



单位代码 10635
学 号 112012321001600

西南大學

硕士学位论文

复杂网络中节点重要度评估算法的研究

论文作者：杜雨弦

指导教师：邓勇

学科专业：计算机应用技术

研究方向：信息融合与智能计算

论文提交日期：2015 年 4 月 18 日

论文答辩日期：2015 年 5 月 23 日

学位授予单位：西南大学

中国●重庆
2015 年 5 月

独创性申明

学位论文题目：复杂网络中节点重要度评估算法研究

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：

签字日期：

年 月 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：

导师签名：

签字日期：

年 月 日

签字日期：

年 月 日

目 录

摘要	I
ABSTRACT	III
第一章 绪论	1
1.1 研究背景和意义	1
1.2 国内外研究现状	4
1.2.1 复杂网络研究现状	4
1.2.2 失效模式及影响分析技术研究现状	5
1.3 本文的研究方法和内容	7
1.4 论文组织结构	7
第二章 相关理论与技术基础	9
2.1 复杂网络基本静态几何特征	9
2.1.1 网络的图表示	9
2.1.2 网络直径与平均距离	9
2.1.3 集聚系数	10
2.1.4 真实网络统计性质实例	11
2.2 复杂网络节点中心性理论	11
2.2.1 度中心性	12
2.2.2 介数中心性	12
2.2.3 接近中心性	12
2.2.4 特征向量中心性	13
2.2.5 流介数中心性	13
2.3 失效模式及影响分析理论	14
第三章 基于有效距离的接近中心性研究	17
3.1 有效距离理论	17
3.2 基于有效距离的接近中心性	18

3.2.1 算例	19
3.3 数值实验与分析	20
3.3.1 实验数据	20
3.3.2 Susceptible-Infected 疾病传播模型	21
3.3.3 实验与分析	21
3.4 本章小结	25
第四章 基于 TOPSIS 模型的节点重要度识别算法研究	29
4.1 TOPSIS 算法	29
4.2 基于 TOPSIS 模型的节点重要度识别算法	30
4.2.1 传统中心性问题描述	30
4.2.2 TOPSIS 中心性评估算法	31
4.2.3 算例	32
4.3 数值实验与分析	35
4.4 本章小结	38
第五章 基于 FMEA 的节点重要度识别算法研究	41
5.1 网络节点的失效模式及影响分析模型	41
5.2 数值实验与分析	42
5.3 本章小结	44
第六章 总结与展望	51
6.1 本文工作总结	51
6.2 未来工作展望	52
参考文献	53
发表文章目录	61
简历	63
致谢	65

复杂网络中节点重要度评估算法的研究

计算机应用技术 专业 硕士研究生 杜雨弦

指导教师 邓勇 教授

摘要

近年来, 复杂网络系统已经融入到人们生产生活的方方面面。作为一个新兴且活跃的科学研究领域, 复杂网络早已引入到在现实世界网络的实证研究。目前, 在计算机科学、社会科学、生物科学、管理科学等众多领域得到了越来越多的人的重视。一方面, 伴随着复杂网络的不断发展, 人类的生产生活质量有了大幅度的提高和升华, 并且为之带来了极大的便利。但是另一方面, 复杂网络系统的运行也对人类的生产生活带来了一定的负面冲击, 比如疾病的快速传播、大面积的停电事故、以及交通运输的瘫痪等等。因此, 我们需要对各种复杂网络系统有着更为深刻的认识和分析, 以便对可能造成的负面影响进行预测、避免、控制等等。在众多复杂网络研究方向中, 节点重要度评估已经成为其研究发展中一个较为深远的方向。

虽然目前已经许多的中心性方法被提出来度量节点的重要度, 但不同的中心性在各个方面或多或少都存在着一定的不足和局限性。由于不同的中心性的机制不同, 而且有着不同的不足, 因此, 当对同一个网络使用不同的中心性进行节点重要度评估时, 往往会得到不同的结果。为此, 我们有必要对现有的中心性进行改进, 从而能全面有效的对复杂网络节点重要度进行评估。

本文主要提出了三种不同的中心性算法对节点进行重要度评估。首先将有效距离引入节点最短路径的应用中, 用其代替传统的测地线和地理距离来衡量网络节点的距离, 并利用改进后接近中心性对节点进行重要度评估。然后提出了一种基于 TOPSIS 算法的多属性决策模型的中心性算法, 该算法将多个中心性作为多属性进行融合来评估节点重要度。最后, 我们基于失效模式及影响分析模型, 将复杂网络的节点信息进行建模来刻画发生频度、严重程度、检测难易程度, 并通过风险顺序数来对节点进行重要度评估。为了体现出本文提出的算法的有效性和实用性, 我们都将把这些算法应用到真实的网络中进行对比实验。

本文的工作主要包括以下几个方面:

(1) 提出基于有效距离的接近中心性算法

在真实的网络中, 往往会有孤立节点和单向边, 这会导致部分节点对的距离是无穷大, 在此情况下则利用传统接近中心性来评估节点重要度是无效的。针对此问题, 我们引入有效距离, 来代替传统的测地线和地理距离来度量网络节点的距离。该模型不但解决了传统接近中心性失效的问题, 还能广泛应用于加权网络中, 而且更加合理的表示了网络节点信息流的传输过程。

(2) 提出基于 TOPSIS 模型的中心性算法

TOPSIS 算法是一种被广泛应用的多属性决策算法, 它能有效地融合多个有差异的属性, 并得出一组接近理想最优解的排序。由于各种中心性算法都存在着种种缺点, 并且不同中心

性会产生不同的评估结果，因此我们认为有必要提出一种折衷的算法来融合这些差异以及克服单一中心性所存在的缺点。我们将度中心性、接近中心性和介数中心性这三个最为基础的中心性作为多个属性引入到 TOPSIS 多属性决策模型中，将融合后的中心性作为网络节点的重要度评估算法。该算法不仅解决了这三个中心性各自存在的缺陷，还有效地将它们差异进行折衷融合。并且首次将工程评估中的 TOPSIS 算法引入到复杂网络系统中，对跨学科研究领域研究有着积极的影响。

(3) 提出基于失效模式及影响分析模型的中心性算法

失效模式及影响分析是一种可靠性设计的重要方法，它通过由发生频度、严重程度、检测难易程度得出的风险顺序数来对模式进行评估。我们利用网络的结构和节点的信息进行建模来刻画发生频度、严重程度、检测难易程度。我们认为如果一个节点的入度越大，则表明其他节点发生故障时影响到该节点的机会就越大，那么这个节点发生“失效”的概率就越高。同时，倘若一个节点到其他所有节点的有效距离越短，则表明该节点失效后影响的传播就越广，那么这个节点失效的严重程度就越大。在此模型中，我们定义了网络节点的熵的概念。因为信息熵表示的是系统或者个体的不确定性，因此我们认为如果一个节点的熵值越大，则这个节点在网络中所处的结构也就越复杂，那么对这个节点进行失效探测的难易程度也就越难。最后，我们根据新的模型得出的风险顺序数对节点进行重要度评估。风险顺序数值越大，则节点越重要。

关键词： 复杂网络，识别重要节点，失效模式及影响分析，SI 疾病传播模型

The study of identifying influential nodes in complex networks

Yuxian Du

Supervised by Prof. Yong Deng

ABSTRACT

Over the years, complex network systems have been all around in people's production and living. As a young and active field of scientific research inspired largely by empirical studies of real-world networks, it have been paid attention by more and more people in many fields, such as computer sciences, social sciences, biological sciences, management sciences and so on. On the one hand, with the constant development of complex networks, people's production and living have been improved significantly, and brought great convenience. One the other hand, there are some negative effects caused by complex networks. Such as the rapid spread of the disease, widespread blackouts accidents, as well as transportation paralysis. In order to predict, prevent and control the failures caused by complex networks, we need understand and analysis the complex network more profoundly. In the so many complex network researches, identifying influential nodes has been a popular area.

Nowadays, although many centrality measures have been proposed to identify influential nodes, the rankings of identifying influential nodes may be discrepant by using a different centrality measure. Therefore, with different centrality measures, it may obtain different results. In order to address this issue, it is necessary for us to modified the existing centrality measures of identifying influential nodes.

In this paper, we propose three different centrality measures to identify influential nodes. In the first measure, we introduce effective distance into the application of the shortest path problem. Effective distance is applied to replace the conventional geographic distance or binary distance. A new closeness centrality measure based on effective distance is proposed to identify influential nodes. Then, a new centrality measure based on TOPSIS is proposed. Degree centrality, closeness centrality and betweenness centrality are taken into account as the multi-attribute in TOPSIS application. This new centrality measure compromise the different results by Degree centrality, closeness centrality and betweenness centrality. In our last measure, we use node information to model the assessments of occurrence, severity and detection in the application of failure mode and effects analysis (FMEA). According to the value of risk priority number (RPN), the identification of influential nodes can be determined. To show the effectiveness and practicability of the proposed measures, we apply four real networks by Susceptible-Infected (SI) model in each centrality measure.

The main work of this paper is summarized as follows:

(1) Proposing of a modified closeness centrality measure based on effective distance

In many real networks, there may exist some isolated nodes and unidirectional edge. It may lead to unreachable distances. In this case, the traditional closeness centrality is useless. Inspired effective distance, a new closeness centrality based on effective distance is proposed. Effective distance is applied to replace the conventional geographic distance or binary distance.

(2) Proposing of a new centrality measure based on TOPSIS

As a widely used multiple attribute decision making method, TOPSIS is very useful to aggregate the differences between attributes. Due to the fact that there are many shortcomings and deficiencies in these centrality measures, it is necessary for us to propose a new centrality to compromise the results by different centrality measure. In our model, degree centrality, closeness centrality and betweenness centrality are taken into account as the multi-attribute in TOPSIS application to identify influential nodes.

(3) Proposing of a new centrality measure based on FMEA

Failure mode and effects analysis (FMEA) is an engineering and management technique, and it is widely used to define, identify, and eliminate known or potential failures, problems, errors, and risk from the design, process, service. A traditional FMEA is determined by the risk priority number, which is obtained by multiplying the scores of the occurrence, severity and detection. We use node information to model the assessments of occurrence, severity and detection. We think that more in-degrees a node has, higher probability a node suffer effects from other nodes. For this node, the probability of occurrence is high by the way. Meanwhile, if a node has more shorter effective distances between other nodes, we think that the node has a higher spreading ability of failure effects, which means that the node has higher failure severity. In our model, we define the entropy of nodes. Entropy can be described as the nondeterminacy of a system or unit. So we think that if the entropy score of a node is high, it means that this node is in a high complex structure. By this way, it can be deemed that it is difficult to detect the node's failure. According to this proposed model, we can identify influential nodes by the new risk priority number. The higher RPN of a node, the more important degree it should be assigned.

Keywords: Complex networks, Identify influential nodes, Failure mode and effects analysis, Susceptible – Infected model

第一章 绪论

自二十世纪九十年代以来，以因特网为代表的信息技术的迅速发展使人类社会大步迈入了全新的网络时代。从因特网到万维网，从城市电力网络到全球城市交通网络，从群体的交际网络到社会关系网络等等，可以说，人类已经处在一个充满着各种各样的复杂网络的世界中。随着各个领域复杂网络的不断发展，人类的生产生活质量有了大幅度的提高，并且带来了极大的便利。但是在另一方面，复杂网络也对人类的生产生活带来了一定的负面冲击，比如传染病的快速传播、大面积的停电事故、以及交通运输网络的瘫痪等等。因此，我们需要对各种复杂网络系统有着更为深刻的认识和分析，以便对可能造成的负面影响进行预测、避免、控制等等。

由于网络中的节点存在着许多的全局信息和局部信息，因此对有影响力节点的识别显得尤为重要。而在现实领域的各个方面，节点重要度的识别得到了广泛的应用，包括网页排名、用户评价、疾病控制、关键词抽取、标签推荐、攻击防御等方面。而如何将各种前沿、有效的算法思想及模型引入到复杂网络分析中，如何有效融合各中心性算法的差异，对复杂网络的节点重要度分析和评估都是本文解决的主要问题。

1.1 研究背景和意义

随着科学技术的飞速发展，人们在工作生产中对数据的分析、处理、加工技术在不断地提高和创新，这在一定程度上推动了网络分析方面的研究发展。根据研究结果显示，现实生活中所存在的大多数大型网络都不属于传统理论分析中说研究的完全随机网络和规则网络，而是一些呈现高度复杂性的网络。其复杂性主要表现为结构复杂、节点多样性、连接多样性、网络进化、动力学复杂性、多重复杂性融合等方面。这就是复杂网络。

人们自从认识复杂网络研究的重要性和实用性以来，在各领域越来越多的学者更加的关注复杂网络分析。根据 Web of Science 中 Science Citation Index Expanded 和 Social Sciences Citation Index 数据库显示，从 1988 年到 2014 年期间，以“Complex Networks”为检索主题共检索出 8755 篇论文，其中包括来自于世界顶级杂志和期刊，例如《Science》、《Nature》、《PNAS》、《PLoS One》、《Europhysics Letters》等。图 1.1 显示了 1988 年到 2014 年期间以“Complex Networks”为主题的论文每年被 SCI 数据库收录的情况。从中我们可以发现，研究复杂网络的论文基本上是逐年递增，这意味着各领域的研究学者们对复杂网络

的研究热情越来越高。并且从表 1.1 中可知，复杂网络研究涉及到各个领域，其中包括物理、数学、计算机科学等学科领域，此外还包括众多交叉学科。

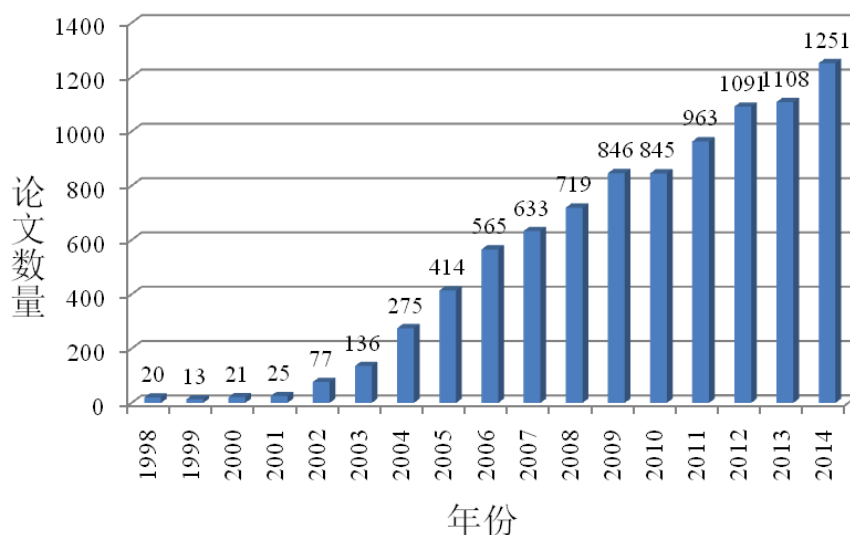


图 1.1: SCI 数据库 1998 - 2014 年收录复杂网络论文情况

表 1.1: SCI 数据库中 1998 - 2014 年收录复杂网络论文的前十个研究方向

研究方向	论文数量	百分比
Physics	4807	53.399%
Mathematics	1090	12.108%
Computer Science	1057	11.742%
Science Technology other Topics	929	10.320%
Engineering	687	7.632%
Mathematical Computational Biology	377	4.188%
Biochemistry Molecular Biology	364	4.044%
Mechanics	304	3.377%
Neurosciences Neurology	188	2.088%
Environmental Sciences Ecology	180	2.000%

在复杂网络飞速发展的同时，国内的许多学者也日益关注着该方向的研究与应用。目前，复杂网络研究已列入国家未来长远重大发展规划战略中，并且在 973、863 等国家重大科研项目上加大了对复杂网络相关研究的资助力度。图 1.2 统计了从 2003 年到 2014 年期间，国家自然科学基金对复杂网络相关研究项目的资助情况^[1]。

通过研究表明，复杂网络中节点往往蕴含着许多的网络全局信息和局部信息。所以为了有效地对网络系统进行分析研究，网络节点重要度的识别也就成为

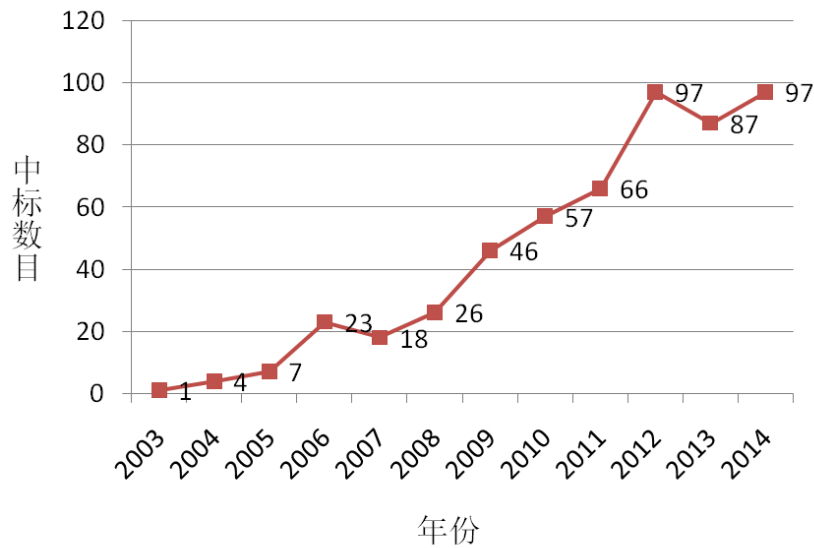


图 1.2: 2003 - 2014 年, 国家自然科学基金对复杂网络研究项目的资助情况

一项十分有意义的工作。虽然目前已有多种中心性算法被提出来研究节点的重要度评估, 但是已有的这些算法通常存在着较大差异, 以导致会产生不同的评估结果, 并且不同的中心性算法或多或少都有一些缺陷和局限性。因此, 我们致力于提出较为有效的融合模型对不同中心性算法进行折衷融合。

最近在《Science》杂志上的一篇文献中, Brockmann 和 Helbing 提出了一种名为有效距离的概念来刻画网络节点中的距离。他们认为在以信息流为交互的网络中, 直观的地理位置距离和测地线边数无法有效的衡量出节点之间的距离。他们通过利用有效距离准确地对 H1N1 病毒和 SARS 病毒爆发根源地进行了回溯和预测, 验证了其在应用上的有效性。由此看来, 有效距离能够在网络中更准确地度量出节点间的距离。利用有效距离来构建一种新的中心性算法模型是一个切实可行的方案。此外, TOPSIS 算法作为一种多属性决策方法, 已经在医学、工程管理、金融评估等领域得到了广泛的应用。它是通过检测评估目标与理想最优解、最劣解的距离来进行排序, 若评估目标最靠近最优解同时又最远离最劣解, 则为最好; 否则不为最优。利用该模型可以对多个中心性算法进行融合, 能最大可能的消除各中心性算法的缺陷, 使得得出的评估达到理想最优。因此建立该模型来对网络节点的重要度评估是一个有效的方案。在不断的研究中我们发现, 充分利用节点在网络中的全局信息和局部信息能够对复杂网络分析产生较为积极的影响。将节点的信息转化为失效模式及影响分析模型中的不同风险因素, 把模式的失效映射到节点重要度的研究中同样具有非常大的可行性。将不同领域的前沿的有效的模型思想引入到复杂网络分析研究中, 对网络节点重要度的研究都具有重要的意义, 而且将具有广阔的应用前景, 并且对跨领域研究有着较为积极的启发和影响。

1.2 国内外研究现状

1.2.1 复杂网络研究现状

随着人类社会生活发展的多元化以及多样化, 复杂网络系统已经时时刻刻融入到我们的生活、生产之中。因此, 作为新兴领域, 复杂网络逐步地得到了广泛的认识和研究。最早的网络是以研究图论为开始的。实际网络的图表示方法可以追溯到 18 世纪伟大的数学家欧拉对注明的“Konigsberg 七桥问题”的研究。欧拉对七桥问题的抽象和论证思想, 开创了数学中图论的研究。20 世纪 60 年代, 由两位匈牙利数学家 Erdős 和 Rényi 建立的随机图论被公认是在数学上开创了复杂网络理论的系统性研究^[2]。在 20 世纪的后 40 年中, 随机图论一直是研究复杂网络的基本理论, 但是绝大多数实际的复杂网络结构并不是完全随机的。在 20 世纪即将结束之际, 对复杂网络的科学探索发生了重要的转变, 复杂网络理论的研究再也不仅限于数学领域。人们开始考虑节点数量众多、连接结构复杂的实际网络整体特性, 在从物理学到生物学、社会学等众多学科中掀起了研究复杂网络的热潮, 甚至于被称为“网络的新科学”。

1998 年 6 月, 美国康奈尔大学理论和应用力学的博士生 Watts 及其导师 Strogatz 教授在《Nature》杂志上发表了一篇名为《“小世界”网络的集体动力学》的文章^[3], 这篇文章揭示了复杂网络的小世界特征, 即大多数的节点拥有少量的边, 少数的节点拥有大量的边。在文献^[3]中, Watts 和 Strogatz 引入了一个小世界网络模型, 称为 WS 小世界模型。由于 WS 小世界模型构造算法中的随机化过程有可能破坏网络的连通性, 因此 Newman 和 Watts 在随后又提出 NW 小世界模型^[4]。该模型是通过用“随机化加边”替代 WS 小世界模型中的“随机化重连”过程而得到的。1999 年 10 月, 时任美国圣母大学物理系教授 Barabási 及其博士生 Albert 在《Science》杂志上发表了题为《随机网络中标度的涌现》的文章^[5], 他们揭示了复杂网络的节点度服从幂律分布, 并且随之提出了一个无标度网络模型, 被称为 BA 模型。通过这一模型的提出, 他们考虑到了以前许多网络模型都忽略的实际网络的增长特性和优先连接特性。这些模型的提出, 为复杂网络的研究提供了较为丰富和扎实的基础理论。在这些模型的建立基础上, 许多的研究学者对复杂网络的其他特性和性质进行了较为深入和广泛地研究, 比如复杂网络的鲁棒性和脆弱性研究^[6-9]、复杂网络的传播动力学研究^[10-13]、复杂网络的博弈演化研究^[14-18]、以及复杂网络节点重要度评估研究^[19-24]等。

在节点重要度的研究中, 最早最为基础的方法就是度中心性 (Degree centrality)^[25]、介数中心性 (Betweenness centrality) 和接近中心性 (Closeness centrality)^[26]。度中心性描述的是一个节点邻居的个数, 它反映的是如果一个节点邻居的数量越多, 那么这个节点影响力也就越大。这一方法最早是在无权网络中进

行研究的,因此在有权网络中,度中心性被刻画为节点的强度,其定义为与节点相连的边的权重值和^[26]。介数中心性表示的是网络中所有节点对的最短路径中,如果经过一个节点的最短路径数越多,那么这个节点就越重要^[27]。介数中心性主要适用于网络流是以最短路径来传输的网络,它能有效刻画节点在网络中传输网络流的控制力。与介数中心性类似,接近中心性也考虑的是网络节点间在最短路径上的距离,它利用计算节点与其他所有节点的最短距离之和的倒数来描述节点的重要度。一个节点与其他节点的平均距离越小,该节点的接近中心性就越大^[26]。

考虑到节点在网络中位置的影响, Kitsak 等人提出了 k -核分解法(k-shell decomposition)用于确定节点在网络中的位置,并将节点在层数上进行区分,越处于内层的节点越重要。该方法认为倘若一个节点处于网络的核心位置,即便它的度比较小,但是它可能也会有较大的影响力。魏代俊等人^[28]将证据理论引入复杂网络系统中,利用节点的度或者权重信息来确定基本概率分配,最终利用证据理论组合规则融合基本概率分配的结果来对节点重要度进行评估。证据理论已被广泛应用于决策和评估模型的各个领域^[29-34],因此该方法在理论上具有很强的应用基础。1998年,Google的创始人Lawrence Page和Sergey Bein^[19]提出了PageRank算法,它是一种基于网页之间相互的超链接计算的技术。PageRank算法是通过网页间的链接关系来确定网页的排序,它的核心思想是网络中一个页面的重要性取决于指向它的其他页面的数量和质量,如果一个页面被很多高质量页面指向,则这个页面的质量也高。页面的数量是指这个页面本身的邻居数量,而质量是指指向它的其他页面的邻居的数量。PageRank算法作为谷歌搜索引擎的核心算法,它已经被广泛研究在各个领域^[35-37]。由于PageRank算法无法对存在悬挂节点的网络进行排序,以及算法中每个节点的随机跳转概率是相同的等这些缺陷,吕琳媛等人^[24]提出LeaderRank算法。在LeaderRank中,添加了一个与所有节点都双向连接的背景节点,这样就保证了初始网络的强连通性,即解决了网络中存在悬挂节点的问题,也使得节点跳转的概率和一个网页上的链接数负相关。因此,LeaderRank较PageRank而言,在许多方面都表现的更好^[38]: (1) LeaderRank比PageRank收敛速度更快; (2) LeaderRank能应用于存在悬挂节点的网络之中; (3) LeaderRank能找出传播网络流更快更广的节点,识别出的节点更为重要; (4) LeaderRank比PageRank具有更强的鲁棒性,更能有效地抵抗垃圾用户攻击和随机干扰。

1.2.2 失效模式及影响分析技术研究现状

失效模式及影响分析(Failure mode and effects analysis, FMEA)是由美国国家航空和宇宙航行局最先提出的一项系统工程分析方法,随后在20世纪50年

代初期，美国古拉曼飞机公司第一次将其应用于战斗机操作系统的设计分析上，取得了较好的效果，并且在 60 年代中期，失效模式及影响分析技术已被正式运用于航天工业。1976 年，美国国防部颁布了失效模式及影响分析的军用标准，这使得该技术开始在汽车工业和医疗设备工业得应用。如今，随着科学技术的不断发展，失效模式及影响分析技术已逐步经在航空航天工业、核工业、电子工业、制造工业等领域得到了广泛的发展与应用，并且已经在工程实践中形成了一套科学而完整的分析方法。

传统的失效模式及影响分析是通过计算风险顺序数（Risk priority number, RPN）来确定系统或者部件的失效评估值。风险顺序数则是简单的利用发生频度（Occurrence）、严重程度（Severity）、检测难易程度（Detection）三者相乘得出的，然而这三个风险因素都是由专家直接进行打分评估给出的。因此，传统的失效模式及影响分析具有许多的局限性和不足：(1) 传统的风险顺序数是由严重程度、发生频度、检测难易程度相乘得出的，而目前没有明确的理由或者原理指出数值上的相乘是合理的；(2) 严重程度、发生频度、检测难易程度都是由专家主观打分得出的，在一定程度的定量上都很难区分不同的打分值，比如“好”、“较好”、“中等”等语言变量；(3) 三个风险因素的不同数值组合可能会得出相等的风险顺序数数值，例如严重程度、发生频度、检测难易程度两组不同的打分值是 4、2、1 和 1、1、8，它们的风险顺序数数值都为 8，但是两组表示的意义却完全不同；(4) 风险因素之间的权重是没有被考虑的，或者说在传统的失效模式及影响分析中它们被认识是相同权重的，这在真实系统中是不合理的；(5) 传统的失效模式及影响分析方法仅仅只考虑了严重程度、发生频度、检测难易程度这三个风险因素，而在实际应用中往往可能会有其他风险因素需要被考虑，比如工作时间、生产成本等等。

针对上述种种的缺陷和不足，在各个领域许多的研究学者对传统失效模式及影响分析进行了修正和改进。1993 年，Gilchrist^[39]提出了一种能应用于期望成本模型的失效模式及影响分析。在这个方法中，故障成本、故障发生的概率、故障被检测的程度以及年生产量都被考虑应用于该期望成本模型。随后，Ben-Daya 和 Raouf^[40]对期望成本模型的失效模式及影响分析进行了改进，他们认为故障发生频度和检测难易程度在该模型中不完全是独立的，且是难以被评估的，而严重程度却是应该被忽略掉的。Bowles 和 Peláez^[41]将模糊理论引入失效模式及影响分析方法中的，利用模糊语言来描述风险因素的评估。这些评估可以是模棱两可的、定性的、不精确的，甚至是定量的也可以。这个模型能够提供一个更加灵活的框架结构来对严重程度、发生频度、检测难易程度进行融合。失效模式及影响分析也被发展到多目标、多属性决策的方法之中。在 Franceschini 和 Galetto^[42]提出的方法中，风险因素和失效模式分别被解读为评估标准和可选方案。整个方法将

每一个决策标准考虑为备选方案的一个模糊子集。Chin 等人^[43]提出了一种基于群体证据推理的失效模式及影响分析方法。该方法能处理不完整的、未知的以及区间值的评估信息，有效地解决了在专家打分系统下产生的不确定信息问题。除此之外，大量的失效模式及影响分析研究还在数学规划、人工智能等方面得到了改进^[44-48]。

1.3 本文的研究方法和内容

本文主要致力于研究复杂网络中节点重要度的评估，通过不同的算法思想，引入不同领域的各种模型，进而构建较为合理、较为实用的节点重要度评估模型。本文的研究内容和方法主要包括以下几个方面：

第一，由于在现实网络中可能会存在孤立节点和一定的单向边，这使得部分节点之间是不可达的，各节点的距离表示为无穷大，此时利用传统的接近中心性来评估节点重要度则是无效的。为了解决这一问题，我们将有效距离这一概率进入到复杂网络系统中，进而构建一种新的更加有效合理的接近中心性来，利用该模型来对节点重要度进行评估。

第二，我们发现，当使用不同的中心性算法对节点重要度进行评估时，往往会得到不同的排序。由于这种情况普遍存在，我们认为有必要提出一种折衷的算法模型来对这些冲突的中心性算法进行融合，以达到最准确的节点重要度评估。因此，我们利用 TOPSIS 多属性决策模型，将度中心性、接近中心性和介数中心性作为多个属性来进行融合，利用最终求得的相似紧密度来对节点进行重要度评估。

第三，据我们所知，网络中的节点通常包含许多的信息，比如网络结构、各节点局部信息等等。因此我们对各节点进行失效模式及影响分析建模，分别利用节点的入度、有效距离以及新定义的节点熵来量化节点发生“失效”的发生频度、严重程度以及检测难易度。最终通过建模得出风险顺序数来对评估节点的重要度。

为了能直观高效地展现出我们模型的有效性和实用性，在本文中，我们都利用真实网络并与其他各种中心性算法进行比较实验。

1.4 论文组织结构

本文组织结构如下：

第一章，绪论。阐述本文的研究背景与意义，介绍复杂网络和失效模式及影响分析技术的国内外研究现状，以及总括本文的主要内容。

第二章，相关理论与技术基础。简单介绍了网络表示和复杂网络的基本静态几何特征，并且概述了几种应用广泛的中心算法及思想，最后对失效模式及影响分析方法进行了介绍。

第三章，基于有效距离的接近中心性研究。引入有效距离概率，对传统接近中心性进行建模，提出一种新的接近中心性来评估节点重要度，并利用在真实网络中的对比实验对模型进行了有效性和实用性验证。

第四章，基于 TOPSIS 模型的节点重要度识别算法研究。考虑到不同中心性算法对节点重要度评估时会产生不同的结果，我们把度中心性、接近中心性和介数中心性作为 TOPSIS 多属性决策模型中的基本多源属性进行折衷融合。利用融合后得出的相似紧密度对节点的重要度进行评估，并且在真实网络中与不同的中心性算法进行对比实验，最终验证该模型的有效性和实用性。

第五章，基于 FMEA 的节点重要度识别算法研究。分别将节点的入度、有效距离和新定义的节点熵引入到失效模式及影响分析模型中，并且刻画出发生频度、严重程度和探测难易度。利用最后得出风险顺序数对节点重要度进行评估。在真实网络的实验中验证了该模型的有效性和实用性。

第六章，总结与展望。分析与总结目前已完成的工作，指出下一步研究的主要方向和思路。

第二章 相关理论与技术基础

在本章，我们首先介绍了复杂网络的基本概念和特征，还介绍了复杂网络节点中心性的相关理论，概述了几种应用广泛的中心性算法思想和计算方法。然后对失效模式及影响分析方法进行了介绍。

2.1 复杂网络基本静态几何特征

2.1.1 网络的图表示

一个具体的网络可抽象为一个由点集 V 边集 E 组成的图 $G = (V, E)$ 。网络的节点总数为 $N = |V|$ ，网络的边总数为 $M = |E|$ 。 E 中的每一条边都有 V 中一对节点与之相对应。如果任意的点对 (i, j) 与 (j, i) 对应同一条边，那么该网络称为无向网络（undirected network），否则称为有向网络（directed network）。如果每条边都有相应的权值，则该网络称为加权网络（weighted network），否则称为无权网络（unweighted network）。无权网络可视为权重都为 1 的特殊的加权网络。

在现实的网络中，节点可以表示具体的人、城市、网页等等，而边则可以表示人与人之间的合作关系、城市之间的交通连接、网页间的链接关系等等。图2.1给出了四种不同类型网络的例子。

2.1.2 网络直径与平均距离

在网络中，任意连接点之间经过边数最少的一条路径称为测地线（geodesic）。测地线的边数（ d_{ij} ）称为节点 i 与节点 j 之间的距离，也就是节点 i 到节点 j 所经过的边的最小数目，即最短路径长度。当节点 i 与节点 j 之间没有路径连通时， $d_{ij} = 0$ 。网络的直径定义为所有距离中的最大值^[49]：

$$D = \max_{1 \leq i, j \leq N} d_{ij} \quad (2.1)$$

其中， N 为网络节点的总数。

平均距离是指所有节点对之间距离的平均值，它描述的是网络中节点之间的平均分离程度，即反映的是网络有多小。其定义为^[50]：

$$L = \frac{1}{N^2} \sum_{j=1}^N \sum_{i=1}^N d_{ij} \quad (2.2)$$

其中， N 为网络节点的总数。网络的平均距离也叫做网络的特征路径长度。

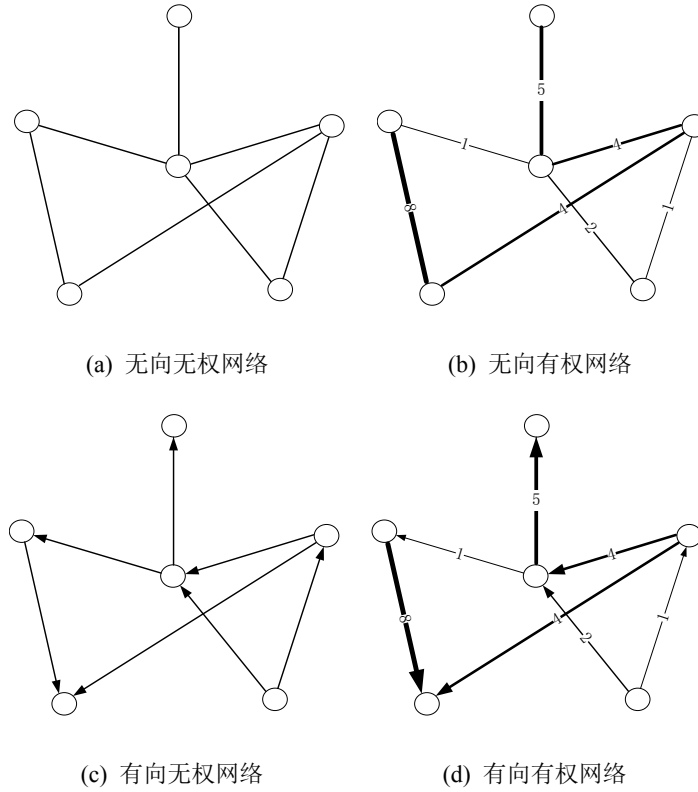


图 2.1: 四种不同类型网络的例子

2.1.3 集聚系数

大多数的真实网络都有着一个共同的性质，集聚性，即一个节点邻居的邻居本身就是该节点的邻居。在真实的社交关系中，它表现为一个人的朋友圈或者熟人圈中的很多人都是相互认识的。假设节点 i 与 k_i 个节点直接相连（即节点 i 有 k_i 个邻居），那么这 k_i 个节点之间可能存在的最大边数为 $k_i(k_i - 1)/2$ 条，若这 k_i 个节点之间实际存在的边数为 M_i 条，则集聚系数可定义为^[3]：

$$C_i = \frac{2M_i}{k_i(k_i - 1)} \quad (2.3)$$

从几何特点看，其等价定义为

$$C_i = \frac{\text{与节点} i \text{ 相连的三角形的数量}}{\text{与节点} i \text{ 相连的三元组的数量}} \quad (2.4)$$

由各个节点的集聚系数，我们可以得出整个网络的平均集聚系数^[3]：

$$C = \frac{1}{N} \sum_{i=1}^N C_i \quad (2.5)$$

显然， $0 \leq C \leq 1$ 。若 $C = 0$ ，则表明该网络中所有节点都是孤立的，没有与其他节点相连；若 $C = 1$ ，则表明网络中所有的节点都是两两相连的，该网络为完全图。

2.1.4 真实网络统计性质实例

随着复杂网络的不断研究，越来越多的真实网络的统计性质被广大学者关注，表 2.1 列出了部分真实网络的部分统计性质^[51]。其中， N 为节点总数， M 为边的总数， $\langle k \rangle$ 为平均度数， L 为平均路径长度， C 为集聚系数。表中空格表示没有可靠的数据。

表 2.1: 部分真实网络的基本统计数据^[51]

网络	类型	N	M	$\langle k \rangle$	L	C
电影演员	无向	449913	25516482	113.43	3.48	0.78
公司董事	无向	7673	55392	14.44	4.60	0.88
数学家合作	无向	253339	496489	3.92	7.57	0.34
物理学家合作	无向	52909	245300	9.27	6.19	0.56
生物学家合作	无向	1520251	11803064	15.53	4.92	0.60
电话呼叫图	无向	47000000	80000000	3.16		
电子邮件	有向	59912	86300	1.44	4.95	0.16
电子邮件地址	有向	16881	57029	3.38	5.22	0.13
学生关系	无向	573	477	1.66	16.01	0.001
WWW(nd.edu)	有向	269504	1497135	5.55	11.27	0.29
WWW(Altavista)	有向	203549046	2.13E+09	10.46	16.18	
引用网络	有向	783339	6716198	8.57		
罗氏词典	有向	1022	5103	4.99	4.87	0.15
单词搭配网络	无向	460902	1.7E+07	70.13		0.44
自治层网络	无向	10697	31992	5.98	3.31	0.39
电力网	无向	4941	6594	2.67	18.99	0.08
铁路网	无向	587	19603	66.79	2.16	0.69
软件包	有向	1439	1723	1.20	2.42	0.082
软件类	有向	1377	2213	1.61	1.51	0.012
电子电路	无向	24097	53248	4.34	11.05	0.03
对等网络	无向	880	1296	1.47	4.28	0.011
代谢网络	无向	765	3686	9.64	2.56	0.67
蛋白质网络	无向	2115	2240	2.12	6.80	0.071
海洋食物网	有向	135	598	4.43	2.05	0.23
淡水食物网	有向	92	997	10.84	1.90	0.087
神经网络	有向	307	2359	7.68	3.97	0.28

2.2 复杂网络节点中心性理论

在众多的复杂网络研究领域，节点重要度评估已成为研究发展较为深远的方向。对于之前所提到的传染病的快速传播、大面积的停电事故、以及交通运输网络的瘫痪等实际复杂网络问题，如果我们能够事先对网络的结构特征有一

定的了解，并且找到关键的节点或节点群以做出必要的预防措施，就可能避免造成巨大的各种损失。而这其中最为核心的问题就是怎样识别或找出这些重要的节点。所谓的重要节点是指较网络其他的节点而言能在更大程度上影响网络的结构与功能的一些比较特殊的节点。在一个网络中，重要节点的数量一般而言都比较少，但是其影响却可以快速地波及到网络中的大部分节点^[52]。识别节点重要度这一研究早已应用在了生物学、管理学、社会学、经济学、医学等各个学科。根据不同学科所研究的具体问题，人们提出了各种不同的网络节点中心性来刻画节点重要度。

2.2.1 度中心性

节点的度中心性是指一个节点邻居的个数之和，它反映的是如果一个节点邻居的数量越多，那么这个节点影响力也就越大。即节点的度越大，其重要性就越大。节点的度中心性是网络中刻画节点重要性最简单、最直观、也是计算复杂度最低的指标。网络中节点 i 的度中心性可定义为^[25]：

$$C_D(i) = \frac{\sum_j x_{ij}}{N-1} \quad (2.6)$$

其中， i 表示所求的节点， j 表示其他所有的节点， N 表示整个网络中节点的总数。 x_{ij} 表示节点 i 与节点 j 之间的连接关系，如果两个节点相连，则为 1；反之，则为 0。

2.2.2 介数中心性

在网络中所有节点对的最短路径中，如果经过一个节点的最短路径数越多，那么这个节点就越重要。根据这个思想，Freeman 提出了网络节点的介数中心性，其定义为^[26]：

$$C_B(i) = \frac{2}{(N-1)(N-2)} \sum_{s \neq i \neq t} \frac{g_{st}(i)}{g_{st}} \quad (2.7)$$

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。 $\frac{2}{(N-1)(N-2)}$ 是用来对介数进行归一化的公式， N 为网络节点的总数。

2.2.3 接近中心性

节点接近中心性表示的是该节点与网络其他所有节点的最短距离之和的倒数，一个节点与其他节点的平均距离越小，则该节点的接近中心性就越大。其定

义为^[26]:

$$C_C(i) = \left[\sum_j^N d_{ij} \right]^{-1} \quad (2.8)$$

其中, d_{ij} 表示节点 i 与节点 j 的最短距离。

2.2.4 特征向量中心性

特征向量中心性不但考虑的是一个节点邻居的数量, 还考虑了该节点每个邻居的重要性。其中心思想为, 指派给网络中每一个节点一个相对得分值, 在每个节点分值的贡献中, 连到高分值节点的连接比连到低分值节点的连接大。其本质为一个节点的分值是与它相连的邻居分值之和, 节点可通过连接很多其他重要的节点来提高自身的重要性, 分值比较高的节点可以和大量一般节点相连, 也可以和少量其他高分值节点相连。令 x_i 为节点 i 的重要性分值, 则特征向量中心性定义为^[53]:

$$C_E(i) = \frac{1}{\lambda} \sum_{j=1}^N a_{ij} x_j \quad (2.9)$$

其中, λ 为常数, N 为网络节点总数。记 $x = [x_1, x_2, x_3, \dots, x_n]^T$, 则公式 (2.9) 可写为特征向量方程:

$$Ax = \lambda x \quad (2.10)$$

该式中各个特征向量解对应不同的特征值 λ 。计算向量 x 的基本方法是给定一个初始值 $x(0)$, 然后采用幂迭代算法。

2.2.5 流介数中心性

传统的介数中心性认为网络流在网络中是以最短路径传输的, 而在现实网络中, 如果利用最短路径来传输网络流, 在某些情况下往往会增加传输时间、降低传输效率。流介数中心性认为网络中所有不重复的路径中, 若经过一个节点的路径的比你越大, 那这个节点就越重要。其定义为^[54]:

$$C_{FB}(i) = \sum_{s < t} \frac{\tilde{g}_{st}(i)}{\tilde{g}_{st}} \quad (2.11)$$

其中, \tilde{g}_{st} 为网络中节点 s 与节点 t 之间的所有路径数, $\tilde{g}_{st}(i)$ 为节点 s 与节点 t 之间经过节点 i 的路径数。传统的介数中心性和流介数中心性考虑的是两个不同的极端, 前者只考虑了网络流传输的最短路径, 而后者考虑的是传输的所有路径并且认为每条路径的作用相同。

2.3 失效模式及影响分析理论

失效模式及影响分析是一种可靠性设计的重要方法，它实际上就是失效模式分析和失效影响分析的组合。在对系统进行失效模式及影响分析过程中，通过对系统各组成部分潜在的各种故障模式及其对系统功能的影响进行分析，并且把每一个潜在故障模式按它的严重程度进行分类，提出可以采取的预防措施，来提高系统的可靠性。它是一个“事前行为”，而不是“事后行为”。失效模式及影响分析实际是一组系列化的过程，主要包括三大步骤：

1. 找出系统或产品中潜在或存在的失效模式。
2. 根据相应的评估体系，对找出的失效模式进行风险量化评估
3. 列出故障原因和机理，寻找出预防和改进的措施。

失效模式及影响分析是通过求得风险顺序数来确定模式的失效评估，风险顺序数值越大，模式失效的优先级就应该越高，从而反映出该模式更为重要。风险顺序数是由发生频度、严重程度、检测难易程度的评估数值相乘得出的，其定义为^[55]：

$$RPN = O \times S \times D \quad (2.12)$$

其中， O 表示模式发生失效的频度， S 表示模式失效后产生后果的严重程度， D 表示检测出模式失效的难易程度。三个不同的风险因素都是通过 1-10 的数值量化来表示的，如表 2.2、表 2.3 和表 2.4 所示。

表 2.2: 模式发生失效的频度标准。

等级	发生失效的频度	标准
10	非常高	$\geq 1/2$
9	很高	$1/3$
8	高	$1/8$
7	较高	$1/20$
6	中等偏高	$1/80$
5	中等	$1/400$
4	中等偏低	$1/2000$
3	低	$1/15,000$
2	很低	$1/150,000$
1	几乎不	$\leq 1/150,000$

失效模式及影响分析已在美国三大汽车制造公司（戴姆勒—克莱斯勒、福特、通用）得到广泛应用，并且已经作为汽车零部件生产行业的可靠性设计分析方法。其中作原理为^[56, 57]：(1) 确定潜在的失效模式，对失效会产生的后果进行

表 2.3: 模式失效后产生后果的严重程度标准。

等级	失效后产生的后果	标准
10	非常危险	对系统造成安全性危害
9	严重	系统无法运行
8	有很大影响	系统需要修复才能运行
7	有较大影响	对系统正常运行有影响
6	有一定影响	对系统产生影响
5	中等	对系统产生轻微影响，需要修复
4	影响比较小	带来较小影响，建议修复
3	影响很低	带来微小不便，可进行修复
2	影响非常小	带来微小不便，不需要修复
1	无影响	毫无影响

表 2.4: 检测出模式失效的难易程度标准。

等级	检测出失效的程度	标准
10	无法检测出	无法发现潜在原因和随后带来的失效
9	非常低	有非常低的可能发现潜在原因和随后带来的失效
8	很低	有很低的可能发现潜在原因和随后带来的失效
7	低	有轻微的可能发现潜在原因和随后带来的失效
6	较低	有较小的可能发现潜在原因和随后带来的失效
5	中等	可能会发现潜在原因和随后带来的失效
4	较高	有较大大的机会能发现潜在原因和随后带来的失效
3	高	有很大机会能发现潜在原因和随后带来的失效
2	很高	有非常大的机会能发现潜在原因和随后带来的失效
1	几乎肯定	几乎肯定会发现潜在原因和随后带来的失效

评分；(2) 客观评估各种原因出现的可能性，以及当某种原因出现时企业能检测出该原因发生的可能性；(3) 对各种潜在的产品和流程失效进行排序；(4) 以消除产品和流程存在的问题为重点，并帮助预防问题的再次发生。

失效模式及影响分析原理的应用主要体现在美国三大汽车制造公司制定的《潜在失效模式和后果分析》表格中，该表的内容包括^[58]：

- (1) 功能要求：填写被分析过程（或工序）的简要说明和工艺描述；
- (2) 潜在失效模式：记录可能会出现的问题点；
- (3) 潜在失效后果：推测问题点可能会引发的不良影响；
- (4) 严重度（S）：评价上述失效后果并赋予分值（1-10分，不良影响愈严重分值愈高）；

(5) 潜在失效起因或机理：潜在问题点可能出现的原因或产生机理；

(6) 频度（O）：上述潜在失效起因或机理出现的几率（1-10分，出现的几率愈大分值愈高）；

(7) 现行控制：列出目前本企业对于潜在问题点的控制方法；

(8) 探测度（D）：在采用现行的控制方法实施控制时，潜在问题可被查出的难易程度（1-10，查出难度愈大分值愈高）；

(9) 风险顺序数（RPN）：严重度、频度、探测度三者得分之积，其数值愈大潜在问题愈严重，愈应及时采取预防措施；

(10) 建议措施：列出“风险顺序数”较高的潜在问题点，并制定相应预防措施，以防止潜在问题的发生；

(11) 责任及目标完成日期：制定实施预防措施的计划案；

(12) 措施结果：对预防措施计划案实施状况的确认。

从此表中可以看出，失效模式及影响分析原理的核心是对失效模式的严重度、频度和探测度进行风险评估，通过量化后的指标来确定高风险的失效模式，并且随后制定预防措施加以控制，从而努力将风险减到可接受的水平甚至完全消除。因此，失效模式及影响分析不仅适用于汽车零部件生产系统的分析体系，也可以应用于其他类似的生产和管理体系。

第三章 基于有效距离的接近中心性研究

本章主要基于有效距离对复杂网络节点的接近中性进行研究。首先，我们介绍了有效距离理论，及其在已有的研究中取得的成果。其次，阐述了传统的节点接近中心性在节点重要度评估应用中的劣势和不足，以及有待解决的问题。然后，我们基于有效距离提出一种新的接近中心性模型来对节点进行重要度评估，并且解决了传统节点接近中性的缺陷和不足。最后，我们利用四个真实网络来验证基于有效距离的接近中心性的有效性和实用性。

3.1 有效距离理论

在传统的网络研究中，节点之间的距离往往是计算连接节点之间的测地线的边数，或者以节点之间的实际地理距离来衡量的。在现实网络中，虽然以经过的边数或以实际地理位置来计算距离，在逻辑结构和物理结构上是有道理的，但是在以信息流为交互的真实网络中却有悖原理。例如，在以城市为节点的航空网络中，北京飞往呼和浩特的国内直飞航线有 444 公里，而北京到广州的国内直飞航线是 1967 公里，从直观上我们肯定认为北京到呼和浩特的距离远小于北京到广州的距离。但是基于有效距离的思想，会得出北京到广州的距离小于北京到呼和浩特的距离，其原因是北京到广州的航空客运量远高于北京到呼和浩特的航空客运量。因为假如在北京爆发某一疾病，如果疾病是由其他城市传染过来的，我们肯定认为是由广州传播过来的可能性要高于由呼和浩特传播过来的可能性。虽然广州较呼和浩特而言离北京的地理距离要远一些，但是其人流量远大于呼和浩特到北京的人流量，所以疾病是由广州传播过来的可能性更大一点，从而可认为北京到广州的“有效距离”比北京到呼和浩特的“有效距离”要小一些。

有效距离是由 Brockmann 和 Helbing^[59]提出的，他们于 2013 年在《Science》杂志上发表了一篇名为《The hidden geometry of complex, network-driven contagion phenomena》的文章。他们认为在以信息流为交互的网络中，直观的地理位置距离和测地线边数无法有效的衡量出节点之间的距离。于是他们提出了一种新的距离，名为有效距离（Effective distance）^[59]。令 F_{ji} 为从节点 i 传向节点 j 的信息流， $P_{ji} \in [0, 1]$ 为从节点 i 传向节点 j 的信息流比例：

$$P_{ji} = \frac{F_{ji}}{F_i} \quad (3.1)$$

$$F_i = \sum_j F_{ji} \quad (3.2)$$

F_{ji} 被用来描述全局流动性网络的传播流量。在不同的网络中, F_{ji} 可用来表示不同类型的信息流。例如, 在电力网中, F_{ji} 表示传输的电量; 在航空网中, F_{ji} 表示客运流量; 在微博网络中, F_{ji} 表示微博的转发量; 在单纯的加权逻辑网络中, F_{ji} 表示边的权重。

令 d_{ji} 为从节点 i 指向节点 j 的有效长度:

$$d_{ji} = (1 - \log P_{ji}) \geq 1 \quad (3.3)$$

值得注意的是, 在真实网络中从节点 i 指向节点 j 的有效长度往往不等于从节点 j 指向节点 i 的有效长度, 即 $d_{ji} \neq d_{ij}$ 。

基于上述有效长度的定义, 可以得出从节点 i 到节点 j 的有效距离:

$$D_{ji} = \min_{\Gamma} \lambda(\Gamma) \quad (3.4)$$

与有效长度一样, 在真实网络中从节点 i 指向节点 j 的有效距离往往不等于从节点 j 指向节点 i 的有效距离, 即 $D_{ji} \neq D_{ij}$ 。

在文献[59]中, Brockmann 和 Helbing 利用有效距离来预测和回溯 H1N1 病毒和 SARS 病毒爆发的根源地, 他们通过该模型模拟实验准确的找出了病毒爆发的根源地, 并表明了有效距离在其应用上的有效性和实用性。

3.2 基于有效距离的接近中心性

首先, 传统以经过的边数或以实际地理位置来计算的距离已经被认为是无法准确衡量网络节点的距离。其次, 在存在有单向边的有向网络中, 接近中心性是无法度量网络节点重要性的。因为, 接近中心性是根据节点之间的距离求出的, 而在存在有单向边的有向网络中, 其他节点到具有单向边的节点的距离是无穷大, 这就会导致所有节点各自的距离之和都为无穷大, 从而使得所有节点的接近中心性都为 0。因此, 针对上述传统接近中心性的种种弊端和不足, 我们基于有效距离提出了一种新的接近中心性模型。定义 $C_{eff}(i)$ 为节点 i 的基于有效距离的接近中心性:

$$C_{eff}(i) = \left[\sum_j^N D_{ji} \right]^{-1} \quad (3.5)$$

其中, D_{ji} 表示从节点 i 到节点 j 的有效距离, N 表示网络节点总数。

与传统的接近中心性相比, 其最大的不同之处就是用有效距离替换测地线和地理距离来度量节点之间的距离。由于基于有效距离的接近中心性是利用自身节点的流入流量和邻居节点的总流出流量来获得的, 因此它比其他一般中心性更好的承载了节点邻居的信息。它不但考虑了网络全局的结构, 还考虑了周围邻居的局部信息。

3.2.1 算例

为了更好地理解我们提出的基于有效距离的接近中心性，本小节给出一个例子来体现有效接近中心性是如何计算求得的。假设有一个有向有权网络，如图 3.1 所示。

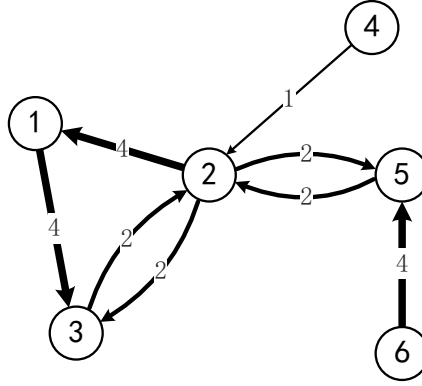


图 3.1: 具有 6 个节点的有向有权网

在这个例子中，我们仅对 2 号节点进行数值计算。根据公式 3.5，我们可以得出 2 号节点的有效接近中心性：

$$C_{eff}(2) = \frac{1}{D_{12} + D_{32} + D_{42} + D_{52} + D_{62}}$$

根据公式 3.1-3.4，可计算出各个有效距离：

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

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

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

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

$$D_{62} = 0$$

由于 5 号节点与 6 号节点是单向链接的，且路径是由 6 号节点指向 5 号节点。因此 6 号节点无法到达 2 号节点，所以 $D_{62} = 0$ 。

通过以上方法，可以求得每两个节点对之间的有效距离，如表 3.1 所示。

综上，2 号节点的有效接近中心性为：

$$\begin{aligned} C_{eff}(2) &= \frac{1}{D_{12} + D_{32} + D_{42} + D_{52} + D_{62}} \\ &= \frac{1}{1.6931 + 0 + 2.3863 + 0 + 2.3863} = 0.1547 \end{aligned}$$

表 3.1: 图 3.1 中每个节点对的有效距离

D_{ji}	1 号节点	2 号节点	3 号节点	4 号节点	5 号节点	6 号节点
1 号节点	0	1.6931	2.6931	2.6931	2.6931	3.6931
2 号节点	2	0	1	1	1	2
3 号节点	1	2.3863	0	3.3863	3.3863	4.3863
4 号节点	0	0	0	0	0	0
5 号节点	4.3863	2.3863	3.3863	3.3863	0	1
6 号节点	0	0	0	0	0	0

表 3.2 列出了图 3.1 中所有节点的有效接近中心性，并且列出了各节点其他不同的中心性值。可以清楚的看出，在该有向图中，传统接近中心性是无法识别出节点的重要度，而我们提出方法却能有效识别出节点的重要度。

表 3.2: 图 3.1 中每个节点不同的中心性值

节点	C_D	C_C	C_B	C_E	C_{FB}	$EDCC$
1	8	0	0	0.1425	0.2000	0.0759
2	13	0	0.6500	0.3047	0.8250	0.0975
3	8	0	0.1500	0.1599	0.1000	0.0842
4	1	0	0	0.0905	0	0.0655
5	8	0	0.3500	0.1599	0.4000	0.0849
6	4	0	0	0.1425	0	0.0696

3.3 数值实验与分析

3.3.1 实验数据

本小节设计了五组对比试验，将我们提出的接近中心性与不同的节点中心性进行对比分析，验证了接近中心性在识别节点重要度方面的有效性和实用性。本小节我们利用四个不同类型的真实的网络来验证接近中心性的性能：(1) Airlines 网络：这是一个航空有向无权网络，其节点表示机场，有向边表示机场之间的航线。该网络数据能在“<http://vlado.fmf.uni-lj.si/pub/networks/data/>”上获得；(2) USAir97 网络：这是 1997 年的美国航班网络，且是无向有权的。网络中的节点表示美国的机场，边表示机场之间的航线，权重表示两个机场在坐标轴上的距离。该网络数据也可在“<http://vlado.fmf.uni-lj.si/pub/networks/data/>”上获得；(3) Netscience 网络：这是一个无向无权网络，他表示的是同一领域

的科学家的合作关系^[60]。每一个节点表示一个科学家，如果科学家之间有合作关系，则其对应的节点以边连接起来。该网络数据可在“<http://www-personal.umich.edu/~mejn/netdata/>”上获得；(4) USairport 网络：这是一个有向有权网络，它表示的是美国机场之间的航空网。其权重表示的是已经预定的航班座位数。该网络数据可在“<http://toreopsahl.com/datasets/#usairports>”上获得。表 3.3 列出了这个四个真实网络的部分统计特征。

表 3.3: 四个真实网络的部分统计特征。 n 表示节点总数， m 表示边的总数， $\langle k \rangle$ 表示节点的平均度数， k_{max} 表示节点的最大度数， C 表示网络的平均集聚系数。

网络	n	m	$\langle k \rangle$	k_{max}	C
Airlines	235	1297	5.52	130	0.2789
USAir97	332	2126	12.81	139	0.3126
Netscience	379	914	4.82	34	0.3706
USairport	1572	28235	17.96	596	0.4892

3.3.2 Susceptible-Infected 疾病传播模型

为了更好的展现出有效接近中心性的与其他不同中心性的性能，我们利用 Susceptible-Infected 疾病传播模型（SI 疾病传播模型）^[13]来评估结点在网络中的传播能力，从而体现节点的重要度。SI 疾病传播模型已在模拟疾病动态传播领域得到了广泛的研究^[61, 62, 28]。在 SI 模型中，每个节点都有两个离散的状态：Susceptible 状态和 Infected 状态。已感染的节点以 β 的概率将疾病传播给该节点易感染的邻居。在疾病传播网络中， β 描绘的是一个节点在网络中可以发挥影响的范围或规模。在加权网络中节点 i 将疾病传播给节点 j 的概率为 $\beta = (\frac{w_{ij}}{w_M})^\alpha$ ，其中 w_{ij} 表示两节点相连的权重， w_M 表示所有权重中的最大值， α 为正常数^[63]。而在无权网络中，其概率表示为 $\beta = (\frac{1}{2})^\alpha$ 。在随后的实验中，我们用 $F(t)$ 表示在 t 时刻网络中被感染的节点总数。

3.3.3 实验与分析

在本小节中，我们将提出的有效距离接近中心性（EDCC）与度中心性（DC）、接近中心性（CC）、介数中心性（BC）、特征向量中心性（EC）、流介数中心性（FBC）进行比较分析。

1) 实验一：前十个最重要节点的比较（见表 3.4）

表 3.4 显示了利用不同的中心性算法在四个真实网络中排在前十最总要的节点。此表中 SI 疾病传播模型感染的节点总数 $F(t)$ 只计算到 $t = 10$ 时的个数。在

Airlines 网络中, 137 号节点拥有最强的传播能力。此外, 在其他所有的中心性算法中, 137 号节点都具有最强的传播能力, 即最重要。在排在前十的节点中, $F(t)$ 与 $EDCC$ 有 9 个完全相同的节点, 这表明 $F(t)$ 与 $EDCC$ 同时认为这 9 个节点都是网络中最重要。 $F(t)$ 与 DC 、 BC 、 EC 、 FBC 在前十个节点中相同节点的个数都是 5, 与 EC 相同的个数是 6。在 Netscience 网络中, $F(t)$ 与 CC 具有最多的相同节点, 紧随其后的是 BC 和 $EDCC$ 。在 USairport 网络中, EC 和 $EDCC$ 都与 $F(t)$ 有 5 个相同的节点, 它们的传播能力是最强的。在 USAir97 网络中, DC 、 EC 、 $EDCC$ 在传播能力上表现得最好, 排在前十的节点中, 它们与 $F(t)$ 都有 8 个相同的节点。而 CC 却是表现得最差的。通过前十个最重要节点的比较, 我们可以看出, 有效距离接近中心性在 Airlines 网络, USairport 网络和 USAir97 网络中都表现得最好, 并且在 Netscience 中有效距离接近中心性几乎和接近中心性表现得一样好。

表 3.4: 四个真实网络用不同中心性排列出的前十个最重要节点

Airlines						Netscience						
DC	BC	EC	FBC	$EDCC$	$F(t)$	DC	CC	BC	EC	FBC	$EDCC$	$F(t)$
137	137	137	137	137	137	4	26	26	26	26	51	26
51	51	201	51	190	201	5	95	51	4	231	44	231
81	81	131	81	155	193	26	51	169	5	51	26	51
131	131	155	201	204	204	16	231	95	95	5	297	44
71	201	81	131	177	131	67	100	67	67	169	100	95
42	71	193	71	201	155	70	52	5	16	52	231	100
155	85	85	42	193	190	95	5	231	51	67	95	233
85	42	173	174	131	164	15	44	100	32	95	308	232
201	193	71	193	81	196	32	234	44	70	100	337	5
193	155	204	155	196	177	51	297	66	8	4	52	297
USairport						USAir97						
DC	BC	EC	FBC	$EDCC$	$F(t)$	DC	CC	BC	EC	FBC	$EDCC$	$F(t)$
97	76	97	76	1219	76	118	172	8	118	118	118	118
1016	584	1016	1219	76	752	261	182	261	261	8	261	261
754	1331	754	97	907	1016	248	198	166	201	261	201	313
335	1433	333	754	1016	333	67	216	182	47	201	47	201
333	207	335	333	1050	1190	255	166	47	67	47	67	67
654	591	654	654	1250	1219	47	206	118	255	182	313	248
606	604	1219	1016	681	606	166	225	152	248	67	255	47
1222	428	606	907	981	184	201	133	67	166	152	144	182
752	892	1222	335	427	1050	182	140	201	182	313	112	255
1064	1520	76	866	606	1192	147	181	13	8	255	182	162

2) 实验二：肯德尔系数差异比较（见图 3.2）

为了较为清晰地衡量出不同中心性与 SI 传播模型的相关性，我们引入肯德尔系数^[64, 65]来作为评估标准。肯德尔系数的思想是，假设两个随机变量分别为 X 和 Y ，它们的元素个数均为 N ，两个随机变量取的第 i ($1 \leq i \leq N$) 个值分别用 X_i 、 Y_i 表示。 X 与 Y 中的对应元素组成一个元素对集合 XY ，其包含的元素为 (X_i, Y_i) ($1 \leq i \leq N$)。当集合 XY 中任意两个元素 (X_i, Y_i) 与 (X_j, Y_j) 的排行相同时，即 $X_i > X_j$ 且 $Y_i > Y_j$ 或者 $X_i < X_j$ 且 $Y_i < Y_j$ 时，这两个元素就被认为是一致的。当 $X_i > X_j$ 且 $Y_i < Y_j$ 或者 $X_i < X_j$ 且 $Y_i > Y_j$ 时，这两个元素被认为是不一致的。当 $X_i = X_j$ 或者 $Y_i = Y_j$ 时，这两个元素既不是一致的也不是不一致的。肯德尔系数 τ 的定义如下^[64, 65]：

$$\tau = \frac{n_c - n_d}{0.5n(n-1)} \quad (3.6)$$

其中 n_c 和 n_d 分别表示一致和不一致集合对的数量。

肯德尔系数能明确的表示出一种排序与另一种排序的相同程度或匹配程度，它能够比较直观的体现两种方法在数值结果中的相关程度。因此我们依次设定 SI 模型的感染概率从 0.01 逐渐递增至 0.1，然后将不同中心性得到的排序与 SI 模型的排序依照感染概率的变化进行肯德尔相关性计算，得到图 3.2 的结果。需要明确的是肯德尔系数 τ 的值越高，表明该中心性与 SI 模型的相关性越高。从图 3.2 中我们能够发现在 Airlines 网络和 USAir97 网络中，与 SI 模型的相关性最高的算法是有效距离接近中心性。而在 USairport 网络中，有效距离接近中心性都只是稍逊于度中心性而位居次席。在 Netscience 网络中，当感染概率低于 0.075 时，有效距离接近中心性与 SI 模型的相关性稍微低于流介数中心性。但在感染概率达到 0.075 之后，有效距离接近中心性就表现得最好。

3) 实验三：随着节点数增加传播能力变化（见图 3.3）

我们将不同节点中心性算法的前 N 个节点的平均感染能力进行比较。随着节点数的增加，前 N 节点的感染能力会逐渐下降。如图 3.3 所示，从 Airlines 网络和 Netscience 网络中，我们可以发现在前 5 个节点之后，随着节点数的增加，有效距离接近中心性一直具有最高的平均感染能力。在 USairport 网络中，有效距离接近中心性明显一直优于介数中心性，并且在前 200 个节点之后，所有的中心性都具有相同的平均感染能力。在 USAir97 网络中，在前 20 个节点之前，有效距离接近中心性具有最强的平均感染能力，并且一直都优于传统的接近中心性。

4) 实验四：前十节点的平均累积感染能力比较（见图 3.4 — 图3.7）

我们计算了不同中心性的排序在前十节点的累积平均感染节点数。我们去除掉排序前十节点中相同的节点，以此更明确的分辨出不同中心性排序前十节点的感染能力。从图 3.4 中可明确地看出有效距离接近中心性的感染变化图一直位于

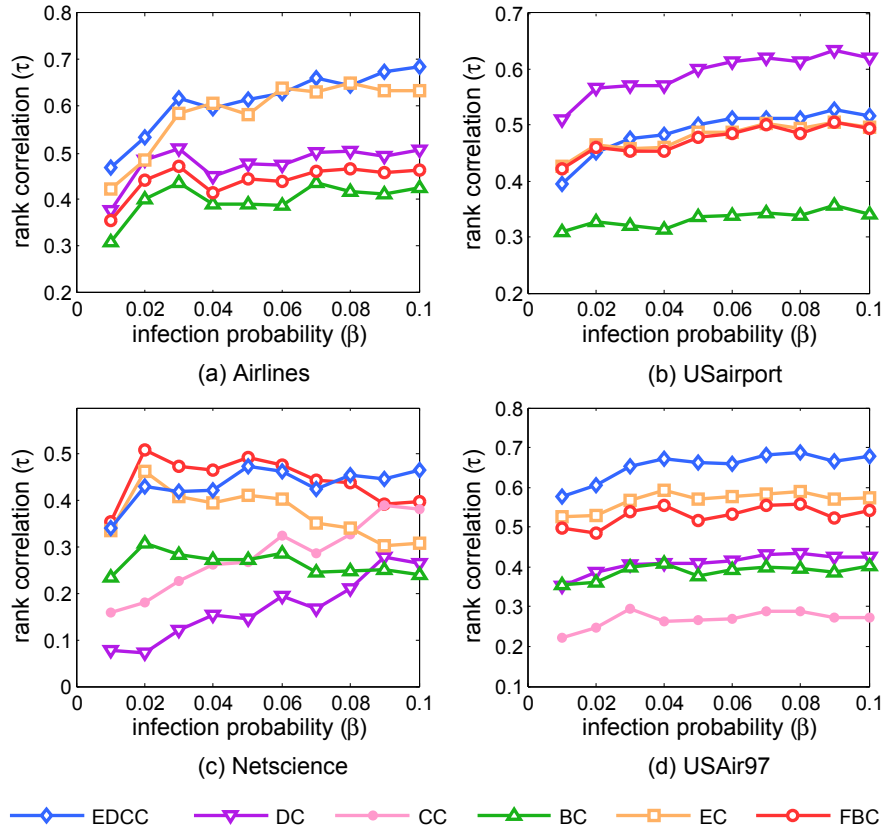
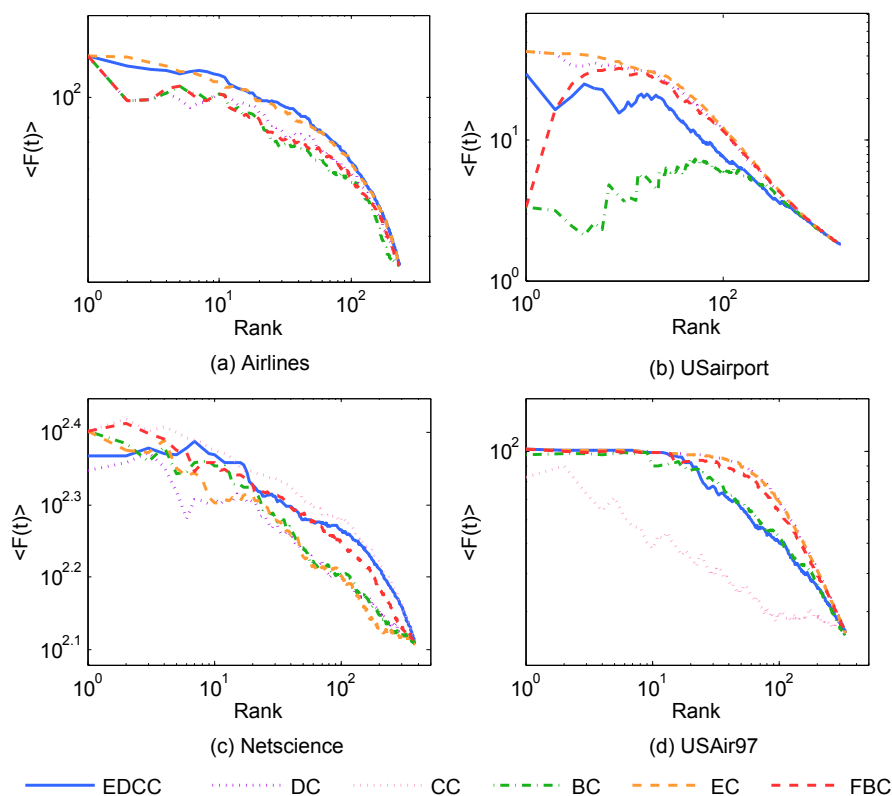


图 3.2: 不同节点中心性算法与 SI 模型的肯德尔系数相关性比较

其他所有中心性感染变化图之上，这表明了，在 t 变化的每一步的节点感染中，有效距离接近中心性所选出的节点的感染数目都要高于其他所有中心性选出的节点感染数目。从侧面则反映了有效距离接近中心性所选出的节点的传播能力更强更广，有效距离接近中心性所选出的节点在网络中更加重要。在 USAirport 网络中，有效距离接近中心性虽然略微比特征向量中心性和流介数中心性差，但是它在每一步的传播中都优于介数中心性，且与度中心性的感染能力几乎一致。在 Netscience 网络中，有效距离接近中心性的传播能力略强于度中心性、介数中心性和特征向量中心性，并且与接近中心性和流介数中心性有着几乎一样的感染能力。从图 3.4 — 图3.7 中可以得出，有效距离接近中心性的传播能力明显的高于接近中心性和介数中心性，并且也略微强于度中心性、特征向量中心性和流介数中心性。

5) 实验五：不同中心性值出现频率比较（见图 3.8）

在一个网络中，具有相同中心性值的节点应该具有相同的排名。在一个特定的中心性算法中，网络节点的排序是按照该中心性值的降序排列的。如果有多个节点具有相同的排名，那么就无法对这些排名相同的节点进行区分。所以如果一个中心性算法给出的排序中具有相同排名的节点有很多个，那么说明这个中心性


 图 3.3: 不同节点中心性算法的前 N 个节点的平均感染能力

算法在节点重要度排序上就不是很有效。图 3.8 则给出了不同中心性算法给出的排序值。从四个网络中我们都能明显的看出，有效距离接近中心性给出的排序中具有相同排名的节点都非常少，并且优于其他所有的中心性算法，尤其是度中心度。这表明了有效距离接近中心性能有效地区分节点之间的重要度。

3.4 本章小结

本章利用有效距离理论提出了一种新的接近中心性，通过真实网络的实验与分析表明我们提出的有效距离接近中心性在节点重要度的评估上明显的优于现有的度中心性、接近中心性、介数中心性、特征向量中心性以及流介数中心性。并且与传统的介数中心性相比，有效距离接近中心性能广泛应用于加权网络等各种类型的网络。

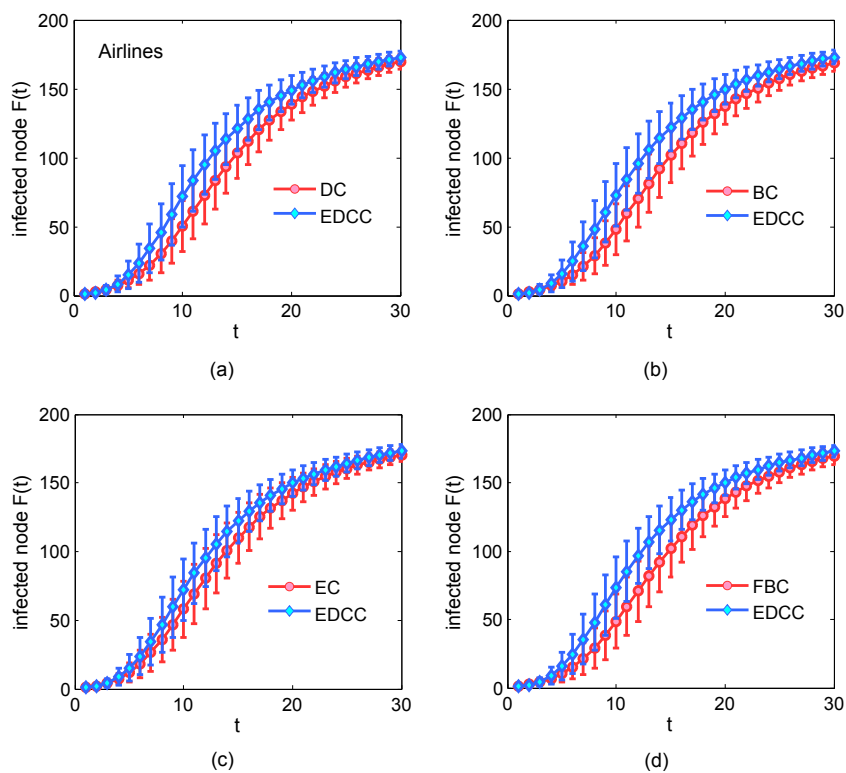


图 3.4: Airlines 网络中不同中心性排序前十节点的累积平均感染节点数

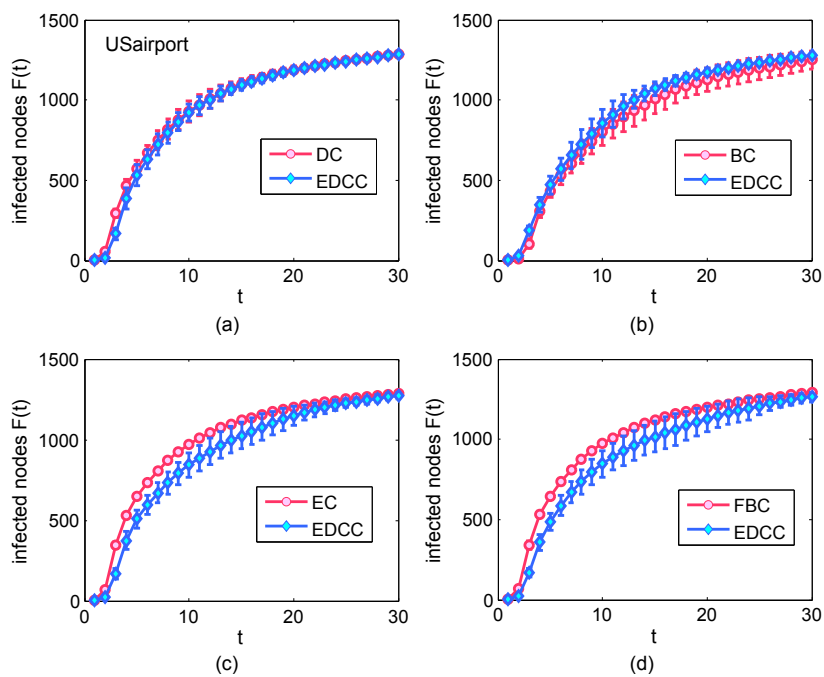


图 3.5: USairport 网络中不同中心性排序前十节点的累积平均感染节点数

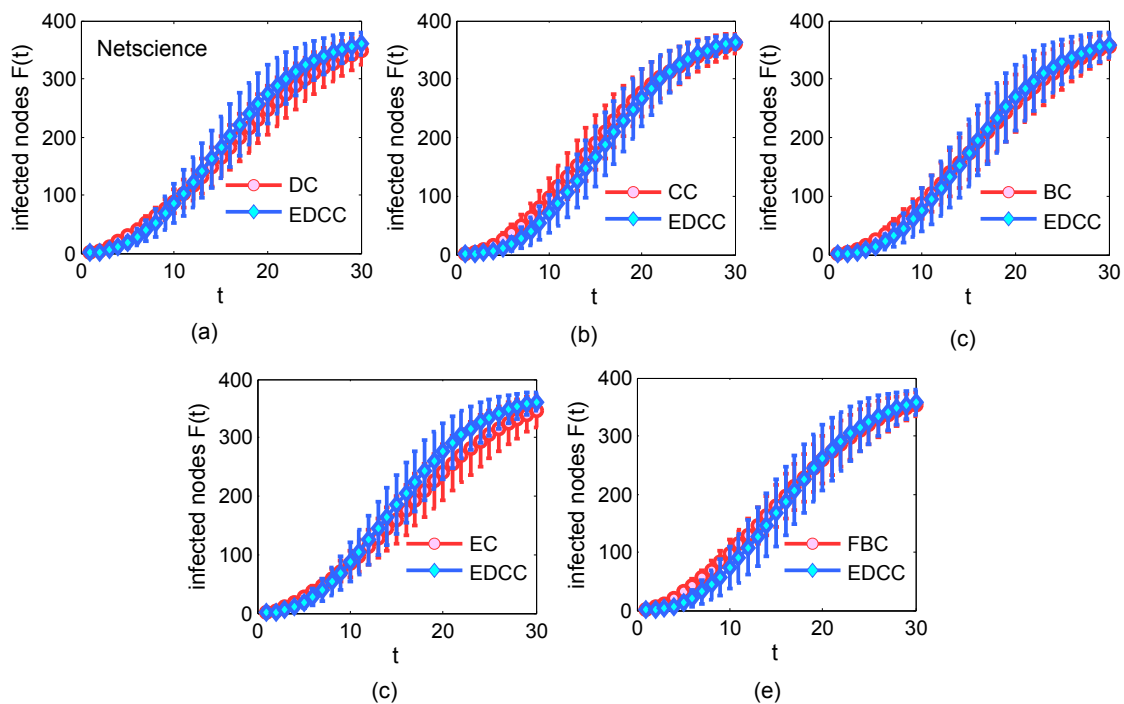


图 3.6: Netscience 网络中不同中心性排序前十节点的累积平均感染节点数

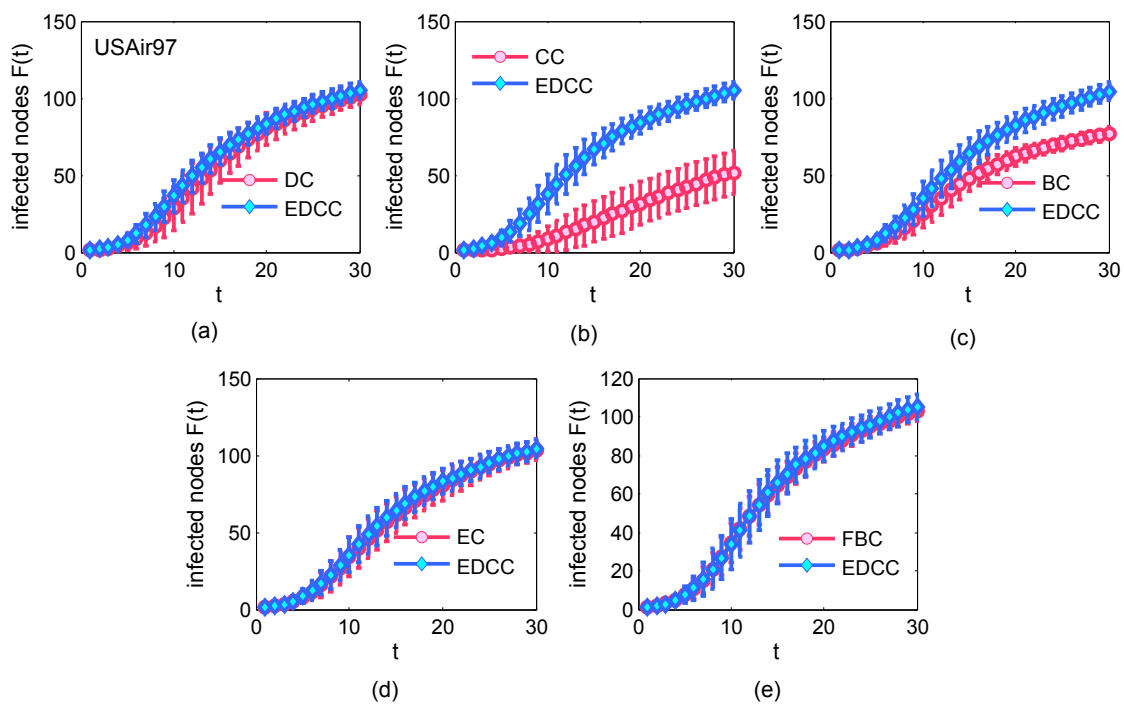


图 3.7: USAir97 网络中不同中心性排序前十节点的累积平均感染节点数

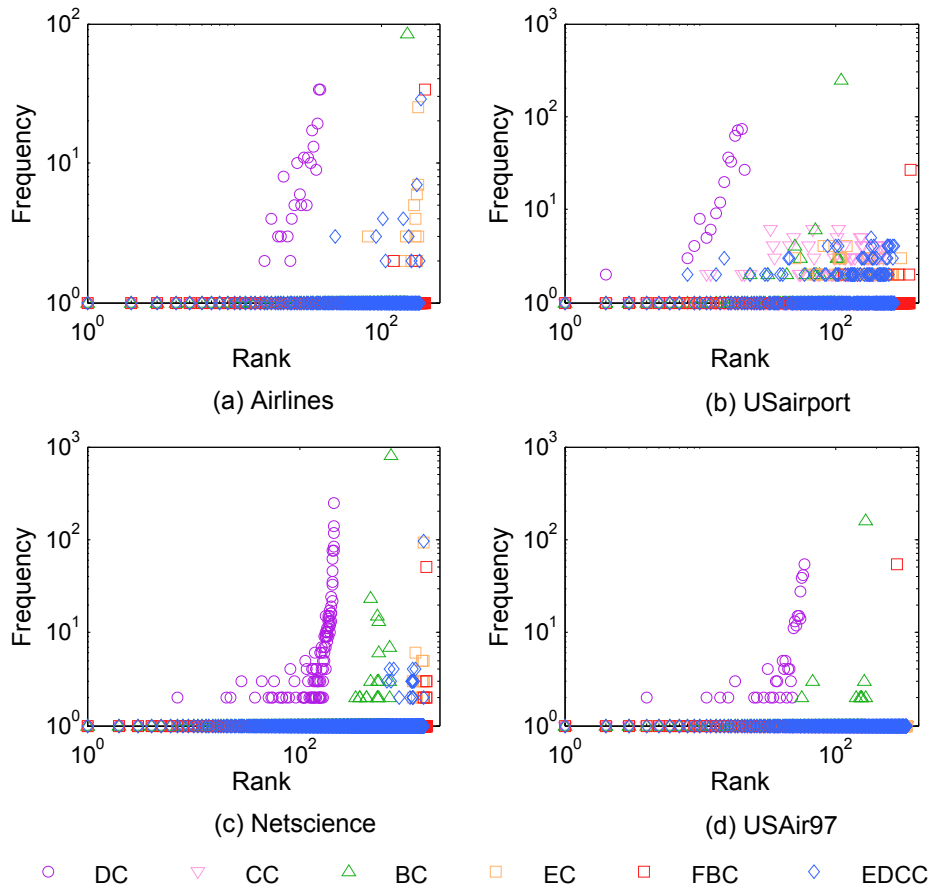


图 3.8: 不同中心性值在网络节点排序中出现的频率

第四章 基于 TOPSIS 模型的节点重要度识别算法研究

本章利用 TOPSIS 法（又称为优劣解距离法）将多个中心性进行有效的融合，来识别网络节点重要度。首先，我们介绍了 TOPSIS 算法的模型和计算思想。然后，我们将 TOPSIS 模型扩展到中心性的节点重要度排序，利用该模型有效的融合了不同中心性表示，甚至是有冲突的中心性表示。最后，我们将建立的此模型运用到真实网络中进行数值试验模拟，研究了该模型的有效性和实用性。

4.1 TOPSIS 算法

TOPSIS 算法是 Technique for order preference by similarity to ideal solution 算法的简称，即逼近最优解的偏好顺序法，它是由 Hwang 和 Yoon^[66] 在 1981 年提出的一种多属性决策方法。其基本原理，是通过检测评价对象与最优解、最劣解的距离来进行排序，若评价对象最靠近最优解同时又最远离最劣解，则为最好；否则不为最优。其中最优解的各指标值都达到各评价指标的最优值。最劣解的各指标值都达到各评价指标的最差值。TOPSIS 算法作为一种有效的多属性决策方法已经在各研究领域得到了广泛的应用^[67-71]。

TOPSIS 充分利用了属性之间的信息，在不需要完全保证属性的独立性的情况下得出一组备选方案的排序。首先，将多个属性信息定义为一个 $m \times n$ 大小的矩阵 $D = (x_{mn})$ ，其列表示的是不同的属性种类，各行表示的是各备选方案。然后，将属性矩阵进行归一化处理^[66]：

$$r_{ij} = \frac{x_{ij}}{\sqrt{\sum_{j=1}^m x_{ij}^2}}, \quad i = 1, \dots, m; j = 1, \dots, n. \quad (4.1)$$

如果各项属性有其各自的权重，则可得到加权的属性矩阵：

$$v_{ij} = w_j \times r_{ij}, \quad i = 1, \dots, m; j = 1, \dots, n. \quad (4.2)$$

其中， w_j 表示的是第 j 列属性的权重值。

依据属性矩阵得出该系统的理想最优解和最劣解。理想最优解和最劣解分别表示各属性值达到各备选方案中的最好的值的组合和最差的值的组合。令最优解为 A^+ ，最劣解为 A^- ，其定义为^[66]：

$$A^+ = \{v_1^+, v_2^+, \dots, v_n^+\} = \{(\max_i v_{ij} | j \in K_b)(\min_i v_{ij} | j \in K_c)\} \quad (4.3)$$

$$A^- = \{v_1^-, v_2^-, \dots, v_n^-\} = \{(\min_i v_{ij} | j \in K_b)(\max_i v_{ij} | j \in K_c)\} \quad (4.4)$$

其中, K_b 表示的是收益准则, K_c 表示的是花费准则。

基于欧几里得距离, 求出各备选方案与理想最优解和最劣解的距离, 分别表示为 S_i^+ 和 S_i^- :

$$S_i^+ = \sqrt{\sum_{j=1}^n (v_j^+ - v_{ij})^2}, \quad i = 1, \dots, m; j = 1, \dots, n. \quad (4.5)$$

$$S_i^- = \sqrt{\sum_{j=1}^n (v_j^- - v_{ij})^2}, \quad i = 1, \dots, m; j = 1, \dots, n. \quad (4.6)$$

最后, 求出各个备选方法与理想最优解和最劣解的相似紧密度^[66]:

$$C_i = \frac{S_i^-}{S_i^- + S_i^+}, \quad i = 1, \dots, m. \quad (4.7)$$

根据最终求得的相似紧密度对各备选方案进行优劣排序。相似紧密度的值越大, 则备选方案越优, 排名就应越靠前。

4.2 基于 TOPSIS 模型的节点重要度识别算法

4.2.1 传统中心性问题描述

尽管目前已经提出了许多的中心性方法来度量节点的重要度, 但是不同的中心性在各个方面或多或少都存在着一定的不足和局限性。虽然度中心性的计算复杂度最低、最易理解, 但是它仅仅只考虑了节点的局部信息, 没有顾全到网络的全局结构。当有两个节点分别属于不同的连通片, 那么用接近中心性度量节点重要度就无法得到合理的结果。这使得接近中心性在有与网络断开连接的连通片中缺乏适用性。虽然介数中心性考虑了网络的全局结构, 并且能用于断连的连通片网络, 但是其仍然有一定的不足。比如, 在真实网络中有很大部分节点都不出于任何两个节点之间的最短路径上, 从而使得这些大部分节点的介数中心性值都为 0。由于不同的中心性的机制不同, 而且有着不同的不足, 因此, 当对同一个网络使用不同的中心性进行节点重要度评估时, 往往会得到不同的结果。图 4.1 描绘了一个简单的人际关系网络^[72]。

表 4.1 中罗列了图 4.1 网络中各个节点的度中心性、接近中心性和介数中心性。从其中我们可以发现, 拥有度中心性最大的节点是 *Diane*。虽然 *Heather* 的度中心性值只有 0.333, 但是他却拥有最大的介数中心性。同时, *Fernando* 和 *Garth* 的接近中心性是最大的。因此, 该网络验证了利用不同的中心性进行节点重要度评估时, 往往会得到较大的差异。

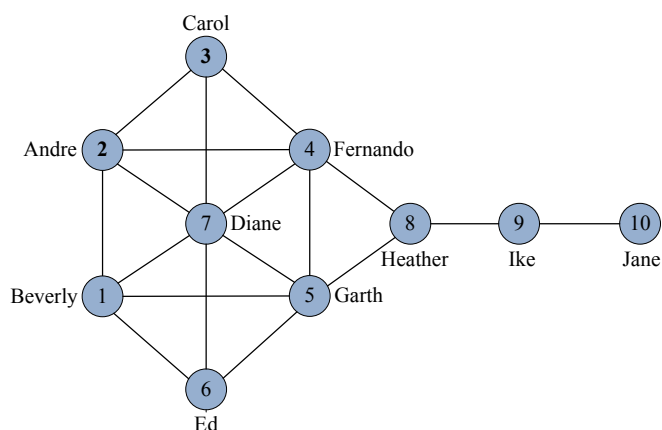
图 4.1: 经典的人际关系网络^[72]

表 4.1: 图 4.1 中各节点的度中心性、接近中心性和介数中心性。

<i>Degree</i>		<i>Closeness</i>		<i>Betweenness</i>	
Nodes	Value	Nodes	Value	Nodes	Value
Diane	0.667	Fernando	0.071	Heather	0.622
Fernando	0.556	Garth	0.071	Fernando	0.370
Garth	0.556	Diane	0.067	Garth	0.370
Andre	0.444	Heather	0.067	Ike	0.356
Beverly	0.444	Andre	0.059	Diane	0.163
Carol	0.333	Beverly	0.059	Andre	0.037
Ed	0.333	Carol	0.056	Beverly	0.037
Heather	0.333	Ed	0.056	Carol	0.000
Ike	0.222	Ike	0.048	Ed	0.000
Jane	0.111	Jane	0.034	Jane	0.000

4.2.2 TOPSIS 中心性评估算法

基于上述中心性的种种缺点，以及不同中心性导致的差异结果等这些问题，我们认为有必要提出一种折衷的算法来融合这些差异以及克服单一中心性所存在的缺点。因此，我们将度中心性、接近中心性和介数中心性这三个最为基础的中心性作为多个属性引入到 TOPSIS 多属性决策模型中，将融合后的中心性作为网络节点的重要度评估算法。该算法不仅解决了这三个中心性各自存在的缺陷，还有效地将它们差异进行折衷融合。并且首次将工程评估中的 TOPSIS 算法引入到复杂网络系统中，对跨学科领域研究有着积极的影响。其模型流程图如图 4.2 所示。

该方法的具体步骤如下：

1. 计算网络节点的度中心性、接近中心性和介数中心性；

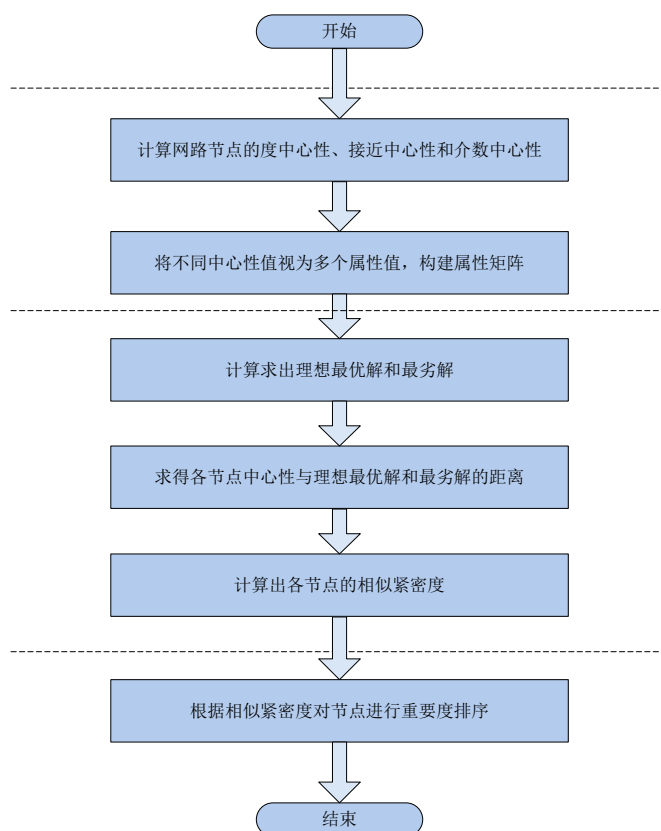


图 4.2: 基于 TOPSIS 中心性算法模型流程图

2. 将不同中心性值视为多属性值，求得其理想最优解和最劣解；
3. 分别计算出每个节点的中心性值与理想最优解和最劣解的距离；
4. 计算每个节点的相似紧密度；
5. 根据节点的相似紧密度对其重要度进行排序。相似紧密度的值越大，则节点越重要。

4.2.3 算例

为了更好地理解 TOPSIS 中心性算法的中心思想，本小节给出一个小网络来体现 TOPSIS 中心性算法是如何计算求得的。假如有一个具有 6 个节点的简单网络，如图 4.3 所示。

首先，计算出各节点的度中心性、接近中心性和介数中心性，如表 4.2 所示：

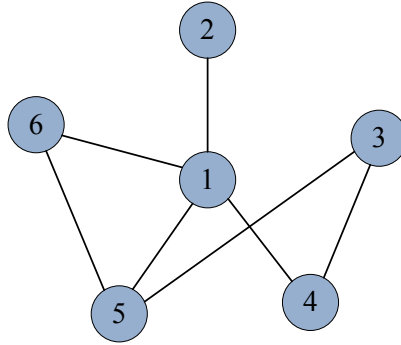


图 4.3: 具有 6 个节点的简单网络

表 4.2: 图 4.3 各节点的度中心性、接近中心性和介数中心性

节点	度中心性	接近中心性	介数中心性
1 号节点	0.8	0.1667	1.1
2 号节点	0.2	0.1000	0
3 号节点	0.4	0.1111	0.1
4 号节点	0.4	0.1250	0.2
5 号节点	0.6	0.1429	0.4
6 号节点	0.4	0.1250	0

首先，得出属性矩阵：

$$D = \begin{bmatrix} 0.8 & 0.1667 & 0.7333 \\ 0.2 & 0.1000 & 0 \\ 0.4 & 0.1111 & 0.0667 \\ 0.4 & 0.1250 & 0.1333 \\ 0.6 & 0.1426 & 0.2667 \\ 0.4 & 0.1250 & 0 \end{bmatrix}$$

随后，根据公式 4.1 将属性矩阵归一化，以第一列的度中心性为例：

$$r_{11} = \frac{x_{11}}{\sqrt{\sum_{i=1}^5 x_{i1}^2}} = \frac{0.8}{\sqrt{0.8^2 + 0.2^2 + 0.4^2 + 0.4^2 + 0.6^2 + 0.4^2}} = 0.2163$$

以此类推，可以得出归一的化矩阵：

$$R = \begin{bmatrix} 0.6489 & 0.5224 & 0.9231 \\ 0.1622 & 0.3134 & 0 \\ 0.3244 & 0.3483 & 0.0839 \\ 0.3244 & 0.3918 & 0.1678 \\ 0.4867 & 0.4478 & 0.3357 \\ 0.3244 & 0.3918 & 0 \end{bmatrix}$$

在该例中我们假定各个指标的权重都相等，即都为 $\frac{1}{3}$ 。那么最终的加权归一化数据如表 4.3 所示。

表 4.3: 最终的加权归一化数据

节点	度中心性	接近中心性	介数中心性
1 号节点	0.2163	0.1741	0.3077
2 号节点	0.0541	0.1045	0
3 号节点	0.1081	0.1161	0.0280
4 号节点	0.1081	0.1306	0.0559
5 号节点	0.1622	0.1493	0.1119
6 号节点	0.1081	0.1306	0

根据公式 (4.3) 和公式 (4.4)，分别得出理想最优解和最劣解：

$$A^+ = \{0.2163, 0.1741, 0.3077\}$$

$$A^- = \{0.0541, 0.1045, 0\}$$

然后，可分别求得 1 号节点与理想最优解和最劣解的距离：

$$S_1^+ = \sqrt{(0.2163 - 0.2163)^2 + (0.1741 - 0.1741)^2 + (0.3077 - 0.3077)^2} = 0$$

$$S_1^- = \sqrt{(0.0541 - 0.2163)^2 + (0.1045 - 0.1741)^2 + (0 - 0.3077)^2} = 0.3547$$

最后，1 号节点与理想最优解和最劣解的相似紧密度为：

$$C_1 = \frac{S_1^-}{S_1^- + S_1^+} = \frac{0.3547}{0.3547 + 0} = 1$$

综上所述，可分别求出所有节点与理想最优解和最劣解的相似紧密度，如表 4.4 所示。具有较大相似紧密度的节点排名应越靠前。

表 4.4: 各节点的相似紧密度和排序

节点	S_i^+	S_i^-	C_i	排名
1 号节点	0	0.3547	1	1
2 号节点	0.3547	0	0	6
3 号节点	0.3055	0.0620	0.1687	4
4 号节点	0.2774	0.0821	0.2283	3
5 号节点	0.2047	0.1619	0.4417	2
6 号节点	0.2047	0.0601	0.1543	5

4.3 数值实验与分析

在本章的实验中，我们仍然利用四个真实的网络来验证基于 TOPSIS 的中心性算法的性能，并且将其与不同的中心性算法进行对比分析，验证了接近中心性在识别节点重要度方面的有效性和实用性。四个真实的网络为：(1) USAirport 网络：它是 2002 年美国客运量排在前 500 位的机场之间的航空网络^[73]。其权重表示的是已经预定的航班座位数；(2) Airlines 网络：这是一个航空网络，其节点表示机场，有向边表示机场之间的航线。该网络数据可以在“<http://vlado.fmf.uni-lj.si/pub/networks/data/>”上获得；(3) Email 网络：这是一个邮件关系网络^[74]，每个节点表示一个上网用户，如果两个用户之间有邮件联系，则用边将两节点连接起来；(4) Football 网络：这是一个美国大学足球俱乐部的网络^[75]。表 4.5 列出了这个四个真实网络的部分统计特征。

表 4.5: 四个真实网络的部分统计特征。其中， n 表示节点总数， m 表示边的总数， $\langle k \rangle$ 表示节点的平均度数， k_{max} 表示节点的最大度数， C 表示网络的平均集聚系数。

网络	n	m	$\langle k \rangle$	k_{max}	C
USAirport	500	2846	11.38	146	0.3097
Airlines	235	1297	11.04	130	0.2789
Email	1133	5451	9.62	71	0.1101
Football	115	613	10.66	12	0.2016

我们将提出的基于 TOPSIS 中心性与度中心性 (DC)、接近中心性 (CC)、介数中心性 (BC)、特征向量中心性 (EC) 进行比较分析。

1) 实验一：前十个最重要节点的比较（见表 4.6）

表 4.6 显示了利用不同的中心性算法在四个真实网络中排在前十最总要的节点。在 US500 网络中，在排序前十的节点中，基于 TOPSIS 的中心性与度中心性

和接近中心性有 8 个相同的节点，并且与介数中心性具有全文相同的 10 个节点。在 Airlines 网络中，四种中心性算法的前三个节点都一样，并且有着相同的排序：137 号节点 \succ 51 号节点 \succ 81 号节点，其中 \succ 表示优于或者更加重要。也就是说，在这四种中心性算法排序中，前三个最重要的节点依次为 137 号节点，51 号节点和 81 号节点。在 Email 网络中，基于 TOPSIS 的中心性与度中心性在前十的节点中全都一样，与特征向量中心性在前九的节点中全都一样，并且其与接近中心性有 7 个相同的节点。在 Football 网络中，度中心性、接近中心性、特征向量中心性与基于 TOPSIS 的中心性排序前十的节点相同的个数分别为 8 个、4 个、9 个。通过该比较试验反映了我们提出的基于 TOPSIS 的中心性算法与其他中心性算法在结果上没有较大差异。

表 4.6: 四个真实网络用不同中心性排列出的前十个最重要节点

USAirport				Airlines			
DC	CC	BC	TOPSIS	DC	CC	BC	TOPSIS
1	7	56	56	137	137	137	137
2	1	7	7	51	51	51	51
3	2	6	6	81	81	81	81
7	3	14	14	131	131	201	131
6	6	3	1	71	71	131	201
8	8	1	3	42	42	71	71
18	11	21	21	155	155	19	42
21	14	2	2	85	201	174	155
11	10	10	8	201	85	42	85
10	17	8	10	193	193	119	193
Email				Football			
DC	CC	BC	TOPSIS	DC	CC	BC	TOPSIS
105	233	105	105	1	59	6	89
333	23	23	333	2	81	2	7
16	105	333	23	3	89	4	1
23	42	41	42	4	107	1	16
42	41	42	16	6	7	7	105
41	76	16	41	7	1	105	4
196	233	233	196	8	16	16	2
233	52	196	233	16	17	3	54
21	135	21	21	54	25	89	3
76	378	355	76	68	93	8	8

2) 实验二：前十节点的累积平均感染能力比较（见图 4.4 — 图 4.7）

在下列实验中，我们利用 SI 疾病传播模型计算了不同中心性的排序在前十

节点的累积平均感染节点数。为了更加明确的分辨出不同中心性排序前十节点的感染能力，我们去除掉排序前十节点中相同的节点。USAirport 中，TOPSIS 中心性算法所选出的前十节点的平均感染能力要略微弱于度中心性算法，但与介数中心性算法评估相同，并且也略优于接近中心性。在 Airlines 网络中，提出的该方法评估出节点的传播能力与度中心性和接近中心性完全相同，并且要强于介数中心性。通过在 Email 网络中的实验发现，其与度中心性表现出相同的前十节点感染能力。虽然比接近中心性差一点，但是明显强于特征向量中心性。并且在 Football 网络中展现出的感染能力也是一样的。

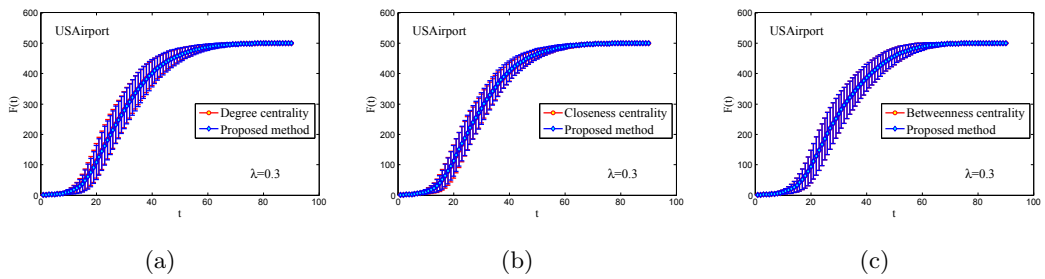


图 4.4: USAirport 网络中不同中心性排序前十节点的累积平均感染节点数

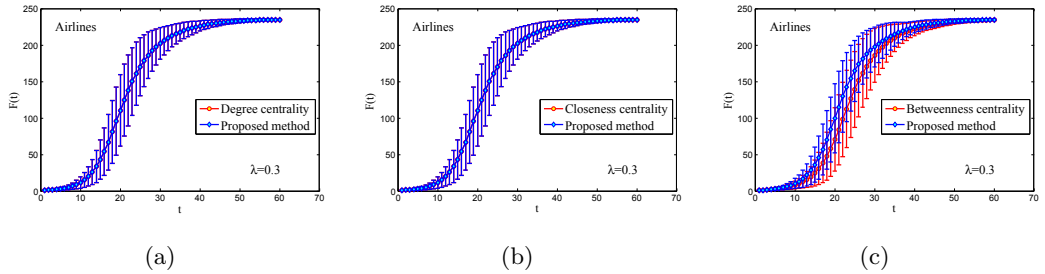


图 4.5: Airlines 网络中不同中心性排序前十节点的累积平均感染节点数

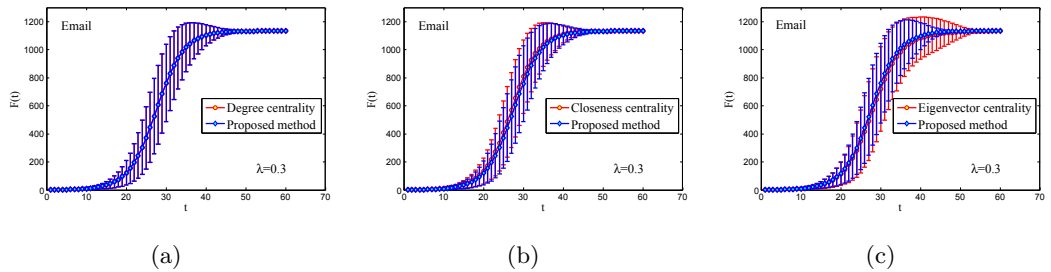


图 4.6: Email 网络中不同中心性排序前十节点的累积平均感染节点数

由于在不同网络中，TOPSIS 中心性与有的中心性算法在前十节点的平均感染能力上有这相同的表现，因此我们单独分开比较前十中排序不同的节点，如图

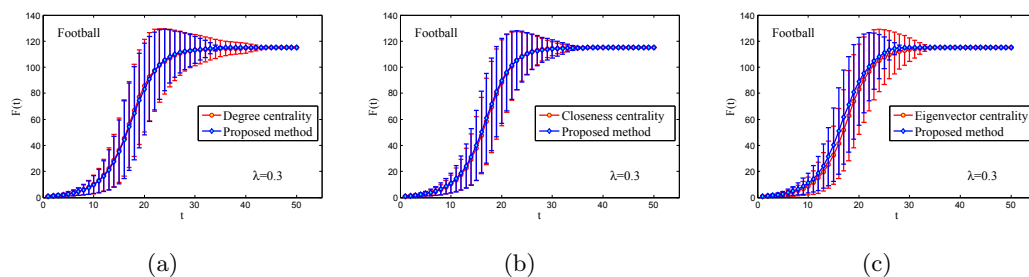


图 4.7: Football 网络中不同中心性排序前十节点的累积平均感染节点数

4.8 — 图 4.10所示。在此组实验中较强感染能力的节点是 1 号节点、8 号节点、201 号节点和 23 号节点，这些节点都是 TOPSIS 中心性算法所评估出来的较为重要的节点。

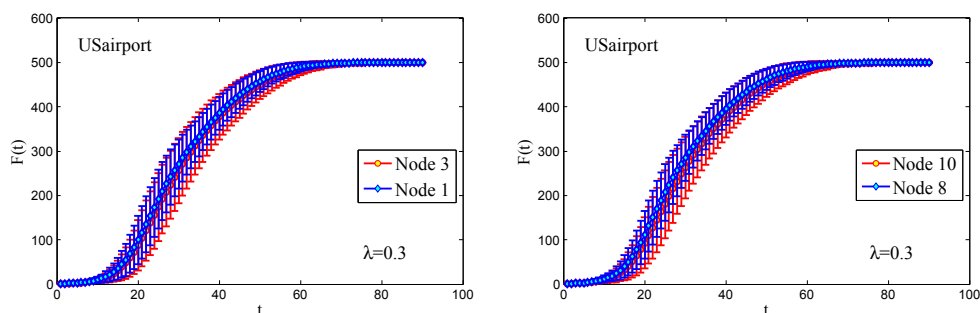


图 4.8: USAirport 网络中 3 号节点与 1 号节点以及 10 号节点与 8 号节点的传播感染能力比较

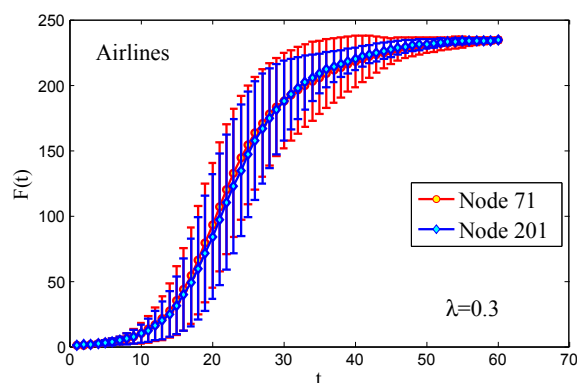


图 4.9: Airlines 网络中 71 号节点与 201 号节点的传播感染能力比较

4.4 本章小结

本章将 TOPSIS 算法引入到复杂网络系统的节点重要度评估中，把度中心性、接近中心性和介数中心性作为 TOPSIS 算法中的多属性进行融合。最终该算

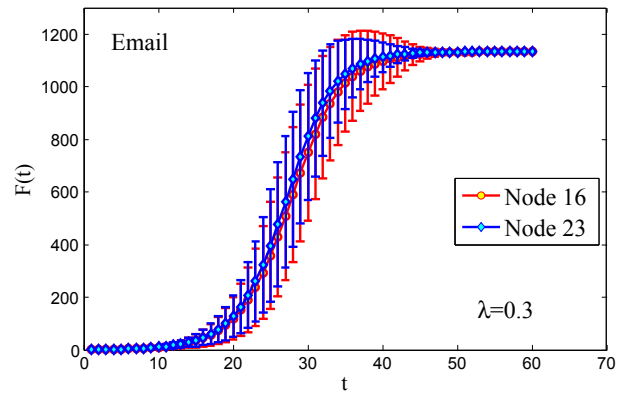


图 4.10: Email 网络中 16 号节点与 23 号节点的传播感染能力比较

法解决了这三个中心性各自存在的缺陷，有效地将它们的差异进行折衷融合，并且对跨学科领域研究有着积极的影响。

第五章 基于 FMEA 的节点重要度识别算法研究

本章我们基于失效模式及影响分析模型，将复杂网络的节点信息进行建模来刻画发生频度、严重程度、检测难易程度。我们分别将节点的入度、有效距离以及新定义的节点熵与发生频度、严重程度以及检测难易程度相关，并通过风险顺序数来对节点进行重要度评估。

5.1 网络节点的失效模式及影响分析模型

我们认为，在一个网络中，如果一个节点的入度越大，即其他节点指向该节点的边越多，那么可以表明其他节点发生故障时传播给该节点的机会就越大，从而反映的是这个节点发生“失效”的概率就越高。这里的“失效”在不同类型的网络中，可以表现为遭受攻击、被感染等行为。因此，我们定义一个节点“失效”的发生频度为：

$$O(i) = \left(\frac{k_i^{in}}{N-1} \right)^\alpha \quad (5.1)$$

其中， k_i^{in} 为节点 i 的入度之和， N 为网络节点总数， α 为正参数， $O(i)$ 表示节点 i “失效”发生频度。

根据第三章的工作可知，如果一个节点到其他所有节点的有效距离越短，那么这个节点就越处于网络结构的中心位置。当该节点发生“失效”后，它就能快速的波及到网络中的其他节点，那么我们就认为这个节点失效的严重程度就越大。因此，我们利用第三章提出的基于有效距离的接近中心性来度量节点“失效”的严重程度，如公式 5.2 所示。

$$S(i) = C_{eff}(i) = \left[\sum_j^N D_{ji} \right]^{-1} \quad (5.2)$$

其中， D_{ji} 表示从节点 i 到节点 j 的有效距离， N 表示网络节点总数， $S(i)$ 表示节点 i “失效”的严重程度。

在信息论中，熵表示的是平均不确定性的量度^[76]，熵的理解可以为不确定性的量度而不是确定性的量度，因为越随机的信源的熵越大。基于信息熵理论，我们对网络提出了一种节点熵，利用节点的熵来表示节点在网络中的不确定性。如果我们越不了解这个节点的信息，那么我们对这个节点进行探测的难度也就越大。因此我们认为节点的熵值越大，则表明对这个节点的“失效”探测也就越难。我们定义节点 j 为节点 i 的邻居，且节点 i 的邻居集合为 M_i ，则节点 i 所有邻居

的度之和为：

$$T = \sum_{j \in M_i} k_j \quad (5.3)$$

我们定义节点的熵为：

$$N(i) = - \sum_{j \in M_i} \frac{k_j}{T} \log \frac{k_j}{T} = D(i) \quad (5.4)$$

其中， k_j 表示节点 j 的度， $D(i)$ 表示节点 i “失效”的探测难易度。

我们给出一个小例子来展示现节点熵是如何计算求得的。假设有一个具有 8 个节点的简单网络，如图 5.1 所示。我们以节点 1 为例，来实现计算过程。

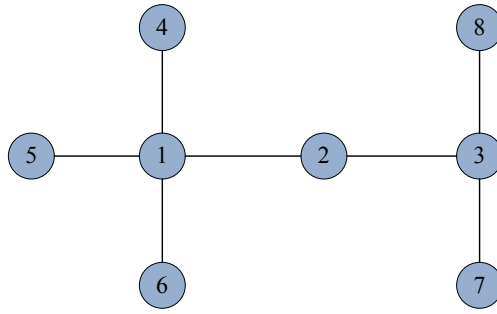


图 5.1: 具有 6 个节点的简单网络

节点 1 的邻居集合为 $M_1 = \{2, 4, 5, 6\}$ ，且其邻居的度之和为：

$$T = 2 + 1 + 1 + 1 = 5$$

节点 1 的节点熵为：

$$N = - \left(\frac{2}{5} \times \log \frac{2}{5} + \frac{1}{5} \times \log \frac{1}{5} + \frac{1}{5} \times \log \frac{1}{5} + \frac{1}{5} \times \log \frac{1}{5} \right) = 1.9219$$

5.2 数值实验与分析

本节中的实验都是利用 SI 疾病传播模型来作为验证节点传播能力的标准，并且应用于四个真实网络。四个真实的网络为：(1) NetScience 网络：这个网络表示的是同一领域的科学家的合作关系^[60]。每一个节点表示一个科学家，如果科学家之间有合作关系，则其对应的节点以边连接起来。该网络数据可在“<http://www-personal.umich.edu/~mejn/netdata/>”上获得；(2) Email 网络：这是一个邮件关系网络^[74]，每个节点表示一个上网用户，如果两个用户之间有邮件联系，则用边将两节点连接起来；(3) Blog 网络：该网络为 2004 美国大选的政治博客网，节点表示博客网页，节点间的边表示博客网页之间的超链接^[77]；(4) Yeast 网络：这个是酵母菌的蛋白质相互作用网，每个节点表示一种蛋白质，

表 5.1: 四个真实网络的部分统计特征。 n 表示节点总数, m 表示边的总数, $\langle k \rangle$ 表示节点的平均度数, k_{max} 表示节点的最大度数, C 表示网络的平均集聚系数。

网络	n	m	$\langle k \rangle$	k_{max}	C
Netscience	379	914	4.82	34	0.3706
Email	1133	5451	9.62	71	0.1101
Blog	1222	16714	27.36	351	0.1601
Yeast	2375	11693	9.85	118	0.1529

如果蛋白质之间有相互作用则用边连接起来^[78]。表 5.1 列出了这个四个真实网络的部分统计特征。

我们将基于 FMEA 中心性算法与度中心性、接近中心性、介数中心性进行对比实验, 令 $F(t)$ 表示在 t 时刻以某节点为感染源传播至全网络时总的被感染的节点总数。

(1) 实验一: 各中心性算法与 SI 模型相关性比较 (见图 5.2 — 图 5.5)

从图 5.2 中可发现, 利用 SI 模型模拟出的高传播能力节点的 FMEA 中心性数值也高, 并且在此方面明显优于度中心性、接近中心性和介数中心性。在 NetScience 网络中, 我们提出的 FMEA 中心性与 SI 模型有着很强的正相关性。在 Email、Blog、Yeast 网络中, 虽然在与 SI 模型进行相关性比较时, 接近中心性要优于其他中心性, 但是 FMEA 中心性与 SI 模型也有较高的相关性。同时, 不难发现在度中心性和介数中心性与 SI 模型的相关性图中, 节点散落都比较杂乱无章, 没有成聚集状。

(2) 实验二: 肯德尔系数差异比较 (见图 5.6)

肯德尔系数能明确的表示出一种排序与另一种排序的相同程度或匹配程度, 它能比较直观的体现两种方法在数值结果中的相关程度。在该实验中, 我们设定 SI 模型的感染概率从 0.01 逐渐递增至 0.1, 然后计算在不同感染概率下 SI 模型模拟的数值与各中心性的肯德尔系数。肯德尔系数 τ 的值越高, 则表明该中心性与 SI 模型的相关性越高。如图 5.6 所示, 在四种网络中, 与 SI 模型的肯德尔系数最高的都是 FMEA 中心性, 并且直观的说明了 FMEA 中心性的数值结果与 SI 疾病传播模型有着很高匹配程度和相关程度。

(3) 实验三: 随着节点数增加传播能力变化 (见图 5.7)

随着节点数的增加, 前 N 个节点的平均传播能力会逐渐降低。在 NetScience、Email 和 Blog 网络中, 本章提出的 FMEA 中心性曲线都始终在其他中心性算法之上, 其说明 FMEA 中心性在前 N 节点的平均传播能力表现的是最强的。观察 Yeast 网络中的前 30 个节点的平均传播能力会发现, FMEA 中心性仅仅略逊于度

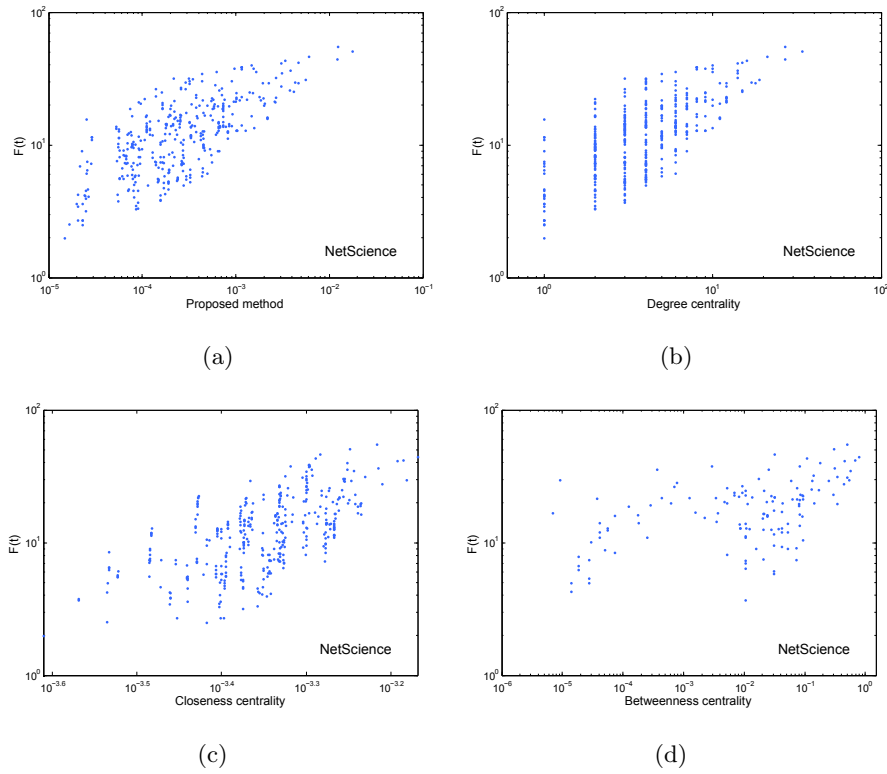


图 5.2: NetScience 网络中, 不同中心性值与 SI 模型的 $F(t)$ ($t = 10$)相关性

中心性, 但是在此之后 FMEA 中心性就具有最强的传播能力。

(4) 实验四: 前十节点的平均累积感染能力比较 (见图 5.8 — 图 5.11)

我们通过去除掉前十节点中相同的节点来比较不同算法排序出的最有影响力节点的传播能力。通过所示的图中可以看出, FMEA 中心性算法体现出了是具有最高的传播能力的中心性算法, 在每一步的传播中, FMEA 所选出节点的感染数目都要略高于甚至是高于其他中心性算法所选出来的节点。

(5) 实验五: 不同中心性值出现频率比较 (见图 5.12)

图 5.12 显示, 度中心性和接近中心性得出的排序有很多相同序值的节点。因此利用这两种中心性算法就不能很好的对这些相同序值的节点进行区分和重要度评估。同时, 不难发现, FMEA 中心性最能体现出节点的不同序值和重要度。

5.3 本章小结

本章利用失效模式及影响分析模型对网络节点的进行建模, 将风险顺序数引入到复杂网络的节点重要度评估方法中。在该模型中, 分别利用节点的入度、有效距离以及新定义的节点熵来刻画发生频度、严重程度以及检测难易程度。该模型不但考虑了网络节点的局部信息, 还将全局结构纳入建模之中, 在理论上具有

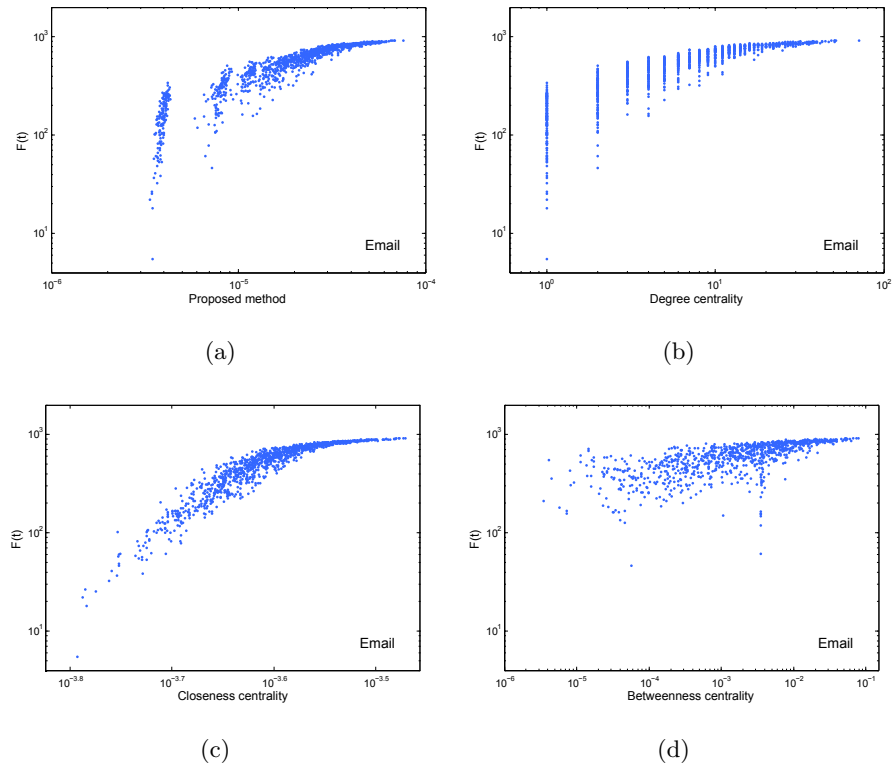


图 5.3: Email 网络中, 不同中心性值与 SI 模型的 $F(t)$ ($t=10$) 相关性

明显的可行性和有效性。通过实验分析最终表明了, 该模型能有效地在真实网络中对节点进行重要度评估。

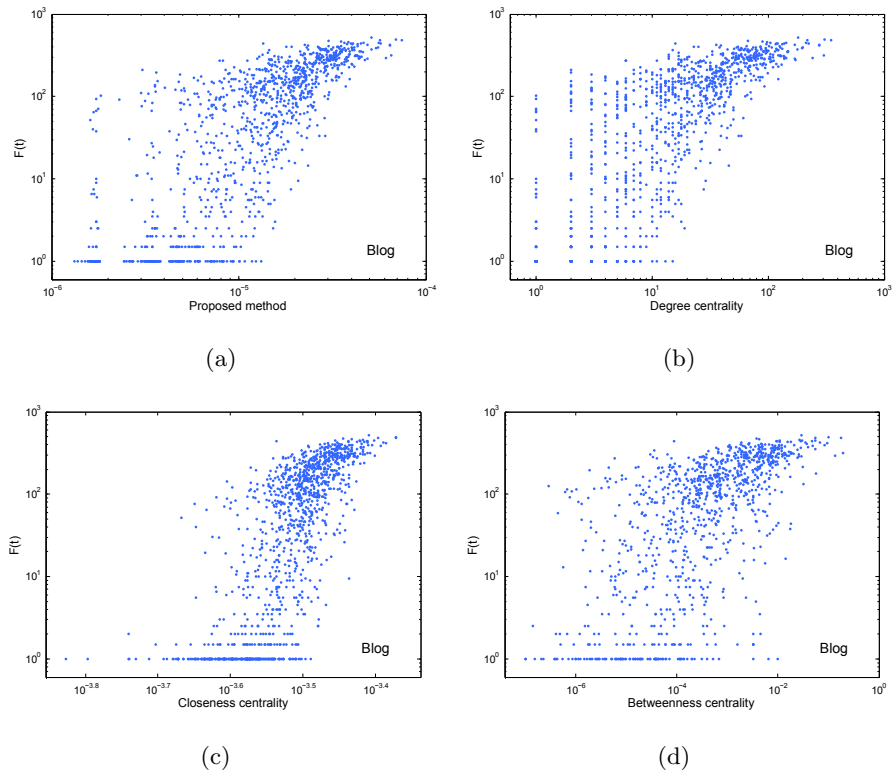


图 5.4: Blog 网络中, 不同中心性值与 SI 模型的 $F(t)$ ($t = 10$) 相关性

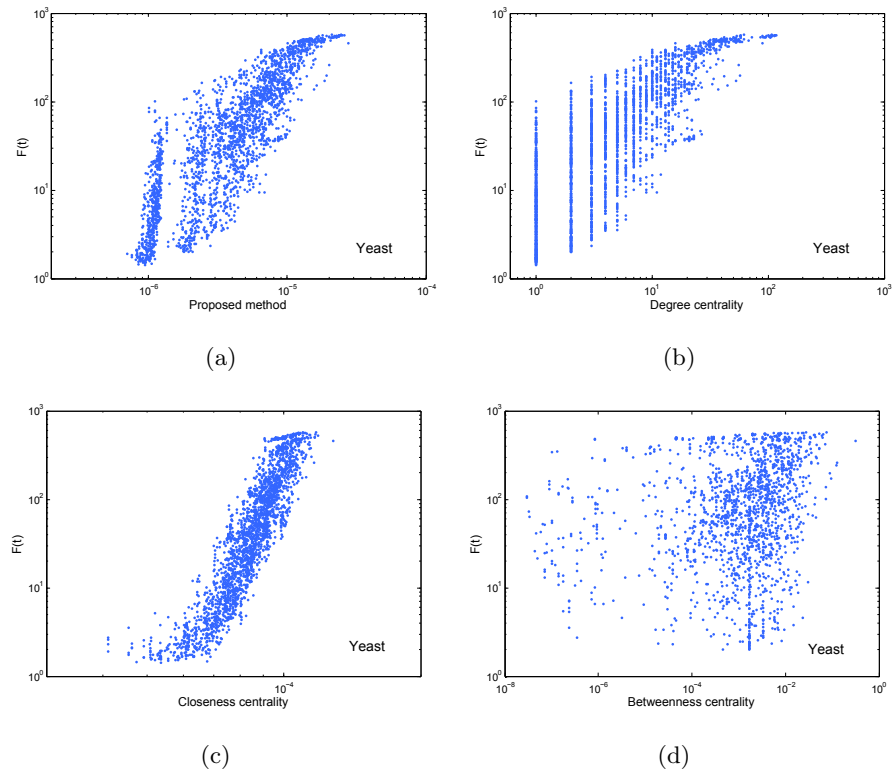


图 5.5: Yeast 网络中, 不同中心性值与 SI 模型的 $F(t)$ ($t = 10$) 相关性

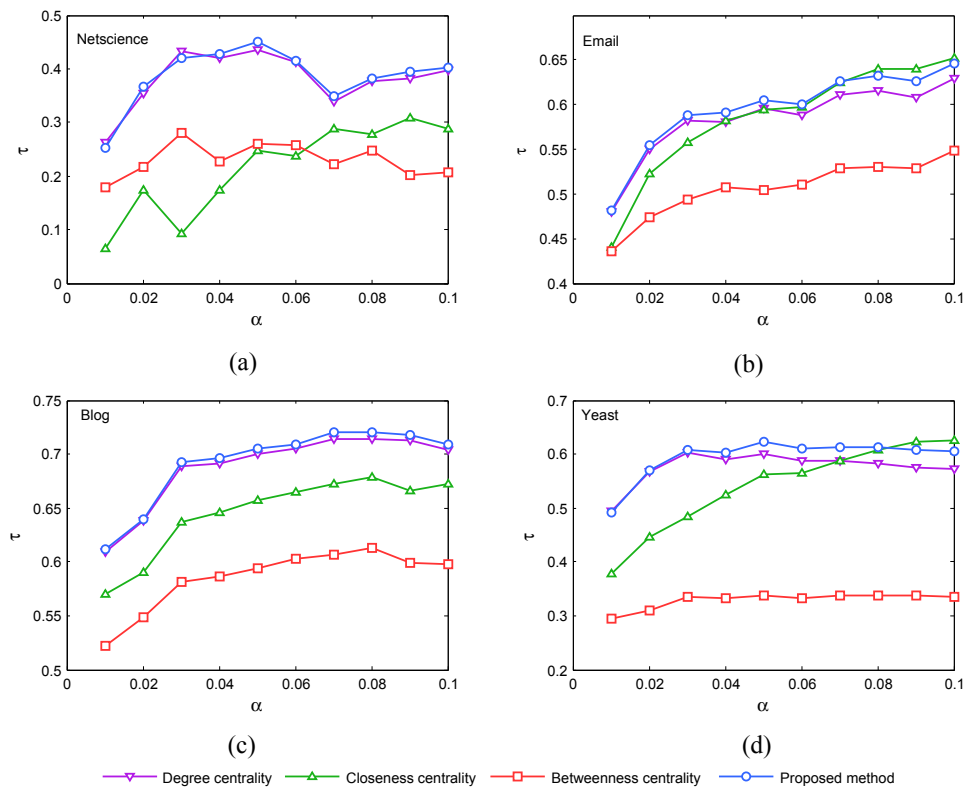


图 5.6: 不同节点中心性算法与 SI 模型的肯德尔系数相关性比较

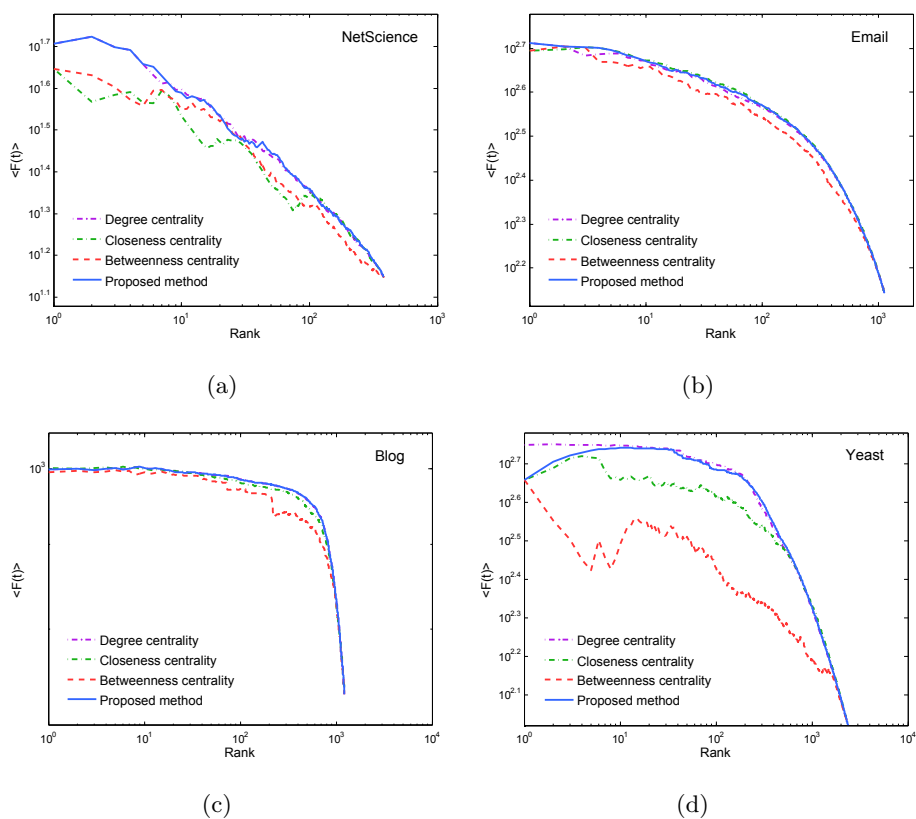


图 5.7: 不同节点中心性算法的前 N 个节点的平均感染能力

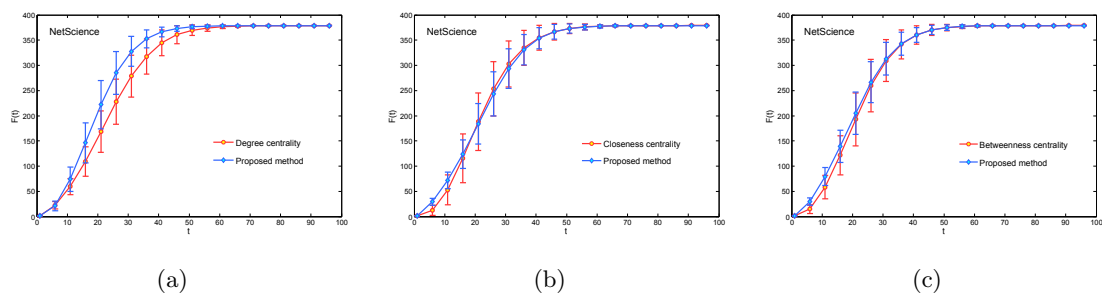


图 5.8: NetScience 网络中不同中心性排序前十节点的累积平均感染节点数

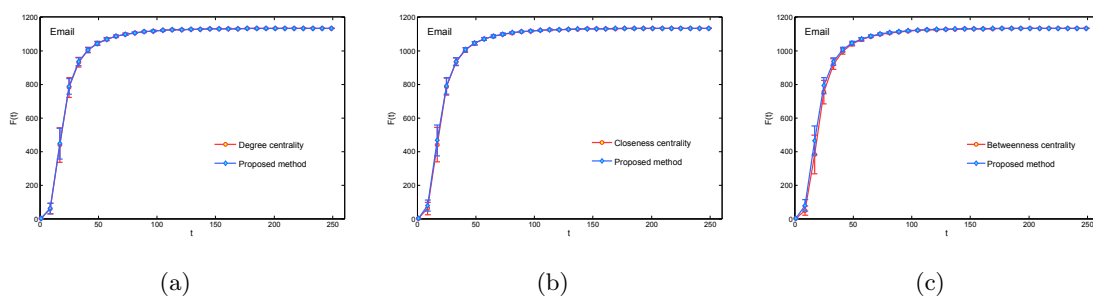


图 5.9: NetScience 网络中不同中心性排序前十节点的累积平均感染节点数

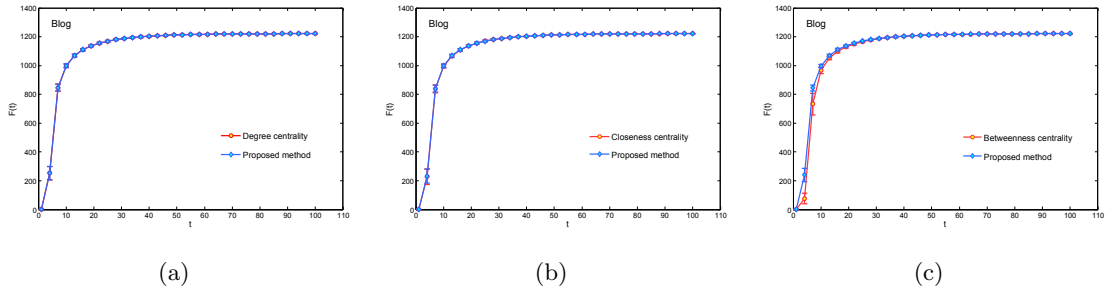


图 5.10: NetScience 网络中不同中心性排序前十节点的累积平均感染节点数

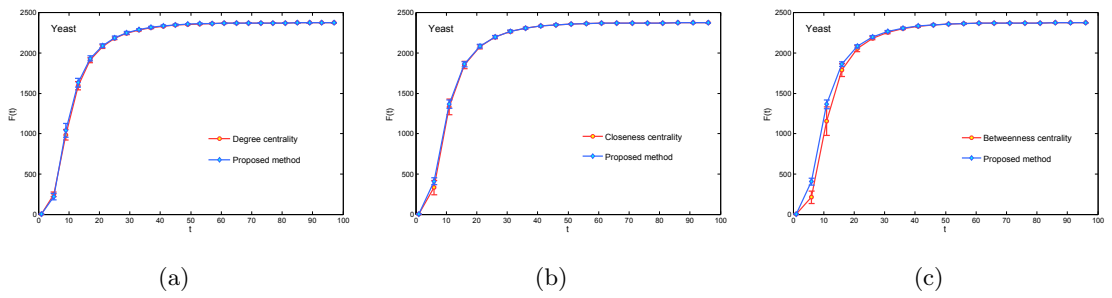


图 5.11: NetScience 网络中不同中心性排序前十节点的累积平均感染节点数

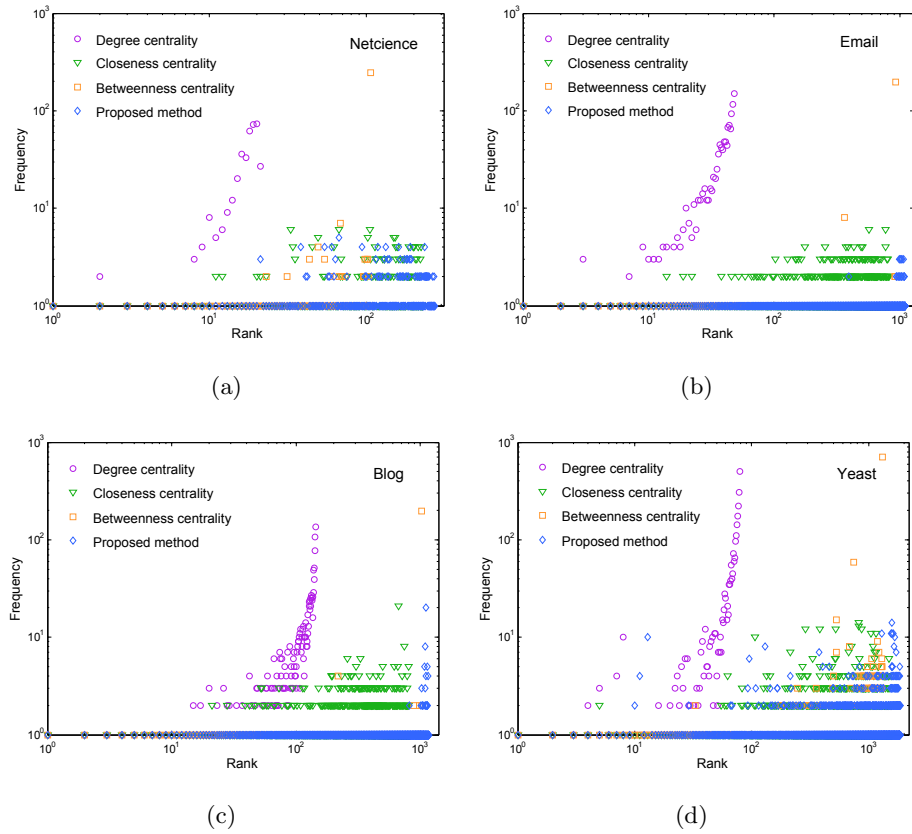


图 5.12: 在四个网络中，不同中心性值在网络节点排序中出现的频率

第六章 总结与展望

6.1 本文工作总结

复杂网络的相关研究自提出以来，一直是各领域的热点研究方向之一，其主要任务就是分析各种复杂网络的特性以便有效的处理和控制在复杂网络系统。长时间的研究发现，复杂网络中的各节点具有丰富大量的网络全局信息和局部信息，因此对网络节点的重要性识别能对复杂网络分析起到非常重要的作用和意义。节点重要度的评估虽然已得到广大的研究学者的关注和研究，但是目前已有的中心性评估算法却在一定程度上有着各种的不足和局限性。针对不同的问题和模型，本文致力于提出多种中心性算法模型以对节点进行有效的重要度评估。

本文所完成的工作及相关理论主要为以下三点：

(1) 由于在传统的复杂网络节点重要度分析中，节点之间的距离是按节点之间的测地线或者以实际地理位置来衡量，这使得在距离结构中忽视了节点周围的信息及结构。并且当网络中存在孤立节点或单向边时，传统的距离计算会导致部分节点的距离是无穷大的，在这种情况下利用接近中心性来评估节点重要度则是无效的。对于此问题，我们将有效距离代替传统距离，利用节点之间的信息流交互过程来确定不同节点对之间的距离，构建新的接近中心性模型。除此之外，我们验证了该模型的有效性和实用性，并且表明该模型也能运用于有向网络中。通过在真实网络中与度中心性、接近中心性和介数中心性的比较，我们发现无论是所选节点的传播能力还是所得排序值的区分度方面，我们提出的基于有效距离的接近中心性算法明显优于其他中心性算法。

(2) 考虑到不同的中心性算法存在各种差异，并且会得出各不相同的评估结果，我们希望构建出一种能有效处理各中心性算法冲突的模型，将不同中心性的评估进行折衷融合以达到最优评估。我们借鉴了 TOPSIS 多属性决策模型，把不同的中心性作为多个属性进行融合决策，试图达到各中心性的理想最优评估。通过该模型的提出，我们解决了多个中心性算法的差异评估问题，在冲突利益最小化的情况下融合得到最优的评估标准。在真实网络的验证中，较其他的中心性算法而言，该模型表现出了强大的优势，得出的各节点重要度有着较强的信服力。

(3) 网络中的各节点具有丰富大量的网络全局信息和局部信息，倘若我们能有效加以利用，定能对网络分析起到巨大的作用。通过分析，我们认为，如果一个节点的入度越大，即其他节点指向该节点的边越多，那么可以表明其他节点发生故障时影响到该节点的机会就越大，从而反映的是这个节点发生“失效”的概率就越高。这里的“失效”在不同类型的网络中，可以表现为遭受攻击、被感染

等行为。同时我们认为，如果一个节点到其他所有节点的有效距离越短，那么这个节点就越处于网络结构的中心位置。当该节点发生“失效”后，它就能快速的波及到网络中的其他节点，那么我们就认为这个节点失效的严重程度就越大。在信息论中，熵表示的是平均不确定性的量度，熵的理解可以为不确定性的量度而不是确定性的量度。因此我们认为节点的熵值越大，我们就越不了解这个节点的信息，则表明对这个节点的“失效”探测也就越难。基于以上思想，我们分别利用节点的入度、有效距离以及新定义的节点熵来刻画发生频度、严重程度以及检测难易程度。通过最终求得的风险顺序数来作为节点重要度评估的依据。通过在真实网络中的测试，及与其他中心性算法的比较，验证了该模型在节点重要度的评估方面的有效性和实用性。

6.2 未来工作展望

本文所提出的模型虽较其他中心性模型而言，具有很强的有效性和实用性，但是也存在着一一定的局限性，并且有待进一步的研究。其未来的工作主要可以从以下几个方面来开展：

(1) 在基于有效距离的中心性算法模型中，可以考虑多维的网络信息流来衡量节点之间的距离。比如在真实人际关系网中，同一个朋友圈中的人们可能存在多种类型的关系，包括生意合作关系、亲朋好友关系等等。我们可以将不同类型的信息作为节点对之间的信息流来改进网络节点的有效距离。

(2) 在基于 TOPSIS 算法的中心性模型中，我们还可以考虑其他的中心性算法作为多个属性引入到模型中，并且可分析不同的网络类型应该引入的中心性算法。比如在分析基于最短路径来传输信息流的网络中，引入度中心性、接近中心性和介数中心性是比较合理的；而在分析信息流是基于随机游走或扩散的网络时，度中心性接近中心性和特征向量中心性应该是合理的。

(3) 在基于 FMEA 模型的中心性算法中，我们不仅可以考虑失效频度、严重程度、和探测难易度来作为节点“失效”的因素，还可以在不同的网络中考虑能多的节点“失效”因素，比如“失效”的作用时间、周期、花费等等。另外，在以后的工作中还可以考虑到不同“失效”因素之间的权重分配。

参考文献

- [1] 2002 - 2014 年历年复杂网络方向国家自然科学基金项目中标情况表. <http://www.lailook.net/jctj/05/2015-02-24/43514.html>, 2014.
- [2] Paul Erdős and A Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci.*, 5:17–61, 1960.
- [3] Duncan J Watts and Steven H Strogatz. Collective dynamics of ‘small-world’ networks. *Nature*, 393(6684):440–442, 1998.
- [4] Mark EJ Newman and Duncan J Watts. Renormalization group analysis of the small-world network model. *Physics Letters A*, 263(4):341–346, 1999.
- [5] Albert-László Barabási and Réka Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [6] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.
- [7] Petter Holme, Beom Jun Kim, Chang No Yoon, and Seung Kee Han. Attack vulnerability of complex networks. *Physical Review E*, 65(5):056109, 2002.
- [8] Jianhua Zhang, Xiaoming Xu, Liu Hong, Shuliang Wang, and Qi Fei. Attack vulnerability of self-organizing networks. *Safety Science*, 50(3):443–447, 2012.
- [9] Qing Shuang, Mingyuan Zhang, and Yongbo Yuan. Node vulnerability of water distribution networks under cascading failures. *Reliability Engineering & System Safety*, 124:132–141, 2014.
- [10] Marc Barthélemy, Alain Barrat, Romualdo Pastor-Satorras, and Alessandro Vespignani. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Physical Review Letters*, 92(17):178701, 2004.
- [11] Yukio Hayashi, Masato Minoura, and Jun Matsukubo. Oscillatory epidemic prevalence in growing scale-free networks. *Physical Review E*, 69(1):016112, 2004.

- [12] Alain Barrat, Marc Barthélemy, and Alessandro Vespignani. *Dynamical processes on complex networks*. Cambridge University Press, 2008.
- [13] Tao Zhou, Jian-Guo Liu, Wen-Jie Bai, Guanrong Chen, and Bing-Hong Wang. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity. *Physical Review E*, 74(5):056109, 2006.
- [14] 王文旭. 复杂网络的演化动力学及网络上的动力学过程研究[D]. PhD thesis, 中国科学技术大学, 2007.
- [15] 杨阳, 荣智海, and 李翔. 复杂网络演化博弈理论研究综述. *复杂系统与复杂性科学*, 5(4):47–55, 2009.
- [16] Craig D Parks, Jeff Joireman, and Paul AM Van Lange. Cooperation, trust, and antagonism how public goods are promoted. *Psychological science in the public interest*, 14(3):119–165, 2013.
- [17] Andrew Belmonte et al. Spatial pattern dynamics due to the fitness gradient flux in evolutionary games. *Physical Review E*, 87(6):062138, 2013.
- [18] Marianne Marcoux and David Lusseau. Network modularity promotes cooperation. *Journal of theoretical biology*, 324:103–108, 2013.
- [19] Sergey Brin and Lawrence Page. The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1):107–117, 1998.
- [20] 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.
- [21] Tore Opsahl, Filip Agneessens, and John Skvoretz. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*, 32(3):245–251, 2010.
- [22] Duanbing Chen, Linyuan Lü, Ming-Sheng Shang, Yi-Cheng Zhang, and Tao Zhou. Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*, 391(4):1777–1787, 2012.
- [23] 胡庆成, 尹龔燊, 马鹏斐, 高旻, 张勇, and 邢春晓. 一种新的网络传播中最有影响力的节点发现方法. *物理学报*, 62(14):140101–140101, 2013.

-
- [24] Linyuan Lü, Yi-Cheng Zhang, Chi Ho Yeung, and Tao Zhou. Leaders in social networks, the delicious case. *PloS one*, 6(6):e21202, 2011.
- [25] Phillip Bonacich. Factoring and weighting approaches to status scores and clique identification. *Journal of Mathematical Sociology*, 2(1):113–120, 1972.
- [26] Linton C Freeman. Centrality in social networks conceptual clarification. *Social networks*, 1(3):215–239, 1979.
- [27] Ulrik Brandes. A faster algorithm for betweenness centrality*. *Journal of Mathematical Sociology*, 25(2):163–177, 2001.
- [28] 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.
- [29] 邓勇 and 施文康. 一种改进的证据推理组合规则. 上海交通大学学报, 37(8):1275–1278, 2003.
- [30] 朱大奇 and 刘永安. 故障诊断的信息融合方法. 控制与决策, 22(12):1321–1328, 2007.
- [31] 邓勇, 施文康, and 朱振福. 一种有效处理冲突证据的组合方法. 红外与毫米波学报, 23(1):27–32, 2004.
- [32] Ying Ming Wang, Jian Bo Yang, Dong Ling Xu, and Kwai Sang Chin. On the combination and normalization of interval-valued belief structures. *Information Sciences*, 177(5):1230 – 1247, 2007.
- [33] Thierry Denœux. Conjunctive and disjunctive combination of belief functions induced by nondistinct bodies of evidence. *Artificial Intelligence*, 172:234 – 264, 2008.
- [34] 邓勇, 朱振福, and 钟山. 基于证据理论的模糊信息融合及其在目标识别中的应用. 航空学报, 26(6):754–758, 2005.
- [35] Peng Chen, Huafeng Xie, Sergei Maslov, and Sidney Redner. Finding scientific gems with google’s pagerank algorithm. *Journal of Informetrics*, 1(1):8–15, 2007.

- [36] Dylan Walker, Huafeng Xie, Koon-Kiu Yan, and Sergei Maslov. Ranking scientific publications using a model of network traffic. *Journal of Statistical Mechanics: Theory and Experiment*, 2007(06):P06010, 2007.
- [37] Ying Ding, Erjia Yan, Arthur Frazho, and James Caverlee. Pagerank for ranking authors in co-citation networks. *Journal of the American Society for Information Science and Technology*, 60(11):2229–2243, 2009.
- [38] 任晓龙 and 吕琳媛. 网络重要节点排序方法综述. *科学通报*, 59(13):1175–1197, 2014.
- [39] Warren Gilchrist. Modelling failure modes and effects analysis. *International Journal of Quality & Reliability Management*, 10(5), 1993.
- [40] Mohamed Ben-Daya and Abdul Raouf. A revised failure mode and effects analysis model. *International Journal of Quality & Reliability Management*, 13(1):43–47, 1996.
- [41] John B Bowles and C Enrique Peláez. Fuzzy logic prioritization of failures in a system failure mode, effects and criticality analysis. *Reliability Engineering & System Safety*, 50(2):203–213, 1995.
- [42] Fiorenzo Franceschini and Maurizio Galetto. A new approach for evaluation of risk priorities of failure modes in fmea. *International Journal of Production Research*, 39(13):2991–3002, 2001.
- [43] Kwai-Sang Chin, Ying-Ming Wang, Gary Ka Kwai Poon, and Jian-Bo Yang. Failure mode and effects analysis using a group-based evidential reasoning approach. *Computers & Operations Research*, 36(6):1768–1779, 2009.
- [44] Pauli AA Garcia, Roberto Schirru, et al. A fuzzy data envelopment analysis approach for fmea. *Progress in Nuclear Energy*, 46(3):359–373, 2005.
- [45] Ying-Ming Wang, Kwai-Sang Chin, Gary Ka Kwai Poon, and Jian-Bo Yang. Risk evaluation in failure mode and effects analysis using fuzzy weighted geometric mean. *Expert Systems with Applications*, 36(2):1195–1207, 2009.
- [46] Gülşen Aydın Keskin and Coşkun Özkan. An alternative evaluation of fmea: fuzzy art algorithm. *Quality and Reliability Engineering International*, 25(6):647–661, 2009.

-
- [47] Heeralal Gargama and Sanjay Kumar Chaturvedi. Criticality assessment models for failure mode effects and criticality analysis using fuzzy logic. *Reliability, IEEE Transactions on*, 60(1):102–110, 2011.
- [48] Annibal Parracho Sant’Anna. Probabilistic priority numbers for failure modes and effects analysis. *International Journal of Quality & Reliability Management*, 29(3):349–362, 2012.
- [49] Réka Albert, Hawoong Jeong, and Albert-László Barabási. Internet: Diameter of the world-wide web. *Nature*, 401(6749):130–131, 1999.
- [50] Fan Chung and Linyuan Lu. The average distances in random graphs with given expected degrees. *Proceedings of the National Academy of Sciences*, 99(25):15879–15882, 2002.
- [51] Mark EJ Newman. The structure and function of complex networks. *SIAM review*, 45(2):167–256, 2003.
- [52] Stefania Vitali, James B Glattfelder, and Stefano Battiston. The network of global corporate control. *PloS One*, 6(10):e25995, 2011.
- [53] Phillip Bonacich and Paulette Lloyd. Eigenvector-like measures of centrality for asymmetric relations. *Social networks*, 23(3):191–201, 2001.
- [54] Linton C Freeman, Stephen P Borgatti, and Douglas R White. Centrality in valued graphs: A measure of betweenness based on network flow. *Social networks*, 13(2):141–154, 1991.
- [55] Anand Pillay and Jin Wang. Modified failure mode and effects analysis using approximate reasoning. *Reliability Engineering & System Safety*, 79(1):69–85, 2003.
- [56] 中国汽车技术研究中心编译. 潜在的失效模式及后果分析. 2002.
- [57] Dean H Stamatis. *Failure mode and effect analysis: FMEA from theory to execution*. ASQ Quality Press, 2003.
- [58] Ford Motor Company. *Potential failure mode and effects analysis (FMEA) reference manual*. 1998.
- [59] Dirk Brockmann and Dirk Helbing. The hidden geometry of complex, network-driven contagion phenomena. *Science*, 342(6164):1337–1342, 2013.

- [60] Mark EJ Newman. Finding community structure in networks using the eigenvectors of matrices. *Physical review E*, 74(3):036104, 2006.
- [61] Wen-Jie Bai, Tao Zhou, and Bing-Hong Wang. Immunization of susceptible–infected model on scale-free networks. *Physica A: Statistical Mechanics and its Applications*, 384(2):656–662, 2007.
- [62] Marton Karsai, Mikko Kivelä, Raj Kumar Pan, Kimmo Kaski, János Kertész, A-L Barabási, and Jari Saramäki. Small but slow world: How network topology and burstiness slow down spreading. *Physical Review E*, 83(2):025102, 2011.
- [63] Yan Gang, Zhou Tao, Wang Jie, Fu Zhong-Qian, and Wang Bing-Hong. Epidemic spread in weighted scale-free networks. *Chinese Physics Letters*, 22(2):510, 2005.
- [64] Maurice G Kendall. A new measure of rank correlation. *Biometrika*, 30:81–93, 1938.
- [65] Maurice G Kendall. The treatment of ties in ranking problems. *Biometrika*, 33:239–251, 1945.
- [66] Ching-Lai Hwang and K Paul Yoon. *Multiple Attribute Decision Making: Methods and Applications*. Springer-Verlag, New York, 1981.
- [67] Hsu-Shih Shih. Incremental analysis for MCDM with an application to group TOPSIS. *European Journal of Operational Research*, 186(2):720–734, 2008.
- [68] Cengiz Kahraman, Orhan Engin, Özgür Kabak, and İhsan Kaya. Information systems outsourcing decisions using a group decision-making approach. *Engineering Applications of Artificial Intelligence*, 22(6):832–841, 2009.
- [69] Zhi-Ping Fan and Bo Feng. A multiple attributes decision making method using individual and collaborative attribute data in a fuzzy environment. *Information Sciences*, 179(20):3603–3618, 2009.
- [70] Anjali Awasthi, Satyaveer S Chauhan, and SK Goyal. A fuzzy multicriteria approach for evaluating environmental performance of suppliers. *International Journal of Production Economics*, 126(2):370–378, 2010.
- [71] Haixin Zhang, Yong Deng, Felix TS Chan, and Xiaoge Zhang. A modified multi-criterion optimization genetic algorithm for order distribution in collaborative supply chain. *Applied Mathematical Modelling*, 37(14):7855–7864, 2013.

-
- [72] David Krackhardt. Assessing the political landscape: Structure, cognition, and power in organizations. *Administrative Science Quarterly*, 35(1):342–369, 1990.
- [73] Vittoria Colizza, Romualdo Pastor-Satorras, and Alessandro Vespignani. Reaction–diffusion processes and metapopulation models in heterogeneous networks. *Nature Physics*, 3(4):276–282, 2007.
- [74] Roger Guimera, Leon Danon, Albert Diaz-Guilera, Francesc Giralt, and Alex Arenas. Self-similar community structure in a network of human interactions. *Physical review E*, 68(6):065103, 2003.
- [75] Michelle Girvan and Mark EJ Newman. Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, 99(12):7821–7826, 2002.
- [76] Claude Elwood Shannon. A mathematical theory of communication. *ACM SIGMOBILE Mobile Computing and Communications Review*, 5(1):3–55, 2001.
- [77] Lada A Adamic and Natalie Glance. The political blogosphere and the 2004 US election: divided they blog. In *Proceedings of the 3rd international workshop on Link discovery*, pages 36–43. ACM, 2005.
- [78] Dongbo Bu, Yi Zhao, Lun Cai, Hong Xue, Xiaopeng Zhu, et al. Topological structure analysis of the protein–protein interaction network in budding yeast. *Nucleic acids research*, 31(9):2443–2450, 2003.

发表文章目录

- [1] Yuxian Du, Cai Gao, Xin Chen, Yong Hu, Rehan Sadiq, Yong Deng, A new closeness centrality measure via effective distance in complex networks, *Chaos*, 2015, 25(3): 033112.
- [2] Yuxian Du, Cai Gao, Yong Hu, Sankaran Mahadevan, Yong Deng, A new method of identifying influential nodes in complex networks based on TOPSIS, *Physica A*, 2014, 399(1): 57-69.
- [3] Yuxian Du, Xi Lu, Xiaoyan Su, Yong Hu, Yong Deng, New failure mode and effects analysis: An evidential downscaling method, *Quality and Reliability Engineering International*. DOI: 10.1002/qre.1753.
- [4] Yuxian Du, Hongming Mo, Xinyang Deng, Rehan Sadiq, Yong Deng, A new method in failure mode and effects analysis based on evidential reasoning, *International Journal of System Assurance Engineering and Management*, 2014, 5(1): 1-10.
- [5] Yuxian Du, Xiaoyan Su, Yong Deng, New risk evaluation in failure mode and effects analysis, *ICIC Express Letters, Part B: Applications*, 2013, 4(6): 1543-1550.
- [6] Yuxian Du, Xinyang Deng, Yong Deng, Failure mode and effects analysis using Dempster-Shafer theory and D numbers methodology, *ICIC Express Letters*, 2014, 8(6): 1731-1736.
- [7] Yuxian Du, Shiyu Chen, Yong Deng, Risk evaluation in failure mode and effects analysis based on Dempster-Shafer theory and prospect theory, *Journal of Information and Computational Science*, 2014, 11(4): 1153-1161.

简 历

基本情况

杜雨弦，男，四川省平昌县人，1988 年 12 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

2008 年 9 月至 2012 年 7 月，西南大学计算机与信息科学学院，本科，专业：网络工程。

2012 年 9 月至 2015 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

信息融合与智能计算、风险分析及评估、复杂网络与复杂系统。

联系方式

通讯地址：重庆市北碚区天生路2号西南大学计算机与信息科学学院

邮编：400715

E-mail: duyuxianswu@163.com

致 谢

时光转眼即逝，在科研的学习道路上我已前行了三年。三年中，有导师的引领，有师兄师姐的帮助，还有同学的相伴。在这转瞬即逝的时光中，我感受了许许多多，也收获了许多许多。回望三年，感恩无限。

首先，我要感谢我的导师邓勇，他严谨细致、一丝不苟的作风一直是我工作、学习中的榜样，给了我起到了指明灯的作用。他循循善诱的教导和不拘一格的思路给予我无尽的启迪，让我很快就感受到了科研学习的快乐并融入其中。直到如今我都还依稀记得第一次见到邓老师的场景，与他的谈话交流中，我不断感受到与老师心灵间碰撞擦出的火花，这点星星之火早已点燃了我科研学习的熊熊火焰。在我第一篇论文的初稿完成后，邓老师一遍又一遍地给我仔细修改，让我至今感动难忘，每一次的修改都是老师对我的一次教导。俗话说的好，“磨刀不误砍柴工”，当每次遇到不懂的问题时，我都会第一时间记在本子上面，然后等答疑的时候问老师，邓老师对于我提出来的问题都一一解答，从来都不会因为我的问题稍过简单加以责备，而是一再的告诫我做设计该注意的地方，从课题的选择到项目的最终完成，邓老师都始终给予我细心的指导和不懈的支持，他真正起到了“传道、授业、解惑疑”的师德，让人油然而生的敬佩。多少个日日夜夜，邓老师不仅在学业上给我以精心指导，同时还在思想、生活上给我以无微不至的关怀，除了敬佩老师的专业水平外，他的治学严谨和科学研究的精神也是我永远学习的榜样，并将积极影响我今后的学习和工作。在此谨向邓老师致以诚挚的谢意和崇高的敬意。

其次，我要感谢实验室的同门我的帮助和指点，没有他们的帮助和提供资料，没有他们的鼓励和加油，这次毕业设计就不会如此的顺利进行。我非常感谢邓鑫洋师兄，还有已毕业的章雅娟师姐和康兵义师兄在科研上对我的帮助，他们作为邓老师在西南大学的第一届硕士生，为我们下面的师弟师妹们做了很好的榜样。在此，我也非常感谢上一届的硕士生张晓革师兄、张海鑫师姐、刘娟师姐以及博士生魏代俊师兄、李娅师姐、王庆师兄，他们的科研能力毋庸置疑，并且他们会无私奉献的帮助我、指导我，在他们身上我学到了许多的生活、科研等各方面的知识。另外我还要感谢同级的高才、兰鑫、陈诗语以及博士生潘东波，有了他们的相伴我不但在科研道路上能奋勇前进，也在生活中收获了许多许多的快乐和感动。此外，我还要感谢其他所有帮助过我、关心我的老师、同学和朋友！

在毕业即将到来之际，我的心情无法平静，从开始进入课题到论文的顺利完成，有多少可敬的师长、同学、朋友给了我无言的帮助，在这里请接受我诚挚的

谢意！最后我还要感谢计算机与信息科学学院和我的母校西南大学四年来对我的栽培。

杜雨弦

2015年3月