
电 子 科 技 大 学
UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

硕士学位论文

MASTER THESIS



论文题目 复杂网络建模的研究与应用

学 科 专 业 计算机软件与理论

学 号 201221060356

作 者 姓 名 吴诗雯

指 导 教 师 屈 鸿 副教授

分类号_____密级_____

UDC^{注1}_____

学 位 论 文

复杂网络建模的研究与应用

(题名和副题名)

吴诗雯

(作者姓名)

指导教师

屈 鸿

副 教 授

电子科技大学

成 都

(姓名、职称、单位名称)

申请学位级别 硕士

学科专业 计算机软件与理论

提交论文日期 2015.03.26

论文答辩日期 2015.05.11

学位授予单位和日期 电子科技大学 2015 年 6 月 28 日

答辩委员会主席

评阅人

注 1: 注明《国际十进分类法 UDC》的类号。

THE RESEARCH AND APPLICATION OF COMPLEX NETWORK MODELING

**A Master Thesis Submitted to
University of Electronic Science and Technology of China**

Major: **Computer Software and Theory**

Author: **Shiwen Wu**

Advisor: **Hong Hu**

School: **Computer Science and Engineering**

独创性声明

本人声明所呈交的学位论文是本人在导师指导下进行的研究工作及取得的研究成果。据我所知，除了文中特别加以标注和致谢的地方外，论文中不包含其他人已经发表或撰写过的研究成果，也不包含为获得电子科技大学或其它教育机构的学位或证书而使用过的材料。与我一同工作的同志对本研究所做的任何贡献均已在论文中作了明确的说明并表示谢意。

作者签名：____ 吴诗雯 _____

日期：2015 年 5 月 11 日

论文使用授权

本学位论文作者完全了解电子科技大学有关保留、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。本人授权电子科技大学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应遵守此规定）

作者签名：____ 吴诗雯 _____

导师签名：____ 屈鸿 _____

日期：2015 年 5 月 11 日

摘 要

上个世纪末,随着以因特网为代表的信息技术的兴起,复杂网络的研究逐渐引起了人们的注意力。然而复杂网络研究面临着巨大的挑战,其根本原因是实际网络的自身存在着多样性和混乱性特征。面对这个挑战,研究学者相继提出了两个重要的模型——WS 小世界模型和 BA 无标度网络模型,这使得复杂网络的研究取得了重大的突破,获得了业界的广泛关注。如何提出一种合理的模型来描述实际的网络系统以及如何应用复杂网络建模方法对现实世界中的问题进行分析建模以揭示其规律是其中最关键的两个问题。

目前,复杂网络科学中经典模型,虽然能模拟现实中的部分复杂网络系统,但利用经典模型所构建的网络并不符合现实网络的特征。其中的无标度网络模型在建模中,加边规则和加节点规则都是匀速增加,不符合实际网络中节点和边都是非匀速增加的特点。由此可见,提出符合实际的复杂网络模型,以及将已有的复杂网络模型及其拓扑特性运用于实际的复杂网络系统是具有重大意义的工作。

本文着重研究复杂网络的建模方法以及运用复杂网络理论对现实网络进行分析与研究。主要工作如下:

1、提出两种新的网络模型。针对现实世界中的网络非匀速增加的特征,基于常见的更新计数过程,提出两种新的网络模型。通过大量的仿真实验以及数学方法计算该模型的度指数以证明本文所提出两种模型是符合复杂网络的特性。

2、对现实世界中的电力交易网络建模。运用现有的网络模型,对现实世界的实际应用——电力交易过程,进行网络建模得到电力交易网络,然后利用复杂网络的研究方法对该网络进行理论分析并且通过仿真实验得出网络的拓扑特性。

3、将复杂网络理论应用于自然语言并提出基于复杂网络的关键词提取方法。根据文本中词与词之间的相邻关系进行建模得到文本单词网络,对该网络进行拓扑分析,发现如果以节点介数和节点的度同时作为关键词提取的依据,将提高关键词提取的准确性

关键词: 复杂网络模型, 更新计数过程, 电力交易系统, 文本分析

ABSTRACT

At the end of last century, with the development of information technology such as the Internet, research of complex network gradually aroused people's attention. Network has its own diversity and disruptive making it become one of the most tremendous challenges in complex system research. In recent years, the small world model and scale-free network model based on statistical physics of complex networks makes the complex network research scholars upsurge again. How to put forward a reasonable model to describe the actual network system and utilize complex network modeling methods to analyze the problems in the real world to reveal its law are the most critical two questions.

At present, the classic model in complex network science can simulate the real part of the complex network system, however, not conform to the characteristics of real network. Edge connection and node adding rules for the scale-free network model are increasing at a constant speed, against the actual situation increasing at a variable speed. Therefore, proposing a model increasing at a variable speed of the complex network model, and applying the existing complex network model and its topological characteristics in the actual work are of great significance.

This paper focuses on the study of complex network modeling method and employs the research methods for complex network to analyze the real network. The main works is as follows:

1. Proposition of two new network models. According to the characteristics of the nonlinear increasing in the real world network, we put forward two kinds of new network model based on the different common renewal counting processes, simulation experiments and calculation the index of the model by using mathematical method can prove that the presented two kinds of model are in accordance with the characteristics of complex networks.

2. Modeling of the practical power transaction network. Apply the existing network model to a practical application of the real world: power transaction process, and obtain the electric trading network. Then, we use the complex network knowledge to analyze the electric network and conclude the characteristics of network by simulation experiments.

3. Application of the complex network theory on natural language and proposition of a new keyword extraction method based on complex network. The adjacent relation between the words in the text are on the ground of modeling complex network. By analyzing the topology of the network, we find out that using node betweenness and node degree as the basis of keyword extraction will improve the accuracy of the keyword extraction.

Keywords: Complex network model, renewal counting process, the electric power trading system, text analysis

目 录

第一章 绪 论	1
1.1 研究背景与意义	1
1.2 研究现状	2
1.3 应用与挑战	3
1.4 论文主要工作	4
1.5 论文组织结构	5
第二章 复杂网络建模相关知识	7
2.1 复杂网络的基本概念	7
2.1.1 网络的图表示	7
2.1.2 平均路径长度	8
2.1.3 聚类系数	8
2.1.4 度与度分布	8
2.1.5 幂律分布	9
2.2 复杂网络的基本模型	9
2.2.1 规则网络	9
2.2.2 E-R 随机图.....	10
2.2.3 小世界网络模型	11
2.2.4 BA 无标度网络模型.....	13
2.3 本章小结	15
第三章 基于更新计数过程的复杂网络建模	16
3.1 基于均匀更新计数过程的无标度网络	16
3.1.1 网络演化算法	17
3.1.2 网络演化实例	18
3.1.3 理论分析	20
3.2 基于高斯更新计数过程的无标度网络	22
3.2.1 网络演化算法	22
3.2.2 网络演化实例	23
3.2.3 理论分析	24
3.3 仿真实验与分析	25
3.3.1 网络建模过程	25

3.3.2 网络度分布	28
3.3.3 网络可靠性分析	32
3.4 本章小结	36
第四章 基于复杂网络的电力交易模型	37
4.1 电力市场的网络模型	37
4.1.1 电力交易过程	37
4.1.1.1 传输过程	38
4.1.1.2 买卖双方的利益	40
4.1.2 电力交易市场的建模算法	40
4.2 理论分析	42
4.3 仿真实验及分析	43
4.3.1 网络模型实例	43
4.3.2 网络拓扑分析	44
4.3.2.1 度分布	44
4.3.2.2 评估指标	46
4.3.3 鲁棒性分析	48
4.3.3.1 随机与蓄意攻击	48
4.3.3.2 网络规模与攻击强度的关系	49
4.4 本章小结	51
第五章 基于复杂网络的自然语言模型	52
5.1 文本网络研究方法	52
5.1.1 文本网络模型	52
5.1.2 节点介数	53
5.1.3 节点重要性	53
5.2 构建文本单词网络	54
5.2.1 文本预处理	54
5.2.2 构建方法	55
5.3 仿真实验与分析	56
5.3.1 网络拓扑分析	56
5.3.1.1 度分布	57
5.3.1.2 最小二乘法求幂指数	58
5.3.1.3 鲁棒性	60
5.3.2 关键字提取实验	61
5.3.2.1 算法实验一	61
5.3.2.2 算法实验二	62
5.4 本章小结	64

第六章 全文总结与展望	65
6.1 全文总结	65
6.2 展望	66
致 谢	67
参考文献	68
攻读硕士学位期间取得的成果	71

第一章 绪 论

1.1 研究背景与意义

随着科技文明的高速发展，网络信息技术与现代生活产生了越来越紧密的联系，复杂网络已出现在人们日常生活中的方方面面。近年来，基于信息技术的网络科学已经成为了人类生活与娱乐中必不可缺的工具。网络不仅拉近了人与人之间的距离，而且使得人与人、人与物、人与自然产生的复杂关联都得以用复杂网络来描述^[1,2]。许多现实中的复杂关系比如：社交关系、电力传输、病毒传播、论文引用等等，恰恰这些关联网络都有着相似与无标度网络模型中的特性。由于实际网络的多样性以及变化性，针对现实网络的研究也面临着越来越严峻的挑战。因此，创建一个能最大近似的模拟实际网络的模型变成了现在研究的主要难点。

20 世纪中期，数学家 Erdős^[3]开创了随机图理论^[4]的建立，给复杂网络理论的研究奠定了坚实的基础^[5]。实际的复杂系统通过用一个点表示一个现实对象，点与点的连线表示对象之间的相互关系的方式抽象出一个无向图。在科学家对这些网络图的研究时，忽略图中的点的位置，边的长度、形状以及是否有重叠等，针对这样抽象出的结构我们称之为网络拓扑结构，而其所表现出的特征成为网络拓扑特性。现代网络模型的建立初始于美国哈佛大学 Stanley Milgram 所做的闻名世界的小世界实验，后来也被称为“六度分离”实验^[6]。这个实验揭示了世界上任意两个人平均只要通过 5.2 个人就能产生关联，同时也揭示了较短的平均最短路径是衡量实际网络的一个重要性质。

网络系统由其网络结构、网络中的节点以及其内部各种因素共同作用与影响表现出其复杂性^[7]。国内对复杂网络的定义普遍认同的是我国著名科学家钱学森所提出的，具有自组织性、自相似性、吸引子、小世界、无标度中的部分或全部性质的网络成为复杂网络。正是由于这些特性恰恰能描述实际的复杂系统的结构的复杂性，用复杂网络的理论应用于现实世界的复杂网络结构便顺理成章的成为世界学者的研究热点。

目前，复杂网络发展历史中主要有两个模型——WS 小世界和 BA 无标度，被研究学者作为最基础的模型进行研究，再此基础上衍生出了许许多多的网络模型。然而，这些研究表明，虽然 BA 网络弥补了 WS 网络中没有考虑到实际网络的规模会随着时间增长以及新加入的节点会优先选择度数大的节点进行连接等特性，但仍然有其自身的局限性。

本文试图提出两种新的复杂网络模型，以更好的切合实际网络所具有的特征。其中一种模型开始于一个低连接，低聚合的小节点组，用均匀更新计数过程拟合

生成顶点过程以及用均匀分布产生新节点的连接边数。另一种模型结合了实际网络节点生成时间间隔并不是始终保持相等的这个特点,使用高斯更新计数过程拟合网络顶点增加过程。这两种模型的提出很好的弥补了 BA 网络模型中顶点的增加率和边的增加率均是不变这一缺点。

随着复杂网络逐渐广泛的应用于各行各业的复杂系统,其中对电力交易系统的研究大部分都旨在提出双边交易的策略,很少专注研究其真正的交易市场结构。本文基于复杂网络模拟真正的双边交易市场结构,构建一个适当的模型,提出了一种新的研究电力交易市场的方法。通过分析该双边交易市场模型的网络特征,帮助我们深刻和正确理解真正的市场状态,得出有助于市场管理者和参与者做决策的结论。

复杂网络不仅可以用于描述实际的复杂系统,也可用于与其他学科的交互应用。本文提出一种新的关键词提取算法,有别于机器学习中的自然语言处理中大部分都是以统计学方式计算出每个词出现的频率。该算法利用文本词与词之间的联系,构建文本词网络,根据词网络的拓扑特性提取关键词,对关键词提取的准确度有明显的提高。

综上所述,本文不光在理论上对建模方法进行了改进,也在实际应用上对现实中的实际网络进行了分析与研究,这均对复杂网络的研究具有十分重要的研究价值和意义。

1.2 研究现状

网络科学伴随着图论被作为各种领域的研究基础孕育而生。1736 年,数学家欧拉(Leonhard Euler)在图论的萌生之初,便解决了著名的“哥尼斯堡七桥”问题。首篇关于随机图的论文,是由匈牙利数学家 Erdős 于 1948 年发表的,这为复杂网络理论研究开启了新的诗篇,奠定了复杂网络的基本模型。随着对实际网络的深入研究,随机图往往不符合实际问题,继而更贴合实际的网络新模型相继出现。

20 世纪后期,两个重要的模型——WS 小世界^[8]和 NW 小世界^[9]分别被提出。其中 WS 是由 Watts 和 Strogatz 提出的,该模型首次同时结合网络具有规则性和随机性两个特性,另一个是由 Newman 和 Watts 在发现了 WS 网络的不足点后提出的。经过他们的研究,发现实际的网络既不是完全常规图也不是完全的随机图,而应该是介于两者之间的,那么可以通过随机化重连或者随机化加边的方式得以实现。这样得到的网络模型的平均路径长度较短,使得网络的小世界特性得以模拟实现。自此,由 Erdős 提出的 ER 模型在复杂网络科学中的统治地位彻底被改写。

进入 21 世纪后,越来越多的研究表明 ER 随机图的网络度分布曲线图相似于

泊松分布曲线图，对于 WS 网络也具有相同的特性。通过大量分析这两种网络的度分布发现网络平均节点度始终有一个上限值，人们把拥有以上特征的网络称为均匀网络。终于在 1999 年，美国圣母大学的物理学家 Barabási 和 Albert 共同提出了 BA 无标度网络模型^[10]，他们结合实际网络节点增长和新增节点优先连接到网络中度较大节点这两个特性成功的构建了符合幂律分布的网络模型。

BA 无标度网络的提出为复杂网络研究提供了全新的思路，迈出了关键的一步。尽管可以用 BA 模型描述拥有高组织性，服从幂律分布的实际网络，但 BA 由于其自身不考虑时间因素的这一局限性，仍然不能完美的描述现实世界中的网络。众多学者受到 Barabási 的 BA 无标度网络模型的启发，在建模时加入各种实际中的因素，大量的复杂网络模型如雨后春笋般接连被提出。

网络科学研究不仅要求构建出对应实际网络的模型，还要分析其理论特性。其中对度指数的计算方法是最为关键的问题，Barabási 提出 BA 模型后，在 1999 年利用动力学方程提出了平均场方法。但平均场方法虽然易于理解，在理论上还不足以令人信服，2000 年 Krapivsky 基于数学的微分方程提出率方程法。同样针对上述平均场方法的不足，同年，Dorogovtsev 发表了利用主方程法精确计算出无标度网络模型度指数的论文^[11]。

综上所述，随着研究者不断的努力，网络科学已经获得了迅猛的发展。然而对复杂网络的建模方法还有待进一步的研究与创新，怎样使得网络科学能进一步切合实际应用为人类创造更为丰富的生活仍然是网络科学研究的难点与热点。

1.3 应用与挑战

复杂网络以其多样性及适应性，展现了其应用前景的广阔与光明，不仅可以用于复杂系统，而且可以和各个学科领域进行交叉应用^[12]。

复杂网络最显著的一方面的应用就是维护现实复杂系统，使其健康、稳健的正常运作。因特网、电力网络、交通网络、经济网络等，这些与人类生活密切相关的网络，任何一个出现问题，都将导致重大的灾难。可以运用复杂网络对这些网络进行鲁棒性的研究，提出合理的网络设计策略，在其遭到意外或蓄意攻击时能做出及时的保护措施。

除上文提到的应用外，生物学研究上也广泛的运用了网络科学理论。人类历史上爆发了多次的传染病，每一次传染病的爆发都带给人类惨痛的损失。如果能对这些传染病病毒的传播机制得以认识的话，那么就能得以这些病毒进行很好的防御工作。复杂网络对传播临界值以及传播动力学上都有着深入的研究，其中对均匀网络、非均匀网络都取得显著的研究成果。

同样, 社会网络的研究也离不开网络科学理论。由哈佛大学的心理学家 Stanley Milgram 做的关于世界很小的实验中提出的“六度分离”理论^[13]突显了真实世界网络具有较短的平均路径以及连通性强的特征。在目前的搜索研究中, 其中的基于图论的最短路径的搜索, 关键路径的查找都取得了一定的成果。复杂网络在最短路径算法、数据搜索、网络链接的搜索都能得到合理的应用。其中 Kleinberg 已经提出了 Kleinberg 网络模型^[14,15], 网络搜索策略也趋于多元化。

网络科学也显现出其在自然语言研究中优势。人们通过文字获取信息, 而处于信息爆炸时代, 人类怎样快速的获得自己想要的结果这将是最关键的问题。基于网络科学模型具有的小世界特性, 在语言网中, 可以快速的捕捉相关联的信息。在浩瀚的书籍中, 通过关键字的提取可以迅速的抓住文本的主题。

复杂网络既然体现了其宝贵的应用价值, 那么对其的研究更加成为当今的迫切需求。目前, 网络科学虽然获得了迅猛的发展, 但仍然存在许多问题没有研究出很好的解决。例如以下几点:

- (1) 概念和规律不够明确。复杂网络的研究现呈现高热度, 但对于网络的特征, 还有很多方面没有深入挖掘研究。很多网络科学中的概念和规律都没有统一的理论对其进行阐述。
- (2) 网络特征不够全面。目前, 网络结构上的拓扑特性研究虽然已经获得了大量的研究成果, 但近些年在这方面的研究已趋于瓶颈阶段。现在已经发现的网络特征可能还不足以全面的概括网络所含的特性, 所以发现更多的现实网络间共性提取新的网络特征是当今迫切需要解决的问题。
- (3) 理论依据不够充分。近几年推出的网络模型, 都是基于统计以及近似的数学推导进行研究与分析的。在对实际网络建模的分析上虽然已经取得了巨大的进步, 为使复杂网络模型应用于更多的工程项目, 就需要更加严谨的数学方法加以分析与阐述。
- (4) 实际应用不够深入。目前, 复杂网络在动力学上的应用才刚刚开始, 都是一些基于理想状态模型的研究。实际的网络要比现在已推出的模型更复杂, 更多变。如何使用复杂网络理论对实际网络系统进行全面的研究, 得出有利于网络设计与保护的结论, 有待学者更加深入的探索与研究。

1.4 论文主要工作

本文对复杂网络建模方法进行分析与研究, 根据实际网络的情况, 提出新的

模型，对新的模型进行分析及阐述。复杂网络已被大量应用于各行各业，本文分别在电力交易网络以及文本分析中基于复杂网络建模方法进行研究与应用。本文的主要完成一下工作：

- (1) 提出两种新的复杂网络建模方法。原有的无标度网络模型在建模过程中，加边规则只是单纯的在单位时间内增加固定的节点和连接边数，这不符合现实网络非固定时间间隔增长的实际特征。针对建模过程的加边规则，本文应用随机理论中的更新计数过程，提出两种新型的无标度网络模型。通过仿真实验分析网络的度分布以及用不同的方法计算度指数，证明所提出的模型是符合传统复杂网络的特性的。
- (2) 对现实世界中的电力交易问题建模。本文针对现实生活中的一个实际应用问题电力交易过程，本文运用复杂网络中的方法，对成功的交易的买卖双方建立起一个网络。并在所建立的模型的基础上，研究部分的复杂网络特性（同步、鲁棒性等），深入挖掘网络结构与功能之间的内在联系，为今后更进一步的实际应用提供有益的理论指导
- (3) 在自然语言分析中运用网络科学理论。本文通过对文本中词与词之间的关系进行复杂网络建模，对所构建的单词网络进行网络特性的分析，揭示文本单词网络也具有复杂网络的特性。并且提出一种新的关键词提取方法，不再是依据单词在文本中出现的次数，而是根据所得网络拓扑图中介数高的词作为关键词提取的依据进行关键词的提取。

1.5 论文组织结构

本文首先对网络模型进行创新，提出了两种新的网络模型，并且提出两种基于网络科学理论的应用。论文的组织结构如下：

第一章、绪论，本文首先对网络科学的研究发展历史，同时对其研究意义进行详细的阐述，然后概括了网络科学已经取得的主要成就，根据研究现状提出复杂网络中的所存在的挑战与机遇，最后总结出复杂网络所存在的不足以及整个论文的组织结构。

第二章、复杂网络相关知识，本文首先介绍研究网络科学的所要具备的基础知识，接着详细的介绍其发展历史中的几个最具现实意义的模型，为读者更容易的理解本文所撰写的内容打下一定的理论基础。

第三章、基于更新计数过程的复杂网络建模，本章首先提出构建新网络模型的理论依据，然后分别对提出的两种无标度模型——基于均匀分布和高斯分布的建模，进行构造方法的描述、理论分析及证明以及仿真实验的分析。最后得出本

文所提出的网络模型是具备现实意义以及符合无标度网络的基本特性的结论。

第四章、基于复杂网络的电力交易模型，本章首先介绍了双边电力交易的基本过程以及在交易过程中所受到的各种限制条件，然后对交易过程进行建模，对所构建的电力交易网络进行理论的分析，并且针对该网络模型进行仿真实验，对实验进行分析。最后总结出电力交易网络的特征，并且证实电力交易网络具有无标度网络特征的网络。

第五章、基于复杂网络的自然语言模型，本文首先对网络节点所具有的特性进行简明的阐述，根据文本单词之间的关系进行复杂网络的建模，然后对文本单词网络进行网络结构的拓扑分析，同时利用网络节点的特征进行文本关键词的提取。最后得出文本单词网络是一种无标网络，并且依据节点的介数进行关键词提取的方法具有重要的意义。

第六章、全文总结与展望，本章首先对全文进行全面的总结，然后指出本文所写内容中的不足点以及今后可能的研究方向。

第二章 复杂网络建模相关知识

近年来网络科学的研究越来越受到业界的关注，其中揭示网络结构复杂性及其与网络行为之间的关系成为其重要的课题之一，然而现实中存在着各种结构不一致的复杂网络，如何从这些结构不一致的复杂网络中提取它们的共性，首先要找到一种能准确描述这些各种不同结构网络的方法。经过专家、学者的研究发现使用数学上的图可以完美的描述各种结构的复杂网络^[16]，并且随着历史研究的发展，所得到的模型也越来越符合实际网络的特征了。

2.1 复杂网络的基本概念

2.1.1 网络的图表示

最早应用图理论对实际网络进行研究的人是 18 世纪伟大的数学家 Erdős，他对“Königsberg 七桥问题”的研究开启了图论的萌芽，并由此开创了数学中的一个新的分支——图论。

定义 2-1: 一个图 G 可以表示为由图中的有序对 (V, E) 组成, 记为 $G=(V, E)$, 其中

(1) V 是图中顶点的集合，集合中至少需要包含一个元素，用 $|V|$ 表示顶点数 ($|V| \geq 1$);

(2) E 是由图中的任意两个顶点以及它们之间的连线构成的边的集合，集合中可以没有元素。用 $|E|$ 表示边的数目 ($|E| \geq 0$)。

图中的边 e 记为 (i, j) ，也可记 $e=(i, j)$, 其中的 i 和 j 都可以称为边 e 的端点，节点之间的连接关系也称为相邻关系。而边与边的连接关系需要这两条边拥有一个相同的端点。如果两条边的两个端点都是相同的，那么把其中一条称为另一的重边。如果一条边的两个端点相同，那么把这样的边成为环。若把图中的顶点抽象为图形上的一个点，边抽象为一条线，那么可以将一个图用几何中的“图形”抽象表示出来。

将图论中图的定义应用于复杂网络中，则任意一个复杂网络可以描述为一个图的形式，图中的顶点可以代表复杂网络结构中的一个个体，可以代表一个人、一台机器、一个地点等等，并且图中边可以代表实际网络系统中个体之间的联系，例如：人与人的相识，机器与机器的通信，地理位置的相邻等。如果网络图中的边没有起点和终点的区别那么可以称之为无向网络，如果网络图中任意一条边在网络中的都具有相同的权重，那么可以称之为无权网络。

2.1.2 平均路径长度

在图论中把从节点 i 出发到达节点 j 所需要经过的图中最少边数称为两个节点之间的最短路径，也称为这两个节点之间的距离，表示为 d_{ij} 。取网络任意节点对距离中的最大值为网络的直径 D ，数学表达式为：

$$D = \max_{i,j} d_{ij} \quad (2-1)$$

取网络任意节点对之间距离的平均值为网络的平均路径长度，即表示为：

$$L = \frac{1}{\frac{1}{2}N(N+1)} \sum_{i \geq j} d_{ij} \quad (2-2)$$

其中 N 为网络的节点数。

2.1.3 聚类系数

各种实际网络中节点之间存在着间接的关系，例如人际关系网络中你的朋友可能是你另外一个朋友的朋友。以此类推，那么不同结构的网络中这种节点与节点之间的联系就有紧密的也有松散的区别，比如在电力交易网和人际关系网这两种网络中，节点关系的紧密程度就肯定不同。在网络科学中把这种代表一个节点与其他节点相关联的能力表示为该节点的聚类系数。

网络中任意节点 i 的聚类系数 C_i ，定义为：假定网络中节点 i 在网络中与该节点产生连接的节点个数为 m_i ，而网络中这 m_i 个邻居节点之间最多可能有 $m_i(m_i-1)/2$ 条边。那么这 m_i 个节点之间在网络中真实存在的边数 E_i 所占可能产生的最多边数 $m_i(m_i-1)/2$ 的比例即为该节点的聚类系数，其计算公式为

$$C_i = \frac{2E_i}{(k_i(k_i-1))} \quad (2-3)$$

用几何知识分析这个公式，该计算公式等价于：

$$C_i = \frac{\text{与点 } i \text{ 相连接的三角形数量}}{\text{与点 } i \text{ 相连接的三元组的数量}} \quad (2-4)$$

其中，三元组由节点 i 相邻的两个节点与其自身所构成。

2.1.4 度与度分布

网络科学研究中，节点的度是其拓扑特性中比不可少的特点。节点 i 的度等于网络图中边的端点含有节点 i 的数目，记为 k_i 。节点度可以在某种程度上作为其在网络中的重要程度的依据。统计网络中每个节点的度，对这些数据求平均值即为

网络节点平均度，用 $\langle k \rangle$ 表示。网络中每个节点度的分布情况可以抽象为分布函数 $P(k)$ 。 $P(k)$ 表示为随机选定的一个节点的度恰好为 k 的概率，也可以理解成度数为 k 的节点数量在网络所含节点总数中所占的比例大小。

2.1.5 幂律分布

研究表明，实际网络中的度较低的节点数量很多，在网络中所占比例很大，而度较高的节点数在网络中存在的数目非常少，在网络中所在比例很小。然而泊松分布表明的是度数较极端——很低或很高的节点数所占比例较低，大部分顶点的度都处于一个较中等的数。正因为实际的网络度分布与 Poisson 分布存在着巨大的差异，那么用 Poisson 分布拟合实际的网络是相当不合理的。针对现实网络中这种度分布特征，幂律^[17]分布较之 Poisson 指数分布，前者曲线没有中间的峰值，并且下滑的坡度要小的多，这种特性恰恰可以用来描述实际网络。那么可以用幂律形式 $P(k) \propto k^{-\gamma}$ 来更加合理的描述实际网络的度分布，其中的参数 γ 也称为网络幂指数。幂律分布通常也被称为无标度分布，正由于幂律分布具备的无标度特性，因此通常把具有幂律特性的网络称为无标度网络。通过大量的实验表明，许多现实网络系统例如：电影演员、论文引用、自制层 Internet、生物代谢等，都符合幂律分布的特点，并且其幂指数 γ 均介于 2 与 3 之间^[18]。

2.2 复杂网络的基本模型

研究实际网络的结构与其行为之间的关联，进而对网络规模的增加或减少节点间连线等的变化进行改进，首先需要在特别熟悉现实网络的构造前提下，然后再此基础上对其的结构进行模拟，最后才能得出符合实际特征的网络模型。以下将简要说明研究历史上出现的几个经典模型。

2.2.1 规则网络

首先介绍规则网络模型，其中有三种类型分别——全局耦合、最近邻耦合、星形。下面将对这三种一一介绍。

网络中所有节点中任意选取其中的两个，若每一对这样挑选出的节点都是彼此相邻的，那么把拥有这样性质的网络成为全局耦合网。通过对全局耦合网络进行拓扑特性分析，可得出其平均路径长度 L 等于 1，节点聚类系数的平均值 C 也等于 1。相对于其他含有相同个数节点的网络，该模型虽然同时显示出聚类系数较高，网络连通性很好的特点，但是仍然完全不符合现实世界中的网络，实际网络通常是较为稀疏的，没有全局耦合网络这么高的聚类性，节点的度也都没有如此

高。

根据实际网络通常较为稀疏，所含边数较少，那么考虑最近邻耦合网络，此模型的构建方法极为简单，不同于每个节点都要与其他节点相连接，只是需要与其附近的邻居节点产生连边。具体可以描述为：假设网络中共有 N 个节点，每个节点都与其自身左右两边各 K 个邻居产生连边。那么计算其网络的聚类系数：

$$C_{nc} = \frac{3(K-2)}{4(K-1)} \approx \frac{3}{4}$$

由此可见该模型的聚类系数较高。计算该网络的平均路径长度：

$$L_{nc} \approx \frac{N}{2K} \rightarrow \infty (N \rightarrow \infty)$$

该网络的平均路径长度 L 随着 N 的增大而增加，当 N 极大时 L 的值也会相当大，那么表明该模型不符合复杂网络中小世界特性。

现在考虑第三个常用的规则模型——星形耦合，该网络有一个特殊的顶点，所有其他节点只与该中心节点产生连边，并且其他节点之间没有任何连边。计算该模型的平均路径长度 L ：

$$L_{star} = 2 - \frac{2(N-1)}{N(N-1)} \rightarrow 2 (N \rightarrow \infty)$$

该模型的聚类系数 C ：

$$C_{star} = \frac{N-1}{N} \rightarrow 1 (N \rightarrow \infty)$$

从此可以看出该模型具备了小世界性质，但是依据节点聚类系数的计算方法，如果一个节点的度只为 1，那么其的聚类系数等于 0，由此该模型的所有节点的聚类系数都是 0，即整个网络的聚类系数为 0。那么该模型不具备无标度的特性。

2.2.2 E-R 随机图

随着复杂网络的发展，规则网络模型已经被证实不符合复杂网络模型的构建。那么由数学家欧拉提出的随机图，使网络科学取得了重大的突破，网络科学的研究迈入了全新的阶段，

随机图与规则网络模型完全不同，随机图中所有的节点与节点之间的连线都是随机生成的。随机图的生成规则是：假定网络图中共有 N 个节点，其中的任意一个与除自身之外所有其他节点相连的概率为 p 。那么以概率 p 在网络图中加边

为规则的随机图，总共能产生 $pN(N-1)/2$ 条边^[19]。

图 2-1 演示了 10 个节点的随机图随着时间的变化情况。

下面计算随机图的几个拓扑特性。由于随机图是按照概率 p ($p \in [0,1]$) 对任意点对之间连线。则以 p 的取值，分别进行计算。

若 $p=0$ ，那么表明，随机图中所有的点都是孤立节点，那么所有的节点的度都为 0，随机图的节点度的平均值 $\langle k \rangle = 0$ ，平均路径长度 $L=0$ ，聚类系数 $C=0$ ；

若 $0 < p < 1$ ，那么根据概率论，可得该模型中任意一个节点度的期望均为 $p(N-1)$ ，即该模型的节点平均度 $\langle k \rangle = p(N-1)$ 。

平均路径长度 $L \approx \frac{\ln(N)}{\ln(k)}$ ，聚类系数 $C = p$ ；

若 $p=1$ ，表明图中任意节点都跟所有其他的节点连，也即该随即图为一个完全图，那么节点度的平均值 $\langle k \rangle = N-1$ ，平均路径长度 $L=1$ ，聚类系数 $C=1$ ；

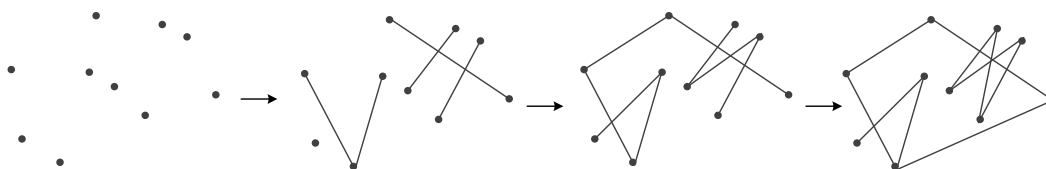


图 2-1 随机图的演化

最后，考虑随机图的度分布，因为图中每个点与其他点是否相连是独立的事件，那么可以根据 Poisson 分布来描述随机图节点的度分布，假定节点的个数 $N \rightarrow \infty$ 。那么可得：

$$p(k) = \left(\frac{N}{k} \right) p^k (1-p)^{N-k} \approx \frac{\langle k \rangle e^{-\langle k \rangle}}{k!} \quad (2-5)$$

其中 $\langle k \rangle$ 是节点度的平均值，从等式可以看出，该模型的度分布符合 Poisson 分布的特点。

虽然随机图的产生规则也不符合实际网络，但是对于后期复杂网络的研究产生了深远的意义，研究者们通过对随机图的扩展，可以变换出很多模型。直到现在许多复杂网络模型的提出，也缺少不了它的理论支持。

2.2.3 小世界网络模型

从“六度分理论”提出以来，大量的研究表明，实际的网络往往具备连通性强，节点间的关联紧密等特点。然而现实中存在的网络通常同时拥有一定的规则

性和某种程度的随机性的特点，如上文所分析的，不能单纯的用一种模型对实际网络结构进行模拟。根据实际网络的特点，在 1998 年，首次出现了同时考虑到把规则性和随机性相结合的模型，该模型被 Watts 和 Strogatz 提出，并且以他们两人名字首字母命名——WS 小世界网络。

WS 小世界模型的构造方法如下：

- (1) 初始网络：初始网络的构造方式如同最近邻耦合网络模型，假定网络总有 N 个节点，网络中任意一个节点都与其左边和右边各 $K/2$ 个的最近邻居节点连接。
- (2) 随机化重连：将第一步所构造的初始网络中的每一条连边，都以概率 p 重新连接，即保持所选中的一条边中的一个顶点，在网络中随机挑选一个其他的节点，以概率 p 将这两个节点连接起来。其中保证两个节点之间没有重边，并且网络中不能有自环。

根据以上的构造规则生成的 100 个节点网络，如图 2-2 所示。

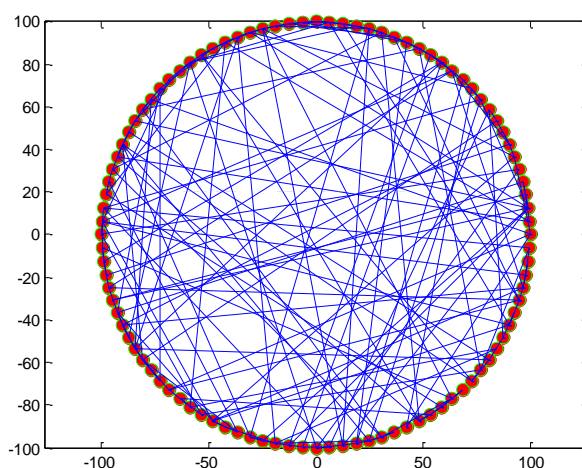


图 2-2 WS 网络模型

构造方法的第一步就保证了所构造的初始网络具有一定的规则性，第二步保证了某种程度上的随机性。用以上的方法构造出的网络整体是高的连通性的，且节点间的距离都比较短，关系都比较紧密。但是该模型也有其局限性，考虑令连接概率 $p=0$ ，那么所构造的网络为最近邻耦合网络，相反，若 $p=1$ 时，那么所构造的网络就成为随机图。就算是 p 介于 0 和 1 之间，重新连边的过程中很有可能会破坏整个网络的连通性。

针对 WS 的重新连边的局限性，二十世纪末 NW 小世界网络被 Newman 和 Watts 提出。该模型较之于上文介绍的 WS 模型，最大的不同点为网络中存在的边的总

数目会增加而不是不变，这保证了该模型始终具有连通性。

网络规模为 100 个节点的 NW 网络模型，如图 2-3 所示。

比较图 2-2 和图 2-3 可以观察到，在网络节点数相同的情况下，NW 较之 WS 网络，前者含有边的数目较多，且 NW 较为密集，这正是由于 NW 模型的随机加边的构造规则所造成的。

NW 小世界模型构造方法：

- (1) 初始网络：其模型的初始网络构造方法与 WS 的初始构造规则相同，假定网络总有 N 个节点，网络中任意一个节点都与其左边和右边各 $K/2$ 个的最近邻居节点连接。
- (2) 随机化加边：与随机化重连不同，对任意选取的两个节点，以概率 p 为这对点连边。

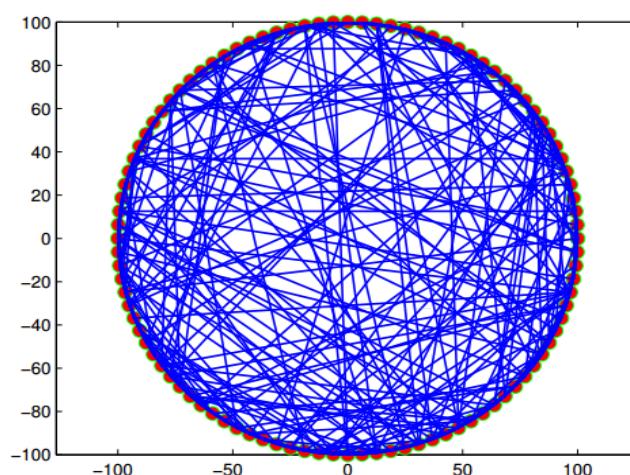


图 2-3 NW 网络模型

同样，该模型也有其自身的局限性，当 p 取值为 0 时，则是一个最近邻耦合网络，当 p 取值为 1 时，则是一个完全邻耦合网络。

在这两种即有规则网络特点又有随机特征的网络提出后，还出现了在此基础上的进行改进的模型^[20,21]。

2.2.4 BA 无标度网络模型

上文所介绍的网络模型都是在固定节点的个数下再对网络的边进行构建，而实际网络规模往往都是由小变大的过程，并且实际网络的节点往往与网络中度数高的点相连接的概率高，WS 与 NW 两个种网络均未考虑实际网络的这两个特性。为了满足实际网络所独有的性质，Barabási 和 Albert 在 1999 年首次提出了 BA 无

标度网络模型。

该模型的提出，突破了网络科学研究的瓶颈，使网络科学的研究再次受到外界的强烈关注^[22]。

BA 无标度网络的构造方法如下：

- (1) 增长：假定网络的初始规模为 n_0 ，这 n_0 个节点可以是一个完全图，网络规模随着时间的流逝而发生变化，每单位时间内在网络中新增一个节点，新生成的节点与 $n(n \leq n_0)$ 个网络中已有的顶点相连接。
- (2) 优先连接：新增加的节点与任意选取的网络中已有的节点 i ，按照概率 p_i 决定是否与之产生连线。按照该规则增加连边，直到新增加的节点的度为 n 为止。

$$p_i = \frac{k_i}{\sum_j k_j} \quad (2-6)$$

根据这种构造方法网络中节点度大的点更容易与新加的节点产生连边，而节点度小的点更难增加其节点度。从而导致，度小的节点占大多数，而度大的节点极少，这恰恰符合了实际网络中出现的“马太效应”。该模型的演化过程图如 2-4 所示：

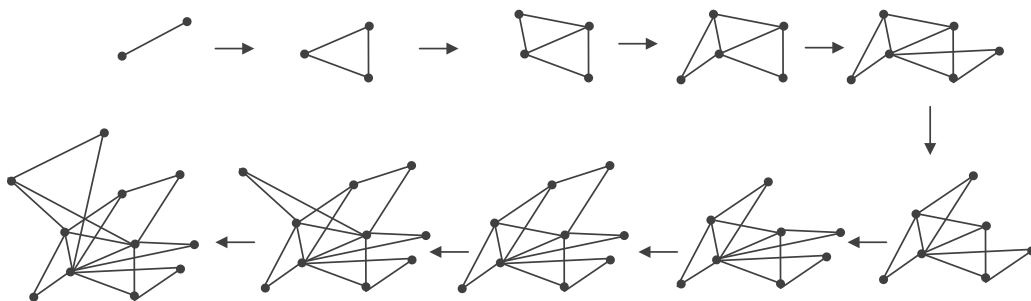


图 2-4 BA 模型的演化过程

其中网络的初始顶点数为 2，每单位时间增加一个顶点，每个顶点与网络中已存在的两个节点连边。

由 Barabási 验证，BA 无标度网络的度分布 p_k 服从幂律分布 $2m^2 k^{-3}$ ，也即 BA 模型的幂律指数为 3。

BA 网络模型如图 2-5 所示。

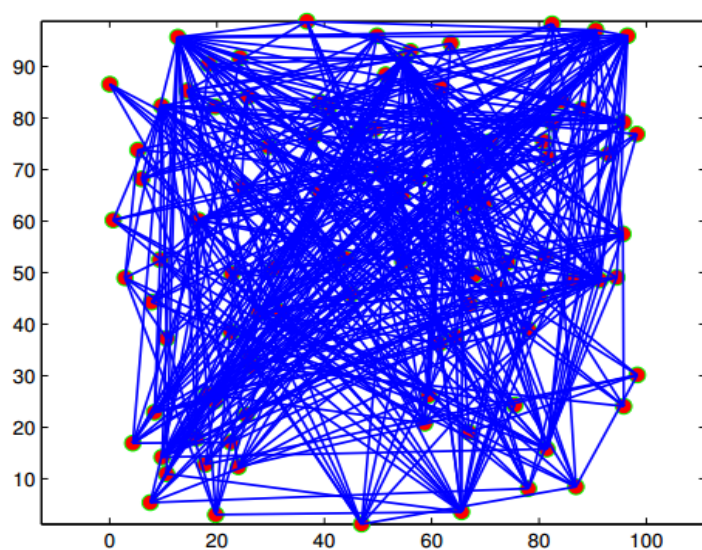


图 2-5 BA 无标度网络

2.3 本章小结

本文所涉及的复杂网络的基础知识主要包括图论的基础知识及复杂网络建模方法两部分。其中为证明本文所提出的两种新的网络模型是无标度网络，分别应用了平均场方法和主方程法，这两种方法会在下文证明时给出详细的介绍。通过这些复杂网络知识的介绍，为理解本文所提出的网络模型以及应用复杂网络理论对实际网络的研究做好了铺垫工作。

第三章 基于更新计数过程的复杂网络建模

BA 无标度网络是第一个考虑了实际网络增长特性的模型，而后提出的局域世界演化模型以及一系列相似的网络模型都是遵循单位时间内增加一个点的规则。虽然这种方法较之前的模型已取得了很大的进步，然而实际网络中网络节点的增加时间间隔并不一定是固定值。比如，因特网的里每个网站的增长时间间隔会随着科学发展的影响，早期落后的因特网增长所需时间相对于现代时间会更长。人口网的增长时间间隔也会受到外界间隔影响，战争和贫穷等因素会导致增长间隔变长。同时，也存在一些按照固定时间间隔增长的网络，例如，一个实验室的社交网，每年会增加固定的学生数，很明显，该网络的生长的时间间隔固定为一年。上述例子说明，现实中很多网络保持增长的过程中，时间间隔却是可变化的，我们可以采用更新计数过程来模拟这种时间间隔。

除此之外，很多无标度网络在增加一个点时便固定新增 m 条边数，而现实情况是每个新加入的节点，新增加的边数是不同的。例如，人际交流网络，每个人在加入一个新的社团，能交到的朋友数是不固定的；如，在文献引用网络中，新发表的文章的引用率也是不同的。由此可以看出，新增节点所带来的新边数应该是在一定的范围之内。假设新增节点都是正常、健康的，可以用均匀分布产生的离散值表示每个新节点的新增加的边数。

复杂网络的建模，除了网络的增加规则之外，初始网络的设定也起着至关重要的作用，BA 无标度模型对初始网络没有明确的设定^[23]。考虑现实网络，初始时都是由随机的小组织构成的^[24]。例如，早期交通不发达的年代，人口规模不大，而交通也并不发达，所以交流的可能性比较小。综上可知，网络初始阶段是具备低聚类，平均路径长等特征，而随机图恰恰满足了这些要求，因此，本文使用小规模 ER 随机图作为建模的初始网络。

3.1 基于均匀更新计数过程的无标度网络

针对上文提出的实际网络节点增加时间间隔是不固定的，本文首先提出基于均匀更新计数过程的无标度网络模型。该模型的网络节点增加时间间隔和每个新增节点在网络中连接的边数是服从均匀分布。下面将详细介绍此网络模型的构建方法。

3.1.1 网络演化算法

具体构造方法如下：

算法名称：均匀更新计数网络构建规则

1. 初始化：生成 N 个节点，以概率 p 在任意一对节点间添加一条边，要求两两节点之间不能有重边，也不能有自环。
 2. 增长：按时间间隔 T_i ($i=1,2, \dots$) 增加节点 i ，每个间隔只增加一个节点，每个新增节点与 m_i 个网络中已有的节点相连，即新增 m_i 条边，其中 T_i 和 m_i 服从均匀分布。
 3. 优先连接：每个新增的节点与已经生成了的任意一个节点相连接的概率为 Π_i 。
 4. 结束：当时间到达 $T=T_1+T_2+\dots+T_N$ 时，输出所构造的网络。
-

在初始化阶段时，可以通过设定初始 E-R 随机网络，通过 N 来调节网络大小， p 来调节聚集程度。

新产生的点与网络中已有的顶点 i 的产生连边的概率为 Π_i ，表达式如下：

$$\Pi(k_i) = \frac{k_i}{\sum_j k_j} \quad (3-1)$$

其中， k_i 表示节点 i 的度，用轮盘赌算法选择现存节点与新节点连边， T 作为终止时间。连接概率的表达式通常也用来评判一个节点的适应程度，该计算公式表明适应度与节点度的值成正比。

上文所提出的网络模型最关键的部分在于网络的增长时间间隔 T_i 服从更新计数过程，其具体定义如下：

定义 3-1：若每次新产生的节点 $N(t)$ ($t \geq 0$) 满足下列条件：

- (1) t_1, t_2, \dots 是新产生节点 $N(t_1), N(t_2), \dots$ 达到的时间；
- (2) $T_i = t_i - t_{i-1}$ ；
- (3) 时间间隔序列 $\{T_1, T_2, \dots\}$ 服从独立同分布。

那么节点增加的时间间隔 $N(t)$ 就被称为更新计数过程，而 T_i 则被称为更新间距。

根据以上定义，因为每个节点生成的时间，只与当下有关，与之前后都无关，

故满足拥有独立条件；在本小节中，时间间隔序列都服从均匀分布，那么很明显该模型的节点增加 $N(t)$ 是符合更新计数过程。

如上节提出的，节点增加的时间间隔和每次新增节点的边均服从离散均匀分布，而均匀分布的定义如下：

定义 3-2： 每个新增节点 i ，新增边数 m_i 在区间 $[a, a+1, \dots, b]$ 上，如果其概率满足：

$$P(m_i = x) = \frac{1}{b-a}, x \in [a, a+1, \dots, b] \quad (3-2)$$

则称新增边数满足离散的均匀分布。

下文将详细阐述基于均匀更新计数过程的无标度网络。

3.1.2 网络演化实例

在本节中，给出一个具体的基于均匀更新计数过程的无标度网络以详细阐述网络的演化过程。

根据上节提出的建模方法，为了能清晰描述，初始化所需要的各个参数设定如表 3-1 所列：

表 3-1 参数演示

Parameter	N	p	$[a, b]$	$[c, d]$	T
Value	4	0.3	$[1, 3]$	$[1, 2]$	12

说明： N 代表初始网络节点数， p 是互相连接的概率， $[a, b]$ 表示更新计数过程中的时间间隔， $[c, d]$ 是新增节点的增加边数的范围， T 代表终止时间。



图 3-1 初始网络

假定初始网络有 4 个节点，任意两个节点之间有 0.3 的概率相连。最后结果如图 3-1 所示。

根据上述所得的初始网络，进行进一步的演化，以 $[1, 3]$ 为时间间隔增加节点，每个新增节点连接到一个或两个已存在的节点上，在最终成为含有 10 个节点的网络。图 3-2 演示了网络的增长、连接和终止的过程。

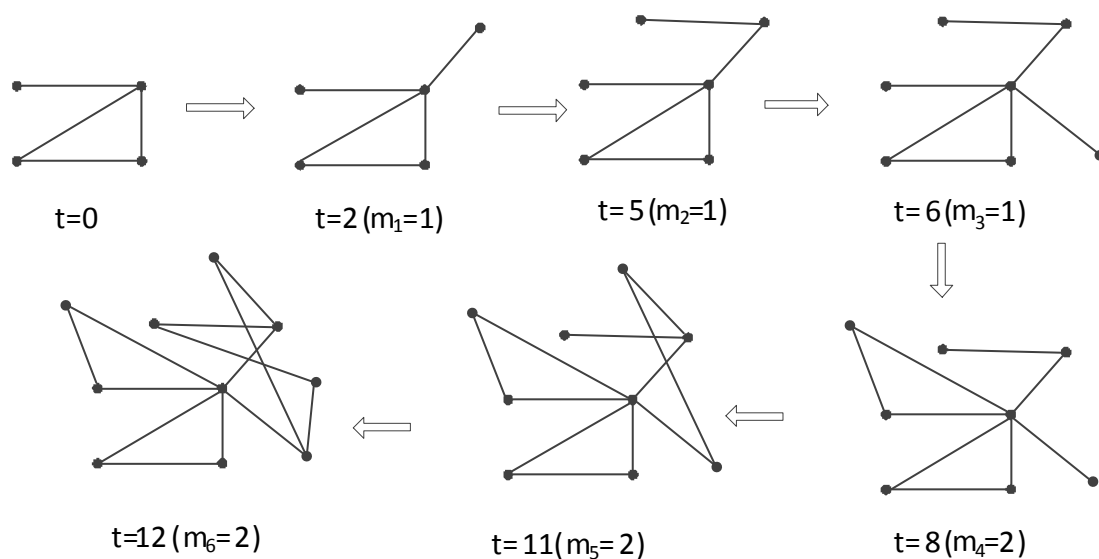


图 3-2 网络增长过程

由图 3-2 分析可知，每个节点的到达时间分别为 $t_1=2$, $t_2=5$, $t_3=6$, $t_4=8$, $t_5=11$, $t_6=12$ ，时间间隔为 $\{3, 1, 2, 3, 1\}$ 服从 $[1, 3]$ 内的离散均匀分布。而每个新节点增加的边数是 $\{1, 1, 1, 2, 2, 2\}$ ，服从 $[1, 2]$ 范围的离散均匀分布。最终达到所设定的时长 $T=12$ 时，演化结束。通过分析图的拓扑结构，可以算出，图中每个节点的度为 $\{2, 2, 2, 2, 2, 2, 2, 3, 3, 6\}$ ，该图的度分布如表 3-2。

表 3-2 度分布

Degree	2	3	6
Number	7	2	1

可更直观的为表示为直方图 3-3 的形式。

由 3-2 和 3-3 图可以很直观的看出，度低于 3 的节点占多数，度高于 3 的节点只有 1 个，从网络的度分布形状看，近似于长尾分布，这都说明了该网络符合幂律分布“马太效应”特点。根据公式(2-3)，要计算整个网络图的聚类系数 C ，首先要求得所有节点的 C 的值 $\{1, 1, 1, 1, 2/15, 0, 0, 0, 0, 0\}$ ，然后这个集合的平均值即可。由此，可以求得图的聚类系数为：

$$\bar{C} = \frac{\frac{2}{15} + 1 + 1 + 1 + 1 + 0 + 0 + 0 + 0 + 0}{10} \approx 0.4133 \quad (3-3)$$

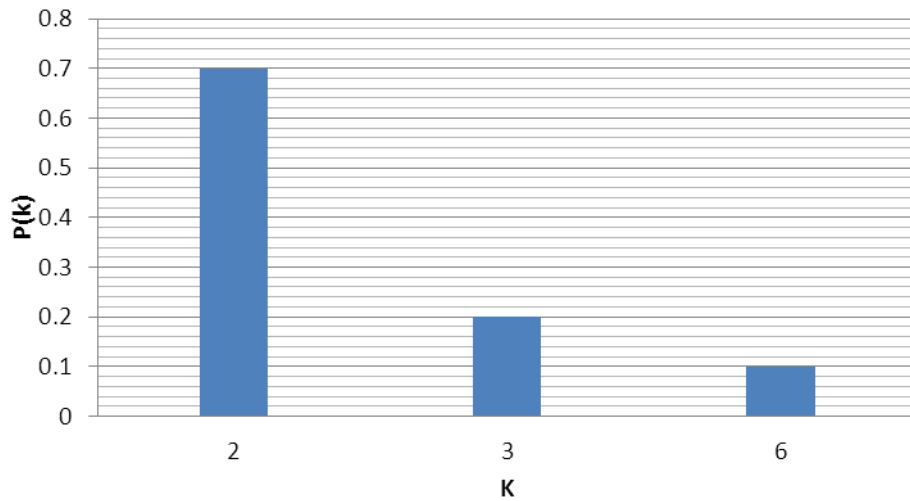


图 3-3 网络度分布

3.1.3 理论分析

在本小节中，我们的目标是计算出本文所提出的新网络模型的度指数，从而证明该网络是一个与时间无关的无标度网络。为了得出该结论，本文主要运用基于物理学的平均场方法。

当 $t \rightarrow \infty$ 时，初始的小网络对今后网络发展的度分布的影响不大，因此，在后面的分析中，均忽略初始网络。

平均场方法是通过物理学中的动力学方程式来求得网络在稳定状态下的分布。下面介绍使用该方法推导的步骤。

假设用 $k_i(t)$ 是在 t 时刻节点 i 在此刻的度，那么此时网络的平均总度数是：

$$\sum_j k_j = \frac{2(c+d)t}{a+b} \quad (3-4)$$

其中， a, b 为时间间隔的上下界，而 c, d 是新增边数的上下界。

每当有一个新产生的节点，便有 $(c+d)/2$ 条新增的边，那么在选取网络已有的顶点 i 进行连边的过程中，节点 i 至少有一次被选中的几率为：

$$P = \binom{(c+d)/2}{1} [\Pi(k_i)] [1 - \Pi(k_i)]^{\frac{(c+d)}{2}-1} \approx \frac{c+d}{2} \Pi(k_i) = \frac{(a+b)k_i}{4t} \quad (3-5)$$

其中 $\Pi(k_i)$ 是根据公式(3-1)得来。

那么平均一瞬间产生一个度的概率为 $2P/(a+b)$ ，由连续性理论，即把 $k_i(t)$ 看

作连续动力学函数，近似满足如下动力学方程：

$$\begin{cases} \frac{\partial k_i(t)}{\partial t} \approx \frac{2P}{a+b} = \frac{k_i}{2t} \\ k_i \left(\frac{i(a+b)}{2} \right) = \frac{c+d}{2} \end{cases} \quad (3-6)$$

求解方程得：

$$k_i(t) = \frac{(c+d)}{2} \left(\frac{2t}{(a+b)i} \right)^{\frac{1}{2}} \quad (3-7)$$

其中的 $1/2$ 是动力学指数。

在计算网络度分布时，节点是从网络已存在的 i 个节点中随机地选择的，即节点的选取服从均匀分布，所以 $p(i) = (a+b)/2t$ 。结合动力学方程的解，网络瞬时度分布推导如下：

$$P\{k_i(t) < k\} = P\left\{i > \frac{2t}{a+b} \left(\frac{c+d}{2k} \right)^2\right\} = 1 - \left(\frac{c+d}{2k} \right)^2 \quad (3-8)$$

对 k 求偏导得：

$$P_k(t) = \frac{\partial P\{k_i(t) < k\}}{\partial k} = \frac{(c+d)^2}{2} k^{-3} \quad (3-9)$$

由上式进一步可以推算出网络的稳态分布：

$$P(k) = \lim_{t \rightarrow \infty} P_k(t) = \frac{(c+d)^2}{2} k^{-\gamma} \quad (3-10)$$

其中 $\gamma=3$ ，可见，基于均匀更新过程程的无标度模型的稳态分布是独立于时间的。

通过以上的推导可以得出结论，若时间间隔是服从独立的离散均匀分布，那么更新过程程的无标度网络是符合 Barabási 提出的无标度网络中独立于时间的特性的。

除了均匀的时间间隔，更常见的是正常分布的时间间隔，也就是长时间不增长的情况和短时间就增长的情况都是比较罕见，间隔一般在一个平均范围之内。比如互联网、经济网等的增长。下面我们就给出基于高斯分布更新计数过程的网络模型。

3.2 基于高斯更新计数过程的无标度网络

上节提出的网络模型增长所需的时间间隔是服从均匀分布，而实际中很多复杂网络的增长时间间隔都在不断的变化。在不同的时期，增长间隔也是不同的，例如互连网，增长间隔在信息传递高度发达的现代肯定比上个世纪低很多，但总体是趋于平均的。如果我们这些现象看作更新计数过程，那么在现实复杂网络中出现的这种现象恰恰符合随机过程中的高斯分布。

故在本小节中将提出一种新的扩展网络模型——基于高斯更新计数过程的无标度网络。

3.2.1 网络演化算法

该模型的构造方法很接近与前面的更新计数过程的无标度网络的构造方法，也是分为四个步骤：

算法名称：高斯更新计数过程网络构建规则

1. 初始化：生成 N 个节点，以概率 p 在任意一对节点间添加一条边，要求两两节点之间不能有重边，也不能有自环。
 2. 增长：按时间间隔 T_i ($i=1,2, \dots$) 增加节点 i ，每个间隔只增加一个节点，每个新增节点与 m_i 个网络中已有的节点相连，即新增 m_i 条边，其中 T_i 服从高斯分布， m_i 服从均匀分布。
 3. 优先连接：每个新增的节点与已经生成了的任意一个节点相连接的概率为 Π_i 。
 4. 结束：当时间到达 T 时，输出所构造的网络。
-

该过程符合前面的定义 3-1，唯一之前提出的模型构造方法的不同之处在于，该更新计数过程中的增长时间间隔符合更有普遍性的高斯分布。

接下来阐述高斯分布的定义：

定义 3-3：新增节点所需时间间隔 T_i 是遵循参数为 μ 和 σ 的高斯分布，数学表达式为：

$$f(T_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, 0 < x < +\infty \quad (3-11)$$

其中 T_i 必须是一个整数。

这里的高斯分布需要离散化，因为我们的时间间隔必须是整数，其分布函数近似于离散的二项分布或者泊松分布。确定该函数最重要的参数是期望 μ 与方差

σ^2 ，下面我们详细演示基于高斯计数过程的无标度网络的演化过程。

3.2.2 网络演化实例

该模型的初始网络如上节的图 3-1 所示，增长时间间隔服从期望为 3，方差为 1 的高斯分布，而新增边数设为服从[1, 3]之间的均匀分布，同时令网络终止时间 $T = 20$ ，该过程如图 3-4。

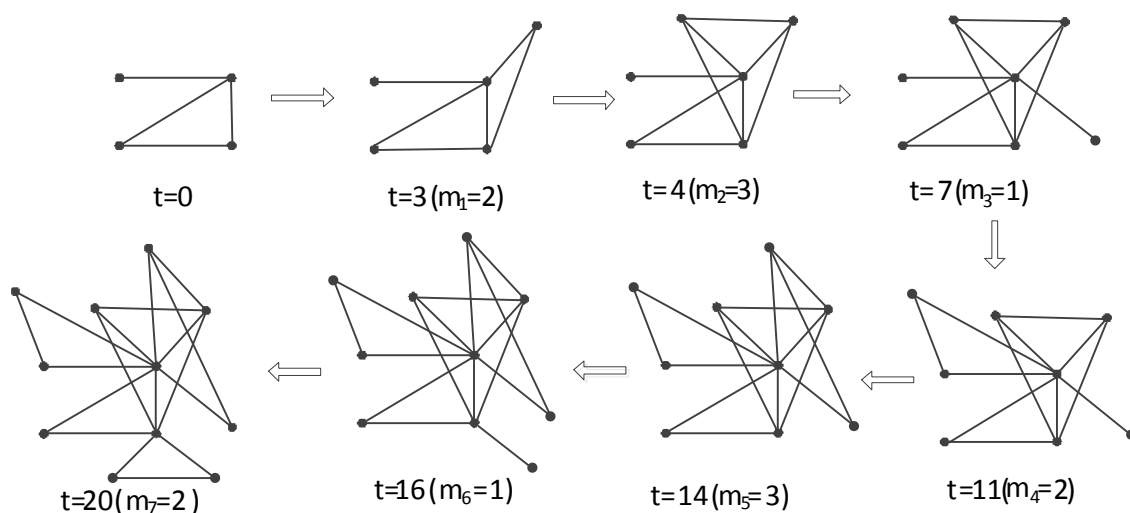


图 3-4 网络增长过程

从图中我们可以看出，时间间隔分别为{3, 1, 3, 4, 3, 2, 4}，符合期望为 3，方差为 1 的高斯分布；而新增节点连接边数分别为{2, 3, 1, 2, 3, 1, 2}，服从[1, 3]之间的均匀分布。该网络的度分布如表 3-3。

表 3-3 度分布

Degree	2	3	4	6	7
Number	6	2	2	1	1

直方图表示如图 3-5。

从图 3-5 中可以看出，网络中度小于 5 的节点占 83.33%，而度大于 5 以上的节点只有 16.67%，满足无标度网络中幂律分布特性。节点的聚类系数分别为{1, 1, 1, 1, 1, 1/3, 2/3, 1/3, 1/3}，由此可得平均聚类系数为：

$$C = \frac{1+1+1+1+1+1+\frac{1}{3}+\frac{1}{3}+\frac{1}{3}+\frac{2}{3}}{11} \approx 0.7879 \quad (3-12)$$

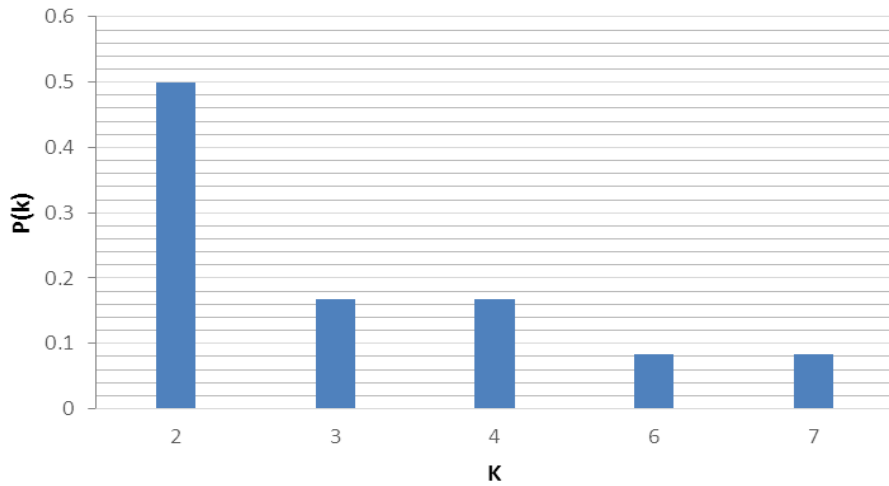


图 3-5 度分布

3.2.3 理论分析

本节通过平均场方法求网络度指数，忽略初始网络，令每个单位时间新增网络边数的期望为 $(\alpha + \beta)/2$ 的均匀分布，时间间隔服从期望为 μ 方差为 σ 的高斯分布。

令 $k_i(t)$ 代表节点 i 在 t 时刻的度，那么在该时刻网络拥有的平均总度数为：

$$\sum_j k_j \approx (\alpha + \beta) \frac{t}{\mu} \quad (3-13)$$

网络中已经产生的节点 i 被新增加的节点选中，增加一条连边的概率为：

$$P = \binom{(\alpha + \beta)/2}{1} [\Pi(k_i)] [1 - \Pi(k_i)]^{\frac{(\alpha + \beta)}{2} - 1} \approx \frac{\alpha + \beta}{2} \Pi(k_i) = \frac{\mu k_i}{2t} \quad (3-14)$$

那么平均一瞬间产生一个度的概率为 P/μ ，根据连续动力学方程理论，节点 i 的度满足以下方程：

$$\begin{cases} \frac{\partial k_i(t)}{\partial t} \approx \frac{k_i}{2t} \\ k_i(\mu i) = \frac{\alpha + \beta}{2} \end{cases} \quad (3-15)$$

上式方程的通解是

$$k_i(t) = t^{\frac{1}{2}} C \quad (3-16)$$

结合边界条件，最后得出的解为：

$$k_i(t) = \frac{\alpha + \beta}{2} \left(\frac{t}{\mu i} \right)^{\frac{1}{2}} \quad (3-17)$$

因为节点 i 是从所有的 μ 个顶点中随机选出的，则节点 i 在时间 t 内服从均匀分布。由高斯分布的性质得： $p(i) = \mu/t$ ，进一步推出网络瞬时度分布为：

$$P\{k_i(t) < k\} = P\left\{i > \frac{t}{\mu} \left(\frac{\alpha + \beta}{2k} \right)^2\right\} = 1 - \left(\frac{\alpha + \beta}{2k} \right)^2 \quad (3-18)$$

对 k 求偏导：

$$P_k(t) = \frac{\partial P\{k_i(t) < k\}}{\partial k} = \frac{(\alpha + \beta)^2}{2} k^{-3} \quad (3-19)$$

由上式可得网络稳态分布：

$$P(k) = \lim_{t \rightarrow \infty} P_k(t) = \frac{(\alpha + \beta)^2}{2} k^{-3} \quad (3-20)$$

其中 $\gamma = 3$ 为网络度指数，表明网络是独立于时间。

通过以上讨论我们可以看出，不论增长时间间隔符合均匀分布还是更一般的高斯分布，只要节点增长率是连续的，增长机制确定，即网络度分布始终会到达一个稳态。我们可以进一步推断，即使时间间隔服从比高斯更复杂的分布，基于更新计数过程的网络模型依然是服从幂律分布的无标度网络。

3.3 仿真实验与分析

本节利用 Matlab 对本章提出的两种模型进行模拟建模，重点在于建模过程及模型的拓扑分析，并通过模拟对实际网络的蓄意攻击或随机攻击，实现对网络模型鲁棒性的分析。

3.3.1 网络建模过程

分析网络模型的基础在于网络模型的建立，而建立模型的关键在于随机过程的实现。下面将详细描述模拟更新计数过程的方法。根据上文所介绍的算法，两种更新计数过程建网流程图如 3-6 所示：

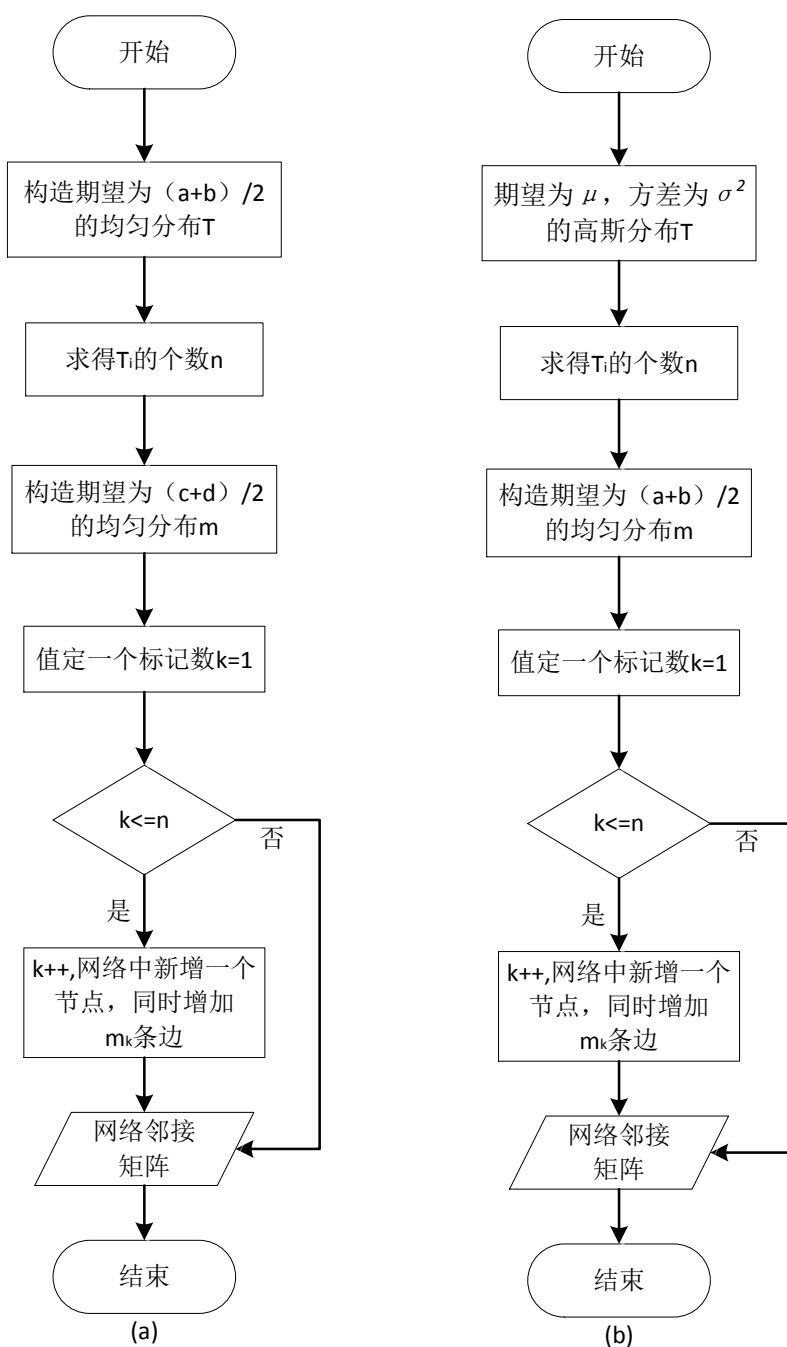


图 3-6 更新计数过程建网流程图

(a)基于均匀更新过程的建网流程图；(b)基于高斯更新过程的建网流程图

模拟基于均匀更新计数过程的无标度网络的难点在于如何模拟时间间隔服从均匀分布的时间序列。本文首先构造一个期望为 $(a+b)/2$ 的均匀分布的随机序列，然后构造服从期望为 $(c+d)/2$ 的均匀分布随机序列，具体步骤如下：

(1) 输入 a 、 b 和终止时间 T_{\max} ；

- (2) 通过 `unifrnd(a, b, 1, 1)` 产生一个 $[a, b]$ 均匀分布的随机数 t ;
- (3) 若累计时间 $T_i \leq T_{\max}$, 那么使 $T_i = T_i + t$, 否则输出这个时间序列。该时间序列的时间间隔服从期望为 $(a+b)/2$ 的均匀分布, 那么该时间序列符合均匀分布更新计数过程。
- (4) 计算这个时间序列 T_i 共含有多少个元素, 即为该网络模型共含有的网络节点数 n 。
- (5) 输入 c, d , 通过 `unifrnd(c, d, 1, n)` 构造一个期望为 $(c+d)/2$ 的均匀分布, 其作为每个新加入的节点所增加的边数序列。

通过以上步骤, 可以模拟出节点的增加服从均匀分布更新过程, 新增边数服从均匀分布的网络模型, 即基于均匀分布更新计数过程的网络模型。

基于高斯分布更新计数过程的无标度网络重点则在于如何生成时间间隔服从高斯分布的时间序列。首先构造出期望为 μ , 方差为 σ^2 的高斯分布的随机序列, 然后构造服从期望为 $(a+b)/2$ 的均匀分布随机序列具体步骤如下:

- (1) 输入 μ, σ^2 和终止时间 T_{\max} ;
- (2) 通过 `normrnd($\mu, \sigma, 1, 1$)` 产生一个服从期望为 μ , 方差为 σ^2 的高斯分布的随机数 t ;
- (3) 若 t 是整数, 则保留, 否则重新选取;
- (4) 若累计时间 $T_i \leq T_{\max}$, 那么使 $T_i = T_i + t$, 否则输出这个时间序列。该时间序列间隔服从期望为 μ , 方差为 σ^2 的高斯分布, 那么该时间序列符合高斯分布更新计数过程。
- (5) 计算这个时间序列 T_i 共含有多少个元素, 即为该网络模型共含有的网络节点数 n 。
- (6) 输入 a, b , 通过 `unifrnd(a, b, 1, n)` 构造一个期望为 $(c+d)/2$ 的均匀分布, 其作为每个新加入的节点所增加的边数序列。

通过以上步骤, 可以模拟出节点的增加服从高斯分布更新计数过程而新增边服从均匀分布的网络模型, 即基于高斯分布更新计数过程的网络模型。

根据时间序列得到的累计节点数以及新增节点所对应增加的网络边数, 可以分别构建这两个模型的网络。两个模型均以节点个数为 10 的 ER 随机网络作为最初始的构建图形, 其中任意两个节点以 0.2 作概率加边, 从而构建初始网络。基于均匀更新计数过程的模型, 设定其期望 $(a+b)/2=0.25$, 终止时间设为 50。图 3-7(a) 为根据该模型的构建方法得出的网络。

基于高斯更新计数过程的模型, 设其期望为 0.5, 方差为 0.25。以下为这个模型所构建的网络, 如图 3-7(b)所示。

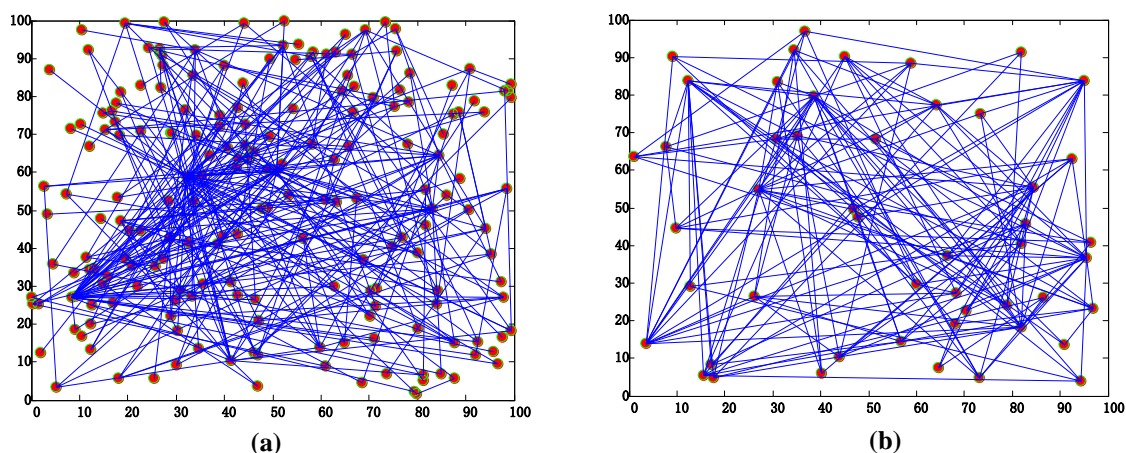


图 3-7 基于更新计数过程的网络

(a)基于均匀更新计数过程的网络；(b)基于高斯更新计数过程的网络

3.3.2 网络度分布

节点的度是网络中简单但极其重要的特性，网络的度分布对于整个网络和其拓扑而言又是至关重要的性质。本节主要使用统计学方法分析网络度分布以及从实验数据中得到度指数。

符合幂律分布的网络中，度较低的节点所占比例较高，少数节点的度比较高，把那些度数高的节点称作为网络的“hub”。幂律分布函数用来表示网络度分布，如上式所表达的含义，可以看出 $f(x)$ 与 x 成反比。

在本节中，通过模拟两种模型在不同参数下的网络，可以统计网络的度分布。如图 3-8 所示。

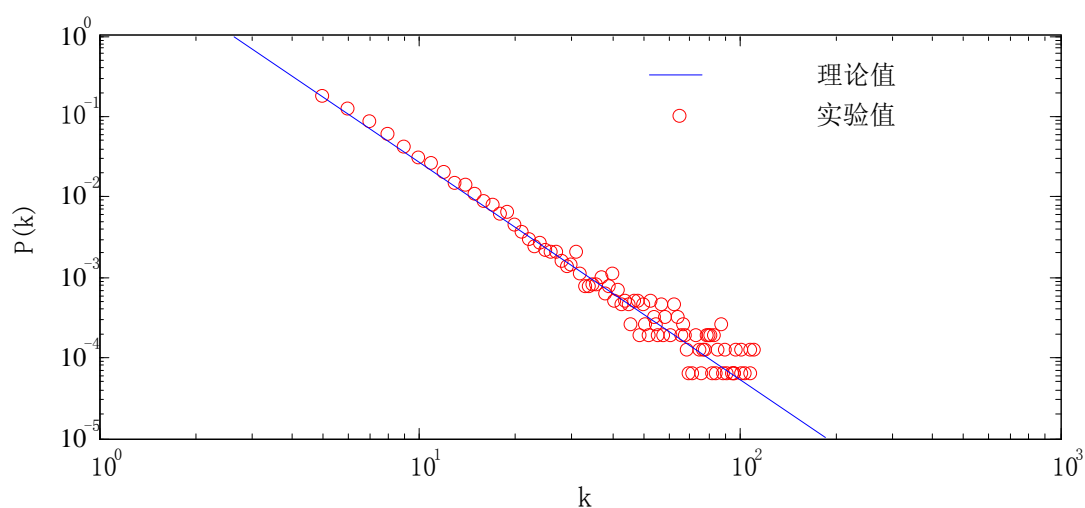


图 3-8 基于高斯更新计数过程的网络度分布

图 3-8 中小空心圆点表示网络的度分布， x 轴表示节点度数， y 轴表示图中点对应 x 轴上的度数所占总节点数的比例。其中的直线表示拟合所有图中的散列空心圆点后得到的直线。然后现实中的网络构建过程必然要比简单的幂律分布复杂的多，故在图中，能看到很多在直线外的点，这属于正常的误差范围。目前研究中，通常有对数盒子法^[25,26]，最大似然估计^[27]等方法进行网络度分布和度指数的统计。

通过分析图可知，高度数的节点肯定是偏离度分布斜率的点，由于实验数据的统计偏差，长尾分布中接近尾端的值也通常是有偏差的。因此，网络中频数小于等于 1 的度一定是噪音。为了消除误差，给数据设定一个阈值，然后从实验数据中删除若干个噪声数据。除了这些高度数节点外，由于网络模型构建中每个节点新增的边数符合均匀分布，那么低于期望值的度是会造成误差，故这些度也应该作为误差值删除。

消除误差后，接下来对图中的点进行拟合得到一条直线，通过对数坐标法进行拟合，方程可改写为：

$$k \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = b \quad (3-21)$$

其中 $k = (\log x, 1)$, $x_1 = -\gamma$, $x_2 = \log \beta$, $b = \log y$ 。

因此，一旦有了实验数据 $\log x, \log y$ ，把数据带入等式 (3-21) 中，可以得到 $\bar{x} = (x_1, x_2)$ 的值。有了 \bar{x} 的值后，运用向量的欧几里得范数求得真正的 x ，这种方法通常被称为最小 2 乘法或方差和方法。

为了拟合出最贴合实验数据的直线，需要得到其最小的方差，据此，可以画出网络度分布的直线。

根据以上介绍的方法，在双对数坐标下绘制两种模型的度分布，并求得拟合出的直线斜率，该斜率也称为度指数。在仿真实验中，影响建模的参数有三个，分别是节点增长时间间隔、边增长数目以及终止时间。具体来说，均匀更新计数过程网络所需设定的参数分别是 α （即时间间隔的期望值 $(a+b)/2$ ）， β （即加边数的期望值 $(c+d)/2$ ），以及 t 。高斯更新计数过程网络所需设定的参数分别是 μ （即时间间隔期望，这里将方差 σ^2 统一设定为 0.25）， γ （即加边数的期望值 $(c+d)/2$ ）以及 t 。在仿真实验中，固定这三个参数中的两个，改变另外一个参数，观察各个参数对度指数的影响。表格 3-4 以及表格 3-5 分别展示了两种网络类型具体参数的变化以及不同参数对应的度指数。

分析两个表 3-4 与 3-5 数据可得，不论三个参数怎么变化，两个模型的度指数始终是在 2.7—3.0 这个范围内的，可见这两个模型的度指数与这三个参数无关。

更确切的说，虽然度指数有一定的浮动，但是变化过程与参数的变化并没有直接的联系，可以推断人为的消除噪声是导致度指数变化的主要原因。

表 3-4 基于均匀分布更新计数过程网络度指数与参数的关系

α	β	t	γ	α	β	t	γ	α	β	t	γ
0.25	10	1000	2.9	0.25	10	1000	2.9	0.25	10	1000	2.7
0.25	10	2000	2.8	0.25	20	1000	2.8	0.20	10	1000	2.8
0.25	10	3000	2.9	0.25	30	1000	2.8	0.15	10	1000	2.8
0.25	10	4000	2.9	0.25	40	1000	2.7	0.10	10	1000	2.9
0.25	10	5000	3.0	0.25	50	1000	2.8	0.05	10	1000	2.9

表 3-5 基于高斯分布更新计数过程网络度指数与参数的关系

μ	γ	t	γ	μ	γ	t	γ	μ	γ	t	γ
0.25	10	1000	2.9	0.25	10	1000	3.0	0.25	10	1000	2.8
0.25	10	2000	2.7	0.25	20	1000	2.8	0.20	10	1000	3.0
0.25	10	3000	2.8	0.25	30	1000	3.0	0.15	10	1000	2.9
0.25	10	4000	2.8	0.25	40	1000	2.8	0.10	10	1000	2.9
0.25	10	5000	3.0	0.25	50	1000	2.9	0.05	10	1000	3.0

同样，在模拟构建基于这些参数的网络过程中，记录了不同参数下网络的度分布。在图 3-9 中应用三种不同的颜色代表一个参数取不同值时的网络度分布以及度指数。其中，圆圈代表实验数据， x 轴表示节点度， y 轴表示对应的度数在网络中出现的概率，直线代表拟合所得的图形，其斜率代表度指数（为了便于清晰观察，图中把拟合的直线，平行转移在图的空白处）。如图 3-9 以及图 3-10 所示。

从图 3-9 中可以看出，无论 α ， β 或是 t 怎么变化，所构建的网络模型都遵从幂律分布，具有典型的无标度网络的特征。在 β 值相同， α 和 t 不同的情况下，基于相同模型下的网络，随着这两个参数的增大，网络节点数相应增加，显示出相似的结果。再者，在 α 和 t 相同，而 β 值不同的情况下，网络度分布与 β 的值成正比，随着 β 的增加其曲线变宽。这是因为 β 决定了网络每次增加节点时新增边的数量，也就是说 β 值越大导致节点的度越大，从而扩大了网络中最大和最小的度差距，这使得网络越平缓。

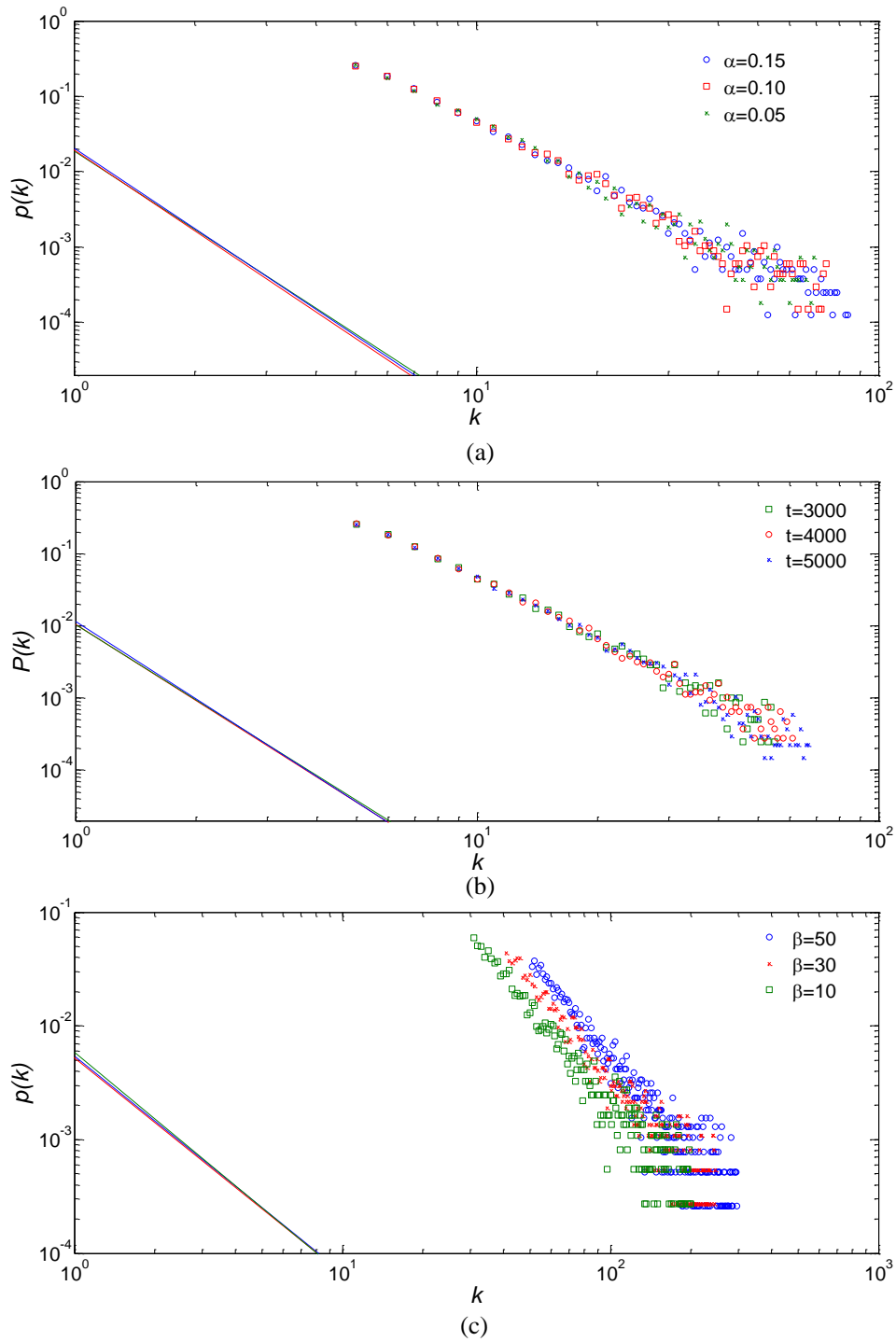


图 3-9 基于均匀更新计数过程的网络参数变化对比图

(a) 参数 α 变化; (b) 参数 t 变化; (c) 参数 β 变化

综上所述，基于两种模型构建的网络，在消除实验数据中的头尾噪音后，均服从指数在 2.7-3.0 的幂律分布。

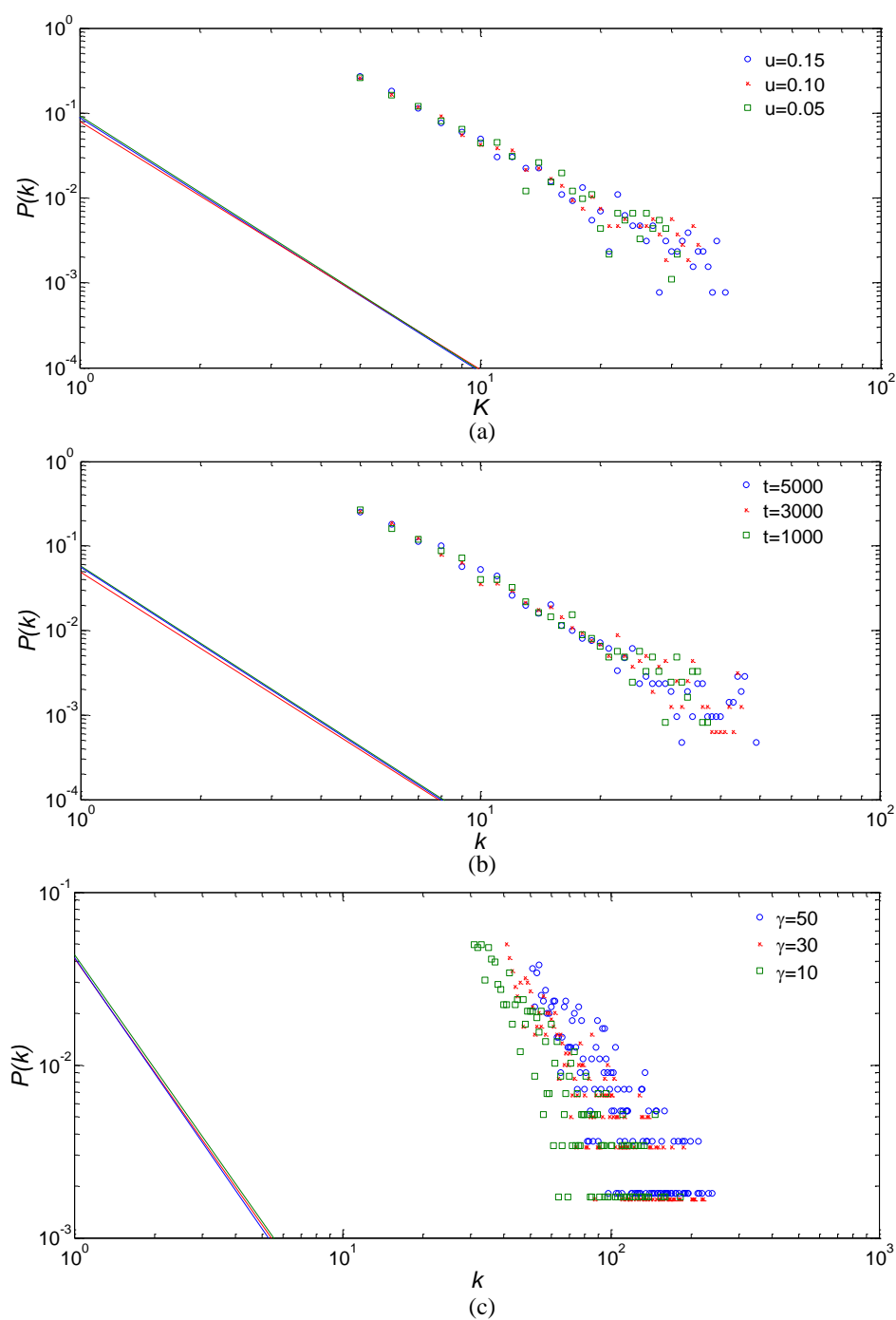


图 3-10 基于高斯更新计数过程的网络参数变化对比图

(a)参数 μ 变化; (b)参数 t 变化; (c)参数 γ 变化

3.3.3 网络可靠性分析

复杂网络的可靠性表现在当其受到自然灾害和人为攻击时, 保持健康、稳定

的能力。网络可靠性一般通过网络任意节点到其他节点的距离长短以及节点之间关系的紧密程度进行评估。本章实验的目的是验证我们提出的模型满足“鲁棒但又脆弱”的特性^[28,29]

平均路径长度 L 是形容网络连通性的重要指标，对绝大多数实际情况而言，如果有较好的连通性，那么其任意两个节点之间的平均搜索步数也将较少。所以，可以对网络被攻击前和被攻击后的平均路径长度 L 进行比较，推出这次攻击的强度和危险系数。

本文中令 l_0 表示被攻击前网络的平均路径长度值，用 l_1 表示遭受到攻击后的值，那么

$$L_{01} = \frac{l_1}{l_0} \quad (3-22)$$

代表这次攻击的强度。如果网络没有遭受到攻击的干扰，那么 $L_{01}=1$ ，表示该网络是稳定的；如果受到攻击后 $L_{01}<1$ ，则表明网络的连通性反而提高了，说明该网络有很强的自我保护和修复能力；如果 $L_{01}>1$ ，意味着受到攻击后网络的连通性降低，其可靠性也降低了。

同样的，聚类系数 C 表示网络中一个节点的连通性，那么计算出聚类系数的平均值则能表示整个网络的连通性。同理，令 c_0 表示被攻击前网络的平均聚类系数值， c_1 表示被攻击后的其值，那么

$$C_{01} = \frac{c_1}{c_0} \quad (3-23)$$

代表此次攻击的强度和危害程度。如果 $C_{01}=1$ ，那么该次攻击并没有对网络造成伤害，表示该网络是稳定的；如果 $C_{01}<1$ ，则表明网络受到攻击后连通性反而提高了，说明该网络有很强的自我保护，修复能力；如果 $C_{01}>1$ ，表明此次攻击对网络产生较大的影响，使网络的连通性降低，变得更加不可靠。

第一个实验是基于之前所提出的两种模型构建出的网络进行蓄意与随机攻击的比较。实验中，两种模型的终止时间均设为 500，对于均匀更新计数过程无标度网络，设其网络节点增长时间间隔 α 为 0.25，而高斯更新计数过程无标度网络，设其增长时间间隔 μ 也为 0.25。对于这两种模型，均以 l_0 和 C_0 表示受到攻击前平均路径长度和聚类系数，而遭到攻击之后这两个值分别用 l_1 和 C_1 表示，根据公式 (3-23) 和公式 (3-24) 计算网络变化率。用攻击率 m/n 代表此次攻击的强度，其中 n 代表网络总节点数， m 代表本次攻击删除的节点数。例如，若一次攻击的攻击

率为 0.05，网络规模为 2000 个节点，那么，本次攻击需要删除 100 个节点。由以上可知，随着攻击率的增加，蓄意攻击将使网络的平均路径长度增大，平均聚类系数迅速降低，严重影响网络的连通性。而对于随机攻击而言，攻击率的增大，对网络平均路径和聚类系数均没有大的影响，网络对随机攻击具有很好的免疫能力。

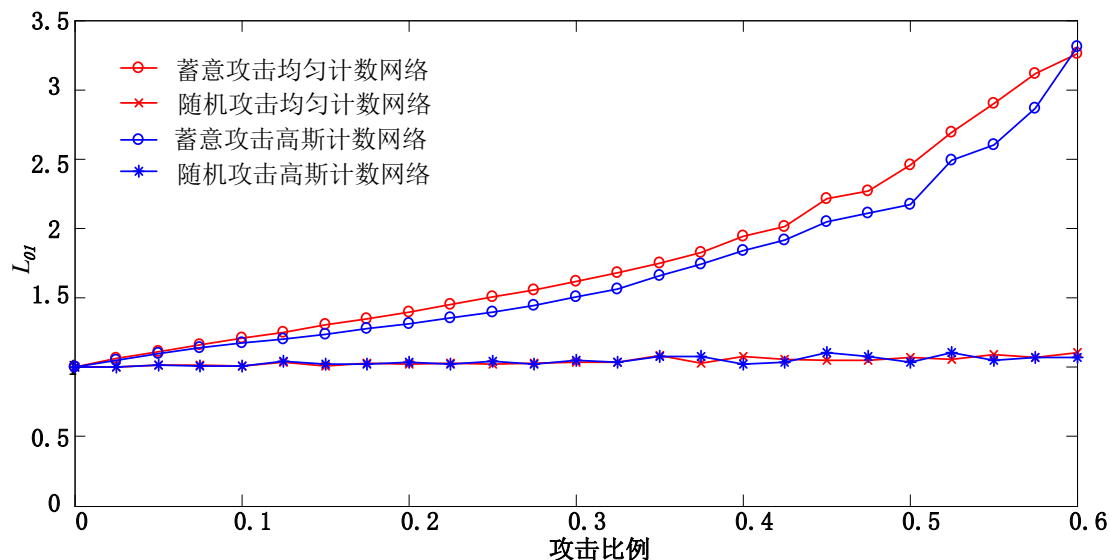


图 3-11 不同攻击形式及不同攻击率下平均路径长度的比较

通过仿真实验，记录下两种模型在进行不同类型的攻击之后其平均路径长度和聚类系数。为了便于观察，用图 3-11 和图 3-12 记录这些数据。

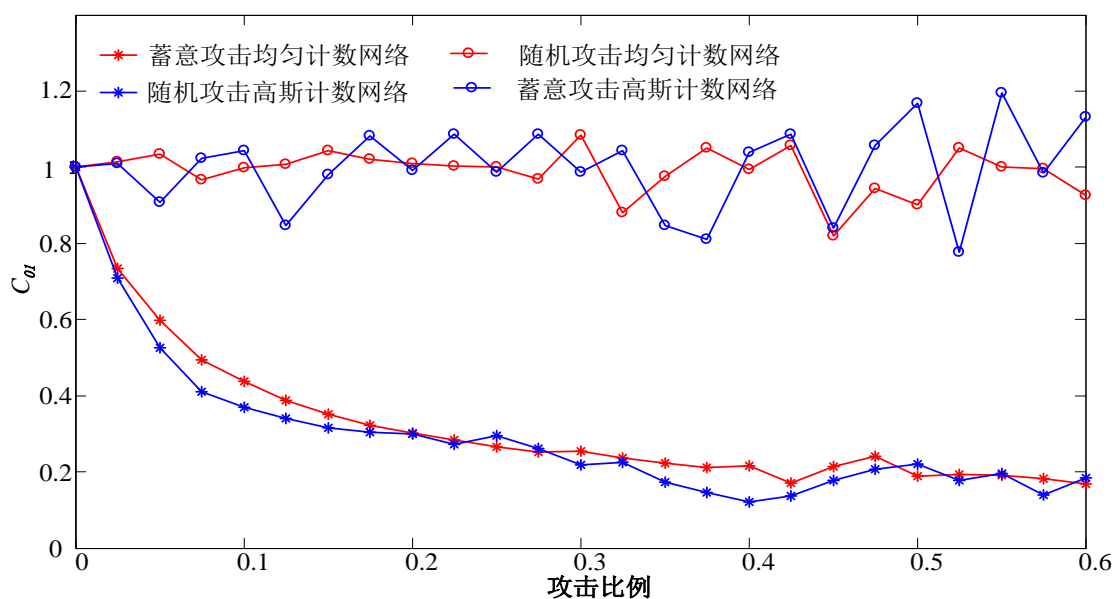


图 3-12 不同攻击形式不同攻击率下网络聚类系数的比较

从图 3-11 和图 3-12 中,可以清晰的看出,两种网络对蓄意攻击都表现出巨大的脆弱性,平均路径长度 L 与受到的攻击强度成正比,随着攻击强度的加大,其值产生了明显的上升趋势,于此相反,聚类系数者与之成反比,随着攻击强度的加大,其值产生了明显的下降趋势。潜在的原因是由于网络度分布的不均匀性,较少的节点拥有较大的节点度。一旦这些高度数的节点被删除,与其相连的边也被删除,那么网络的连通性将大大降低,从而导致网络的平均路径长度增加,平均聚类系数降低。相反的对于随机攻击,在图中,两种模型对其都显示出较强的鲁棒性,无论攻击的强度如何变化,网络的平均路径长度以及其聚类系数都保持相对稳定。这是因为网络中大部分节点的度都处于较低的水平,就单个节点而言,在网络中并不是处于重要的地位。由于随机攻击是随机选择节点删除,往往选中的会是度比较小的节点,删除这样的节点对整个网络而言并没有太大影响,故随机攻击对网络并不会造成很大的威胁。

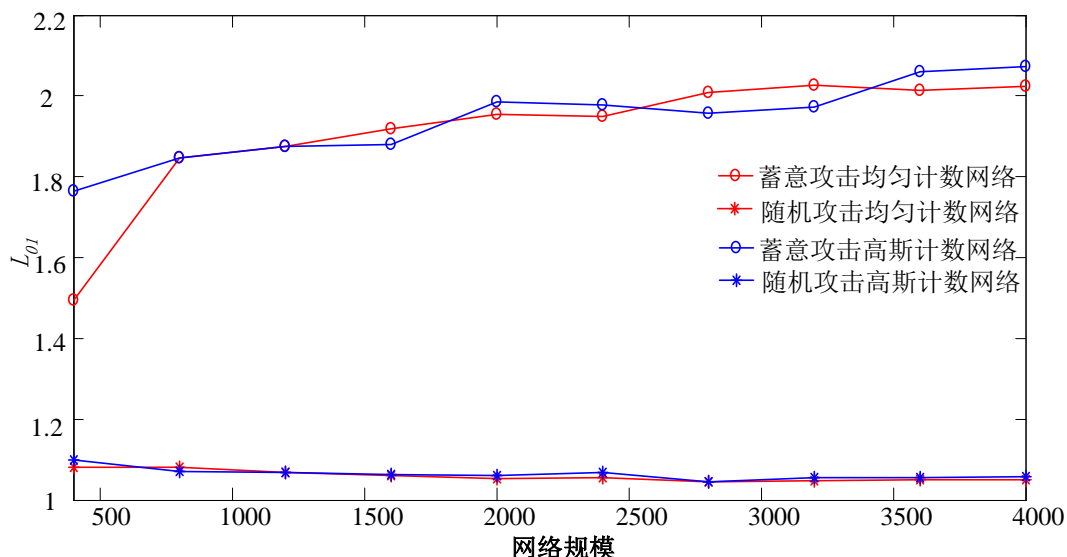


图 3-13 不同攻击形式不同规模网络平均路径长度的比较

另一个实验试图发现网络可靠性与其规模之间的关系。同样是通过参数的设定,模拟不同网络模型下不同规模下的两种攻击,记录对网络连通性的影响。对于均匀更新计数过程和高斯更新计数过程,网络节点增长间隔设置为 0.25,而网络的规模节点数分别取 400,800,1200,1600,2000,2800,3200,3600 以及 4000,攻击率均设为 0.5,这表示网络中一半的节点将被删除。如上一个实验,同样以 l_0 和 c_0 表示受到攻击前网络的平均路径长度和平均聚类系数,而遭到攻击后这两个值分别用 l_1 和 c_1 表示。从图中可以看出,两种模型,在蓄意攻击的攻击下,网络平均路径长度迅速加大而聚类系数迅速减小,表明网络遭受到了严重的伤害。与此相反的是,这两种模型对随机攻击均具备较强的容错能力,其中大规模网络更加明显。

同样为了便于观察,本文用图的形式直观的进行表示这些数据,如图 3-13 和图 3-14 所示。

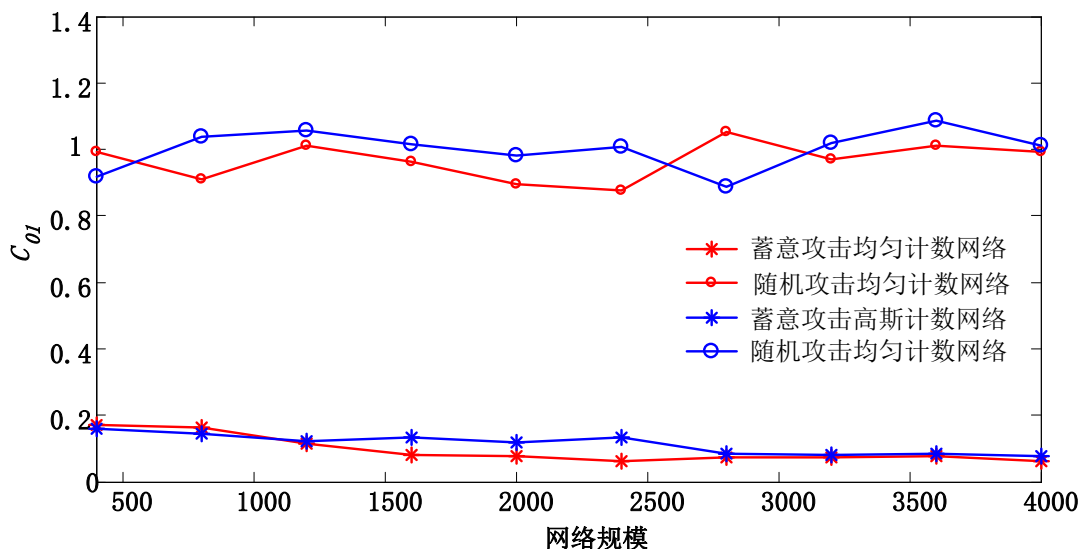


图 3-14 不同攻击形式不同规模网络聚类系数的比较

从图 3-13 以及图 3-14 两图中可以看出,对于蓄意攻击而言,增加网络规模并没有使网络的抗攻击能力增强,反而抵抗能力降低。此外,对图进行仔细观察发现,具有大规模的网络对随机攻击有较好的免疫能力。这是因为,在构建模型的过程中,规模大的网络的高度数节点较多,一旦这些节点受到攻击,平均路径长度将增大,聚类系数将降低。同样,随着网络规模的增加,随机攻击将更加难以选中度高的节点删除,所以当网络是大规模时,对随意攻击有着较强的鲁棒性。

3.4 本章小结

本章中提出了两种新的网络模型:基于均匀更新计数过程的无标度网络模型和基于高斯更新计数过程的无标度网络模型。本章采用物理上的平均场方法在理论上分析其度分布,计算其度指数。最后,通过模拟网络建模过程,统计网络度分布,分析网络可靠性,证明所提出的模型具有无标度网络的基本特性。

本章研究表明,无论网络增长时间间隔和连接边数如何变化,网络的构建机制一旦确定,那么网络的拓扑特性也就确定了。正如,实际中的社会或经济网络的特性并不会因为世界发生战争或经济危机而改变。此外,本章在构建网络的过程中通过设定不同的增长时间间隔,进一步证实了 Barabási 对复杂系统及网络所提出的“独立于时间、功能、范围”的特征^[30]。

第四章 基于复杂网络的电力交易模型

电力交易模式是典型的市场与电力工业重组的产物，目前电力交易的研究大多数都是针对交易过程中双方交易的策略^[31]，而没有对交易产生后电力供应方和使用方所组成的网络进行研究^[32]。然而由于电力交易网络的复杂性，目前的研究结果中还没有一个很好的模型能对电力交易网络进行合理的分析及研究。近年来，大量科学研究表明，复杂网络可以很好的模拟现实世界中的复杂系统^[33]，那么使用复杂网络理论研究电力交易网络系统也必然成为一个值得尝试的方法^[34]。基于复杂网络对电力交易进行建模将更清晰的反映电力在市场中真正的价值，以及交易双方的供求关系。

本章主要工作为提出双边电力交易模型^[35]，通过仿真构建电力交易双方所构成的网络，对模型进行可行性的研究。通过分析该模型的几个网络拓扑特性以验证模型的有效性。此外，通过对网络鲁棒性的测试，得出有益于电力交易双方市场决策的结论。在本章研究中，忽略市场管制及环境等外界因素，仅仅考虑电力交易双方均是自由、平等的买卖双方。

4.1 电力市场的网络模型

在双边电力交易模式中，可以把交易抽象为，买家和卖家两种角色，若买卖双方都同意对方所提出的交易条件那么本次交易成功。通过这种方式，可以把双边电力交易市场抽象为一个网络。用符号如下表示：

$$G = (V, E)$$

其中 $V = \{v_{bi}, v_{sj}\}$ 代表买卖双方，在网络中表示成一个节点， $v_{bi} (1, 2, 3 \dots n)$ 代表买家 i ， $v_{sj} (1, 2, 3, \dots m)$ 代表卖家 j ； $E = \{e_{i1}, e_{i2}, \dots e_{ij}\}$ 代表买家 i 与卖家 j 一次交易成功，在网络中体现为节点 i 与节点 j 之间有一条边。故 G 代表整个交易完成后，所构成的网络图。

在本文提出的交易模型中，买卖双方的交易的关键在于卖家的价格的优势大小。不同于其他的商品交易，电力交易还要考虑物理层的运输方面，只有成功将电力输送出去，才是完成了本次交易。

4.1.1 电力交易过程

双边电力交易是一个动态变化的过程，通过交易是否成功逐渐建立买卖双方的关系进而形成一个网络。在整个买卖谈判中，首先需求方选择一个供应方尝试

进行交易，这个过程中买卖双方的交易不仅仅只是考虑双方利益最大化因素，还需要考虑局部电力不平衡以及电力生产与环境污染之间的矛盾等问题。只有解决这些问题后，此次交易才算是成功，否则进行新的一次的交易。下面给出交易时所需要受到的具体限制。

4.1.1.1 传输过程

只有卖家通过传输网把电成功传输给买家，那么这次交易就可以达成。也就是说电力交易需要受到传输线路上的物理限制。

当买卖双方预先达成交易的情况下，能不能完成此次交易，重点在于此次交易的电量是不是超过传输总线上的限额。总线是传输网中进行电力的传输和分发的设备。在产生交易之前，电力传输网络就给买家和卖家在传输总线上预先分配了各自的传输份额。在尝试交易时，交易网与传输线路之间的关系如图 4-1 所示。

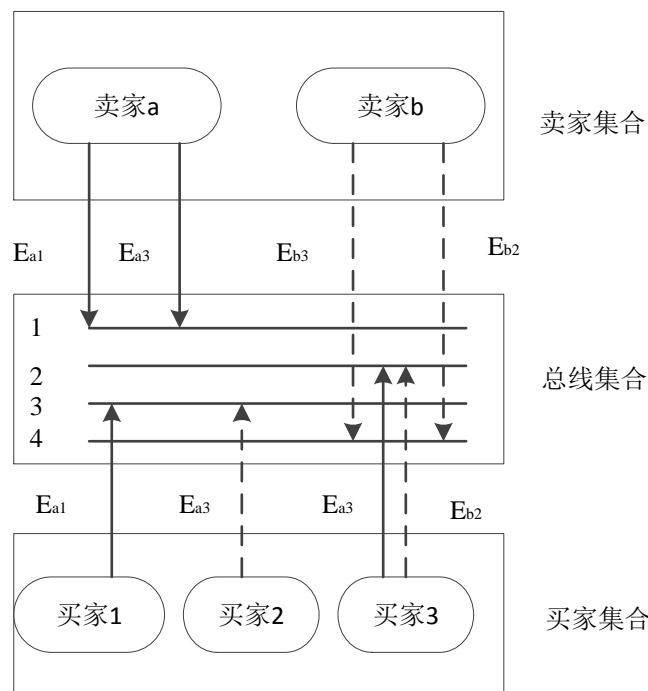


图 4-1 预交易网与电力传输网之间的关系

其中，卖家 a , b ，以及买家 1, 2, 3 对应于电力网中的节点，他们之间的连边为 E_{a1} , E_{a3} , E_{b2} , E_{b3} 。对应于实现电力传输图中，总线可以看做一条直线，图中有 4 条总线。卖家 a 挂在总线 1 上，卖家 b 挂在总线 4 上，买家 1, 2 都挂在总线 3 上，卖家 3 挂在总线 2 上，它们之间的连边代表有买卖双方有交易产生。

下面具体列出，交易所受到传输线方面的物理和经济上的约束：

- (1) 电缆负载限制

买卖双方的交易量是以功率流的形式在电力传输网中体现的，故交易量不能超过传输网中每条总线的最大功率负荷量。若某次交易超过总线的最大传输量，那么必须减少此次的交易量或是该买家选择另外的卖家进行交易。直接物理约束可以表示为：

$$w_i < W_i, i \in I \quad (4-1)$$

其中 i 是传输线的编号， w_i 是 i 此时的负荷量， W_i 是 i 的最大负荷量。

从等式 (4-1) 中可以得出，如何计算 w_i 是直接约束条件中的关键问题。利用直电压潮流模型，电力可以通过电力传输分配系数进行建模。线路 i 的功率传输量可以表示为：

$$w_i = \sum_k M(i)_k (c_k - s_k), i \in I \quad (4-2)$$

其中 k 表示总线编号， i 表示传送线路编号， c_k 表示总线 k 的现有的负载量， s_k 代表此次交易的电量， $M(i)_k$ 代表传送线 i 的电力传输分配系数矩阵中对应的系数值，可以用以下关系式表示：

$$M(i)_k = \frac{o(m,k) - o(n,k)}{o(m,n)} \quad (4-3)$$

其中 m 是传输线 i 的头结点编号， n 是 i 的尾节点编号。 $o(m,k), o(n,k)$ 均是传输阻抗矩阵中的元素值， $o(m,k)$ 对应该矩阵的 m 行 k 列上的值， $o(n,k)$ 对应矩阵 n 行 k 列的值。 $o(m,n)$ 是传输线 i 的所存在的阻抗。

那么基于上述理论，以图 4-1 中的传输图为例，计算其中传输线 k 的流量为：

$$w_i = M(i)_1(E_{a1} + E_{a3}) + M(i)_4(E_{b2} + E_{b3}) - M(i)_3(E_{a1} + E_{b2}) - M(i)_2(E_{a3} + E_{b3}) \quad (4-4)$$

其中 E_{a1} 表示卖家 a 与买家 1 的交易量， E_{a3} 表示卖家 a 与买家 3 的交易量， E_{b2} 表示卖家 b 与买家 2 的交易量， E_{b3} 表示卖家 b 与买家 3 的交易量。

(2) 传输成本限制

间接约束主要是指电力传输的成本，也称为转运费。在电力交易市场中，买卖双方均要负担这笔转运费，而且这笔费用与双方在传输网中的交易量有关。只有承担足够的转运费，卖家的电力才能通过电力传输网输送给买家使用。

本文使用逆流法进行转运费评估，如下：

$$WD = \sum_{i \in I} C_i \frac{w_i}{W_i} \quad (4-5)$$

其中的 i 是传输网中的线路的编号， C_i 是 i 的安装成本。

4.1.1.2 买卖双方的利益

除了上述提出的物理传输层面的约束外，另一个重要的因素是经济层面的约束。像其他的物品买卖一样，买卖双方都能获得一定的经济利益是达成此次交易的关键条件。电力交易同样需要满足买卖双方均能接受的利益，才能最终达到双赢的效果。本文把这满足双方利益的过程作为经济限制条件。根据市场经济的普遍性原则，交易中的利润和成本是经济限制中的核心问题。本文假定双边电力交易的经济限制基于两个条件：只要卖家能从这次交易中获得收益，那么同意与买家产生交易；同样，只要买家能够承担此次交易中卖家所提出的价格，那么也可以看成买家获得了收益，同意与卖家产生交易。

以上两点可用数学公式简要说明，对卖家而言只要满足：

$$[\frac{1}{2}(Q_s + Q_b)F_{sb} - \frac{1}{2}WD] - Q_s F_{sb} > 0 \quad (4-6)$$

对买家而言只要满足：

$$Q_b F_{bs} - [\frac{1}{2}(Q_s + Q_b)F_{bs} - \frac{1}{2}WD] > 0 \quad (4-7)$$

其中 Q_s 表示卖家提出的价格， Q_b 代表买家提出的价格， F_{sb} 代表此次买卖双方的交易量， w_i 表示此次交易在传输线 i 上所产生的电流负载， WD 如上节中公式(4-5)中所示。

4.1.2 电力交易市场的建模算法

通过上面的描述，双边电力市场的交易过程已经确定，而对电力交易市场的建模流程需要进一步地介绍。

首先建模需要明确初始网络，在构建网络的第一步中，即买卖双方产生交易前，本文假设所有的卖家已经在交易市场中存在了，卖家与卖家之间有着一个优先连接机制，这个机制和上文提到的模型的构建方法中的根据节点度优先连接一样机制以这种方式，卖家之间首先建立一定的关系，构建初始网络。

- (1) 初始网络，根据随机图理论构建初始卖家网络；
- (2) 预交易，对网络中的买家基于价格优先选择一个卖家进行预交易，记

录该与交易量；

- (3) 条件验证，检查预交易过程中是否符合上文提出的传输线上的两个约束以及买卖双方是否都有利益；若有其中任何一个条件不满足，该买家重新选择一个卖家，直到满足条件所有条件记录能满足的交易量，或没有卖家与该买家能达成交易此时该买家交易量为 0；
- (4) 结束条件，是否所有买家都尝试过预交易；
- (5) 结束，根据交易构建电力交易网络。

下面给出建模的具体过程，如图 4-2 所示。

从图 4-2 可知，构建初始网络中，开始时所有卖家节点的度均为 0，那么根据优先连接机制，无法获得各个卖家的连接概率。本文在初始时，随机选择在卖家节点之间添加若干条边，然后其他卖家根据优先连接机制，使卖家节点之间有一定的关系。如流程图所示，买家是依次加入市场中与卖家达成交易，直到没有新的买家进入市场，则完成所有交易，从而构建整个网络。

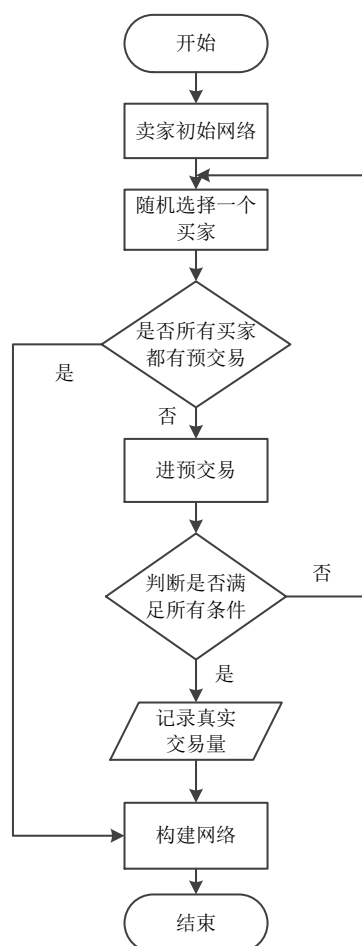


图 4-2 建网流程图

交易的产生过程是：每加入一个买家节点，首先选择一个价格最低的卖家尝试进行交易，交易过程中需要考虑到物理层上的约束及经济层上的约束，只有当满足这些约束的条件下买卖双方才能达成一致。一旦双方达成交易，便在交易网络中给相应的买卖双方节点增加一条连线。若在交易中存在任何一个约束不满足时，便尝试下一轮的交易，不停的循环这个过程，直到市场中没有新的买家加入，那么此时双边电力交易网络达到稳定状态。

4.2 理论分析

如上一节所提出的交易策略，买卖双方的电力交易主要取决于卖方的价格优势，本文称把这种优势称为基于价格优势的连接机制。基于这种机制其实是符合两个现实条件的：卖家价格与买家选择该卖家进行交易的概率成反比，即卖家价格越高，与买家产生交易的可能性越低，这符合现实交易的情况，人们总是希望低买高卖；再者交易的网络模型的特征不会受交易价格度量单位变化的影响。

故，可以用 $Q^{-\lambda}$ 表示这种以价格为优势的连接机制。

假设 Q 是卖家提出的一个价格， λ 是一个常量，那么显然随着 Q 的增加， $Q^{-\lambda}$ 会减小。这便符合了卖家价格与买家选择该卖家进行交易的概率成反比的条件。

因为网络构建过程中与交易的价格单位无关，可以把它符号化为：

$$f(mx) = nf(x) \quad (4-8)$$

其中 $f(x)$ 是一个任意的函数， a, b 均是常量。若有 $f(1)f'(1) \neq 0, x=1$ ，那么对任意的 a 有：

$$x \frac{df(mx)}{d(mx)} = \frac{f(x)}{f(1)} \frac{df(m)}{dm} \quad (4-9)$$

把 $a=1$ 带入等式中可得：

$$x \frac{f(x)}{d(x)} = \frac{f'(1)}{f(1)} f(x) \quad (4-10)$$

等式 (6) 的解为：

$$\ln f(x) = \frac{f'(1)}{f(1)} \ln x + \ln f(1) \quad (4-11)$$

即：

$$e^{\ln f(x)} = e^{\frac{f'(1)}{f(1)} \ln x + \ln f(1)} \quad (4-12)$$

最终可得：

$$f(x) = f(1)x^{-\lambda} \quad (4-13)$$

其中 $\lambda = -f'(1)/f(1)$ 。故本节所提出的以 $Q^{-\lambda}$ 表示价格优先的连接机制是符合幂律分布的特点即等价于等式 (4-13)。

同理基于价格优势的连接机制可以表示为与度优先连接机制类似如公式 (2-6) 的表达式：

$$\Pi_i = \frac{Q_i^{-\lambda}}{\sum_{j=1}^m Q_j^{-\lambda}} \quad (4-14)$$

其中等式中 m 为卖家个数， i, j 为卖家编号， Π_i 为买家与卖家 i 产生交易的概率。

从连接机制的定义中可知，双边电力交易成功可能性的大小首先与由卖家的价格优势有关，即卖家价格越低，双方交易的可能性越大。因此基于价格优势的连接机制可以用 (4-14) 等式表示。

根据相关领域知识，本文设 $\lambda=3$ 。

4.3 仿真实验及分析

在本节中，基于上文提出的建模流程，构建双边电力交易市场网络模型。并对该模型几个基本的网络结构拓扑特性进行分析。鲁棒性是复杂网络中对实际网络最具有现实意义的性质，故最后对电力交易网络的鲁棒性进行详细的阐述及分析。

4.3.1 网络模型实例

在本实验中构建了三个不同规模的双边电力交易网络，以获得更加全面和准确的双边电力交易市场的结构信息，证明文中所提模型的可行性和合理性。这三个网络中，第一个是由 112 个市场参与者基于 IEEE 118-bus 电力传输网构建；第二个是由 268 个市场参与者基于 IEEE 300-bus 电力传输网构建；对三个是由 1024 个参与者基于 IEEE 2383-bus 电力传输网构建。其中，构建双边电力交易网络实验所需的数据是从论文^[36-40]参考所得。下面列出模拟出的电力交易网络模型图，以

268 个节点为例，如图 4-3 所示。

在图 4-3 中建立的双边电力交易市场中，有 68 个卖家，200 个买家，总供电能力为 7250 兆瓦，总需求电量为 7109.67 兆瓦，电力传输网中每条传输线的最大负荷为 40 兆瓦，建设传输线所需的成本也为每兆瓦 6 元。在 268 规模的交易市场中，受到物理层制约影响的交易有 8 笔，共减少 54.44 兆瓦的交易量；受到经济层制约影响的交易有 27 笔，共减少 189.97 兆瓦的交易量。最后在 1024 个参与者的市场中，有 326 个卖家，698 个买家，总供电能力为 31730 兆瓦，总需求电量为 31303.28 兆瓦，电力传输网中每条传输线的最大负荷为 80 兆瓦，建设传输线所需的成本与前两个规模的代价一样。整体交易完成后，有 42 笔和 108 笔交易分别受到物理层制约和经济层制约的影响，减少的交易总量分别为 141.04 兆瓦和 594 兆瓦。

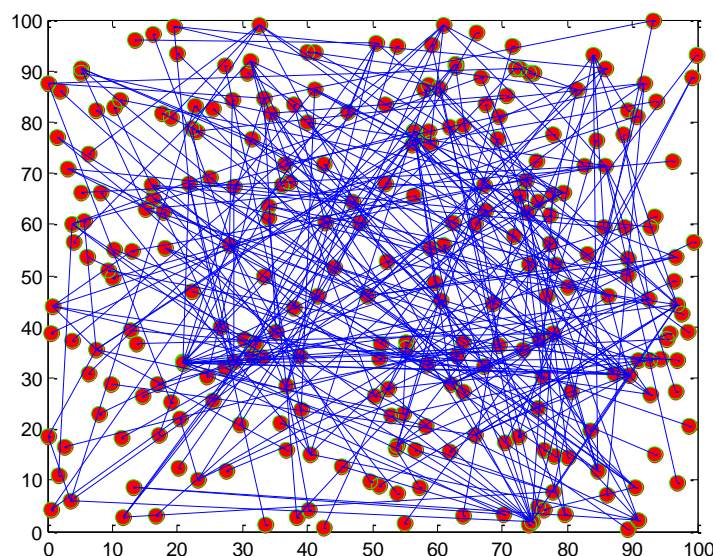


图 4-3 电力交易 268 个节点网络图

4.3.2 网络拓扑分析

4.3.2.1 度分布

在双边电力交易网络中，如果一个节点具有较大的度，意味着该节点与市场中多个其他的参与者发生了交易，那么它对整个市场网络的价格及交易量的波动都有着极其重要的影响。

对上小节中三种规模所建的网络，进行度与度分布的分析。从上至下分别表示 118 买卖双方，268 个买卖双方和 1024 个买卖双方。结果如图 4-4 所示。

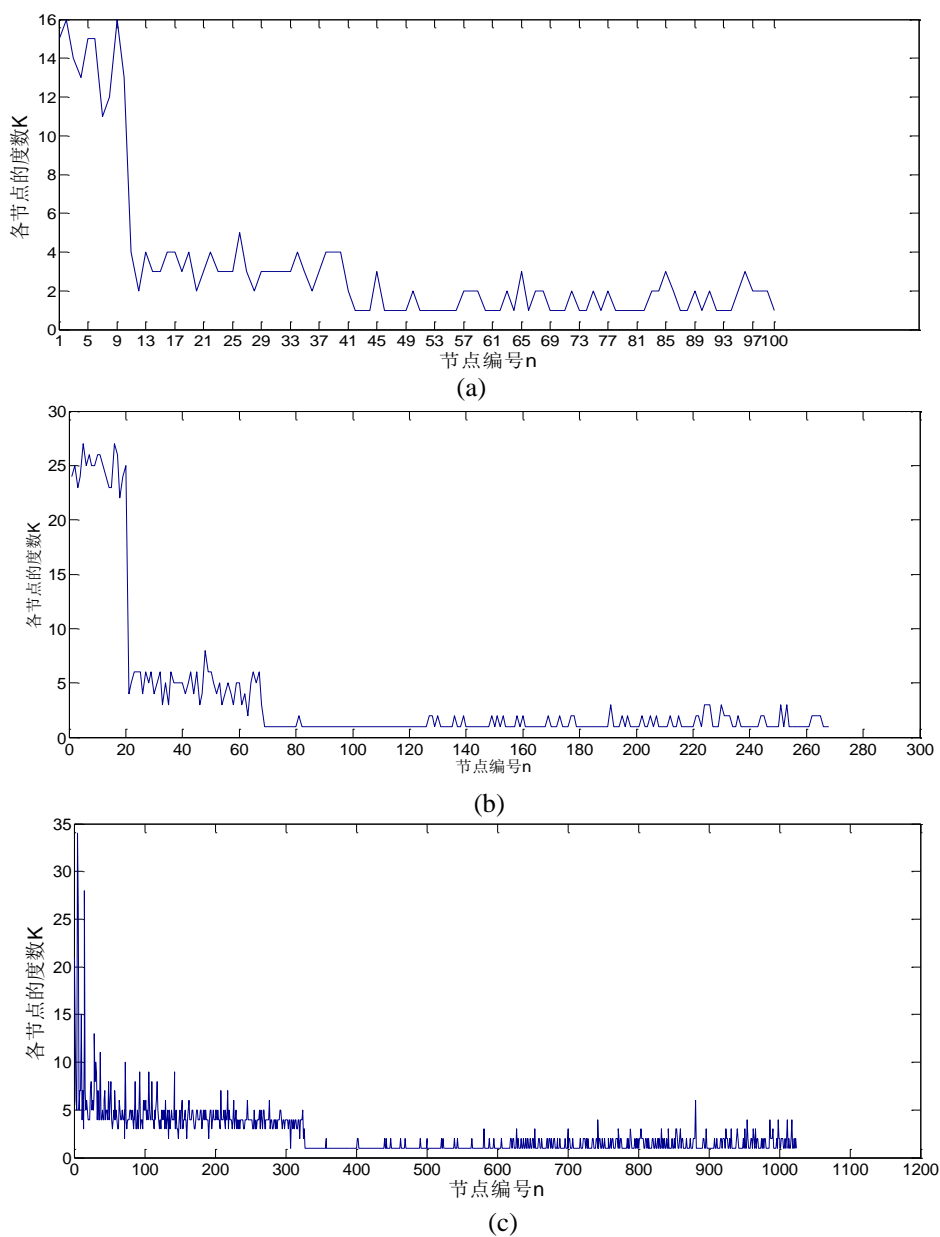


图 4-4 不同规模网络的度分布

从图 4-4 中可以看到，电力交易网中大部分顶点的度都是较低的，节点的度较高的顶点所占比例很小。还可以看出，三种不同规模的电力交易网其度分布都具有长尾分布的特点，也就是说都遵从幂律分布。

为了进一步分析双边电力交易网络的度分布特点，以上所得的实验数据绘制在双对数坐标的图形中。使用最小二乘法对其中的散列点将其模拟成一条直线，计算该直线的斜率，该斜率即为该网络的度指数，如图 4-5 所示。

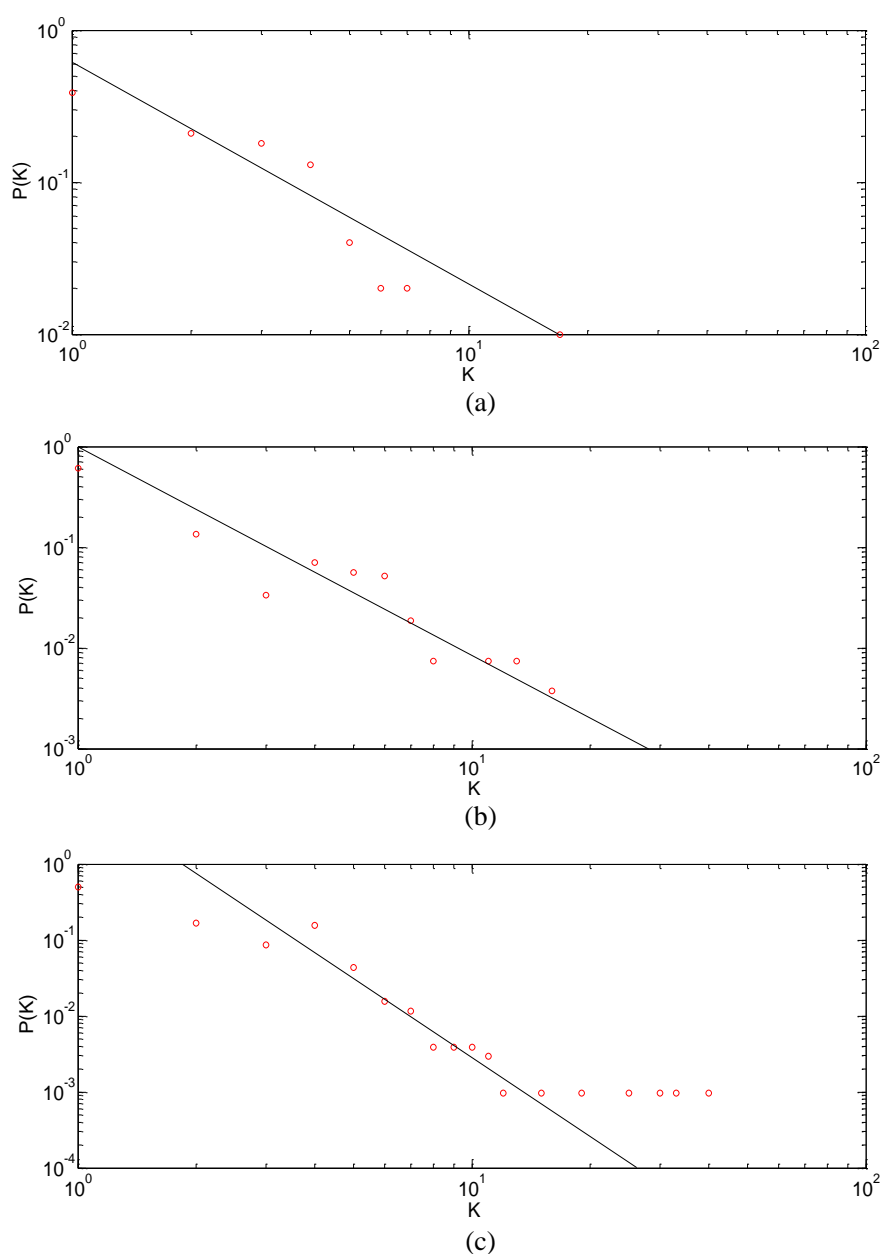


图 4-5 不同规模电力交易网络的度指数

由图 4-5 可以很清楚的看出，三种不同规模的网络度指数均在 2~3 之间，符合复杂网络中幂律分布的特性。

综上所述，双边电力交易模型具备明显的无标度特性。电力交易市场网络的无标度性体现在网络中的大多数成员节点度都处于相同的较低水平对市场影响较小，只有少数成员因为其价格优势，使其他成员都优先与其进行交易而具有较高，较强的市场影响力。由此更加说明该模型符合真实的双边电力交易市场。

4.3.2.2 评估指标

聚类系数体现整个网络的传递性，具体表现在两个同时连接到相同顶点的节点，它们之间彼此互相连接的概率。在双边电力交易网络模型中，聚类系数表示买卖双方成员关系的紧密，也就是电力交易市场中成员交易的相关性。具体的计算方式在第三章已经介绍过，此处不重复介绍。

双边电力交易模型的平均路径长度 L 表示市场中买卖双方的成员想到达对方，需要经过其他成员的个数的平均值。其值得计算方法参考第三章中的介绍。

三种不同规模的网络其聚类系数和平均路径长度如表 4-1 所示。

从表中可以看出，当网络规模增长到原来的两倍或十倍时，网络平均聚类系数分别下降 42.95% 和 77.4%，而平均路径长度分别增长 14.9% 和 36.1%。由此可以说明，聚类系数与规模成反比，相反平均路径长度与规模成正比。

表 4-1 不同规模下聚类系数及平均路径长度

网络规模 \ 特征值	聚类系数 C	平均路径长度 L
112 个节点网络	0.0284	4.22
268 个节点网络	0.0162	4.85
1024 个节点网络	0.0064	5.74

为了得出更精确，可信的结论，本实验还对比了分别采用完全随机连接和完全度优先连接机制在相同规模 1024 个节点下的两种网络。具体对比数据如表 4-2 所示。

表 4-2 不同网络模型其聚类系数及平均路径长度

网络模型 \ 特征值	聚类系数 C	平均路径长度 L
随机网络模型	0.0082	5.12
BA 网络模型	0.017	4.24
电力交易网络模型	0.0064	5.74

表 4-2 中结果表明，基于双边电力交易网络模型相比基于纯随机连接网络模型的聚类系数低 22.5%，而比纯优先连接机制所构成的网络模型低 62.3%。而对于平均路径长度，电力交易网络模型比随机连接模型多 12.1%，比纯优先连接机制网络模型多 35.3%。

上述实验结果表明，根据双边电力交易过程所构建网络模型的平均聚类系数较小，而平均路径长度较大。导致这种结果的原因主要有两点，其一，电力交易

的交易范围较小，只有少部分的供应商和输送方参与市场交易，而大部分普通用户根本不涉及这种交易；其二，电力交易受电力传输网负载量的限制，导致部分交易的受阻。

双边电力交易网络具有低聚类系数表明对于市场网络中节点度较大者存在潜在的机会，而平均路径长度较大会阻碍双边电力交易的效率。

4.3.3 鲁棒性分析

任何复杂系统都具有很强的抗攻击能力，而网络的鲁棒性是用来描述这一能力的指标。若果一个网络拥有较强鲁棒性，便表明它能很好的应对网络某些节点的故障，保持自身的稳定。

本节中，通过模拟对双边电力交易网络进行攻击，以测试的其鲁棒性。在本实验中，通过随机挑选网络中的若干顶点并撤销其在网络的位置完成随机攻击，而蓄意攻击是依据节点度的大小从高到低挑选网络中的顶点并撤销其在网络的位置。另外，攻击的强度以删除网络中节点的个数来体现，实验中分别删除 2%，4%，6%，8%，10% 的节点数作为攻击规模进行对比。

为了全面的分析双边电力交易网络的鲁棒性，实验中分别对三种规模的网络进行强度不同的随机和蓄意攻击，取每种规模攻击前与攻击后网络中所拥有的最大连通子图，对其进行平均路径长度的比较，两者所得到的比值作为评估鲁棒性的指标。如果是鲁棒性很高的网络，那么图中的平均路径长度的比值在受到随机攻击时将不会有太大的变化。该比值用以下公式表示：

$$L = \frac{l_a}{l_b} \quad (4-16)$$

4.3.3.1 随机和蓄意攻击

为了全面的测试双边电力交易网络的鲁棒性，本实验在三种规模下分别进行随机和蓄意两种攻击，攻击强度分别设为 1% 到 10% 之间，比较攻击所受影响 L 的大小。如图 4-6 所示。

在图 4-6 中，可以清晰的看出，当攻击强度从 1% 逐步上升到 10% 时，对于三种规模的 L 指数在随机攻击下，没有太大的影响。而对于蓄意攻击，在攻击强度增大时，平均路径长度有较大波动。原因是其遭受到蓄意攻击时，节点度高的顶点大部分被删除，导致模型被分为几个小的孤立的连通子图。也就是说，双边电力交易网络在受到人为蓄意的攻击下将遭到严重的破坏，即其对此种类型的攻击抵抗性很差。

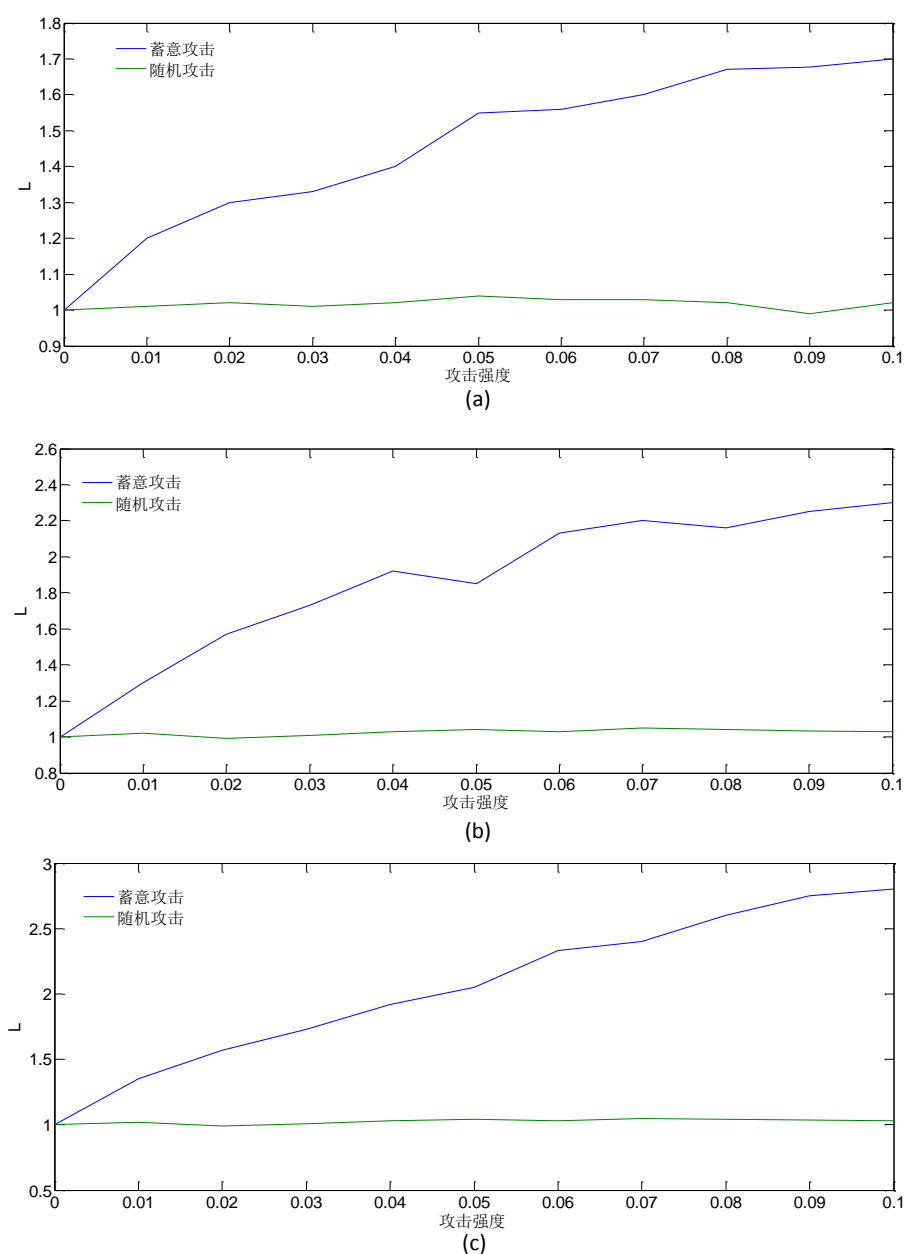


图 4-6 不同规模在不同攻击强度下的平均路径长度比

综上所述，双边电力交易网络模型对随机攻击有较强的鲁棒性，相反的对于蓄意攻击则显示其明显的脆弱性。

4.3.3.2 网络规模与攻击强度的关系

上述实验中，证明双边电力交易网络对蓄意攻击没有很好的抵抗能力，而这恰恰显示出双边电力交易市场中的风险，需要对蓄意攻击行为进行更进一步的分析及预防。本实验中，除了分析网络对两种攻击的鲁棒性，还针对规模大小

与攻击强度之间的关系做了一定的程度的研究。

以下做了两组实验对比，一组是蓄意攻击程度为 1% 时，受攻击影响指数 L 与网络规模的关系，另一组是蓄意攻击程度为 10% 时，受攻击影响指数 L 与网路规模的关系。如柱状图 4-7 和图 4-8 所示，图中纵坐标为指数 L ，横坐标为网络规模。

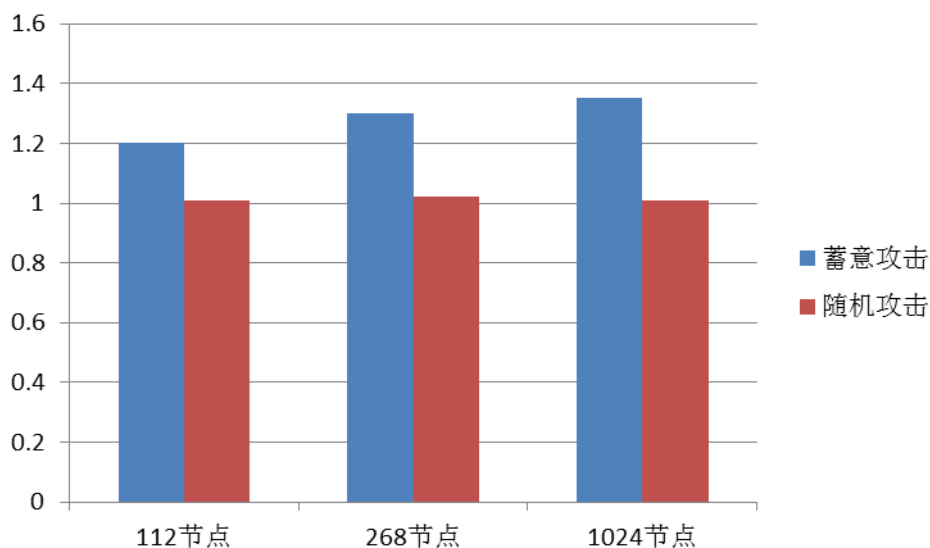


图 4-7 不同规模网络遭受强度为 1% 攻击后平均路径长度比

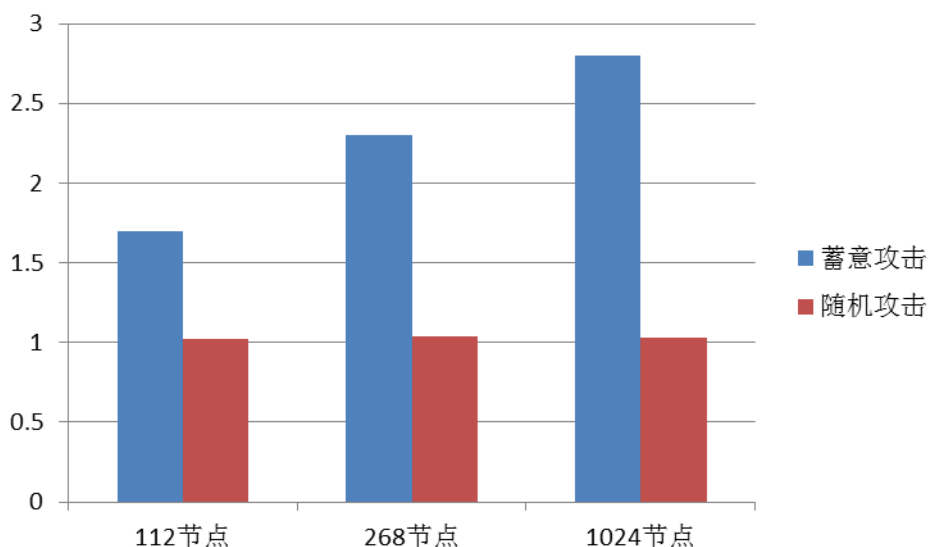


图 4-8 不同规模网络遭受强度为 10% 攻击后平均路径长度比

比较 4-7 以及 4-8 两图，可以发现，当蓄意攻击强度上升时，网络规模越大，受到的影响就越大。也就是说，越大规模的双边电力交易网络受到高强度的蓄意攻击时，对攻击越没有抵抗能力，越脆弱。产生这一现象的原因是，在大规模的双边电力交易网络中，度数高的节点个数必然要比小规模网络中的个数多。这会

导致大规模的网络中所含的那些节点度高的顶点其邻居节点数目比小规模的网络中要多的多。而度数高的顶点往往是网络结构图中所含的最大连通子图的关键节点。当蓄意攻击强度增加时, 这些度数高的节点就成为了攻击首要目标。当度数高的那些节点被删除时, 与规模小的网络相比, 规模大的网络中更多的节点的相连关系遭到破坏, 这导致其连通性被破坏的更彻底。而对于随机攻击不管强度多少, 对任何规模的网络都没有太大的影响。

综上, 双边电力交易网络对随机攻击有着良好的鲁棒性, 也就是说随机出现的突发状况对整个市场的价格以及交易量不会产生太大的影响。相反地, 双边电力交易网络对蓄意攻击显示出其脆弱性, 特别对大规模的网络来说, 更加凸显了这一点。那么为了提高双边电力网络对蓄意攻击的鲁棒性, 可以考虑使用设备冗余、多方采购或增加市场中介等方法。

4.4 本章小结

现代生活已经离不开电力, 人们的生活也已深深地对电能产生了依赖, 对电力交易网络的研究必然成为当前研究的热门问题。研究基于复杂网络的电力市场网络, 发掘电力交易网络的特点, 保持健康、平稳的发展具有重要的意义。本文基于复杂网络分析方法, 综合考虑交易过程中存在的物理层的约束以及经济层的约束, 对双边电力交易进行网络建模, 有助于发现双边电力交易市场的特征。

本章根据所提出的模型, 对三种规模的双边电力交易网络进行重复模拟实验, 通过对比不同规模下的实验数据得出一致的结果, 验证所提出模型的可行性及有效性。同时, 证明本文所提出的网络模型在某种程度上适用于模拟双边电力交易市场。

基于双边电力交易市场的仿真结果的分析, 双边电力交易网络显示出真实双边交易市场的某些特征, 网络度分布具有长尾分布的性质, 其平均聚类系数较低, 平均路径长度较大, 对随机攻击表现出强大的鲁棒性, 对蓄意攻击有着明显的脆弱性。那么从这些特征中, 可以看出网络中度数高的节点对应的市场成员对整个交易市场的价格及交易量起着重要的作用。网络的鲁棒性特征则对交易市场的决策以及规避市场风险都有着重要的指导意义。

第五章 基于复杂网络的自然语言模型

目前,网络科学已被越来越多的人了解与熟知,并且已经成为许多的科学家进行跨领域研究的工具,其中运用网络科学进行自然语言分析也是研究者所热衷的课题。然而,关键字提取^[41,42]采用的方法是基于机器学习,如支持向量机、TF、TF-IDF^[43],或者基于统计的方法,如加权平均标准版偏差法等等^[41,42]。随着研究的深入发展,越来越多的人发现,文本可以构成一个复杂网络^[44,45],这为网络解决自然语言处理问题提供了基础。

本文将文本中的单词看作一个节点,单词之间的相邻关系作为节点之间的连接边。利用这种方式,建立一个文本的网络模型,对其的拓扑性质进行分析,并根据所构建的网络,计算每个词对应节点的介数和节点度,最后计算节点重要程度,挑选其中最大重要性的若干个词作为该文本中的关键词。

5.1 文本网络研究方法

利用网络科学的研究方法对实际问题进行具体实验与研究,首先要抽象出实际问题的特征,提取出模拟的方法,然后根据所提出的方法进行建模。在文本中相邻两个单词往往会构成一个词组或者表示一种常用的单词,本文首先抽象出单词之间的相邻关系进行网络建模,其中结合网络模型中节点的度和中心介数两个特性,提出一个新的关键词提取算法。

5.1.1 文本网络模型

构建网络需要明确网络中的节点和节点之间的连边,把现实世界中的复杂系统映射成一个网络,那么在对文本的建模的过程中,把文本中的单词作为网络的节点,把单词与单词之间的位置关系作为节点之间的连边。可以用以下关系表示一个网络:

$$TN(t_i) = \{N(t_i), E(t_i, f(t_i))\}$$

其中 t_i 是相应的文档, N 表示网络顶点的集合, E 是顶点间边的集合, f 表示对应的映射函数。

根据图论的表示方法,可以把本文单词网络表示为图的形式:

$$G = (V, E), \quad V = \{v_i | i = 1, 2, \dots, N\}, \quad E = \{(v_i, v_j) | v_i, v_j \in V\}$$

其中 G 表示整个文本对应的单词网络， V 表示其所包含的单词集合， E 表示词与词之间的联系。通过这种方式，在文本中出现频繁的单词并且这些词往往又可以构成多种短语，那么必然会在网络中拥有较高的节点度。

5.1.2 节点介数

介数体现节点在网络中的重要程度，衡量节点在网络通信中所起作用大小，在实际应用起着十分关键作用。具体来说，在求任意两个节点最短路径的过程中，所有经过某节点的路径总数与全部最短路径总数之比即为该节点的介数。抽象成数学公式为：

$$B_i = \sum_{m \neq n \neq i} \frac{p_{mn}(i)}{p_{mn}} \quad (5-1)$$

其中 B_i 代表网络中节点 v_i 的介数值， p_{mn} 代表网络中所有的两个节点 v_m, v_n 之间最短路径的总条数， $p_{mn}(i)$ 代表网络中任意两个节点 v_m, v_n 之间的需要经过该点 v_i 的最短路径的条数。

通过介数定义可以看出，介数相对较高的节点具有重要的价值，它对整个网络的通信流畅度起着无比关键的影响，对网络中各个节点之间的连通具有重大的价值。

5.1.3 节点重要性

由文本第二章中介绍的节点度表示一个节点的局部重要程度，如果单以节点度作为提取关键词的指标，那么只能提取出局部网络最重要的节点，这将大大减少网络中度数不高但对整个网络的影响较大的节点被提取出的可能性。同样的，若单单以节点介数较高的节点作为文本的关键字，只是考虑到了对网络中连通性产生重要影响的这一因素，就没有很好地体现网络局部上的特征。综合这两种因素的考虑，本文将两个因素结合起来考虑，不仅考虑网络整体的信息传输性，而且将节点在局部网络中的重要地位进行结合，提出了一种新的衡量节点重要程度的指标，可以用以下公式表示：

$$DB_i = \alpha \frac{D_i}{\sum_k D_k} + (1 - \alpha) B_i \quad (5-2)$$

其中 k 表示有多少个不同的节点度， D_i 表示顶点 v_i 的度， B_i 表示顶点 v_i 的介数， α 是一个可调节的参数，它的取值范围为 $0 \leq \alpha \leq 1$ 。

该方法是从 PageRank 算法中得到的启发。PageRank 算法是利用页面被引用的次数和那些引用页面的自身的 PageRank 值求得这个页面的 PageRank 的值，这个数越大说明其在网络中越重要。对本文而言，节点的度可以被看作其他节点对其引用，节点的度越高，该节点在局部网络中占有更重要的地位。对于介数越高的节点，网络中最短路径所通过这个点的次数越多，那么从这个点出发更容易达到其他点，由此可见这个点对网络整体而言具有相对较高的意义与价值。故本文综合考虑网络的局部特性和整体特性，结合节点的度与介数两个重要指标，作为关键词提取的依据。

5.2 构建文本单词网络

本文提出基于网络科学对文本进行建模，验证该网络是符合复杂网络的无标度特性，进而利用该网络的拓扑特性完成关键字的提取。下面给出文本建立单词网络模型的详细过程。

5.2.1 文本预处理

本文主要针对英文文本进行算法的研究，在构建网络之前需要对文本中的单词进行预处理，否则如果直接对文本构建网络，所提取的关键词将不能准确表述文章的内容。预处理阶段的步数：

- (1) 断句、分词处理，根据文本中的标点符号以及单词与单词之间的空格进行单词切分。
- (2) 大小写变换，由于文本中有单词有大小写的不同，为了避免单词不受大小写影响，把文本中的字母全部改为小写形式；
- (3) 词态变换，英文单词中，同一个单词存在多种形式，为了不错分单词，必须对文本中的单词统一变换成该单词的原型模式；
- (4) 去停用词，文本中存在很多无意义的单词，比如：the、a、and 等等，预先收集好大部分的停用词，实验中去掉这些停用，减少无关词的干扰，提高关键词提取的准确率；

步骤(4)中的停用词是语言中的过渡词，通常是一些介词、量词、转折词等，对文章所要的表达真实意义没有任何作用。故停用词的存在，将对实验数据的采集造成一定的干扰，如保留这些停用词，实验中提取出的关键词很有可能会是这些在文本中出现频率很高的无实际意义的词。对停用词的采集，通常依赖于大量的统计，因此，实验中采用互联网中业界所认可的停用词表，包含了大多数文本中的停用词。

5.2.2 构建方法

大量关于人类自然语言的研究表明，由自然语言所构成的网络显示出复杂网络中的拓扑特性，Cancho 和 Sole 发现，自然语言网络是一个小世界网络模型。因为单词网络可看作人类自然语言网络中的一种，它具备自然语言网络所含的相同特性，故单词网络也符合小世界网络模型。本文依据 Cancho 和 Sole 提出的连边方法，利用单词跨度为 1 的关系构建网络。众多的研究表明，如果采集的单词过于密集，那么单词之间相距较长的关系则被忽略；如果跨度过长，采集到单词的关联过于琐碎、繁重，将掩盖真正的重要特性，对数据的处理也将变得繁琐。故使用这样的方法作为构建文本网络的连边规则是比较合理的。

利用网络科学理论进行关键词提取算法，首要任务是将经过预处理的文本映射成为一个网络，对网络进行拓扑分析，计算网络中各节点的度数和介数，利用度的重要性指标计算综合重要性，最后对节点综合重要性 DB 值进行排序，提取其中值最大的前 k 个节点，找出这个节点所对应的单词，即提取出了文本的关键词。

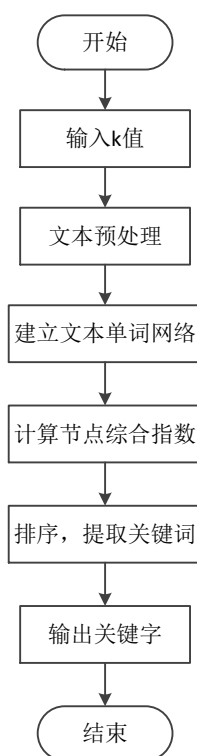


图 5-1 基于复杂网络的关键词提取算法流程图

网络科学理论进行关键词提取的算法具体步骤如下：

- (1) 预处理，根据上节描述的预处理步骤对原始文件进行处理；
- (2) 构建文本单词网络，将文本中单词看作为网络中的节点，与该单词左

右相邻的单词对应的节点之间加边，从而构建出一个网络。

- (3) 计算节点综合重要性指数，对文本单词网络进行拓扑分析，计算各个节点的度数和介数。
- (4) 提取关键词，将单词介数值按递增顺序进行排序，输出前 k 个作为文本的关键词。

用图 5-1 表示这个算法的整个过程。

5.3 仿真实验与分析

本节根据以上提出的算法，进行网络模型的建立，验证本文所提出的节点重要性指标的计算方法的合理性及算法有效性。

首先，对仿真实验的环境进行介绍，本文中的实验是在 Windows7 32bit 操作系统，2G 内存，编程工具为 Matlab 环境下进行实验。下面以一篇小文章为例说明建网的效果与得到的关键词。实验中设所提取的关键词的个数为 6， α 的值设为 0.15。主要是因为 PageRank 算法中的一般取的阻尼系数为 0.85，故在本文中把 α 的默认设为 0.15。

下面使用的三篇文章来自 Internet，除去后面的参考文献，其中一篇最短的文章总的字数是 700 个单词左右。根据网络科学理论的关键词提取方法，分别对它们进行建模，得到文本单词模型。这里选择其中一篇较短的文章的文本单词网络作为示例，如图 5-2 所示。

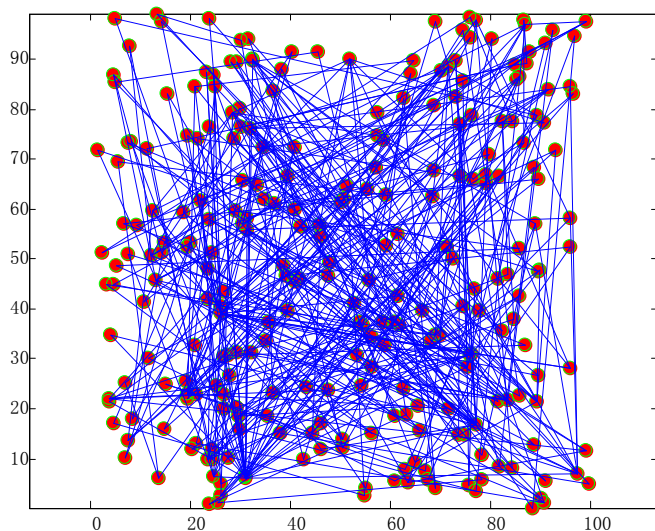


图 5-2 文本单词网络图

5.3.1 网络拓扑分析

本文提出了一种文本单词网络模型，那么研究该网络的拓扑特性，便是其中

必不可少的部分。下面分别从该网络的度分布特性以及网络鲁棒性两方面说明，文本单词网络模型具有无标度网络的特性。

5.3.1.1 度分布

由网络的度定义可知，文本单词网络总节点的度越高，则与其相邻的单词就越多，该节点所对应的单词就越容易和其他单词构成短语，在文本网络中就越重要。本节通过构建三种不同规模的文本单词网络，分别对网络中的各个节点的度进行统计分析。如图 5-3 所示。

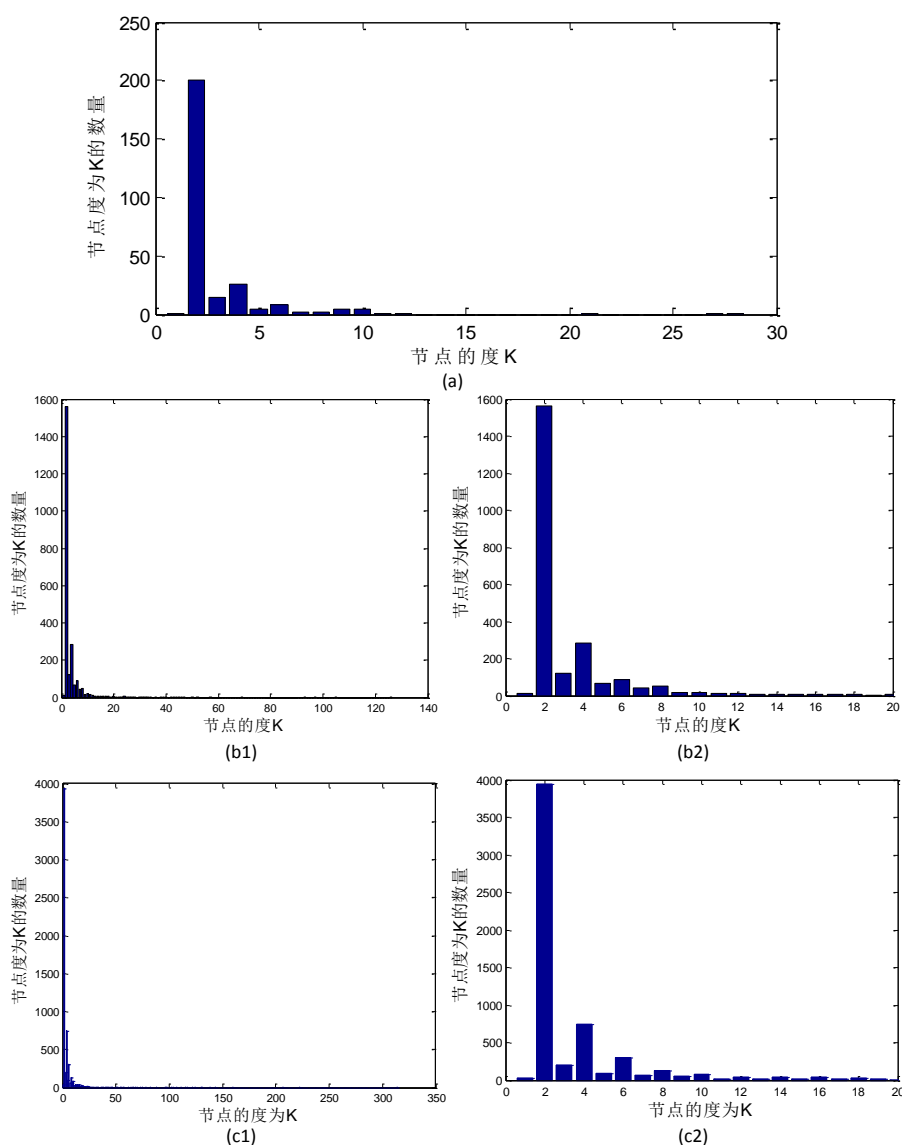


图 5-3 三种不同规模文本单词网络的度分布

(a)800 字文本；(b1)6000 字文本；(b2)6000 字文本细节图；(c1)30000 字文本；
(c2)30000 字文本细节图

从图 5-3 中可以看出，不论网络规模是多大，网络中度较低的节点占了很大的比例，而网络中只有少部分节点拥有较高的度。特别是网络中度为 2 的顶点占有很大比例，而度为 1 的顶点却很少。这是由构造网络的连接机制为单词位置相邻的节点之间才能连边，而文本中大部分的单词都有两个左右相邻的节点。

根据图 5-3，文本单词网络的度分布是具有“马太效应”的特点，符合长尾分布的。

为了进一步说明网络度分布符合幂律分布，利用最小二乘法拟合度分布的散列点，形成一条直线，其中该直线的斜率即为其幂律指数，求得度指数验证其的无标度特性。

5.3.1.2 最小二乘法求幂指数

复杂网络中分析网络度分布，通常是基于统计方法中的对数盒子法，该算法的主要思想是在双对数坐标下。首先取若干个中心点将它们作为一个盒子，然后对图形的散列点按距离中心最近原则，划分散列点属于哪个盒子，最后将每个盒子中所拥有的散列点纵坐标的值的和除以所包含的散列点个数，即得到每个中心点的频率，最后用平滑的直线连接各个中心点，即得到拟合后的直线。

对数盒子法的关键在于中心点的选取，这直接影响了拟合的效果，而中心的选取又没有很好的理论依据可循，往往是根据研究者的经验或者反复试验才能找准中心点。故本文采用最小二乘法将度分布图中的散列点进行直线拟合求得幂律值。

最小二乘法早在 1809 年就由著名数学家高斯提出，主要思想是让散列点的值与拟合所得的直线之间差值的 2-范数最小。

设所要拟合的直线为 $y = ax + b$ ，而实际的散列点表示为 (x_i, y_i) ，那么满足以下关系式即求得 a 和 b 的值：

$$\min_{a,b} \left\| \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} - \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \right\|_2 = \min_b \|Ab - Y\|_2 \quad (5-3)$$

该方程的解为：

$$a = \frac{\sum_i^n x_i y_i - n \cdot \bar{x} \bar{y}}{\sum_i^n x_i^2 - n \cdot \bar{x}^2} \quad (5-4)$$

$$b = \bar{y} - a\bar{x} \quad (5-5)$$

其中 $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ ，为 x 值的算术平方，故上式也可以写成以下形式：

$$a = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad (5-6)$$

下面基于最小二乘法对网络的度分布，进行拟合，求得其幂律指数，获取网络度指数。如图 5-4 所示，从上到下分别是 800 规模文本，6000 规模文本以及 3000 规模文本。

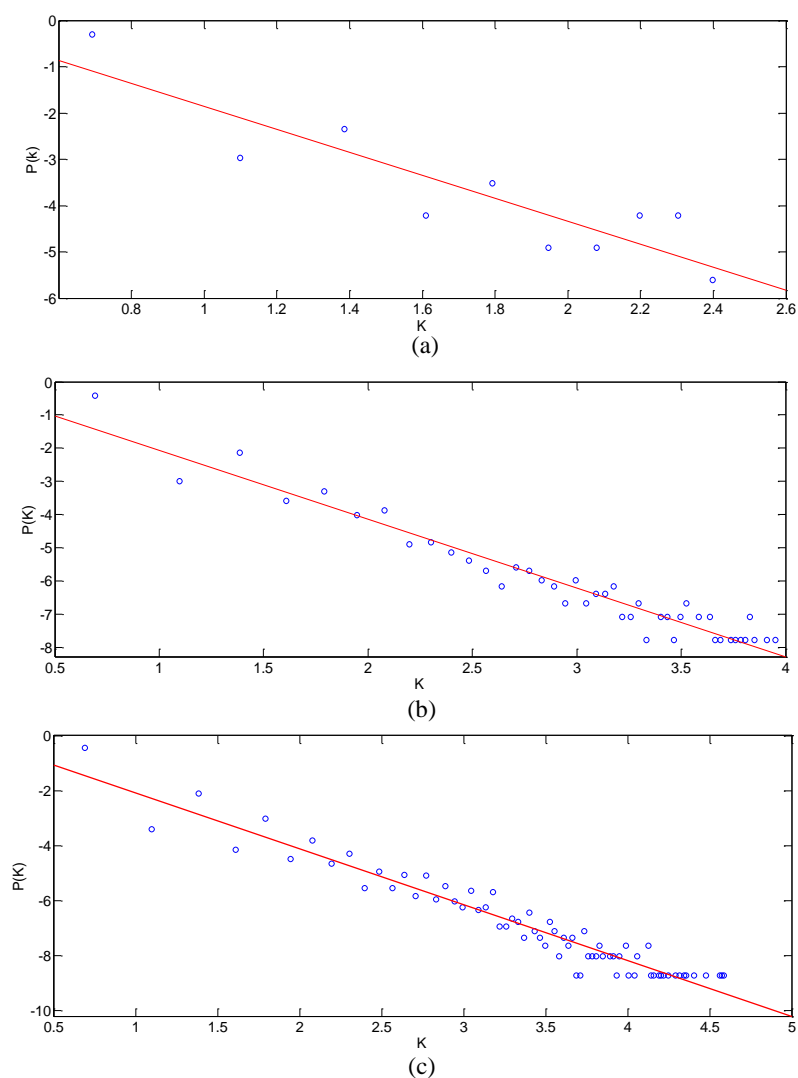


图 5-4 三种规模文本单词网络的度分布

在进行散列点模拟中，因为网络中大部分的单词都能与左右两边的单词产生联系，故我们把节点的度等于 1 的点当作噪声点。然而度很大的点很少，越是度数大的节点出现的频率也越低，会产生较大的稀疏，故实验中把 2% 的尾部散列点也作为噪声点。

从图 5-4 可以得出拟合网络度分布所得的直线，其中图(a)的斜率为 2.5，图(b)中直线的斜率为 2.6，而图(c)中所得直线斜率为 2.8。因为网络度分布所拟合的直线斜率为该网络的幂指数，说明文本单词网络的度指数是属于[2.5,3]这个范围，那么该网络度分布具有长尾分布特性，即无标度的特性。

5.3.1.3 鲁棒性

网络模型进行拓扑分析，必不可少的要对网络的鲁棒性进行分析，那么文本单词网络在面对网络攻击时，其自身调节以及修复能力如何，都是需要进一步探索的问题。遭受攻击的方式分为两种，一种是随意攻击，另一种是蓄意攻击，下面对网络攻击进行模拟实验，其中攻击的强度用删除网络中的节点数表示。如图 5-5 所示。

由图 5-5 可以看出，无论网络规模是多大，文本单词网络在受到蓄意攻击时，其平均路径长度随着攻击强度的增大，不断的增加。这是由于蓄意攻击的强度提升，网络中大度节点被不断的删除，导致连通性降低，平均路径长度逐渐增加。而对随机攻击，则表现平稳，表明就算增强了随机攻击的强度，其网络的平均路径长度也没有太大变化。原因是，由于网络是符合幂律分布，网络中度低的节点占相当大的比例，随机的删除网络中的节点往往选中的是那些度较低的节点，删除这些节点对整体网络没有太大的影响，故其平均路径长度几乎没什么变化，保持在一个较平稳的数值上。

综上所述，文本单词网络遭到蓄意攻击时其自身网络不能有很好的应对能力，一旦遭到强度大的蓄意攻击，整个网络将面临瘫痪的危险。也就是说，若文本删除那些度高的单词，文本将难以读懂，故文本网络对蓄意攻击没有抵抗能力，显示其脆弱性。相反，对于随机攻击文本单词，就算遭到强度很大的攻击，网络能保持其通畅性，有很强的免疫能力。对应文本来说，就是去掉文章中那些无关紧要的词，读者还是能顺利的读懂这篇文章，故文本网络对随机攻击体现出其鲁棒性特征。

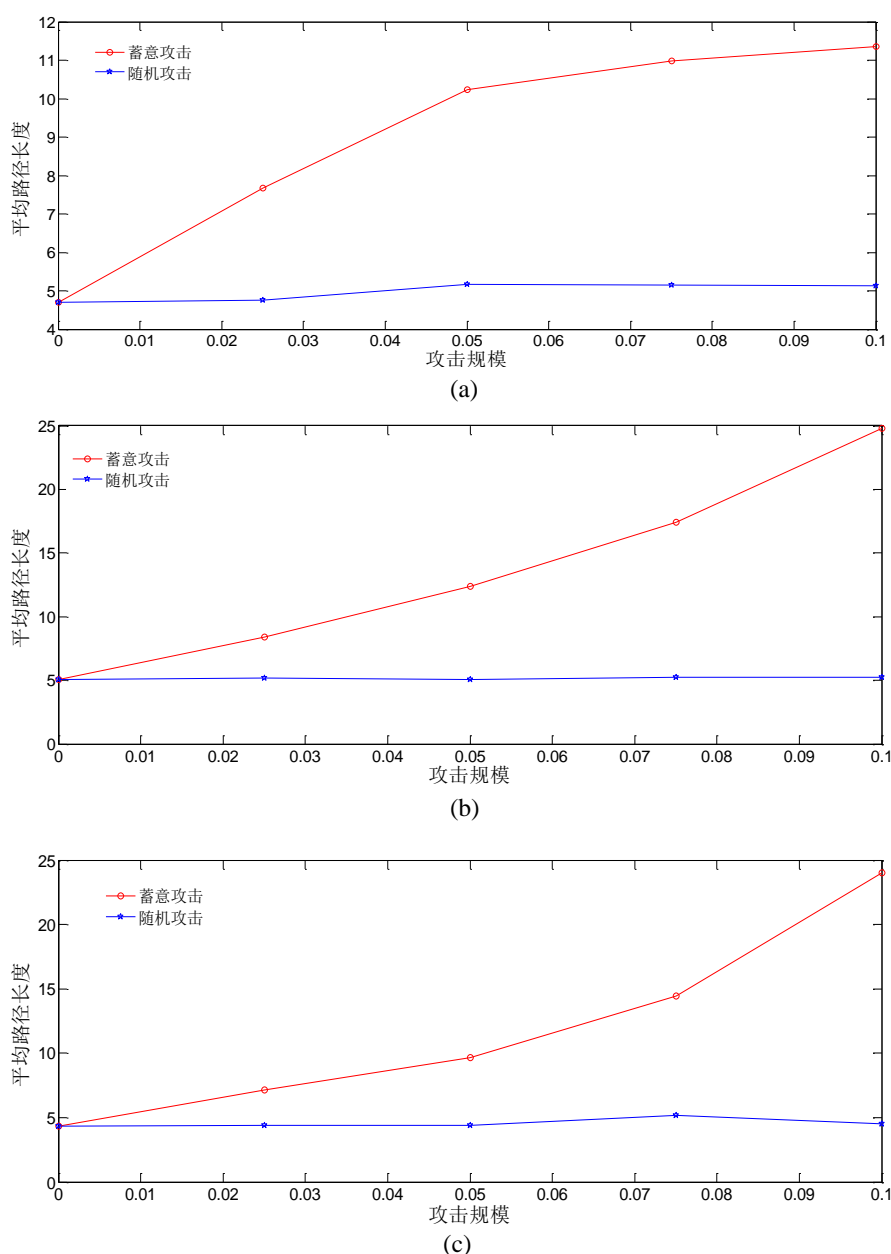


图 5-5 三种规模文本单词网络的鲁棒性分析

5.3.2 关键字提取实验

根据上文所介绍的关键字提取方法，本文分别针对不同规模的文本进行关键词的提取，得出每个文本的关键词以及通过这些所提取的关键词计算出本算法的准确率和召回率。

5.3.2.1 算法实验一

上小节实验主要是对文本单词网络整体进行拓扑分析，本节实验基于构造好

的文本单词网络，利用上述的算法，计算各个节点的综合重要程度。选取文本中的关键词。本实验均取文章中的五个关键词为实验结果，故设算法中的实验参数 $k=5$ 。

一下列出各个规模的文本中关键词所对应节点的度，节点的介数，以及综合重要程度。如表 5-1、表 5-2 以及表 5-3 所示。

表 5-1 800 字文本前 5 个关键词对应节点的拓扑值

特征值 \ 单词	juice	FDA	Brazilian	fungicide	level
节点的度	28	27	9	10	9
节点介数	0.5935	0.4898	0.2359	0.2096	0.1910
综合重要程度	0.5052	0.4171	0.2012	0.1789	0.1631

表 5-2 6000 字文本前 5 个关键词对应节点的拓扑值

特征值 \ 单词	network	node	search	HDLS	random
节点的度	67	55	46	33	16
节点介数	0.2712	0.2280	0.2063	0.0927	0.0903
综合重要程度	0.2306	0.1938	0.1754	0.0789	0.0768

表 5-3 30000 字文本前 5 个关键词对应节点的拓扑值

特征值 \ 单词	user	recommendation	data	network	similarity
节点的度	222	160	130	115	143
节点介数	0.1834	0.1127	0.0831	0.0809	0.0798
综合重要程度	0.1559	0.0958	0.0706	0.0688	0.0678

表中所提取出来的关键词，基本符合人工提取的结果。基于复杂网络的关键词提取方法，除了发现能表达文本的中心思想的单词外，对网络中单词与单词之间的联系，也有很好的体现。实验中，也验证了 α 的取值为 0.15 是一个合理的决策。

5.3.2.2 算法实验二

为了进一步的验证算法的有效性，实验从互联网中下载了 50 篇英文文章，提

取关键字，并引入准确率和召回率作为评判的标准。

准确率以及召回率可以用以下公式计算：

$$P = \frac{|key \cap DB|}{|DB|} \quad (5-7)$$

$$R = \frac{|key \cap DB|}{|key|} \quad (5-8)$$

其中 P 代表准确率， $|key|$ 表示人工提取的结果， $|DB|$ 表示根据本文所提出方法提取的关键词。 $|key \cap DB|$ 表示人工结果和本文提出的算法所得结果之间的交集。

同理，可以得到平均准确率和平均召回率的计算公式：

$$\bar{P} = \frac{1}{N} \sum_{i=1}^N P_i \quad (5-9)$$

$$\bar{R} = \frac{1}{N} \sum_{i=1}^N R_i \quad (5-10)$$

其中 N 代表文本个数， P_i 代表文本 i 的准确率， R_i 代表文本 i 的召回率。

对所选定的文章首先进行格式转换，然后对文本进行预处理，通过实验得出了这些文本的关键词，分别对各个文本进行准确率 P 和召回率 R 的计算，得出所有的数据，进行平均准备率和平均召回率的计算。再对相同的文章基于 TF、TF-IDF 两个算法得出平均准确率和平均召回率。实验结果对比如表 5-4。

表 5-4 实验结果对比

指标值 \ 算法	TF	TF-IDF	本文方法
\bar{P}	0.369	0.387	0.585
\bar{R}	0.635	0.648	0.763

由表 5-4 说明，实验数据比较符合客观事实，与人工提取的关键词有较大的相似性，并且比 TF、TF-IDF 算法提取的精确率分别有 58.5% 以及 51.1% 的提升。说明新的算法的效果还是比较理想的。

但本文所提的算法没有达到预期的效果，造成这样结果的主要原因在于单词与单词之间的语义关系没有很好的办法抽象出来。在自然语言处理中也同样很棘手，目前还没有很好的方法解决这个问题。

5.4 本章小结

本章主要提出了一种基于复杂网络方法对自然语言进行研究与建模的方法，利用构造的模型，基于复杂网络的研究方法对其进行拓扑分析以及运用算法进行关键词的提取，验证文本单词网络具有无标度网络的特性，其幂指数接近于 3，表示具有无标度特征。然后运用复杂网络节点的拓扑特征，提出一种新的关键字提取的方法——利用所构造的网络中的节点，利用节点的度和其中心介数，综合考虑得出衡量节点的综合重要性的指标，据此提取关键词。最后通过基于本文所提新的关键词提取方法对大量的文本进行试验，计算出平均准确率和平均召回率，用数据验证了，节点的拓扑特征中的节点的介数是一个具有很强现实意义的重要特征值，同时也验证了本文所提的基于网络科学理论的关键词提取方法是一个合理的、有效的方法。

第六章 全文总结与展望

6.1 全文总结

本文工作的重点是研究复杂网络的建模方法以及将网络科学理论运用在实际的复杂系统结构中。通过熟悉网络科学相关的基础理论知识，以及研究大量的文献之后，对网络的模型提出更深层次的思考。结合挖掘出的现实世界中的复杂系统结构特点，利用网络科学理论对这些实际网络进行建模与分析。

本文的主要工作总结如下：

简要的介绍了网络科学发展历史，当今学者为网络科学的发展所贡献的力量，以及今后的发展方向。

详细介绍了网络科学相关的基础理论知识、其基本的研究方法以及网络科学发展过程中学者提出的几个经典模型。

在 BA 模型的基础上，本文提出了两种新的无标度网络模型，分别是基于两种常见的随机分布——均匀分布、高斯分布进行构建的。这两种模型较之经典的 BA 模型更加切合实际网络具有的并非匀速增长时间间隔。其中均匀更新计数网络的网络节点个数的增加的时间间隔是均匀分布的，而后者网络节点增长时间间隔是符合高斯分布的。为了验证这两种新的网络模型的合理性，本文通过平均场方法进行理论的推导计算其幂律指数，并且通过仿真构造出这两种网络模型通过分析其拓扑特性，证明所提出的模型确实是符合实际网络的无标度网络。

通过对电力交易的了解与分析，利用复杂网络理论对电力交易问题进行建模，取得电力交易买卖双方的电力交易网，通过对不同规模的电力交易进行仿真实验，分析实验数据，得出电力交易网络的特性，以及通过理论推导验证其幂指数接近 3，验证电力交易网络也具备无标度特性。同时利用所得的结论，为电力交易市场的改进以及基础设施的建设提供合理的依据。

提出了一种基于复杂网络方法对自然语言进行研究与建模的方法，利用文本中单词与单词之间的邻近关系构造文本网络。同时对不同规模的文本进行仿真实验，对文本网络的拓扑特性主要是鲁棒性进行分析与研究，同时利用最小二乘法对文本网络进行幂指数的计算，求得其幂律指数属于[2.5,3]，证明其也是一个无标度网络。并且利用节点中心介数以及节点的度，综合衡量单词在文本中的重要程度，进而提出全新的关键词提取方法。

6.2 展望

目前，网络科学理论的研究仍处于飞速的发展的阶段，怎样提出更加切合现实网络特征的模型是势在必行的问题。本文的工作内容虽然取得了一定的成果，但是还存在几点不足之处需要进一步的研究和改进，如下：

- (1) 实际网络规模不仅仅会增加，相反在一些情况下还会减少。本文所提出的网络模型，只考虑到了实际网络的非匀速增加的特性，对于网络节点个数的减少没有很好的方法对其进行模拟建模。
- (2) 对于电力交易市场，本文只是考虑买卖双方的交易，对于电力的传输，如何分配电力额度到各个地区使得代价最少，以及如果利用电力交易网络的特征对电力市场基础设施的建立提供更加实用的意见。
- (3) 在对自然语言进行建模中，本文所提出的新提取方法，效果并不是如预期的那么理想。如何加入除中心介数以及节点度之外的关键词提取的因素，以获得更加理想的效果。

以上这些点相信随着科学的发展，人们不断的努力都可以加以解决，使得复杂网络理论能更好的服务于人们。

致 谢

研究生三年的求学生涯，是那么的美好与短暂，来不及对它道别就要结束。在这三年里，除了收获了学业上的进步、结识了如家人般的朋友，同时也收获了很多人生中的宝贵经验。在即将离开学校进入新的人生阶段时，心中除了怀有对未来生活的向往，更多的是对于即将告别校园生活以及朝夕相处的老师和同学们的感伤。

在此需要对在这三年中给予我帮助、温暖的老师与朋友们致以满心的感激之情。首先，需要感谢我的导师——****教授，他不光在学术上给予我细心的指导，而且在生活上也予以无微不至的关心。再此也特别需要感谢，我所在研究小组的师兄，是他带领我进入复杂网络科学的大门，同时也在论文的撰写中给我提出了很多宝贵的意见和建议。同时也感谢我所在实验室的全体同学，是他们让实验室有了家的温暖。最后需要感谢我的亲人，在我困惑时给予我温暖；在我迷茫时给予我支持；在我成功时给予我掌声。

参考文献

- [1] 官山, 何大韧, 朱陈平. 跨越多个实际科学领域的合作网络与合作-竞争网络[J]. 力学进展, 2008, 38:827-834.
- [2] 袁娟, 张宁. 企业竞争网络的拓扑结构分析[J]. 上海理工大学学报, 2007, 29(1):37-41.
- [3] 保罗霍夫曼. 数字情种—埃尔德什传. 米绪军[M]. 章晓燕, 廖东卫译. 上海: 上海科技教育出版社, 2000.
- [4] P. Erdős, A. Renyi. On the Evolution of Random Graphs[J]. Pub. Math. Inst. Hung. Acad. Sci. 1960, 5:17-61.
- [5] A. L. Barabási. Scale-free networks: a decade and beyond[J]. Science, 2009, 325:412-413.
- [6] S. Milgram. The small world problem[J]. Psychology Today, 1967, 1:61-67.
- [7] S. H. Strogatz. Exploring complex networks[J]. Nature, 2001, 410: 268-276.
- [8] D. J. Watts, S. H. Strogatz. Collective dynamics of small-world networks[J]. Nature, 1998, 393(6684): 440-442.
- [9] M. E. J. Newman, D. J. Watts. Renormalization group analysis of the small-world network model[J]. Physics Letters A, 1999, 263(4): 341-346.
- [10] A. L. Barabási, R. Albert. Emergence of scaling in random networks[J]. Science, 1999, 286(5439):509-512.
- [11] S. N. Dorogovtsev, J. F. F. Mendes. Effect of the accelerating growth of communication networks on their structure[J]. Physical Review E, 2001, 63: 25-101.
- [12] 胡海波, 王科, 徐玲, 汪小帆. 基于复杂网络理论的在线社会网络分析[J]. 复杂系统与复杂性科学, 2008, 2:1-14
- [13] D. J. Watts. Six Degree: The Science of a Connected Age[C]. New York: Norton, 2003.
- [14] L. Backstrom, D. Huttenlocher, J. Kleinberg, L. Xiangyang. Group formation in large social networks: Membership, growth, and evolution[C]. New York: ACM Press, 2006: 44-54.
- [15] J. Kleinberg. The small-world phenomenon: an algorithm perspective[C]. Proceedings of the thirty-second annual ACM symposium on Theory of computing. ACM, 2000: 163-170.
- [16] 王小帆, 李翔, 陈关荣. 复杂网络理论与应用[M]. 清华大学出版社, 2006.
- [17] 姚灿中, 杨建梅. 幂律拟合的进展及其在产业网络中的应用[N]. 管理学报, 2008, 3:371-375.
- [18] M. E. J. Newman. The structure and function of complex networks[J]. SIAM Review, 2003, 45:50-59.
- [19] B. Bollobás. Random Graps[C]. New York: Academic Press, 2nd ed. , 2001.

-
- [20] X. F. Wang, G. R. Chen, Synchronization in small-world dynamical networks[J]. *Int. J. Bifurcation and Chaos* 2002, 12:187-192
- [21] Y. Fan, M. Li, J. Chen, L. Gao, Z. Di, J. Wu. Network of Econophysicists: A Weighted Network to Investigate the Development of Econophysics[J]. *Int. J. Mod. Physical B*.2004,18: 2505-2511.
- [22] A. L. Barabási, R. Albert. Emergence of scaling in random networks[J]. *Science*,1999, 286(5439):509-512.
- [23] D. H. Shi, H. J Zhou, L. M Liu. A discussion of Barabási and Albert's 1999 paper[J]. *Physics Procedia*,2010,3:1767-1774.
- [24] S. N. Dorogovtsev, J. F. F. Mendes, Scaling behavior of developing and decaying networks[J]. *Europhysics Letters*,2000,52:33-39.
- [25] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, A. L. Barabási. The largescale organization of metabolic networks[J]. *Nature*, 2000, 407: 651-654.
- [26] S. Milojevic. Power-law distributions in information science: Making the case for logarithmic binning[J]. *Journal of the American Society for Information Science and Technology*, 2010, 61, 12: 2417-2425,.
- [27] A. Clauset, C. R. Shalizi, M. E. J. Newman, Power-law distributions in empirical data[J]. *SIAM review*,2009, 51:661-703.
- [28] R. Albert, H. Jeong, A. L. Barabási, Attack and error tolerance of complex networks[J]. *Nature*, 2000, 406:378-382.
- [29] J. M. Carlson, J. Doyle. Highly optimized tolerance: Robustness and design in complex systems[J]. *Physical Review Letters*, 2000, 84: 25~29.
- [30] A. L. Barabási. Scale-free networks: a decade and beyond[J]. *science*, 2009, 325:412-413.
- [31] J. W. M. Cheng, F. D. Galiana, D. T. McGillis, Studies of bilateral contracts with respect to steady-state security in a deregulated environment[J]. *IEEE Trans. Power Syst.*,1998,13:1020-1025.
- [32] A. K. David, F. S. Wen, Bilateral transaction bargaining between independent utilities under incomplete information[J]. *Proc. Inst. Elect, Eng, Gen, Transm*, 2001, 148:448-454.
- [33] S. Soft. Power system economics: designing markets for electricity[C]. Wiley-IEEE Press, New York, 2002.
- [34] W. B. Arthur. Complexity and economy[J]. *Science* 1999,284:107-109.
- [35] J. Bower, D. Bunn. A model based comparisons of pool and bilateral markets for electricity[J]. *Energy J* ,2000,21:1-29.

- [36] State Grid Corporation of China, State Grid Corporation of China 2012 Power Exchange Annual Report[R]. Bei Jing: China Electric Power Press, 2013.
- [37] State Grid Corporation of China, State Grid Corporation of China 2007 Power Exchange Annual Report[R]. Bei Jing: China Electric Power Press, 2008.
- [38] State Grid Corporation of China, State Grid Corporation of China 2009 Power Exchange Annual Report[R]. Bei Jing: China Electric Power Press, 2010.
- [39] State Grid Corporation of China, State Grid Corporation of China 2010 Power Exchange Annual Report[R]. Bei Jing: China Electric Power Press, 2011.
- [40] State Grid Corporation of China, State Grid Corporation of China 2011 Power Exchange Annual Report[R]. Bei Jing: China Electric Power Press, 2012.
- [41] J. Bo, H. T. Teng, Y. J. Shen, F. Z. Qu. Chinese patent mining based on sememe statistics and key-phrase extraction[J]. Proc.ADMA Conference, 2007,56:27-43.
- [42] H. Jiao, Q. Liu, H. B. Jia. Chinese keyword extraction based on N-gram and word cooccurenc[J]. International Conference on Computational Intelligence and Security Workshops,2007,30(3):20-39.
- [43] 施聪莺. TFIDF 算法研究综述[J]. 计算机应用, 2009,29(6):167-180.
- [44] 赵鹏. 一种基于复杂网络特征的中文文档关键词抽取算法[J]. 模式识别与人工智能,2007, 20 (6):827-831.
- [45] 马力. 基于小世界模型的复合关键词提取方法研究[J]. 中文信息学报, 2009,23(3):121-128.

攻读硕士学位期间取得的成果

参与项目：

主研《高性能推理引擎技术》 来源：中电 29 所 项目编号:2013006255.

申请专利：

申请编号：201410305704

发明名称：一种优化训练样本集的 KNN 文本分类方法.

获得奖励：

2012-2013 年度计算机科学与工程学院研究生三等奖.

2013-2014 年度计算机科学与工程学院研究生一等奖，"优秀研究生"称号.

2014-2015 年度计算机科学与工程学院研究生二等奖.