分别取两种方法所得前30个节点进行SIR传播实验的实验结果图。 由图3.10可以观察出: 在SIR传染病模型达到平衡后, 改进的k-shell分解算法比魏等人的方法所得到的传染节点要多。由此可以看出, 改进的k-shell分解算法排序得到的前30个节点的传播能力比魏等人的方法要强。因此可以得出结论: 改进的方法比魏等人的方法在节点重要度排序方面要好。

3.4.4 脆弱性实验

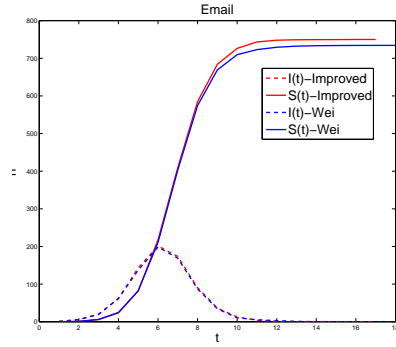
在第二章有提到, 网络中一个节点的脆弱性(vulnerability)定义为 $V_i = \frac{E-E_i}{E}$ 。其中E表示网络效率, 其计算方法在第二章有详解, 而 E_i 表示去掉节点i后的网络效率[74, 75]。网络节点的脆弱性指标可以判断一个节点对网络全局的影响力。下面给出了一个脆弱性计算的例子, 如图3.11是一个含有5个节点8条测地线的的连通图。

1) 未去掉任何节点之前, 网络的效率 $E = \frac{1}{5(5-1)} \times 8 = 0.400$ 。

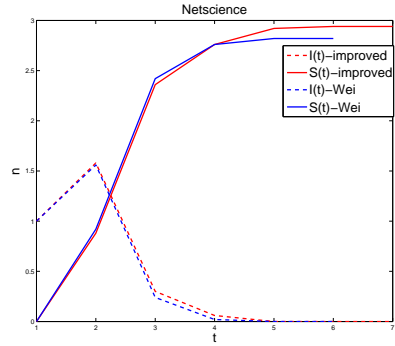
2) 去掉节点1之后, 网络剩下4个节点5条测地线, 则 $E_1 = \frac{1}{4(4-1)} \times 5 = 0.417$ 。



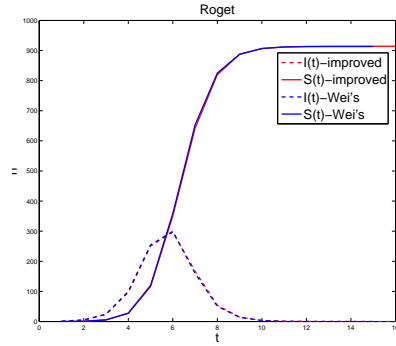
(a) Blogs网络进行传染病实验



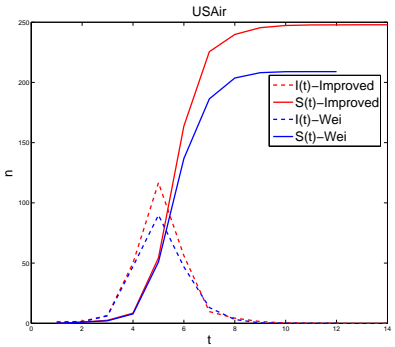
(b) Email网络进行传染病实验



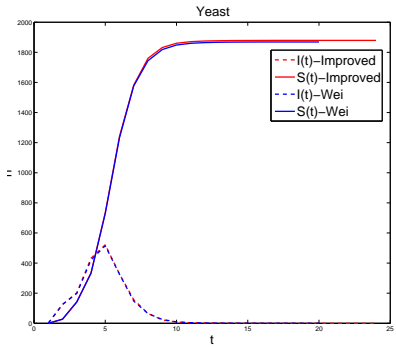
(c) Netscience网络进行传染病实验



(d) Roget网络进行传染病实验



(e) USAir网络进行传染病实验



(f) Yeast网络进行传染病实验

图 3.10: SIR传染病实验:n表示感染节点或恢复节点的个数, t表示迭代步数

- 3) 去掉节点2之后, 网络剩下4个节点4条测地线, 则 $E_2 = \frac{1}{4(4-1)} \times 4 = 0.333$ 。
- 4) 去掉节点3之后, 网络剩下4个节点5条测地线, 则 $E_3 = \frac{1}{4(4-1)} \times 5 = 0.417$ 。
- 5) 去掉节点4之后, 网络剩下4个节点4条测地线, 则 $E_4 = \frac{1}{4(4-1)} \times 4 = 0.333$ 。
- 6) 去掉节点5之后, 网络剩下4个节点6条测地线, 则 $E_5 = \frac{1}{4(4-1)} \times 6 = 0.500$ 。

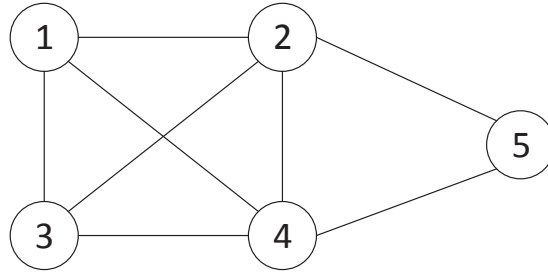


图 3.11: 脆弱性算例

由此可得网络中各节点的脆弱性计算方法如下：

$$\begin{aligned}
 V_1 &= \frac{E - E_1}{E} = \frac{0.400 - 0.417}{0.400} = -0.0425 \\
 V_2 &= \frac{E - E_2}{E} = \frac{0.400 - 0.333}{0.400} = 0.1675 \\
 V_3 &= \frac{E - E_3}{E} = \frac{0.400 - 0.417}{0.400} = -0.0425 \\
 V_4 &= \frac{E - E_4}{E} = \frac{0.400 - 0.333}{0.400} = 0.1675 \\
 V_5 &= \frac{E - E_5}{E} = \frac{0.400 - 0.500}{0.400} = -0.2500
 \end{aligned}$$

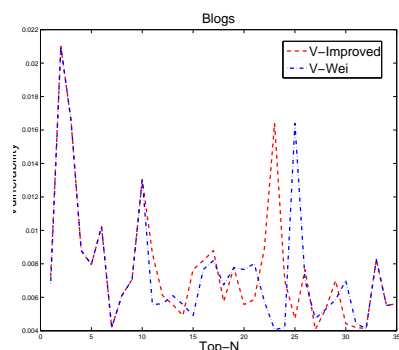
由上述计算可以发现网络节点脆弱性的计算既有正数也有负数，而且所有节点的脆弱性的总和为零。其中，正数表示被删除的网络节点在网络中的连通度较大，删除该节点后，网络整体的连通度变小，具有这种性质的节点影响范围较大，数值越大影响越大。而负数表示被删除的网络节点在网络中的连通度较小，删除该节点后，网络整体的连通度反而变大，具有这种性质的节点影响范围较小，数值越小影响越小。

为了进一步探究改进算法的性能，我们进行了脆弱性实验。实验中调节因子 α 与传播性能实验中的选取是一样的，前30个节点的选取方法也与传播性实验是一样的。脆弱性的实验结果如图3.12所示，实验数据如表3.12所示。

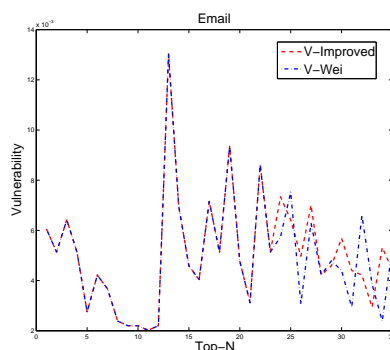
表 3.12: 脆弱性实验结果数据表

网络	V_I	V_w	Δ
Blogs	0.27013295	0.26373182	0.00640112
Email	0.17781447	0.17344658	0.00436762
Net-science	0.18891320	0.18891320	0.0
Roget	0.16063596	0.15762061	0.00301535
USAir	1.0056444	0.9974412	0.0082032
Yeast	0.29598905	0.29598905	0.0

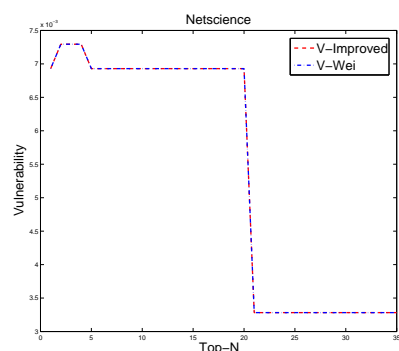
由图3.12和表3.12 可以发现，这六个网络中除了Net-science和Yeast，在其它四个网络中，改进方法排序得到的前30个节点的网络脆弱性之和比魏等人的方



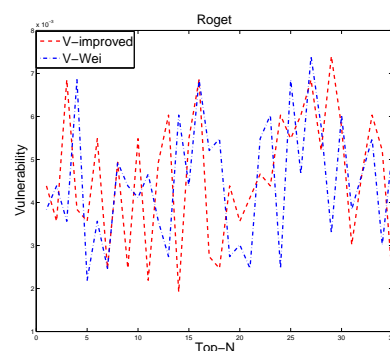
(a) Blogs网络脆弱性实验



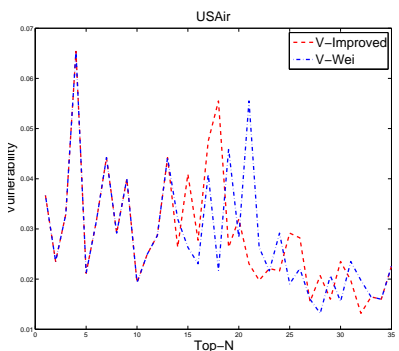
(b) Email网络脆弱性实验



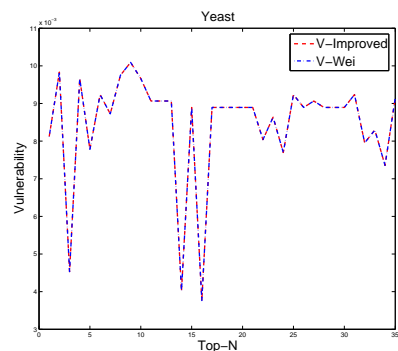
(c) Netscience网络脆弱性实验



(d) Roget网络脆弱性实验



(e) USAir网络脆弱性实验



(f) Yeast网络脆弱性实验

图 3.12: 脆弱性实验: Top-N表示前N个节点的排序顺序, vulnerability表示网络节点脆弱性

法排序得到的前30个节点网络脆弱性之和要大, 这说明改进方法的排序得到的前30个节点对网络的影响力从总体上来说比魏等人的方法排序得到的前30个节点对网络的影响力要大, 这进一步证实了使用改进算法排序所得到的节点的重要度要比魏等人的方法有一定的优越性。

3.5 本章小结

第三章主要对改进的加权k-shell分解算法的进行了阐述。并且在六个真实的网络Blogs、Email、Net-science、Roget、USAir、Yeast分别进行了k-shell 分解实验、传播性能实验、脆弱性实验来验证改进方法的性能。实验结果表明改进的k-shell分解算法比文章的方法在一定程度上有优越性。

第四章 基于有效距离的k-shell分解算法

本章的研究内容主要是基于有效距离的k-shell分解算法。以前对k-shell分解算法的研究主要在加权、无权等方面进行考虑，而未对有效距离这一因素进行分析讨论。本章首先介绍了有效距离理论，然后介绍了基于有效距离的k-shell分解算法，最后对新提出的方法在四个真实网络上进行了实验仿真以验证该想法的有效性及其实用性，并对实验结果进行了分析讨论。

4.1 有效距离理论

2013年，Brockmann 和Helbing等人在science上发表了一篇文章——《The hidden geometry of complex, network-driven contagion phenomena》[76]，该文章中阐述了有效距离理论。Brockmann 等人认为传统的基于节点与节点之间测地线的边数和基于节点与节点之间的距离的研究是不完善的。在现实世界中，影响网络中节点交流的因素不仅包括节点与节点之间测地线的边数和节点与节点之间的距离，还有节点于节点之间信息流传输的大小有关。举个例子，从重庆到北京的航空线距离大约1500公里，大约有40个航班，也就是40条测地线。而从重庆到西宁的航空线距离大约900公里，有2个航班，也就是2个测地线。倘若在重庆突然爆发流感，单单从点与点之间的距离进行考虑，我们可能认为从重庆传播流感到西宁的时间要比从重庆传播到北京的时间短一些。然而从重庆到北京有40多个航班，而从重庆到西宁仅仅两个航班，从这一方面看从重庆到北京传播流感的可能性和传播速度将远大于从重庆到西宁。由此便产生了一个问题，到底是从重庆到北京传播的快呢，还是从重庆到西宁传播的快？

由于传统的基于节点测地线边数和基于距离的理论无法准确衡量节点重要度，Brockmann等人提出了一个新的解决方法——有效距离（Effective Distance）。有效距离不仅考虑节点与节点之间的距离关系，也考虑了节点与节点之间测地线的边数。文献[76]中的实验模拟了H1N1 和SARS病毒的传播，实验结果表明有效距离的模拟结果更加接近真实的疾病传播数据。对于任意两个节点 V_i 和 V_j ，设从节点 V_i 到节点 V_j 的信息流为 F_{ji} ，则节点 i 的信息流定义为：

$$F_i = \sum_j F_{ji} \quad (4.1)$$

从节点 V_i 到节点 V_j 的信息流比例为 P_{ji} 定义为

$$P_{ji} = \frac{F_{ji}}{F_i} \quad (4.2)$$

信息流 F_{ji} 在不同网络中有不同的含义。在交通网络中， F_{ji} 表示节点 V_i 到节点 V_j 的车流量。在水网中， F_{ji} 表示节点 V_i 到节点 V_j 的水流量。在电网中， F_{ji} 表示节点 V_i 到节点 V_j 的输电量。

Brockmann等人将节点 V_i 到节点 V_j 的有效长度 d_{ji} 定义为

$$d_{ji} = (1 - \log P_{ji}) \quad (4.3)$$

由于 $P_{ji} \leq 1$ ，则 $\log P_{ji} \leq 0$ ，所以 $d_{ji} \geq 1$ 。在大多数网络中，从节点 V_i 到节点 V_j 的有效长度 d_{ji} 与从节点 V_j 到节点 V_i 的有效距离 d_{ij} 不相等。

设集合 Γ 表示节点 V_i 到节点 V_j 的所有有效长度的集合，则从节点 V_i 到节点 V_j 的有效距离定义如下

$$D_{ji} = \min \lambda(\Gamma) \quad (4.4)$$

同样的，在大多数网络中， $D_{ji} \neq D_{ij}$ 。

4.2 基于有效距离的k-shell分解方法

到目前为止，k-shell分解算法的研究方法也只停留在传统的基于距离的方法和基于节点之间测地线边数的方法或它们的改进方法中，并未考虑有效距离这一因素。本章提出了基于有效距离的k-shell分解方法。事物都是不断进步发展的，同样的，新提出的基于有效距离的k-shell分解算法可以克服传统方法的一些弊端和不足，考虑问题较为科学。

4.2.0.1 基于有效距离的k-shell分解算法模型

设任意节点 V_i 到任意节点 V_j 的有效距离为 D_{ji} ，节点 i 的度为 DC_i ，则节点 i 的有效距离重要度EDI (Effective Distance Importance) 定义为

$$EDI_i = DC_i^\alpha \left(\sum_{j \in \Psi} \frac{1}{D_{ji}} \right)^{1-\alpha} \quad (4.5)$$

其中， α 为调节因子， $\alpha \in [0, 1]$ 。 Ψ 为节点 i 的所有邻居节点的集合。因有效距离越大与节点的影响力成反比，所以取倒数。

基于有效距离的k-shell分解方法的计算步骤与k-shell分解方法的计算步骤类似，计算方法如下：

第1步 按照公式(4.5)依次计算每个节点有效距离重要度，并向下取整。

第2步 把向上取整后的有效距离重要度按照经典的k-shell算法进行分解，直到所有的点都分解到相应的k-shell中。

4.2.0.2 算例

为了加深对改进方法的理解，在此列出一个简单的算例。网络例子如图4.1所示 首先，我们计算图4.1中所有节点之间的有效距离。以节点2 为例，从节点2流

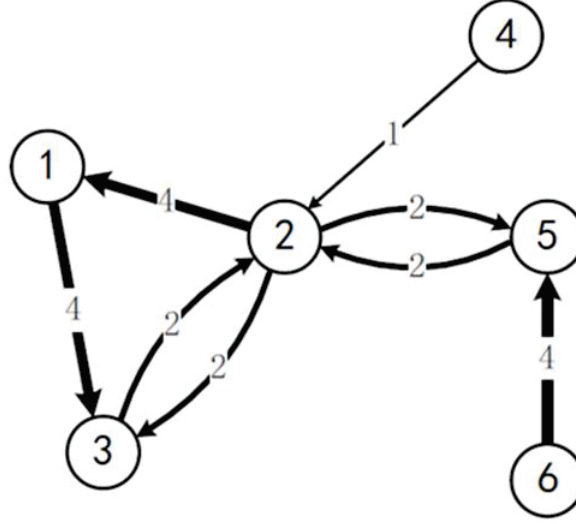


图 4.1: 基于有效距离的k-shell分解方法算例

出的信息流 F_2 的计算方法如下

$$F_2 = F_{12} + F_{32} + F_{52} = 4 + 2 + 2 = 8$$

则节点2到各节点的有效距离计算方法如下：

$$D_{12} = 1 - \log \frac{F_{12}}{F_2} = 1 - \log \frac{4}{8} = 1.6931$$

$$D_{32} = 1 - \log \frac{F_{32}}{F_2} = 1 - \log \frac{2}{8} = 2.3863$$

$$D_{42} = 1 - \log \frac{F_{42}}{F_2} = 1 - \log \frac{0}{8} = 0$$

$$D_{52} = 1 - \log \frac{F_{52}}{F_2} = 1 - \log \frac{2}{8} = 2.3863$$

依次类推各节点之间的有效距离如表4.1所示 计算完图4.1中各节点的有效距离后，令 α ，然后按照公式4.5 进行有效距离重要度的计算。计算结果如表4.2 所示 由表格4.2可以发现，此次计算的EDI结果都为正数，则不必再进行取整运算。下一步进行k-shell分解。先从网络中删除节点4，再进行EDI 计算。依次类推，直到所有节点都分解到相应的核中

表 4.1: 改进k-shell分解算法权重计算表

节点	1	2	3	4	5	6
1	Inf	1.6931	Inf	Inf	Inf	Inf
2	Inf	Inf	1	1	1	Inf
3	1	2.3863	Inf	Inf	Inf	Inf
4	Inf	Inf	Inf	Inf	Inf	Inf
5	Inf	2.3863	Inf	Inf	Inf	1
6	Inf	Inf	Inf	Inf	Inf	Inf

表 4.2: 改进k-shell分解算法权重计算表

节点	1	2	3	4	5	6
EDI	2	4	2	1	2	2

4.3 仿真实验与分析

为探究本章提出的基于有效距离k-shell分解方法的可行性与有效性，我们在四个真实网络中进行了SI传播实验和kendall系数相关性实验。这四个网络的相关统计特性如表4.3所示。

4.3.1 实验数据

本章设计了两组对比实验，将我们提出的基于有效距离的k-shell分解分别与经典k-shell分解、加权k-shell分解进行了实验对比，以验证基于有效距离k-shell分解实验的有效性和实用性。本章在四个真实网络上进行了仿真实验，这四个网络分别是C.elegans、netscience、polblogs和USairport 网络。

- 1) C.elegans，有向加权网络。该网络是秀丽线虫的神经网络，数据详细信息见文章[77].
- 2) netscience，无向有权网络。该网络是科研合作网络，网络中的每个节点表示一个科学家。数据集下载地址为：<http://www-personal.umich.edu/mejn/netdata/>
- 3) polblogs，有向无权网络。该网络是美国政要之间的微博关系网络，网络中每个节点表示一个美国政客。数据集的下载地址为<http://www-personal.umich.edu/mejn/netdata/>。
- 4) USairport，无向无权网络。该网络是美国机场之间航空网络。数据集的下载地址为<http://www.toreopsahl.com/datasets/usairports>。

表 4.3: 四个真实网络的统计特性其中, N 表示网络节点数目, M 表示网络边数, $\langle k \rangle$ 表示网络平均加权重, D 表示网络直径, C 表示平均聚集系数。

网络	性质	N	M	$\langle k \rangle$	D	C
Celegans	有向加权	302	3620	11.987	7	0.625
Netscience	无向加权	1536	4112	5.354	11	0.619
Polblogs	有向无权	1490	19025	12.768	9	0.172
Celegans	有向加权	332	4262	12.807	8	0.318

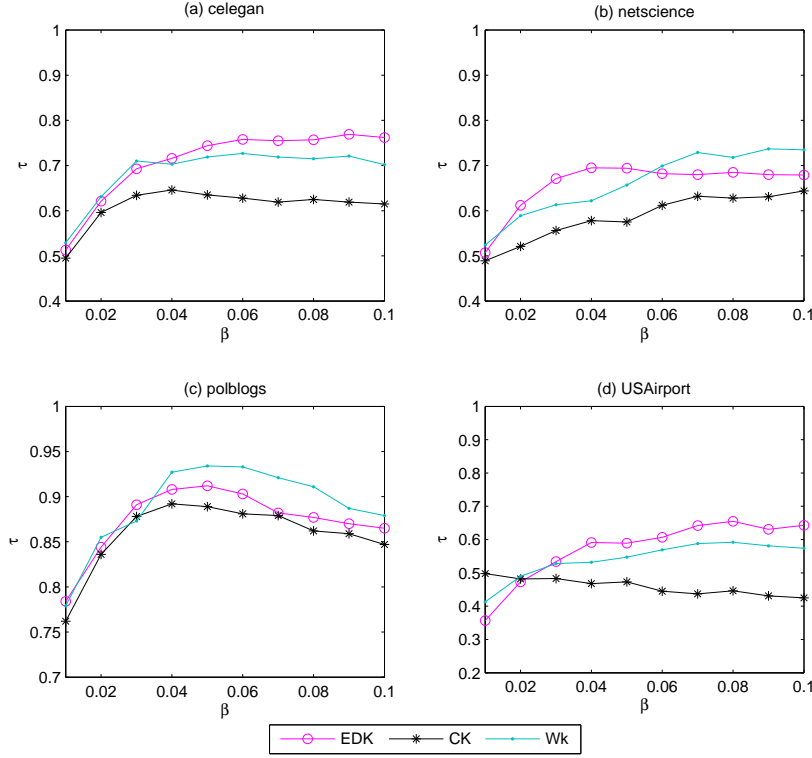
4.3.2 Kendall秩相关性实验

统计学中, Kendall秩相关性(Kendall rank correlation coefficient)通常指的是Kendall tau相关性高[78]。在统计假设检验中, Kendall相关性通常用来进行t检验。对于两个相互独立的变量 $X = x_1, x_2, \dots, x_n$ 和 $Y = y_1, y_2, \dots, y_n$, 则被比较的观察值分别为 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 。对于任意两对观察值 (x_i, y_i) 和 (x_j, y_j) , 如果 $x_i > x_j \wedge y_i < y_j$ 或者 $x_i < x_j \wedge y_i > y_j$,则这两对观察值是不协调的; 如果 $x_i > x_j \wedge y_i > y_j$ 或者 $x_i < x_j \wedge y_i < y_j$,则这两对观察值是协调的; 如果 $x_i = x_j$ 或者 $y_i = y_j$,则这两对观察值无关。由此可以计算Kendall协调因子 τ 的值如下所示

$$\tau = \frac{\text{相互协调观察值的数目} - \text{相互不协调观察值的数目}}{n(n-1)/2} \quad (4.6)$$

其中, τ 值越大, 两组观察值的相关性越高; τ 值越小, 两组观察值的相关性越低。

为了较为清晰地衡量不同分解方法与SI传染病模型的相关性, 我们采用Kendall秩相关系数作为评估标准[79, 80]。Kendall秩相关系数可以较为清楚地表示出两种排序之间的匹配性, 可以直观体现两种方法在数值结果中的相关性。在本小节的Kendall相关性实验中, 我们依次设定SI传染病模型中的传染概率从0.01开始, 以0.01的间隔递增至0.1。然后将不同分解方法得到的排序结果与SI模型得到的排序结果进行Kendall相关性计算, 实验结果如图4.2所示。从图4.2可以观察出在celegan网络和USAirport网络中, 基于有效距离的k-shell分解(EDK)明显优于经典k-shell分解算法(CK)和加权k-shell分解算法(WK)。在netscience网络中, 当 $\beta \leq 0.05$ 时, 基于有效距离的k-shell分解(EDK)明显优于其它两种方法, 而当 $\beta > 0.05$ 时, 基于有效距离的k-shell分解方法略逊色于加权k-shell分解算法。在polblogs网络中, 当 $\beta > 0.04$ 时, 基于有效距离的k-shell分解方法略逊色于加权k-shell分解算法。综上所述, 在大多数情况下基于有效距离的k-shell分解方法与SI传染病模型的相关性比经典k-shell分解方法和加权k-shell分解方法优秀, 初步证明了基于有效距离k-shell分解方法的有效性和实用性。


 图 4.2: Kendall相关性实验: β 表示感染概率, τ 表示kendall相关性

4.3.3 Susceptible-Infected疾病传播实验

4.3.3.1 SI传染病模型

对于传染病模型, 本文中的第三章已经提到过SIR(Susceptible – Infected – Recovery)传染病模型, 而本章中所采用的是SI(Susceptible – Infected)传染病模型[81, 82]. 在SI模型中, 网络中的节点可以分为两种: 易感人群(Susceptible individuals)和已感染人群(Infected individuals). 设网络中节点总数为 N , 网络中易感人群的比例为 $S(t)$, 已感染人群的比例为 $I(t)$, 每个病人每天接触并传染的人数为 Ψ , 初始传染源为 I_0 , 则SI传染病模型的数学形式如下公式:

$$\begin{cases} \frac{dI}{dt} = \lambda I(1 - I) \\ I(0) = I_0 \end{cases} \quad (4.7)$$

通过公式4.7可得SI传染病模型随时间变化的趋势如图4.3

4.3.3.2 SI疾病传播实验

为了验证基于有效距离k-shell分解的有效性和可行性, 我们在四个真实网络上分别进行了SI疾病传播实验. 实验中设传染率 $\lambda = 0.1$, 初始传染源分别

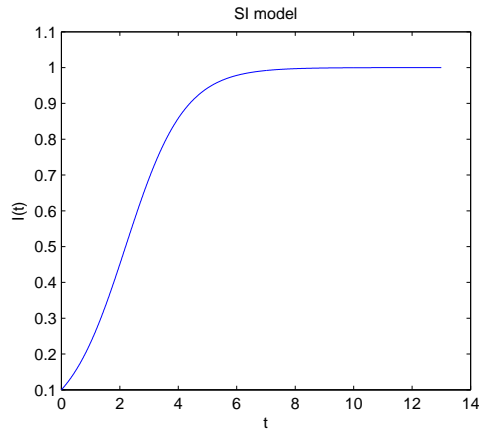


图 4.3: SI传染病模型演化趋势图

为有效距离k-shell分解排序所得的前10个节点、魏等人提出的加权k-shell分解排序所得的前10个节点和经典k-shell分解排序所得的前10个节点。实验结果如图4.4所示，其中CK(Classical K-shell)曲线表示经典k-shell分解实验排序所得前10个节点作为初始传染节点进行SI传播实验的结果，WK曲线表示经典魏等人加权k-shell分解实验排序所得前10个节点作为初始传染节点进行SI传播实验的结果，EDK(Effective Distance K-shell)曲线表示基于有效距离k-shell分解排序所得的前10个节点进行SI传播实验的结果。从表格4.4中可以观察到：celegans网络中，基于有效距离的k-shell分解实验结果明显优于其它两种方法。在netscience网络中，经典k-shell分解比基于有效距离k-shell分解结果要好一些。在polblogs网络中，基于有效距离的k-shell分解实验结果比其它两种方法优秀，但是从图中不太容易观察，要从表格数据进行分析。在USAirport网络中，基于有效距离k-shell分解明显优于其它两种方法。综上所述，基于有效距离的k-shell分解算法比经典k-shell分解算法和魏等人提出的加权k-shell分解算法在大多数情况下具有优越性。

表 4.4: SI传染病实验中被传染节点数目统计表

网络	celegans	netscience	polblogs	USAirport
t	20	30	20	20
N_{EDK}	254.8	60.5	923.7	254.6
N_{WK}	243.9	51.6	902.8	240.6
N_{CK}	234.8	66.3	906.7	230.8

4.4 本章小结

本章主要提出了一种新的基于有效距离k-shell分解方法。首先介绍了有效距离理论，然后提出了基于有效距离的k-shell分解方法。为了验证提出方法的有效性和可行性，又在四个真实网络上分别进行了kendall相关性实验和SI疾病传播实

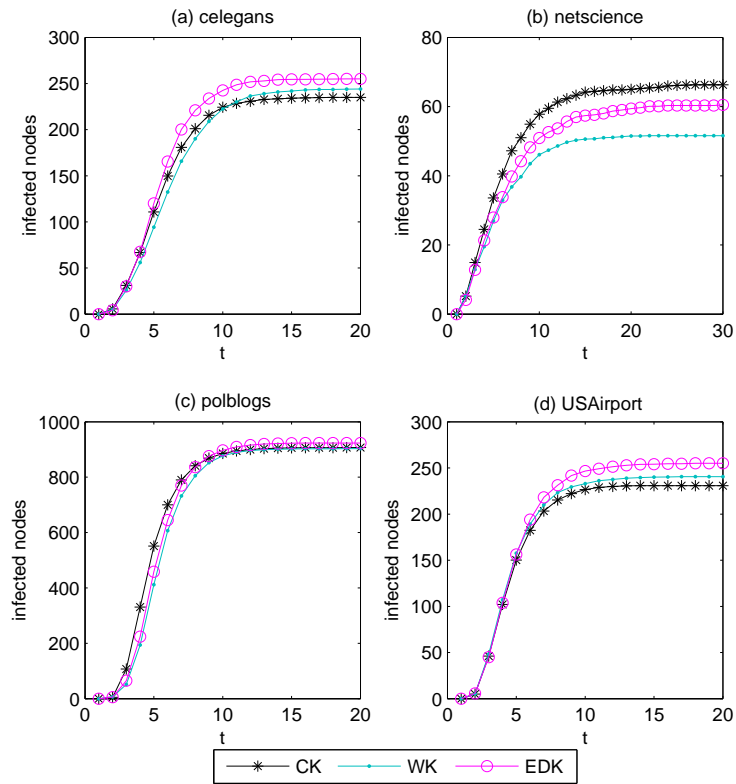


图 4.4: SI传染病实验: t 表示迭代步数, infected nodes表示感染节点的个数

验。实验结果表明本章提出的基于有效距离的k-shell 分解方法在大多数情况性比魏提出的加权k-shell分解方法和经典k-shell分解方法要优越。

第五章 总结与展望

5.1 本文总结

最近几年, 复杂网络作为一门新兴学科受到了学术界的广泛关注, 同时也吸引了众多研究学者参与其中。人们对于复杂网络的研究主要目的是将复杂网络简单化、分析网络特性以便人们更好的处理和控制在网络中正在发生或预防将要发生的灾难, 比如人们通过对社交网络的研究可以阻断疾病传播和谣言散步, 通过对交通网络的研究避免交通堵塞和交通事故等等。如何识别复杂网络中重要节点、阻断或加强节点间的信息交流吸引了众多研究学者。人们对此提出了很多多的解决方法, 但是现有的方法在一定程度上都存在一定的缺陷, 难以满足人们所有的需求。本文跟随各位前辈们的步伐, 主要对复杂网络节点重要度进行了研究, 并提出了一些个人观点和方法。

本文的主要研究内容和相关工作可以总结为以下几点:

(1) 本文主要对复杂网络节点重要度进行研究, 首先对复杂网络节点重要度典型的评估算法进行了阐述。主要包括基于节点度中心性的评估方法、基于路径的评估方法、基于特征向量的评估方法、基于随机游走的评估方法、基于节点位置的评估方法。

(2) 本文在前人的基础上, 对k-shell分解方法进行了改进。K-shell算法中, 加权度的计算将会影响到后续节点k-shell的分配。已存在的放在在加权度的计算上还存在一些不足, 本文通过改进节点加权度的计算对已有的加权k-shell分解方法进行了改进。随后, 我们对改进的k-shell分解方法进行了实验仿真, 以验证改进方法的性能。在实验仿真中, 我们将已有方法和改进方法进行了对比。改进方法在k-shell分解实验、SIR疾病传播实验和脆弱性实验中都具有一定的优势。从而验证了改进方法的有效性和实用性。

(3) 传统的k-shell分解方法中, 节点之间的距离是根据节点与节点之间测地线或者实际地理位置来确定的。这种方法在计算节点间距离时忽略了节点所处网络的结构和邻居节点间的相关信息。当网络中的节点进行单边有向信息交流时, 传统的计算距离的方法会造成某些节点之间的距离变得无穷大。为了解决这一问题, 我们将有效距离理论与k-shell分解方法进行有机结合。在对网络进行分解之前, 通过有效距离计算节点的有效度, 然后根据有效度再对网络进行分解。我们分别在四个真实网络上对基于有效距离的k-shell分解方法进行了实验仿真以验证该方法的有效性。在kendall相关性实验和SI疾病传播实验中, 基于有效距离的k-shell分解方法与传统的k-shell分解方法相比有一定的优越性。

5.2 未来工作展望

复杂网络节点重要度研究主要是从海量节点中找出影响力大、影响范围广的节点，该研究工作对于复杂网络学科的研究和应用拓展有重要的意义和价值。目前的研究工作还有待完善，或许在未来某一天会发现新大陆。就像牛顿经典力学到爱因斯坦广义相对论的发展一样，复杂网络学科在未来会带给人们意想不到的惊喜。当然一切事物的发展都是从量变到质变的过程，只有复杂网络学科的发展达到一定程度时，才能实现质的飞跃。然而目前的复杂网络还存在一些亟需解决的问题，主要可以归纳为以下几点：第一，理论研究跟不上实际工程应用的步伐，有一些工程性的问题得不到理论的指导。第二，对于网络结构的研究常常忽略网络的性能，导致研究结果存在片面性问题。第三，理论与实验验证相分离，古人常说“纸上得来终觉浅，绝知此事要躬行”，然而我们在复杂网络的理论研究中很多时候得不到实验的验证、或者验证过程不科学等，尤其一些研究学者只具备理论知识而不动实际操作，往往会造成理论与实践相背离的现象。

对于未来复杂网络的研究，本文进行了总结，主要有以下几点。

(1) 对于目前的复杂网络节点重要度的算法，大多数算法存在一个明显的问题是：有的算法形式简单、运算速度快，但是结果不是很准确；有的算法可以得到相对准确的结果，但是表述繁琐，计算复杂。比如本文提出的两种基于k-shell分解的算法在计算机运行过程中时间复杂度和空间复杂度都比较高，既消耗时间又消耗内存。能不能出现一种或几种形式简单、运算效率高、结果准确、资源占用率小的算法是未来复杂网络节点重要度研究中需要解决的问题之一。

(2) 目前的研究工作中，常常将网络的结构和网络的性能特征分离。显然无论是网络结构还是网络的性能特征在复杂网络节点重要度的研究中都具有不可取代的地位。比如本文提出的基于有效距离的k-shell分解的方法只考虑了节点在网络中的位置，而没有考虑节点所处网络的性能特征。虽然本文基于有效距离的k-shell方法相比原有的方法具有一定的优越性，但不可否认的是该方法考虑的因素是片面的。如何将网络结构与网络性能进行有机结合是未来需要解决的另一重要问题。

(3) 对于实时变化网络，该怎样判断网络节点重要度呢？由于实时变化网络中，节点参数、性质、位置等因素会随时间而变化。在不断变化的网络中，我们又该如何快速的判断重要节点？如何准确的判断重要节点？本文的理论研究没有考虑时间因素，实验验证也都是在静态网络上进行操作的，诚然，这样的理论研究和实验验证是跟不上时间的步伐的。

学海无涯，复杂网络的研究也是无止境的。鄙人不才，对于复杂网络的认知也仅懂皮毛、能力有限，很多问题考虑不周，对复杂网络存在的问题及日后的发

展也只能表述这么多。所谓勤能补拙，在日后的工作学习中，还需多多思考，多多实践。

参考文献

- [1] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用. 清华大学出版社有限公司, 2006.
- [2] 汪秉宏, 何大韧, 刘宗华. 复杂系统与复杂网络, volume 20. 高等教育出版社, 2009.
- [3] 陆哲明, 郭世泽. 复杂网络基础理论. 科学出版社, 2012.
- [4] 汪秉宏, 刘之景, 严钢, 周涛, 柏文洁. 复杂网络研究概述. 物理, 34(01):0–0, 2005.
- [5] 汪小帆. 21 世纪的网络科学: 从热点论文到社会应用. 科学观察, 5:019, 2007.
- [6] 汪小帆, 方锦清. 略论复杂性问题和非线性复杂网络系统的研究. 科技导报, 22(0402):9–12, 2004.
- [7] Paolo Crucitti, Vito Latora, Massimo Marchiori, and Andrea Rapisarda. Error and attack tolerance of complex networks. *Physica A: Statistical Mechanics and its Applications*, 340(1):388–394, 2004.
- [8] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *nature*, 393(6684):440–442, 1998.
- [9] Paul Erd6s and A Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci*, 5:17–61, 1960.
- [10] Xiang Li and Guanrong Chen. A local-world evolving network model. *Physica A: Statistical Mechanics and its Applications*, 328(1):274–286, 2003.
- [11] Duncan J Watts. The” new” science of networks. *Annual review of sociology*, pages 243–270, 2004.
- [12] Stanley Milgram. The small world problem. *Psychology today*, 2(1):60–67, 1967.
- [13] Charles K Chui and Guanrong Chen. *Kalman filtering: with real-time applications*. Springer Science & Business Media, 2008.

- [14] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *science*, 286(5439):509–512, 1999.
- [15] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [16] Réka Albert and Albert-László Barabási. Statistical mechanics of complex networks. *Reviews of modern physics*, 74(1):47, 2002.
- [17] Sergey N Dorogovtsev and Jose FF Mendes. Evolution of networks. *Advances in physics*, 51(4):1079–1187, 2002.
- [18] Mark EJ Newman. The mathematics of networks. *The new palgrave encyclopedia of economics*, 2(2008):1–12, 2008.
- [19] Ronald S Burt, Michael J Minor, and Richard D Alba. *Applied network analysis: A methodological introduction*. Sage Publications Beverly Hills, CA, 1983.
- [20] Alex Arenas, Albert Díaz-Guilera, Jurgen Kurths, Yamir Moreno, and Changsong Zhou. Synchronization in complex networks. *Physics Reports*, 469(3):93–153, 2008.
- [21] Yamir Moreno, Maziar Nekovee, and Amalio F Pacheco. Dynamics of rumor spreading in complex networks. *Physical Review E*, 69(6):066130, 2004.
- [22] Liang Zhao, Kwangho Park, and Ying-Cheng Lai. Attack vulnerability of scale-free networks due to cascading breakdown. *Physical review E*, 70(3):035101, 2004.
- [23] Daijun Wei, Xinyang Deng, Xiaoge Zhang, Yong Deng, and Sankaran Mahadevan. Identifying influential nodes in weighted networks based on evidence theory. *Physica A: Statistical Mechanics and its Applications*, 392(10):2564–2575, 2013.
- [24] Adilson E Motter and Ying-Cheng Lai. Cascade-based attacks on complex networks. *Physical Review E*, 66(6):065102, 2002.
- [25] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *nature*, 406(6794):378–382, 2000.

-
- [26] Stefano Boccaletti, Vito Latora, Yamir Moreno, Martin Chavez, and D-U Hwang. Complex networks: Structure and dynamics. *Physics reports*, 424(4):175–308, 2006.
- [27] Marián Boguná and Romualdo Pastor-Satorras. Epidemic spreading in correlated complex networks. *Physical Review E*, 66(4):047104, 2002.
- [28] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1978.
- [29] 郭强, 汪秉宏, 刘建国, 任卓明. 复杂网络中节点重要性排序的研究进展. *物理学报*, 62(17):178901–178901, 2013.
- [30] 刘建国, 郭强, 汪秉宏, 任卓明, 邵凤. 基于度与集聚系数的网络节点重要性度量方法研究. *物理学报*, 62(12):128901–128901, 2013.
- [31] 淦文燕, 朱熙, 赫南, 李德毅. 复杂网络中重要性节点发掘综述. *计算机科学*, 34(12):1–5, 2007.
- [32] Karen Stephenson and Marvin Zelen. Rethinking centrality: Methods and examples. *Social Networks*, 11(1):1–37, 1989.
- [33] 吕琳媛, 任晓龙. 网络重要节点排序方法综述. *科学通报*, 59(13):1175–1197, 2014.
- [34] Stephen P Borgatti. Centrality and network flow. *Social networks*, 27(1):55–71, 2005.
- [35] Kurt Bryan and Tanya Leise. The 25,000,000,000 eigenvector: The linear algebra behind google. *Siam Review*, 48(3):569–581, 2006.
- [36] Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, and Tao Zhou. Leaders in social networks, the delicious case. *PloS one*, 6(6):e21202, 2011.
- [37] Pavel Berkhin. A survey on pagerank computing. *Internet Mathematics*, 2(1):73–120, 2005.
- [38] Maksim Kitsak, Lazaros K Gallos, Shlomo Havlin, Fredrik Liljeros, Lev Muchnik, H Eugene Stanley, and Hernán A Makse. Identification of influential spreaders in complex networks. *Nature physics*, 6(11):888–893, 2010.

- [39] Shai Carmi, Shlomo Havlin, Scott Kirkpatrick, Yuval Shavitt, and Eran Shir. A model of internet topology using k-shell decomposition. *Proceedings of the National Academy of Sciences*, 104(27):11150–11154, 2007.
- [40] An Zeng and Cheng-Jun Zhang. Ranking spreaders by decomposing complex networks. *Physics Letters A*, 377(14):1031–1035, 2013.
- [41] Antonios Garas, Frank Schweitzer, and Shlomo Havlin. A k-shell decomposition method for weighted networks. *New Journal of Physics*, 14(8):083030, 2012.
- [42] Jian-Guo Liu, Zhuo-Ming Ren, and Qiang Guo. Ranking the spreading influence in complex networks. *Physica A: Statistical Mechanics and its Applications*, 392(18):4154–4159, 2013.
- [43] Duncan S Callaway, Mark EJ Newman, Steven H Strogatz, and Duncan J Watts. Network robustness and fragility: Percolation on random graphs. *Physical review letters*, 85(25):5468, 2000.
- [44] Reuven Cohen, Keren Erez, Daniel Ben-Avraham, and Shlomo Havlin. Break-down of the internet under intentional attack. *Physical review letters*, 86(16):3682, 2001.
- [45] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [46] Duan-Bing Chen, Hui Gao, Linyuan Lü, and Tao Zhou. Identifying influential nodes in large-scale directed networks: the role of clustering. *PloS one*, 8(10):e77455, 2013.
- [47] Shlomi Dolev, Yuval Elovici, and Rami Puzis. Routing betweenness centrality. *Journal of the ACM (JACM)*, 57(4):25, 2010.
- [48] Xue-Qi Cheng, Fu-Xin Ren, Hua-Wei Shen, Zi-Ke Zhang, and Tao Zhou. Bridgeness: a local index on edge significance in maintaining global connectivity. *Journal of Statistical Mechanics: Theory and Experiment*, 2010(10):P10011, 2010.
- [49] Linton C Freeman. A set of measures of centrality based on betweenness. *Sociometry*, pages 35–41, 1977.

-
- [50] Ernesto Estrada and Juan A Rodriguez-Velazquez. Subgraph centrality in complex networks. *Physical Review E*, 71(5):056103, 2005.
- [51] Mark EJ Newman. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Physical review E*, 64(1):016132, 2001.
- [52] 王延庆. 基于接连失效的复杂网络节点重要性评估. *网络安全技术与应用*, (3):59–61, 2008.
- [53] K-I Goh, B Kahng, and D Kim. Universal behavior of load distribution in scale-free networks. *Physical Review Letters*, 87(27):278701, 2001.
- [54] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of mathematical sociology*, 25(2):163–177, 2001.
- [55] Zhou Tao, Liu Jian-Guo, and Wang Bing-Hong. Notes on the algorithm for calculating betweenness. *Chinese Physics Letters*, 23(8):2327, 2006.
- [56] Gang Yan, Tao Zhou, Bo Hu, Zhong-Qian Fu, and Bing-Hong Wang. Efficient routing on complex networks. *Physical Review E*, 73(4):046108, 2006.
- [57] 姚虹, 田柳, 狄增如. 权重分布对加权网络效率的影响. *物理学报*, 60(2):797–802, 2011.
- [58] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [59] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201, 2001.
- [60] Robert Poulin, M-C Boily, and Benoît R Mâsse. Dynamical systems to define centrality in social networks. *Social networks*, 22(3):187–220, 2000.
- [61] Sergey Brin and Lawrence Page. Reprint of: The anatomy of a large-scale hypertextual web search engine. *Computer networks*, 56(18):3825–3833, 2012.
- [62] Filippo Radicchi, Santo Fortunato, Benjamin Markines, and Alessandro Vespignani. Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80(5):056103, 2009.
- [63] Jon M Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM (JACM)*, 46(5):604–632, 1999.

-
- [64] Ronny Lempel and Shlomo Moran. The stochastic approach for link-structure analysis (salsa) and the tlc effect. *Computer Networks*, 33(1):387–401, 2000.
 - [65] 陈关荣, 汪小帆, 李翔. 网络科学导论. 高等教育出版社, 2012.
 - [66] Bo Wei, Jie Liu, Daijun Wei, Cai Gao, and Yong Deng. Weighted k-shell decomposition for complex networks based on potential edge weights. *Physica A: Statistical Mechanics and its Applications*, 420:277–283, 2015.
 - [67] 邵凤, 胡兆龙, 郭强, 任卓明, 刘建国. 复杂网络中最小k-核节点的传播能力分析. *物理学报*, 62(10):1089–1089, 2013.
 - [68] Bonan Hou, Yiping Yao, and Dongsheng Liao. Identifying all-around nodes for spreading dynamics in complex networks. *Physica A: Statistical Mechanics and its Applications*, 391(15):4012–4017, 2012.
 - [69] 朱志良, 汪小帆, 赵之滢, 于海. 基于网络社团结构的节点传播影响力分析. *计算机学报*, 37(4):753–766, 2014.
 - [70] Sir Ronald Aylmer Fisher, Ronald Aylmer Fisher, Statistiker Genetiker, Ronald Aylmer Fisher, Statistician Genetician, Ronald Aylmer Fisher, and Statisticien Généticien. *The design of experiments*, volume 12. Oliver and Boyd Edinburgh, 1960.
 - [71] Zhou Tao, Fu Zhongqian, and Wang Binghong. Epidemic dynamics on complex networks. *Progress in Natural Science*, 16(5):452–457, 2006.
 - [72] 牛永伟, 王达, 曾燕, 汪秉宏, 周佩玲, 周涛, 傅忠谦. 复杂网络上传播动力学研究综述. *自然科学进展*, 15(5):513–518, 2005.
 - [73] Feng Fu, Daniel I Rosenbloom, Long Wang, and Martin A Nowak. Imitation dynamics of vaccination behaviour on social networks. *Proceedings of the Royal Society of London B: Biological Sciences*, 278(1702):42–49, 2011.
 - [74] Cynthia Phillips and Laura Painton Swiler. A graph-based system for network-vulnerability analysis. In *Proceedings of the 1998 workshop on New security paradigms*, pages 71–79. ACM, 1998.
 - [75] Sushil Jajodia, Steven Noel, and Brian O’ Berry. Topological analysis of network attack vulnerability. In *Managing Cyber Threats*, pages 247–266. Springer, 2005.

- [76] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.
- [77] Jordi Duch and Alex Arenas. Community detection in complex networks using extremal optimization. *Physical review E*, 72(2):027104, 2005.
- [78] Maurice George Kendall et al. The advanced theory of statistics. *The advanced theory of statistics.*, (2nd Ed), 1946.
- [79] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.
- [80] Maurice G Kendall. The treatment of ties in ranking problems. *Biometrika*, 33(3):239–251, 1945.
- [81] Romualdo Pastor-Satorras and Alessandro Vespignani. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 63(6):066117, 2001.
- [82] Xiao-Long Peng, Xin-Jian Xu, Xinchu Fu, and Tao Zhou. Vaccination intervention on epidemic dynamics in networks. *Physical Review E*, 87(2):022813, 2013.

发表文章目录

- [1] 宋起超. 一种新的改进的加权k-核分解方法[J]. 软件工程,2016,01:21-22.

简 历

基本情况

宋起超, 女, 山东烟台人, 1989 年 4 月出生, 未婚, 西南大学计算机与信息科学学院在读硕士研究生.

教育状况

2009 年 9 月至 2013 年 7 月, 重庆邮电大学计算机科学与技术学院, 本科, 专业: 信息安全.

2013 年 9 月至 2016 年 7 月, 西南大学计算机与信息科学学院, 硕士研究生, 专业: 计算机应用技术.

工作经历

无.

研究兴趣

复杂网络, 分形.

致 谢

时光匆匆，三年已去，研究生的生活也接近尾声，心中百感交集。这三年有开心、有失落、有成功、有失败，让我受益匪浅。每个人在成长过程中会遇到很多过客，也会遇到很多贵人。我的老师邓勇教授对我们言传身教，让我领略了一个学者应有的风采和素养。他就想火种，点燃每个学生心中的科研之火，让我们感受到了科研工作的高尚与伟大。虽然我没有达到一个科研工作者应有的素养，但是我对科研工作者确实充满深深地景仰与尊敬。我的老师是我见过最敬业、最正直的老师，每每工作到深夜，加班加点的进行科研工作。讲课时，风趣幽默、生动形象、引人入胜，听其一席话胜读十年书。受到这样一个老师的指导，真的是三生有幸，不枉这三年时光。

这三年有老师的谆谆教诲，也有同学朋友的陪伴与帮助。在此要特别感谢李美竹、张齐、刘洋同学，感谢高才师兄、杜雨弦师兄、兰鑫师兄、陈诗语师姐、张海鑫师姐、刘娟师姐、张晓革师兄、王庆师兄、罗川海师弟、刘洁学妹、郑熙学妹、李娅老师、李云鹏学弟、张荣师弟、武振兴师弟、王洪苹师妹等等。各位师兄姐妹刻苦奋斗、团结一致，共同构建了西南大学25 教1009 良好的科研学术氛围，让我有一个良好的学习环境，深受其益。

最后我也特别感谢我的家人。感谢我的父母给了我健康的体魄，让我健康成长；感谢他们为我提供的良好家庭环境，让我无忧无虑；感谢我的姐姐和弟弟对我的陪伴，让我懂得什么是手足之情。感谢他们在我困难时刻，不计任何回报的帮助我走出困境，在我迷惘彷徨时给我指明方向。

在此论文完成之际，谨在此向多年来给予我关心和帮助的老师、同学、朋友和家人表示由衷的感谢！

宋起超

2016年3月



知网查重限时 7折 最高可优惠 120元

本科定稿，硕博定稿，查重结果与学校一致

立即检测

免费论文查重: <http://www.paperyy.com>

3亿免费文献下载: <http://www.ixueshu.com>

超值论文自动降重: http://www.paperyy.com/reduce_repetition

PPT免费模版下载: <http://ppt.ixueshu.com>

阅读此文的还阅读了:

1. [基于概念格的复杂网络重要节点评估](#)
2. [基于多属性的复杂网络节点重要度识别方法](#)
3. [基于传输效率矩阵的复杂网络节点重要度排序方法](#)
4. [基于复杂网络的信息传播研究](#)
5. [复杂网络在交通网络节点重要度评估中的应用](#)
6. [基于复杂网络重叠社区电网节点重要度研究](#)
7. [复杂网络中最小K-核节点的传播能力分析](#)
8. [基于复杂网络的路由策略研究](#)
9. [基于叠加随机游走的复杂网络节点重要度评估方法](#)
10. [基于复杂网络的卫星重要度评估](#)
11. [网络节点重要度的快速评估方法探索](#)
12. [复杂网络中节点重要度评估的节点收缩方法](#)
13. [复杂网络中节点重要度的一个评估指标](#)
14. [基于k-shell的复杂网络影响力最大化算法研究](#)
15. [基于改进k-shell算法的节点影响力的识别](#)
16. [基于任务的物流保障网络节点重要度分析](#)
17. [基于复杂网络的节点影响力评价模型研究](#)
18. [一种新的复杂网络节点重要度分析方法](#)
19. [复杂网络中节点重要度评估](#)
20. [基于复杂网络理论对长三角雾霾网络的城市节点重要性研究](#)
21. [复杂网络中节点重要度评估研究](#)
22. [基于复杂网络的轨道交通网络节点重要度评价](#)
23. [浅议几种复杂网络节点重要度分析的中心性方法](#)
24. [复杂网络上节点重要度的确定方法](#)
25. [公安复杂网络节点重要度研究](#)

- [26. 基于网络特性的产品制造过程节点重要度评估](#)
- [27. 节点重要度贡献的复杂网络节点重要度评估方法](#)
- [28. 节点重要度在复杂网络鲁棒性中的应用](#)
- [29. 基于复杂网络的装备保障力量体系节点重要度评估](#)
- [30. 基于k-shell分解的复杂网络节点重要度研究](#)
- [31. 基于领域相似度的复杂网络节点重要度评估算法](#)
- [32. 动态融合复杂网络节点重要度评估方法](#)
- [33. 基于节点重要度的公路运输站场建设序列研究](#)
- [34. 基于网络节点相似性的链路预测研究](#)
- [35. 网络节点重要度评价方法研究](#)
- [36. 基于接连失效的复杂网络节点重要性评估](#)
- [37. 基于节点重要度的复杂网络可靠性研究](#)
- [38. 基于节点重要度的公路运输枢纽布局研究](#)
- [39. 铁路沿线联运和换乘节点选择研究——基于节点重要度分析](#)
- [40. 基于复杂网络的城市公交枢纽选址研究](#)
- [41. 基于结构洞指数的网络节点重要度评估](#)
- [42. 基于复杂网络的复杂产品协同制造网络研究](#)
- [43. 基于环介数的复杂作战网络节点重要度评估](#)
- [44. 一种基于K-Shell的复杂网络重要节点发现算法](#)
- [45. 基于概念格的复杂网络重要节点评估](#)
- [46. 基于k-壳分解的集装箱海运网络度分布研究](#)
- [47. 一种定量评估复杂网络节点重要度的算法](#)
- [48. 复杂河流网络节点重要度分析](#)
- [49. 改进的加权复杂网络节点重要度评估方法](#)
- [50. 复杂网络节点重要度评估及网络脆弱性分析](#)