

电子科技大学

UNIVERSITY OF ELECTRONIC SCIENCE AND TECHNOLOGY OF CHINA

# 专业学位硕士学位论文

MASTER THESIS FOR PROFESSIONAL DEGREE



论文题目 针对重要节点的图挖掘算法与技术研究

专业学位类别 工 程 硕 士

学 号 201822080331

作 者 姓 名 游慧明

指 导 教 师 刘 震 副教授

分类号 \_\_\_\_\_

密级 \_\_\_\_\_

UDC <sup>注1</sup> \_\_\_\_\_

# 学 位 论 文

针对重要节点的图挖掘算法与技术研究

(题名和副题名)

游慧明

(作者姓名)

指导教师 刘 震 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 专业学位类别 工 程 硕 士

工程领域名称 计算机技术

提交论文日期 2021.03.20 论文答辩日期 2021.06.01

学位授予单位和日期 电子科技大学 2021 年 06 月

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。

# **Research on Graph Mining Algorithm and Technology for Important Nodes**

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

Discipline: **Master of Engineering**

Author: **Huiming You**

Supervisor: **Prof.Zhen Liu**

School: **School of Computer Science & Engineering**



## 摘要

在现实生活中,存在着丰富多彩的复杂系统,例如航空运输系统,海上运输系统等。为了更好地研究种种复杂系统的特性,我们人为地将复杂系统抽象成便于理解记忆的网络结构,这就是复杂网络。在复杂网络的网络结构中,通常会存在一些对维持网络结构和功能的完整性都有着至关作用的节点。一旦这些重要节点受到攻击,网络将会遭到毁灭性的破坏。基于此,本文的工作将围绕着两点展开。一是如何对网络中的重要节点进行挖掘,二是如何对网络的鲁棒性进行优化。

对于本文的第一项工作,主要的贡献是首次将图神经网络算法运用在节点挖掘上,提出了一种带权的图自编码方法(Weighted Graph Autocoding method, WGA),对网络中的重要节点进行挖掘。WGA 方法挖掘重要节点的过程如下:首先,构建网络模型,使用网络模型的权重矩阵作为输入,通过自编码器重构权重矩阵;其次,计算权重矩阵的编码损失和重构损失,从而计算整体损失;然后,输出每个节点的特征向量,利用节点特征向量与其他节点的特征向量的欧氏距离之和来评价该节点的重要性,得到节点重要性排序序列;最后,使用非关键节点比例(Proportion of Noncritical Nodes, PNN)指标,对 WGA 方法得到的节点重要性排序序列进行评价。作为对比实验,本文还使用了 Node2vec 算法、度中心性指标、介数中心性指标、PageRank 算法、DIL(degree value and the importance of lines)方法和 NL(neighbor nodes and the importance of lines)方法对网络中的重要节点进行了挖掘。最后,本文还在真实情境下分析 WGA 方法所挖掘的重要节点的可靠性。

对于本文的第二项工作,可以看作是第一项工作的延伸,主要的贡献是结合了基尼系数,重新定义了网络鲁棒性的评价指标,并提出了基于贪心和动态规划的增边方法优化网络的鲁棒性。具体优化过程如下:首先,选择最重要节点的邻居节点中重要性排名前 20%的节点构成集合 $H$ ,随机选择 $H$ 中节点构成边的 30%作为候选边集,在原网络中移除候选边集,得到需要进行鲁棒性优化的网络模型;其次,分析基尼系数指标,得出网络增边后导致的基尼系数增量,基尼系数增量越大,网络鲁棒性优化方法越好;最后,通过贪心算法和动态规划算法在限制增边条数或增边容量的条件下在候选边集中选择边对网络进行优化,使得网络整体的基尼系数增长得最快。作为对比,本文选取了在度值和最大的节点间增边的方法以及在介数值最大的节点间增边的方法对网络进行鲁棒性优化。

**关键词:** 复杂网络, 图神经网络, WGA 方法, 重要节点挖掘, 网络鲁棒性优化

## ABSTRACT

In real life, there are colorful complex systems, such as air transportation system, sea transportation system, etc. In order to better study the characteristics of various complex systems, we artificially abstract complex systems into a network structure that is easy to understand and remember, which is a complex network. In the network structure of complex networks, there are usually some nodes which are vital to maintain the integrity of network structure and function. Once these important nodes are attacked, the network will suffer devastating damage. Based on this, the work of this thesis will revolve around two points. One is how to mine the important nodes in the network, and the other is how to optimize the robustness of the network.

For the first work of this thesis, the main contribution is that the Graph neural network algorithm is applied to node mining for the first time, and a Weighted Graph Autocoding method (WGA) is proposed to mine the important nodes in the network. The process of WGA method for mining important nodes is as follows: first, the network model is constructed, the weight matrix of the network model is used as the input, and the weight matrix is reconstructed by the autoencoder; second, the coding loss and reconstruction loss of the weight matrix are calculated, so as to calculate the overall loss; then, the eigenvectors of each node are output, and the importance of this node is evaluated by the sum of Euclidean distances between the eigenvectors of nodes and those of other nodes, and the ranking sequence of the importance of nodes is obtained; finally, the Proportion of Noncritical Nodes (PNN) index is used to evaluate the node importance ranking sequence obtained by the WGA method. As a contrast experiment, this thesis also uses Node2vec algorithm, degree centrality index, intermediate centrality index, PageRank algorithm, DIL (degree value and the importance of lines) method and NL (neighbor nodes and the importance of lines) method to mine the important nodes in the network, and obtains the corresponding PNN results. Finally, the reliability of the important nodes mined by WGA method is analyzed in the real situation.

The second work in this thesis can be seen as an extension of the first work, and the main contribution is that the evaluation index of network robustness is redefined by combining the Gini coefficient, and an edge-adding method based on greedy and dynamic programming is proposed to optimize the robustness of the network. The specific

optimization process is as follows: first, the top 20% of the most important node's neighbor nodes are selected to form the set H, and 30% of the nodes in H are randomly selected to form the candidate edge set, the candidate edge set is removed from the original network to obtain the network model requiring robust optimization; second, the Gini coefficient index is analyzed, and the increment of the Gini coefficient is obtained, the larger the increment of the Gini coefficient is, the better the network robustness optimization method is; finally, the greedy algorithm and dynamic programming algorithm are used to optimize the network by selecting edges from the candidate edge set under the condition of limiting the number of edges added or the capacity of edges added, so that the Gini coefficient of the whole network grows the fastest. As a contrast, this thesis selects the method of adding edges between the nodes with the maximum degree value and the method of adding edges between nodes with the maximum dielectric value to optimize the robustness of the network.

**Keywords:** complex network, graph neural network, WGA method, key node mining, network robustness optimization

## 目 录

第一章 绪 论 .....	1
1.1 研究背景与意义 .....	1
1.2 国内外研究现状 .....	4
1.2.1 重要节点挖掘的研究现状 .....	4
1.2.2 网络鲁棒性优化研究现状 .....	4
1.3 本文主要工作 .....	6
1.4 本文的结构安排 .....	7
1.5 本章小结 .....	8
第二章 相关理论与技术 .....	9
2.1 复杂网络基本特征 .....	9
2.2 复杂网络基本模型 .....	11
2.2.1 规则网络模型 .....	11
2.2.2 随机网络模型 .....	12
2.2.3 小世界网络模型 .....	12
2.2.4 无标度网络模型 .....	13
2.3 网络中重要节点挖掘方法 .....	14
2.3.1 PageRank 算法 .....	15
2.3.2 DIL 方法 .....	16
2.3.3 NL 方法 .....	16
2.4 图神经网络相关理论 .....	16
2.4.1 Node2vec 算法介绍 .....	18
2.4.2 SDNE 算法介绍 .....	20
2.5 增边优化网络的相关理论 .....	23
2.6 方法评价指标介绍 .....	23
2.7 本章小结 .....	25
第三章 基于图神经网络算法的重要节点挖掘方法 .....	26
3.1 WGA 方法对网络中重要节点的挖掘 .....	26
3.1.1 构建重要节点挖掘的网络模型 .....	26
3.1.2 使用 WGA 方法学习网络中节点的特征向量 .....	28
3.1.3 对网络中的节点进行重要度评分 .....	29



3.1.4 评估节点排序序列的可靠性 .....	30
3.2 WGA 方法的实验及实验分析 .....	32
3.2.1 实验数据及实验环境 .....	32
3.2.2 重要节点挖掘的实验结果及分析 .....	34
3.2.2.1 PNN 实验结果及分析 .....	34
3.2.2.2 WGA 方法挖掘重要节点的可靠性分析 .....	38
3.3 本章小结 .....	41
第四章 网络鲁棒性优化方法 .....	42
4.1 网络鲁棒性的优化 .....	42
4.1.1 构建网络鲁棒性研究的模型 .....	42
4.1.2 网络鲁棒性指标的分析 .....	44
4.1.3 基于贪心策略的网络增边优化方法 .....	47
4.1.4 基于动态规划的网络增边优化方法 .....	49
4.1.5 基于节点的度和介数值的网络增边优化方法 .....	49
4.2 增边优化网络鲁棒性的实验及实验分析 .....	50
4.2.1 动态规划增边策略与贪心增边策略的对比 .....	50
4.2.2 动态规划增边策略与其他的增边策略对比 .....	56
4.3 本章小结 .....	59
第五章 全文总结与展望 .....	60
5.1 全文工作总结 .....	60
5.2 本文主要贡献 .....	60
5.3 未来工作展望 .....	61
致 谢 .....	62
参考文献 .....	63
攻读硕士学位期间取得的成果 .....	67

## 第一章 绪论

### 1.1 研究背景与意义

现如今，复杂网络已经很快地形成了一门贯穿多领域的交叉型学科，其相关理论被诸多领域所应用。首先，人们需要明白复杂网络的定义。复杂网络的理论探索最早是来自“七桥问题”，“七桥问题”是图论中的经典问题。这个问题来源于生活中的一件事情：在 18 世纪，有一条河流经过柯尼斯堡，两个岛屿在河流的中间，有七座桥梁连接了两个岛屿和河流的两岸。基于此，欧拉猜想，是否有这样一种方式通过全部的七架桥梁而每架桥梁只能通过一次？在这个问题中，数学家欧拉将两个岛屿和河的两岸抽象成了节点，将连接岛屿和两岸的桥梁抽象成边，节点与连接点的边就组成了网络，如图 1-1 所示。

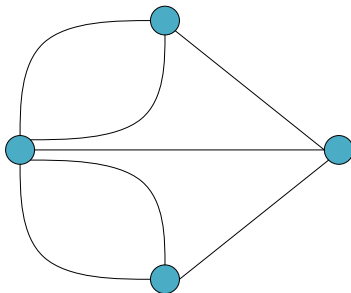


图 1-1 七桥问题

“七桥问题”的提出使得复杂网络理论的研究慢慢进入了科学家研究的视野。复杂网络的研究是具有非常现实的研究价值的，因为在现实生活中，存在着丰富多彩的复杂系统，例如电力网络系统，社交关系网络，航空物流网络，生物学中的食物网系统等等。为了更好地研究种种复杂系统的特性，分析复杂系统的组成结构和变化规则，我们人为地将复杂系统抽象成便于理解记忆的网络结构。但是，几乎所有的网络都是由一些节点和连接这些节点的边组成，其中节点一般用来表示复杂系统中的真实存在，而边则是用来表示这些存在的关系。通常将两个有特定关系的节点使用一条边相连在一起，如果节点之间不存在关系，则为相互独立的两个点。相邻节点在网络中指的是在网络中用边直接相连的节点。在网络中的节点和边可以赋予很多信息，例如网络节点可以有权值，网络的边是有向边，网络就自然就形成了一种带权网络。这样的网络更详尽的表示了复杂系统的组成结构关系。虽然将系统抽象成网络会忽略掉实际的一些细节，但也以一种统一的方式对不同的系统的研究提供了方法。正是由于这种统一的方法，复杂网络这门学科应运而生。复杂

网络学科包含了很多领域的理论知识，特别是在数学、计算机等领域，同样也包含了很多通用的方法，例如图论、随机理论、算法理论分析等等。

而复杂网络有着不同于简单网络的结构特性，简单网络仅仅是用于分析一些规则化的网络的特性，例如随机图等等。而复杂网络几乎都是由真实地复杂系统演变而成的，所以复杂网络的研究与现实世界密不可分。所以，对复杂网络的探索走在了当前科学探索的前沿。

现在，复杂网络领域包含了很多的研究方向，例如针对网络中重要节点的挖掘和鲁棒性优化等方向。大量可靠的研究结果极大地促使了复杂网络科学的快速进步。可以用于实验研究的数据也随着信息科学的快速进步出现了海量化的特点，使得社会走进了大数据时代<sup>[1-2]</sup>。比如，朱涵等<sup>[3]</sup>在很早的时候解释了什么是复杂网络理论，周涛等<sup>[4]</sup>提出了复杂网络研究所涉及的十大问题，何宇等<sup>[5]</sup>从多个方面详细的介绍了复杂网络的演化过程并分类归纳，很多研究人员对复杂网络理论加以系统化的总结归纳<sup>[6]</sup>。

复杂网络作为一门研究性的科学，其相应的也会出现很多用于分析复杂网络的模型。例如，在一个星型网络<sup>[7]</sup>中，存在一个中心节点与其他所有节点相连，而其他节点之间没有直接的边相连；Erdos 与 Renyi 最早根据随机图理论构建了随机网络模型(ER)网络<sup>[8]</sup>，ER 网络是由网络随机删除边而形成的，在 ER 网络中任意两个节点之间关联的概率是相等的。现如今，也有几种非常经典的模型，例如，小世界网络模型<sup>[9]</sup>，NW 模型<sup>[10]</sup>以及无标度网络模型<sup>[11]</sup>等等。

这些网络都是抽象的网络模型。而在现实生活中，随着科技的发展和社会的进步，更大规模的网络应运而生。例如，在信息交流传播中的互联网，日常生活中的电力网，在日常出行方面上发挥关键作用的交通网，在人们沟通交流上发挥重要作用的社交网络。而在其他领域，复杂网络的研究方法也同样适用。例如，生物学中的食物网<sup>[12]</sup>，如图 1-2 所示；生物体内蛋白质错综复杂连接形成的网络<sup>[13]</sup>，如图 1-3 所示等等。

本文的两个主要工作是，挖掘网络中的重要节点和优化网络鲁棒性。

挖掘网络中的重要节点在科学研究中有着巨大的研究价值，在日常生活中也有着很强的现实意义。例如在航空运输网络中，我们需要关心的是飞机常常经过的城市、航空公司和航线；在城市交通网络中，我们主要关心的是出于交通枢纽的公交站点；在社交网络中，我们主要关注的是那些具有核心媒介作用的人或者团体等等。因此，挖掘网络中的重要节点是非常有必要的。

优化网络鲁棒性可以看作是第一个工作的延伸。当我们挖掘到网络中的重要节点时，我们需要关注的就是，由于网络中这些重要节点被攻击而导致网络结构或

者功能的损毁，我们该怎么办。因此，保护重要节点，降低重要节点在网络中的重要性，从而对网络鲁棒性进行优化也是非常值得研究的。

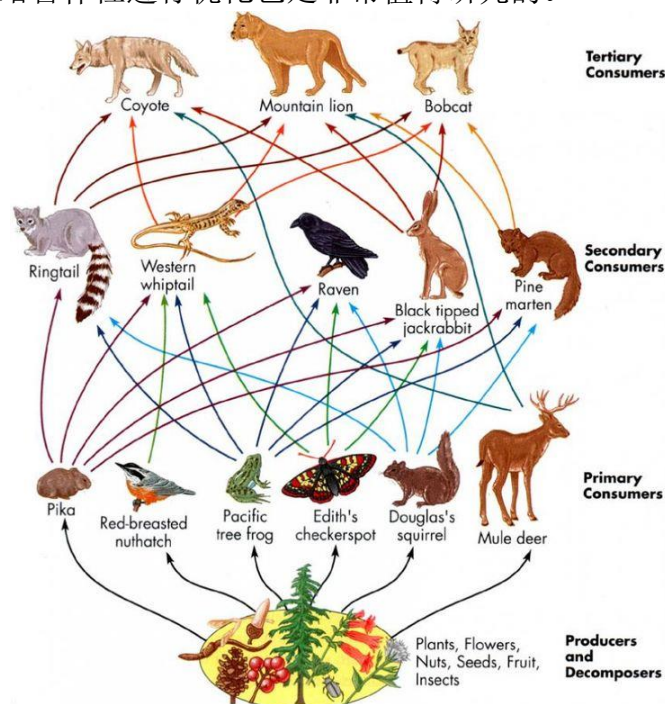


图 1-2 食物网<sup>[12]</sup>

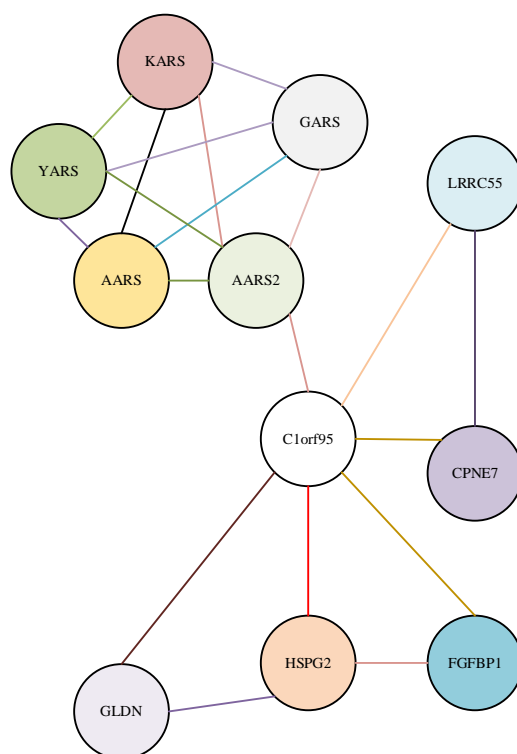


图 1-3 蛋白质网络<sup>[13]</sup>

## 1.2 国内外研究现状

近年来,复杂网络的研究得到了广泛的关注,复杂网络的多个分支研究方向在多种研究领域取得了极大的成果。而在本文中,我们主要关注的点有两个:一是对网络中的重要节点的挖掘,二是保护重要节点,从而优化网络的鲁棒性。

### 1.2.1 重要节点挖掘的研究现状

对于本文研究工作的第一点,对网络中的重要节点的挖掘也是当前复杂网络方向研究的前沿话题,有许许多多的研究人员正投入其中。针对网络中的重要节点的挖掘,就是针对节点进行节点重要度排序。在节点重要度排序的问题上,我们需要找到一种统一的定量分析的方法对不同的网络中的节点进行节点重要度评价。这种统一的方法不受网络结构的影响。我们通常会提出某种指标,用这个指标的数值来作为网络节点的重要度评分。通常数值越大,表示节点在网络中的重要程度也越高。常见的几种指标可以分成三种:第一种是依赖网络结构的局部属性的指标,第二种是依赖网络结构的整体属性的指标,第三种是依赖随机游走算法的指标。

依赖网络结构的局部属性的指标主要考虑的是节点和相邻节点信息,常见的有度中心性指标<sup>[14]</sup>和半局部中心性指标<sup>[15]</sup>等等。简单来说,度中心性(Degree Centrality)指标通常可以很好的解释,假如有这样一个社交圈,某个人和社交圈中的其他很多人都有直接的关系,那么可以说明这个人在这个社交圈中属于核心人物。换句话说,如果某个人在社交圈中的交往范围越广,那么这个人在社交圈中也就越重要。人就相当于网络中的节点。他认识的人的数量反映他在社交圈中的地位。而半局部中心性指标比起度中心性指标来说,增加了节点多阶邻居信息,计算指标更有说服力。所以,半局部中心性指标在评价节点重要性上更为可靠,但是节点的邻居数目是非常庞大的,也会导致计算量的剧增。总而言之,依赖网络结构的局部属性的指标仅仅考虑了网络的局部属性,优点是计算简单,复杂度低,适用于大型网络,缺点是没有考虑网络整体的结构,会出现评价结果的不准确。

依赖网络结构的整体属性的指标主要考虑的是网络结构的整体信息,比如介数中心性<sup>[16]</sup>和特征向量中心性<sup>[17]</sup>等等。介数中心性指标评价的是网络中某个节点的地位的高低。某个节点的介数中心性的值越大,说明这个节点在网络中的地位也就越高。举例来说,在社交圈中,有这么一个人,类似于“交际花”,他与各种各样的人打交道,他人的交流都需要通过此人来传递,那么这个人在社交圈中的作用也就越重要。而特征向量中心性在度中心性指标的基础上,增加了节点的邻居的信息对该节点的作用。比如说,在社交圈中,某个人甲所认识的人乙,在乙的社交圈中是类似于“交际花”的存在,那么甲在他自己的社交圈中也很有可能是很重要的

存在；但是，如果甲认识很多的人，但他所认识的大多数人都是社交圈非常狭窄的人，那么甲在他自己的社交圈中也可能没那么重要。依赖网络结构的整体属性的指标综合了节点自身的信息和网络结构的整体信息，可信度比较高，缺点是计算量大，对于大型网络不太友好。

依赖随机游走算法的指标通常是使用网页排序技术。网页排序技术指的是在我们使用电脑或手机等电子产品浏览网页时，网页之间的跳转和链接。而这体现了网页之间的相互联系，通过这种联系计算网页的重要性。常见的依赖随机游走算法的指标包含了 PageRank<sup>[18]</sup>，LeaderRank<sup>[19]</sup>和 HITS 算法<sup>[20]</sup>。PageRank 和 LeaderRank 算法模拟用户浏览网页的过程，利用网页跳转计算网络中节点的重要度，然后进行排序。而 HITS 算法则是使用某种迭代的方法得到某个具体场景下的最优的网页。依赖随机游走算法的指标需要根据具体的问题模拟具体的场景，所以这种方法具有实际意义，缺点也是很明显的，缺乏通用性，很难用于其他场景中。

针对网络中重要节点的挖掘已经有很多位学者进行了研究。例如，Al-Azim 等<sup>[21]</sup>建立兴趣组和节点排名模型来挖掘重要节点，Liu 等<sup>[22]</sup>对网络节点排名算法进行了综述，包括代表性方法、扩展和应用，王兴隆等<sup>[23]</sup>运用改进的 K-shell 算法对网络中的节点进行排序，卢鹏丽等<sup>[24]</sup>结合度与 H 指数扩展了一种复杂网络节点排序方法。

### 1.2.2 网络鲁棒性优化研究现状

对于本文研究的第二点，通过保护重要节点来提高网络的鲁棒性的研究也是当前复杂网络方向研究的热点，探讨鲁棒性增强策略对于网络结构设计 with 功能改进有着深刻的研究价值。在复杂网络的鲁棒性研究方向上，Albert 等人<sup>[25]</sup>指出，复杂系统对受到的攻击都有强大的修复能力。这表示复杂网络具有强大的稳定性，一般的攻击并不会对网络产生过大的影响。

信息科学技术的蓬勃发展，为全球经济、政治、文化等带来了巨大的便利。互联网是其中的代表。互联网是复杂网络中的极其重要的组成部分。复杂网络科学的进步是人类科学史进步的代表。而复杂网络科学中网络鲁棒性的研究也是受到了不同领域的诸多研究人员的关注。复杂系统的鲁棒性可以理解为复杂系统在受到攻击时保持功能的能力。

鲁棒性在不同的领域，结合实际情况是具有不同的含义的。例如，在生态学的研究上，生物鲁棒性<sup>[26]</sup>表现在生物在生态系统中的适应和调节能力，比如，当面临物种入侵时，生态系统能否还能维持稳定；在建筑领域，结构鲁棒性<sup>[27]</sup>表现在桥梁、隧道、房屋等结构的安全性上；在机器学习领域，鲁棒性又可以表现在算法对

数据的通用性上,当数据发生变化时,算法是否仍然可用<sup>[28]</sup>;在航空运输领域,网络鲁棒性表现在机场、航空公司和航线的运能上<sup>[29]</sup>;在供应链领域,鲁棒性可以理解为保证供应链系统正常有效运行的能力<sup>[30]</sup>。

复杂系统的鲁棒性指的是复杂系统应对外界攻击的能力,是复杂系统本身的固有属性,它在一定程度上取决于复杂系统自身的调节和适应能力。比如,在生物学的研究上,鲁棒性能够用于探索基因重组、基因突变等原理,也能够用于研究蛋白质分子之间连接的过程以及产生的结果;在生态学的研究上,鲁棒性可以用于研究食物链构成的食物网内的每种生物之间的联系,也可以用于研究物种入侵带来的后果;在机器学习领域,鲁棒性被用来研究算法在不同模型数据上的适用程度等等。

信息技术发展推动人类社会加速互连,以复杂网络为代表的网络科学成为复杂系统分析的重要途径。受各种因素影响,网络面临的威胁持续增长。例如,2003年在美国和加拿大许多城市出现的大范围停电事件,这次停电事件的导火索仅仅是极少的线路故障;2010年“震网”病毒攻击电脑主机<sup>[31]</sup>,造成大量主机瘫痪、多个国家地区受到影响;2019年委内瑞拉大范围停电,导致生活出行等全面受到影响<sup>[32]</sup>;2020年初大范围传播的新型冠状病毒<sup>[33]</sup>,造成多人死亡且传染性极其强烈,对全球经济和人类生活造成了巨大的影响。这些频发的大型事故对经济社会造成巨大损失的同时,也暴露出某些关键基础设施应对极端事件的脆弱性,所以,对网络鲁棒性的优化迫在眉睫。

而这些大事件的发生,都是由于级联效应<sup>[34]</sup>导致的。级联效应指的是,一个节点会对它周围的节点造成影响,如果这种影响逐渐传播,就会导致网络产生巨大的影响。比如,在社交网络中,每个人的消息可以被他的亲朋好友不断转发,从而扩大了消息传播的范围和受众,使得信息传播也更为有效。所以,对于级联效应,我们需要重视它产生的有利的和不利的结果。分析级联效应,有利于在生活中避免出现大规模的故障,阻止谣言的扩散,提高信息传播的能力等等。

关于网络鲁棒性在复杂网络中的研究有:Barabasi和Albert研究了万维网的鲁棒性<sup>[35]</sup>并研究了无标度网络的鲁棒性,任广建等<sup>[36]</sup>基于相对熵理论对航路网络鲁棒性进行了研究,裴燕如等<sup>[37]</sup>对生态网络进行了增边优化鲁棒性分析,张振江等<sup>[38]</sup>对铁路快捷货运网络鲁棒性进行了分析,刘迪洋等<sup>[39]</sup>基于社区结构对网络鲁棒性进行了优化,王哲等<sup>[40]</sup>对网络鲁棒性增强方法进行了探索。

### 1.3 本文主要工作

本文工作主要包含两个方面:挖掘网络中的重要节点和优化网络的鲁棒性。

具体的工作安排如下：

1、数据获取和预处理。本文主要使用的数据集是美国航空公开的航班飞行记录数据集<sup>[41]</sup>，作为对比实验，还使用了多个真实数据集进行了实验<sup>[42]</sup>。USAir 网络、UCI 网络、PB 网络和 Email 网络进行了实验。USAir 网络指的是一个标准的美国航空运输网络，UCI 网络指的是加州大学欧文分校网络，PB 网络指的是美国政治博客网络，Email 网络指的是邮件网络。

2、对网络中的节点进行重要性排序。结合图神经网络算法提出的 WGA 方法、Node2vec 算法、度中心性指标、介数中心性指标、PageRank 算法、DIL 方法和 NL 方法对网络中的节点进行重要度评分，每种方法都可以得到一种对应的节点重要性排序序列，再提出 PNN 指标，对每种方法得出的节点序列进行评估。最后在真实情境下分析所挖掘的重要节点的可靠性，表明 WGA 方法在重要节点挖掘上的可行性；

3、对重要节点的部分邻居节点构成的局部网络进行鲁棒性优化。选择在最重要的节点的局部网络上进行增边，使用基尼系数作为网络鲁棒性的指标，采用对网络中的节点进行增边的方法来优化局部网络的鲁棒性。在本文中，分别使用了动态规划增边策略、贪心增边策略、在度值和最大的节点间增边的方法以及在介数值最大的节点间增边的方法在限制增边条数或增边容量的条件下对网络进行鲁棒性优化。最终表明，所提出的动态规划增边策略是最好的网络鲁棒性优化方法。

## 1.4 本文的结构安排

本文总共有五章，具体章节的内容和框架如下所述：

第一章是绪论。本章主要介绍了复杂网络的研究背景和意义，重点介绍了复杂网络研究的重要节点挖掘方向和网络鲁棒性优化的国内外研究现状。1.3 节简要地介绍了本文的几个主要工作。主要贡献是首次将图神经网络算法运用在了网络重要节点的挖掘上，将图神经网络与复杂网络相结合，提出 WGA 方法，运用最前沿的知识，对网络中的重要节点进行挖掘，并与若干种重要节点挖掘方法进行对比，还介绍了网络鲁棒性优化方法。本节简要介绍了本文的整体架构。

第二章是相关理论与技术。本章主要介绍了与本文的相关理论和技术。首先，简要介绍了复杂网络的一些基本特征和三种网络中的重要节点挖掘方法；其次，重点介绍图神经网络研究方向的起源、发展以及两种图神经网络算法的模型；然后，对网络的鲁棒性进行了详细的介绍；最后，给出了节点重要性评价指标，节点挖掘方法的评价指标和网络鲁棒性的评价指标。

第三章是基于图神经网络算法的重要节点挖掘方法。本章详细的提出了 WGA



方法对网络中的重要节点的挖掘，使用节点特征向量的欧式距离之和作为节点重要性的定义，并提出 PNN 指标作为评价网络重要节点挖掘方法的标准，并提出在真实情景下分析 WGA 方法所挖掘的重要节点的可靠性。最后在多个数据集下进行了实验和分析。

第四章是网络鲁棒性优化方法。本章主要是使用增边的方法对网络的鲁棒性进行优化。首先提出结合基尼系数，提出新的基尼系数指标作为网络的鲁棒性指标，然后通过数学证明了基尼系数作为网络鲁棒性指标的可行性，最后提出了基于贪心策略、动态规划增边策略和两种常见的增边策略对网络的鲁棒性进行优化，并在多个数据集下进行了实验及分析。

第五章是全文总结与展望。本章对全文的总体框架及工作进行了总结，探讨了本文的贡献，并且给出了未来的工作展望。

## 1.5 本章小结

本章主要介绍了网络中重要节点挖掘方向以及网络鲁棒性优化方向的国内外研究现状，提出本文的论点所在，并为后文的工作安排奠定基调。

## 第二章 相关理论与技术

本文主要是研究如何对网络中的重要节点挖掘以及如何对挖掘到的重要节点加以保护,从而优化整个网络的鲁棒性。因此,本章首先对复杂网络的基本特征进行介绍;然后,对网络中的重要节点挖掘的方法进行详细的介绍,本文主要的贡献就是首次将图神经网络算法运用在节点挖掘,提出了 WGA 方法对网络中重要节点进行挖掘,以及结合了基尼系数,重新定义了网络鲁棒性的评价指标,并提出了基于贪心算法和动态规划的增边方法优化网络的鲁棒性。

### 2.1 复杂网络基本特征

复杂网络是为了研究现实世界中的复杂系统而形成的一种特殊的网络结构。因此,不同的复杂系统抽象成同一种复杂网络的时候都会有一些基本的相同的特征。例如,节点的度与度分布,网络的平均路径长度,节点的集聚系数和节点的介数。

(1) 节点度是网络中节点最直观的特点。在无向图中,节点度表示的是与节点相连的边数。一般来说,节点度值越大,该节点在网络中的重要性就越高。节点  $i$  的度  $D_i$  具体表述见公式(2-1):

$$D_i = \sum_{j=1}^n a_{ij} \quad (2-1)$$

公式(2-1)中,  $a_{ij} = 1$  表示节点  $i, j$  之间存在连边,  $a_{ij} = 0$  表示节点  $i, j$  之间不存在连边。度的大小体现了节点在网络中的直接影响力。节点的直接邻居越多则表示该节点的度越大。

小世界网络的度分布服从泊松分布,见图 2-1(a),而在大量真实的网络中,节点的度分布大多数服从幂率分布,见公式(2-2)及图 2-1(b):

$$P(D_i) \propto D_i^{-\alpha} \quad (2-2)$$

在图 2-1 中,横坐标表示节点的度,纵坐标表示该度数下的节点占全部节点的比值。

服从泊松分布的特点是,顶峰对应的横坐标表示的是网络中所有节点的平均度,这种类型的节点所占比例最多,以此为分界线,成对称趋势。

而服从幂率分布的节点,则表明网络中大度数的节点在网络中所占的比例是比较少的,但这些节点在网络中发挥着重要的作用。如果这些重要节点遭到破坏,网络很有可能会受到巨大的影响。

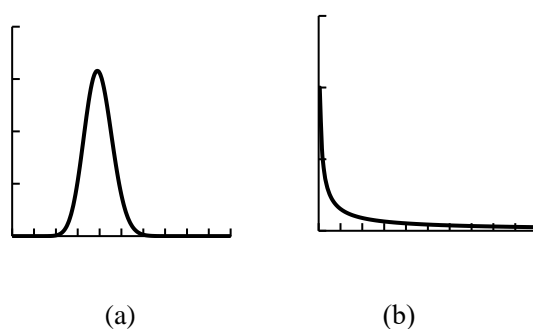


图 2-1 度分布图 (a)泊松分布 (b)幂率分布

(2) 网络的平均路径长度展现的是节点的远近关系。节点对的距离，指的是从一个节点到另一个节点跳转的最少次数。网络中所有节点对的最大距离称作网络直径，见公式(2-3)。网络的平均路径长度是全部节点对的距离的平均值，见公式(2-4)。

$$D = \max d_{ij} \quad (2-3)$$

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij} \quad (2-4)$$

公式(2-3)和公式(2-4)中， $D$ 表示网络直径， $d_{ij}$ 表示节点 $i$ 与节点 $j$ 的距离。 $\langle d \rangle$ 表示网络的平均路径长度。

如图 2-2 所示，图中包含了 5 个节点，6 条边。该图中网络直径为 $D=2$ ，平均路径长度为 $\langle d \rangle=1.4$ 。在一个包含了 $N$ 个节点， $E$ 条边的网络中，使用广度优先搜索的方法计算两个节点之间的平均路径长度，时间复杂度为 $O(NE)$ 。

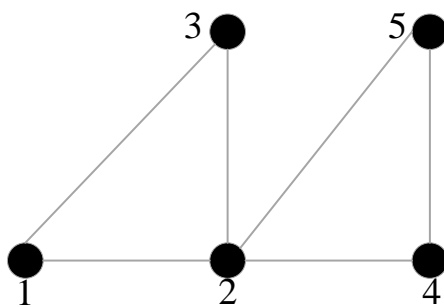


图 2-2 随机子网络

(3) 节点的集聚系数很直观地给出某节点的邻居节点之间的连接程度。例如，在社交圈中，某个人有很多朋友，他的这些朋友也相互认识，这就叫做集聚系数。事实上，在现实世界里，尤其是社交网络，某个人的朋友圈也有很大的可能形成新的朋友圈，也就说明了社交网络的集聚系数普遍偏高。节点的集聚系数通常表示为该节点的邻居节点形成的边的条数和能够存在的最多的边的条数的比值，见公式

(2-5):

$$CC(i) = \frac{E_i}{C_{D_i}^2} = \frac{2E_i}{D_i(D_i-1)} \quad (2-5)$$

公式(2-5)中,  $D_i$ 表示节点*i*的度,  $E_i$ 表示节点*i*的邻居节点形成的边的条数, 换个角度思考, 就是节点*i*和它相邻的两个节点的闭合三角形的个数。

在包含了*N*个节点网络中, 整个网络的集聚系数见公式(2-6):

$$C = \frac{1}{N} \sum_{i=1}^N CC(i) \quad (2-6)$$

显然  $0 \leq C \leq 1$ 。当  $C = 0$  时, 网络中所有节点都是孤立节点, 没有边相连; 当  $C = 1$  时, 该网络构成的是一个完全图的结构, 所有节点之间都有边相连。

在图 2-2 中, 求图中节点的集聚系数和整个图的集聚系数。求解得, 节点 1 的集聚系数  $CC(1) = 1$ , 整个图的集聚系数为  $C = \frac{13}{15}$ 。

(4) 节点的介数直接地表明了节点的在全局网络中的重要程度。在数学上, 节点的介数值表示的是节点出现在网络最短路径上的频率。节点*i*的介数定义见公式(2-7):

$$B_i = \sum_{i \neq u \neq k} \frac{d_{uik}}{d_{uk}} \quad (2-7)$$

公式(2-7)中,  $d_{uk}$ 表示节点u和节点*k*之间的最短路径数目,  $d_{uik}$ 表示节点u和节点*k*的最短路径中经过节点*i*的数目。

节点的介数表明, 如果网络中存在这样一个节点, 这个节点将整个网络分割成了两个或者多个社团结构, 那么, 毫无疑问, 这个节点的介数值是非常大的, 因为不同社团之间的节点的路径必须经过该节点。因此, 节点的介数关乎节点在整个网络上的重要性。

## 2.2 复杂网络基本模型

复杂网络中, 规则网络模型、随机网络模型、小世界网络模型和无标度网络模型是最基本网络模型。每种基本网络模型都具有自身独特的性质, 而又互相关联。接下来对每种基本网络模型进行介绍。

### 2.2.1 规则网络模型

在规则网络模型中, 网络中节点是极其相似的。例如, 完全图网络和星型网络, 如图 2-3 所示。在完全图网络中,  $\langle d \rangle = 1$ ,  $C = 1$ 。而在一个包含 *N* 个节点的星型网络中, 在该网络中有 1 个节点与其他 *N*-1 个节点相连, 而其他 *N*-1 个节点不直接相连, 该星型网络的平均路径长度见公式(2-8), 集聚系数见公式(2-9):

$$\langle d \rangle = 2 - \frac{2}{N} \rightarrow 2 \quad (N \rightarrow \infty) \quad (2-8)$$

$$C = 1 - \frac{1}{N} \rightarrow 1 \quad (N \rightarrow \infty) \quad (2-9)$$

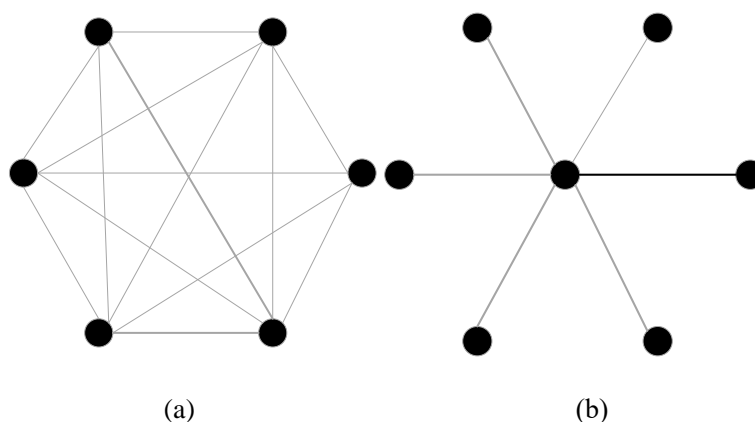


图 2-3 规则网络 (a)完全图网络 (b)星型网络

## 2.2.2 随机网络模型

随机网络模型，也就是 ER 网络。它的构建可以按以下步骤：

- 1、假设有  $N$  个相互独立的节点，节点连接的概率为  $p$ ；
- 2、选择没有相连的节点，以概率  $p$  对节点进行连边；
- 3、直到网络中所有节点对都有边相连为止。

如图 2-4 所示，展示的是一种随机网络。

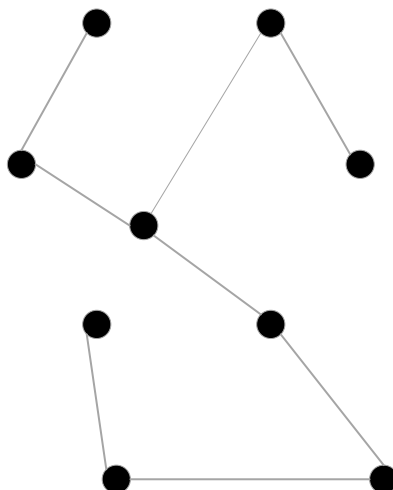


图 2-4 随机网络

## 2.2.3 小世界网络模型

对于一个规则网络，以概率  $p$  对网络中的未连接的两个节点进行相连，就可

以形成一种小世界网络，如图 2-5 所示。

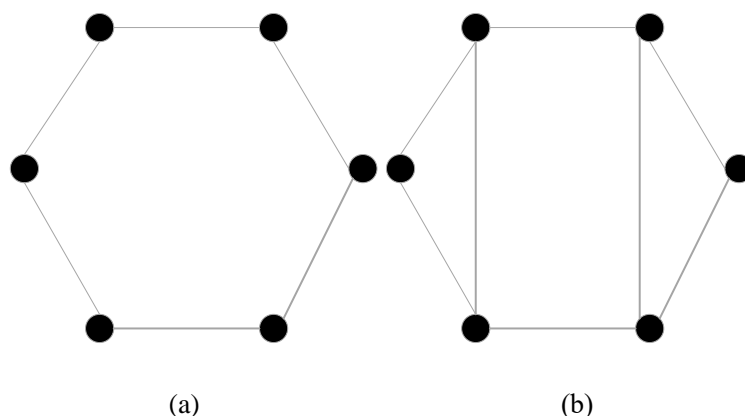


图 2-5 小世界网络形成过程 (a)近邻耦合网络 (b)小世界网络

当  $p=0$  时，为规则网络；当  $p=1$  时，为随机网络； $0 < p < 1$  时，为小世界网络。

#### 2.2.4 无标度网络模型

无标度网络中节点的度分布几乎等同于幂率分布，如图 2-1 所示。在无标度网络中大多数节点都是度值较小的节点，但是往往会存在少量的度值很大的节点，这些节点对维持网络结构和功能上有着巨大的作用。如图 2-6 所示，这是一个包含 1000 个节点的无标度网络，很明显，在图中，只存在着少量比较大的节点，多数节点都是比较小的节点。

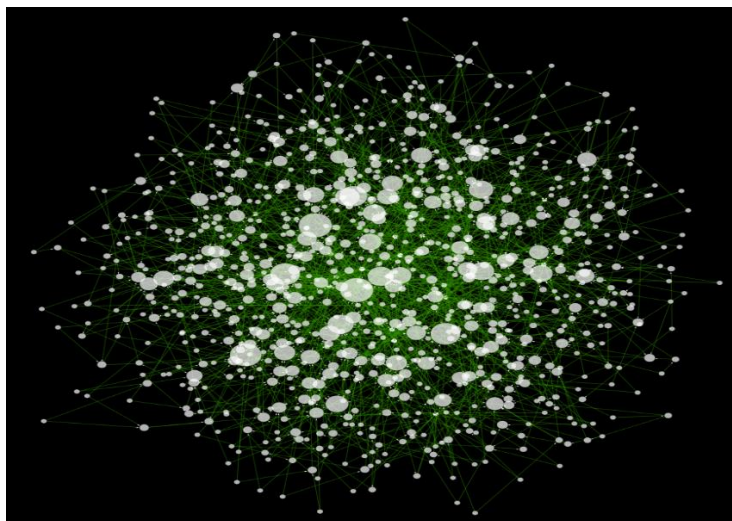


图 2-6 包含 1000 个节点的无标度网络，本图片来自网络

在无标度网络中，时时刻刻都会对网络中的节点和边进行增加或者删除，而新增加的节点以更大的概率连接度值较大的节点，其演化过程如图 2-7 所示。

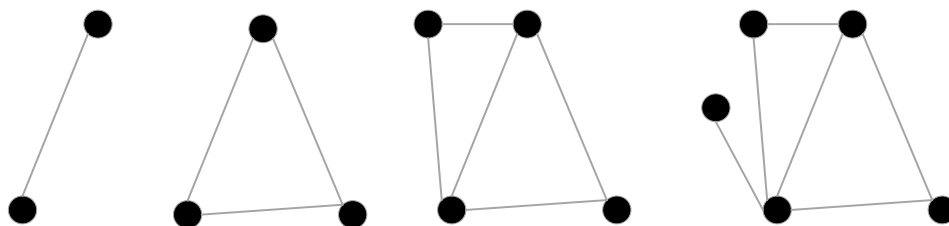


图 2-7 无标度网络演化过程

## 2.3 网络中重要节点挖掘方法

第一章中，对传统的关键节点挖掘方法进行了简要介绍，将网络中重要节点挖掘方法分为了三种：第一种是依赖网络结构的局部属性的指标，第二种是依赖网络结构的整体属性的指标，第三种是依赖随机游走算法的指标。

下面主要介绍这三中重要节点挖掘方法。

1. 第一种是依赖网络结构的局部属性的指标主要包含了度中心性、半局部中心性和聚类系数指标。依赖网络结构的局部属性的指标的方法优点是直观易理解，时间复杂度低，可以用于大型网络；缺点是没有考虑网络整体的结构，会出现评价结果的不准确。关于基于网络局部属性的指标，已经有很多研究学者提出了给予了许多的定义。比如，王建伟等<sup>[43]</sup>提出节点的重要性与节点和它邻居的度有关；Chen 等<sup>[44]</sup>考虑了节点的一阶邻居和二阶邻居信息，给出新的指标；任卓明等<sup>[45]</sup>综合节点邻居和节点的聚集系数，给出一种指标等等。

2. 依赖网络结构的整体属性的指标通常有介数中心性指标，特征向量中心性指标和接近中心性指标等。介数中心性离不开最短路径的计算，如果某个节点在网络中起着“桥梁”一般连接两岸的作用，那么毫无疑问这个节点在网络中起着举足轻重的作用，因为在“桥梁”节点两岸的节点的最短路径上必然经过“桥梁”节点。但是，介数中心性也有它的不足，那就是对最短路径的计算量是非常大的，如果这是一个大型网络，计算量可想而知；特征向量中心性则是依赖邻接矩阵和特征向量进行运作的；接近中心性表示某节点对别的节点的影响程度。数学中，通常将节点到别的节点的平均最短路径的倒数，作为该节点的接近中心性值。举个简单的例子，在星型网络中，核心节点的接近中心性值必然是最大的，这也在一定程度上证明了接近中心性指标的可靠性。还有其他的方法，例如 Travencolo 等<sup>[46]</sup>定义了节点可达性的指标；李鹏翔等<sup>[47]</sup>指出，当删除某节点后，极有可能会出现很多不连通的节点，对这些节点之间距离的倒数求和，作为一种新的指标；谭跃进等<sup>[48]</sup>则定义了一种节点收缩方法等等。

3. 依赖随机游走算法的指标主要使用的是 PageRank 和 LeaderRank 算法。这

两种算法很类似，都是通过网页链接相互指向的方法进行实现的。但是这两种算法是有一点区别的，就是当网络包含孤立节点或社团时，LeaderRank 算法可以很好的进行节点排序，而 PageRank 算法则会出现多个排序结果。另外，依赖随机游走算法的指标还包含了一些其他的算法。例如 Masuda 等<sup>[49]</sup>延伸了拉普拉斯算子指标的度量方法；以及依赖 HITS 算法等。

除了上述的几种针对网络中重要节点的挖掘方法外，最近几年也有很多学者提出了很多种重要节点挖掘方法。例如，Christian 等<sup>[50]</sup>于提出的 2011 年提出的反向目标算法；Kitsak 等<sup>[51]</sup>于 2013 年提出的  $k$ -核分解法；中国矿业大学的闫艳芳<sup>[52]</sup>在 2020 年毕业的硕士论文中提出了融合度与聚类系数的节点重要性排序算法以及基于支持向量机的多层复杂网络节点重要性排序算法等等。

但是这些研究学者几乎都没有将网络中的边考虑进来，他们都忽视了与节点相连的边对网络中的节点重要性的影响。在考虑与节点相连接的边对节点的重要性影响上，有一些学者提出了一些他们自己的看法。例如，WANG 等<sup>[53]</sup>认为需要对网络中的边进行赋值，因为网络中不同的边的重要性是不一致的，但是他还是根据节点的度的值对它所连接的边进行赋值；Liu 等<sup>[54]</sup>则对 WANG 等<sup>[73]</sup>提出的看法发表了一些不同的观点，他认为某条边的重要性不能还从该边所连接的节点的重要性来定义，边应该有其自身独特的特性，比如说连接能力等等，因此他重新对边的重要性进行了定义，提出了一种新的节点重要性评价方法——DIL (degree value and the importance of lines)，DIL 依赖节点的度以及边的重要性重新对节点重要性进行了评价；另外，刘书磊<sup>[55]</sup>等则认为 DIL 方法中仅仅依赖节点的度是不够准确的，他提出将节点的度的信息，改变成节点的多阶邻居的信息，提出新的节点重要性评价方法——NL (neighbor nodes and importance of lines)，NL 方法在一定程度上提高了 DIL 方法在节点重要性评价上的准确性。

本文对 Liu 等<sup>[54]</sup>提出的 DIL 方法和刘书磊<sup>[55]</sup>等提出的 NL 方法进行了运用，并与上述的三种节点重要性评价方法（度中心性，介数中心性和 PageRank）一起，进行了实验。接下来介绍一下 PageRank 算法、DIL 方法和 NL 方法。

### 2.3.1 PageRank 算法

首先，赋予网络中每个节点定义一个 PR 值，初始值设为  $PR_i(0)$ ，使得每个节点的 PR 值之和为 1，见公式(2-10)：

$$\sum_{i=1}^N PR_i(0) = 1 \quad (2-10)$$

然后，设定一个概率  $p$ ，将每个节点的  $PR_i(0)$  变成  $PR_i(0) \times p$ ，那么，所有节点的  $PR_i(0)$  之和就等于  $p$ ，然后将  $1-p$  均分给每个节点 PR 值，使得总和仍为 1。



具体含义就是，节点以概率  $p$  访问下一个连接的节点，同时以  $1-p$  的概率访问网络中的任何一个节点。

假设在一个包含  $E$  条边的网络中，算法收敛的迭代次数为  $I$ ，那么 PageRank 算法的复杂度为  $O(EI)$ ，简单易计算。

### 2.3.2 DIL 方法

DIL 方法在评价节点的重要性上，主要是基于两点来考虑的：一是节点的度，二是结合了与边的重要性。

首先，对网络中节点  $i$  与节点  $j$  相连的边  $e_{ij}$  的重要性进行定义，见式(2-11)：

$$I_{e_{ij}} = \frac{U}{\lambda} \quad (2-11)$$

式(2-10)中， $U = (D_i - p - 1) \times (D_j - p - 1)$ ， $D_i$  和  $D_j$  是节点  $i$  与节点  $j$  的度， $\lambda = \frac{p}{2} + 1$ ， $p$  是网络中所有三角形中包含边  $e_{ij}$  的三角形的数目。

最后，综合节点的度，得出节点  $i$  的 DIL 指标，见式(2-12)：

$$DIL(i) = D_i + \sum_{j \in \zeta(i)} I_{e_{ij}} \times \frac{D_i - 1}{D_i + D_j - 2} \quad (2-12)$$

式(2-12)中， $\zeta(i)$  表示节点  $i$  的邻居节点集合。

### 2.3.3 NL 方法

NL 方法在评价节点的重要性上，用节点的多层邻居的结构取代了 DIL 方法中使用节点的度作为节点重要性评判的根据。

因此，基于 DIL 方法，得出 NL 方法的定义，见式(2-13)：

$$DIL(i) = \sum_{j \in \zeta(i)} N(j) + \sum_{j \in \zeta(i)} I_{e_{ij}} \times \frac{D_i - 1}{D_i + D_j - 2} \quad (2-13)$$

本文所使用的的几种对比性指标是：度中心性指标、介数中心性指标、PageRank 算法、DIL 方法和 NL 方法。

## 2.4 图神经网络相关理论

现如今，信息科学技术在许许多多的领域蓬勃发展。尤其是神经网络<sup>[56]</sup>领域的快速崛起，使得信息科学技术更加的日新月异。神经网络的异军突起，让数据挖掘等研究变得更为的有简便和迅速。现在，许许多多的端到端的深度学习模型被开发出来，使得研究人员在提取数据的特征的时候更为的快速，并且提取的特征也更为可靠，也在一定程度上弥补了手工提取数据特征的缺陷。这种进步得益于以下几点：一是拥有庞大的计算资源，二是训练数据的海量化，三是使用深度学习模型提

取的特征是可靠的。这三点保证了深度学习模型在实际中的可行性和准确性。

可是，传统的深度学习模型在处理非欧式空间数据时，例如提取潜在特征，效果不是很好。比如说，在航空网络中，依赖图的学习系统可以根据城市节点和航空公司节点以及航线边的信息，对航空网络进行处理，但是传统的深度学习方法面对图这种十分复杂的结构时，很难处理得很好。归结原因就是，图是不规则的结构，图中每个节点都可能是大小不一的，图中的边权重也是不一的，图中节点的邻居节点也是千变万化的，这就使得很多操作在图上是难以进行的。另外，传统的深度学习算法往往具有数据独立性，也就是说会人为地认为数据相互独立。但是，对于图这种结构来说，数据独立性肯定是不对的，因为图中的大多数节点必然会与其他节点和边相连。

基于传统的深度学习方法在图上运用的难题，这几年，许多研究学者不断研究可用于在图上的深度学习方法，并取得了丰硕的成果。他们成功地研究出用于处理图的神经网络模型，也就是“图神经网络”<sup>[57]</sup>，并迅速地成为了计算机科学领域的最前沿研究课题之一。

现如今，图神经网络也被细分成了很多分支，主要分为五种，如图 2-8 所示。这五类图神经网络的分支在现实生活中都有着很重要的应用意义。

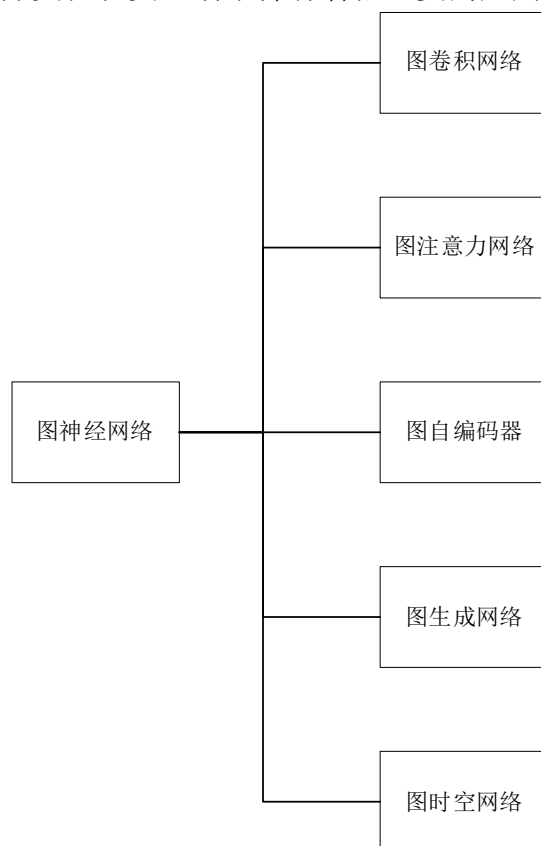


图 2-8 图神经网络分支

而在图神经网络中，有许许多多的图神经网络算法是值得使用的。例如 Node2vec 算法<sup>[58]</sup>和 SDNE 算法<sup>[59]</sup>。而将图神经网络算法运用在重要节点的挖掘上也是值得研究的问题，在刘震等<sup>[60]</sup>就结合图神经网络算法对网络中的重要节点进行了挖掘。

### 2.4.1 Node2vec 算法介绍

网络的表示学习有很多方面的运用，尤其是在节点分类和连接预测等方向上。要学习网络特征，一种学习特征表达的方法是通过定义目标函数然后解决优化问题。但是现有的技术要不就是需要大量训练和运算花费，要不就是效果太差。一个解决的方向是通过局部的邻居关系来制定目标函数，同时这个目标函数可用随机梯度下降(SGD)<sup>[61]</sup>来解决。这样就不用训练同时运算快速。

良好的节点表示算法通常有两个特点：可以把同一网络中的节点映射在附近，以及相似节点有相似的表达方式。

Node2vec 算法定义了一个目标函数，见公式(2-14)：

$$\max_f \sum_{u \in V} \log Pr(N_s(u)|f(u)) \quad (2-14)$$

在公式(2-14)中， $f$ 是节点到特征表达的映射函数 $V \rightarrow R^d$ ， $d$ 为表达空间维度。 $N_s(u) \subset V$ 是节点 $u$ 的网络邻居，通过邻居节点采样方法 $S$ 得到的。

增加条件独立，见公式(2-15)和空间对称假设，见公式(2-16)：

$$Pr(N_s(u)|f(u)) = \prod_{n_i \in N_s(u)} Pr(n_i|f(u)) \quad (2-15)$$

$$Pr(n_i|f(u)) = \frac{\exp(f(n_i) \cdot f(u))}{\sum_{v \in V} \exp(f(v) \cdot f(u))} \quad (2-16)$$

综合目标函数公式(2-14)，条件独立公式(2-15)和空间对称公式(2-16)，可以将目标函数简化为如下公式，见公式(2-17)和公式(2-18)：

$$\max_f \sum_{u \in V} [-\log Z_u + \sum_{n_i \in N_s(u)} (f(n_i) \cdot f(u))] \quad (2-17)$$

$$Z_u = \sum_{v \in V} \exp(f(v) \cdot f(u)) \quad (2-18)$$

式(2-17)所示目标函数的意义就是，在给定节点 $u$ 的邻居节点集合 $N_s(u)$ 下，通过求这个目标函数的最大值的情况下得到 $f(u)$ 的函数形式，从而得到每个节点的特征向量。

对于上面所示的目标函数，存在两个问题：一是找到所有节点 $u$ 的邻居节点集合 $N_s(u)$ ；二是求解这个目标函数。

对于第一个问题采用特殊的二阶 random walk 进行采样，找到每个节点的邻居集合。具体过程可以表述为：给定源节点 $u$ ，进行随机游走，长度是 $l$ 。 $c_i$ 为游走的

第 $i$ 个节点, 起始节点 $c_0 = u$ 。节点 $c_i$ 服从以下分布, 见式(2-19):

$$P(c_i = x | c_{i-1} = v) = \begin{cases} \frac{\pi_{vx}}{Z} & \text{if } (v, x) \in E \\ 0 & \text{otherwise} \end{cases} \quad (2-19)$$

公式(2-19)中 $\pi_{vx}$ 指的是还未标准化的节点 $v$ 到节点 $x$ 的转移概率,  $Z$ 用于对转移概率的标准化。

直接设转移概率为边的权重 $\pi_{vx} = w_{vx}$ 不能很好地综合网络结构信息和搜索不同的邻居, 因此设置偏移概率 $\alpha$ 。该随机游走有两个参数 $p$ 和 $q$ , 考虑刚刚遍历了边 $(t, v)$ , 如图 2-9 所示。

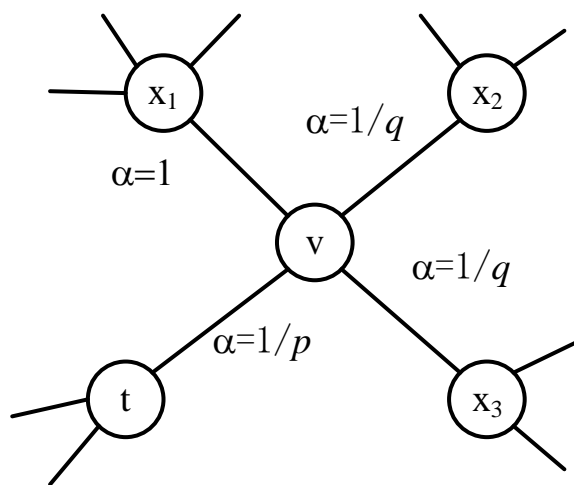


图 2-9 随机游走概率图<sup>[58]</sup>

现在在节点 $v$ 要根据转移概率 $\pi_{vx}$ 访问下一个节点 $x$ 。设 $\pi_{vx} = \alpha_{pq}(t, x) \cdot w_{vx}$ , 其中 $\alpha_{pq}(t, x)$ 的定义见公式(2-20):

$$\alpha_{pq}(t, x) = \begin{cases} \frac{1}{p} & \text{if } d_{tx} = 0 \\ 1 & \text{if } d_{tx} = 1 \\ \frac{1}{q} & \text{if } d_{tx} = 2 \end{cases} \quad (2-20)$$

$d_{tx}$ 表示节点 $t$ 和节点 $x$ 之间的最短距离, 只能属于 $\{0, 1, 2\}$ 。参数 $p$ 和 $q$ 相当于调节 BFS 和 DFS 的程度。

参数 $p$ 和 $q$ 的意义分别如下:

参数 $p$ : 若 $p > \max(q, 1)$ , 那么采样会尽量不往回走, 对应图 2-10 中的情况, 就是下一个节点不太可能是上一个访问的节点 $t$ ; 若 $p < \min(q, 1)$ , 那么采样会更倾向于返回上一个节点, 这样就会一直在起始点周围某些节点徘徊。

参数 $q$ : 若 $q > 1$ , 那么采样将会在节点的周围游走; 若 $q < 1$ , 那么采样将更容易以远离节点的方式游走。

当 $p, q = 1$ 时, 采样的方式就类似于随机游走采样。

所以, Node2vec 算法流程如下:

---

#### Algorithm1 Node2vec

输入:  $G = (V, E, W)$ , 嵌入向量维度  $d$ , 节点游走次数  $r$ , 游走长度  $l$ , 窗口大小  $k$ , 返回概率  $p$ , 出入概率  $q$

输出: 节点嵌入向量矩阵  $f$

```

1: //利用  $p, q$  计算转移概率  $\pi$ 
2:  $\pi = \text{PreprocessModifiedWeights}(G, p, q)$ 
3: 得到新图  $G' = (V, E, \pi)$ 
4:  $walks \leftarrow \emptyset$ 
5: for iter = 1 to  $r$  do
6:   for all nodes  $u \in V$  do
7:      $walk \leftarrow \text{node2vecWalk}(G', u, l)$ 
8:      $walks \leftarrow walks \cup walk$ 
9:   end for
10: end for
11: //使用随机梯度下降训练
12:  $f = \text{StochasticGradientDescent}(k, d, walks)$ 
13: return  $f$ 

```

---

#### Algorithm 2 node2vecWalk

输入:  $G' = (V, E, \pi)$ , 起始节点  $u$ , 游走长度  $l$

输出: 游走序列  $walk$

```

1:  $walk \leftarrow [u]$ 
2: for walk_iter = 1 to  $l$  do
3:    $curr = walk[-1]$ 
4:    $V_{curr} = \text{Neighbors}(curr, G')$ 
5:   //使用 Alias 采样得到下一步游走节点  $s$ 
6:    $s = \text{AliasSample}(V_{curr}, \pi)$ 
7:    $walk \leftarrow walk \cup s$ 
8: end for
9: return  $walk$ 

```

---

### 2.4.2 SDNE 算法介绍

SDNE 算法是非常具有代表性的, 因为 SDNE 算法首先将深度学习与网络表示学习相结合。SDNE 算法依赖自动编码器结构, 通过自动编码器学习特征向量, 优化一阶相似度以及二阶相似度, 并且能够维持网络的局部和整体结构。

接下来对 SDNE 算法的原理进行详细的介绍。

SDNE 算法中主要的工作就是定义一阶相似度和二阶相似度的目标函数。一阶相似度描述的是两个邻居节点的相似程度；二阶相似度描述的是两个节点的邻居节点之间的相似性。所以接下来对一阶相似度和二阶相似度的优化目标进行定义。

假设有一张图 $G$ ，假设 $L$ 是图 $G$ 所对应的拉普拉斯矩阵， $D$ 为图 $G$ 中节点的度矩阵， $S$ 为图的邻接矩阵，分别如下所示。

$$S(G) = \begin{bmatrix} 0 & s_{12} & \cdots & s_{1n} \\ s_{21} & 0 & \cdots & s_{2n} \\ \cdots & \cdots & 0 & \cdots \\ s_{n1} & s_{n2} & \cdots & 0 \end{bmatrix} \quad (2-21)$$

$$D(G) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \quad (2-22)$$

$$d_i = \sum_j s_{ij} \quad (2-23)$$

在公式(2-21)，公式(2-22)和公式(2-23)中， $d_i$ 表示节点 $i$ 的度。对于 $s_{ij}$ ，若节点 $i$ 与节点 $j$ 有边相连， $s_{ij} = 1$ ；若节点 $i$ 与节点 $j$ 没有边相连， $s_{ij} = 0$ 。

计算拉普拉斯矩阵 $L(G)$ ： $L(G) = D(G) - S(G)$ ，见式(2-17)。

$$L(G) = \begin{bmatrix} d_1 & -s_{12} & \cdots & -s_{1n} \\ -s_{21} & d_2 & \cdots & -s_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ -s_{n1} & -s_{n2} & \cdots & d_n \end{bmatrix} \quad (2-24)$$

基于上面几个公式的定义，对一阶相似度和二阶相似度优化目标进行定义。

一阶相似度优化目标的损失函数 $L_{1st}$ 定义，见式(2-25)：

$$L_{1st} = \sum_{i,j=1}^n \|y_i - y_j\|_2^2 = 2tr(Y^T LY) \quad (2-25)$$

式(2-25)中， $L$ 是网络对应的拉普拉斯矩阵， $Y$ 是得到的嵌入向量， $tr(Y^T LY)$ 表示矩阵 $Y^T LY$ 的迹。该损失函数 $L_{1st}$ 可以让网络中的相邻的两个节点对应的嵌入向量在隐藏空间接近。

二阶相似度优化目标函数 $L_{2nd}$ 定义，见式(2-26)：

$$L_{2nd} = \sum_{i=1}^n \|\hat{x}_i - x_i\|_2^2 \quad (2-26)$$

式(2-26)中，对于节点 $i$ ，有 $x_i = s_i$ ，每一个 $s_i$ 都包含了节点 $i$ 的邻居结构信息。 $\hat{x}_i = \hat{s}_i$ ， $\hat{s}_i$ 表示重构后的邻接矩阵元素， $s_i$ 表示输入的邻接矩阵元素，使用网络的邻接矩阵 $S$ 进行输入。邻接矩阵的重构可以让结构相似的节点得到相似的嵌入向量。

然而，图是具有稀疏性的，会导致邻接矩阵 $S$ 中为0的元素是远超过非0元素的数目。这也会导致神经网络全部输出0，这显然是不合理的。所以，我们需要

对非 0 元素使用一个惩罚系数 $\beta$ 。修改目标函数 $L_{2nd}$ 如下，见公式(2-27)：

$$L_{2nd} = \sum_{i=1}^n \|\hat{x}_i - x_i\|_2^2 \odot b_i \quad (2-27)$$

式(2-27)中， $\odot$ 为逐元素积， $b_i = \{b_{i,j}\}_{j=1}^n$ ，若 $s_{ij} = 0$ ，则 $b_{i,j} = 1$ ，否则 $b_{i,j} = \beta > 1$ 。

联合一阶相似度和二阶相似度，得出整体优化目标 $L_{mix}$ ，见公式(2-28)：

$$L_{mix} = L_{2nd} + \alpha L_{1st} + \nu L_{reg} \quad (2-28)$$

$L_{reg}$ 为正则化项， $\alpha$ 为控制编码损失的参数（取值为  $10^{-5}$ ）， $\nu$ 为控制正则化项的参数（取值为  $10^{-4}$ ）。

最后，编码输出即为每个节点的向量表示（通常，特征向量维度都设为 128 维）。所以，SDNE 算法流程如下：

### Algorithm3 SDNE

输入：  $G = (V, E)$ ，邻接矩阵  $S$ ，惩罚矩阵  $B$ ，正则化系数 $\alpha$ 和 $\nu$

输出： 嵌入向量矩阵  $Y$

- 1: 初始化神经网络参数 $\theta$
- 2:  $X \leftarrow S$
- 3: **while** not converge
- 4:     利用  $X$  和 $\theta$ 得到重构矩阵 $\hat{X}$ 和编码器输出  $Y$
- 5:      $\mathcal{L}_{mix} = \|(\hat{X} - X) \odot B\|_F^2 + 2\alpha \text{tr}(Y^T L Y) + \nu \mathcal{L}_{reg}$
- 6:     使用梯度下降法进行反向传播，更新网络参数 $\theta$
- 7: **end while**
- 8: **return**  $Y$

SDNE算法模型的结构如图 2-10 所示。

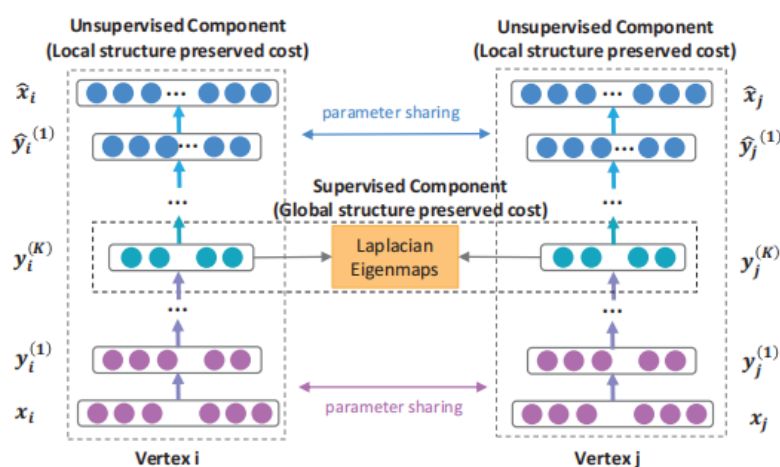


图 2-10 SDNE 算法模型结构<sup>[59]</sup>

在图 2-10, SDNE 算法模型结构图的左部分描述的是自动编码器的结构, 使用图的邻接矩阵作为输入, 输出结果为重构后的邻接矩阵; 然后, 优化邻接矩阵的重构损失保存节点的全局结构。其次, 图中间使用虚线方框框起来的部分,  $y_i^{(K)}$  就是我们需要的特征向量, 模型通过一阶相似度损失函数使得邻接的节点对应的特征向量接近, 从而保留节点的局部结构特性。

## 2.5 增边优化网络的相关理论

本文的工作是针对重要节点的保护来进行网络鲁棒性优化的。当移除少量重要节点时, 网络会快速的瓦解。因此, 我们通过重点保护这些重要节点就能够减缓网络的损毁。因此, 网络鲁棒性优化的前提需要依赖某种重要节点挖掘方法。

例如, Sen 等<sup>[62]</sup>基于布尔代数, 提出评判节点在网络鲁棒性上重要性的方法。Gong 等<sup>[63]</sup>提出四种方法用来保护网络中的重要节点, 评价节点重要性分别依赖节点的度值、介数值、模块中心性和局部中心性。王世锦等<sup>[64]</sup>为了增强航路网络的鲁棒性, 设计了优先配置关键节点保护连边的策略。刘震等<sup>[65]</sup>提出在限制增边条数和容量限制的条件下, 使用动态规划的增边方法对航空网络的鲁棒性进行了优化。

本文的工作也是基于动态规划的增边方法对网络的鲁棒性进行优化。

## 2.6 方法评价指标介绍

### (1) 重要节点挖掘方法评价指标

节点重要性评价指标: 使用 WGA 方法学习网络中节点的特征向量, 用节点与其他节点学习到的特征向量的欧氏距离之和作为评判节点中重要性的标准, 得到一个节点重要性排序序列, 见公式(2-29)。

$$P_i = \sum_{j \in N_i} \text{dist}(v_i, v_j) \quad (2-29)$$

公式(2-29)中,  $P_i$ 表示节点*i*的重要性,  $\text{dist}(v_i, v_j)$ 表示特征向量 $v_i$ 和 $v_j$ 的欧氏距离。

PNN 指标: 采用 PNN 指标来评估网络中节点重要度计算方法的可靠性。PNN 表示删除节点后不会导致严重后果的节点比例, 也就是说, 当我们得到一个节点重要度排序序列, 按照节点排序从高到低, 从网络中依次删除节点, 会得到一条对应的 PNN 值随删除节点变化的曲线。显然, 如果 PNN 值随删除节点变化越快, 则表示删除的节点在网络中越重要, 见式(2-29)和式(2-30):

$$PNN(\theta) = \frac{1}{N} \sum_{i \in V} \delta(i) \quad (2-30)$$



$$\delta(i) = \begin{cases} 1, & \frac{N_{failed(i)}}{N} < \theta \\ 0, & otherwise \end{cases} \quad (2-31)$$

其中 $\theta$ 为容忍阈值， $N$ 表示节点总数，表示 $N_{failed(i)}$ 删除节点 $i$ 导致的失效节点数。

## (2) 网络鲁棒性优化方法评价指标

对于网络鲁棒性的评价可能会存在很多种不同的评价指标，例如采用连通度<sup>[66]</sup>、连通子图相对大小<sup>[67]</sup>和网络效率<sup>[68]</sup>作为鲁棒性的评价标准。本文则使用基尼系数作为网络鲁棒性评价的指标<sup>[65]</sup>。

连通度指的是网络中边的条数与网络可能存在的边的条数的比值，对象是整个网络而不是某个区域。当网络遭受蓄意攻击的时候，网络的连通度还能够维持在一个较大的数值，表明网络的承受能力较大，鲁棒性较强。连通度可以表示为公式(2-32)：

$$\alpha = \frac{|D|}{3|V_d|-6} \quad (2-32)$$

式(2-32)中 $|D|$ 为网络中边的条数， $|V_d|$ 为网络中节点的个数。

最大连通子图，是包含一个图中所有节点而边最少的子图。当网络遭受攻击的时候，最大连通子图的大小和结构也会发生变化。连通子图大小可以表示为公式(2-33)：

$$\beta = \frac{|V_d|}{|V_d|} \quad (2-33)$$

式(2-33)中 $|V_d|$ 为最大连通子图中节点的数量， $|V_d|$ 为网络中节点的个数。

网络中节点对之间效率的平均值称作为网络效率。网络效率越高，网络的连通性也越高，网络的鲁棒性越高。网络效率可以表示为公式(2-34)：

$$\gamma = \frac{1}{|V_d|(|V_d|-1)} \sum_{i,j \in N, i \neq j} \frac{1}{d_{ij}} \quad (2-34)$$

式(2-34)中 $|V_d|$ 为网络中节点的个数， $d_{ij}$ 为节点 $i$ 到节点 $j$ 的最短路径。

在本文中，对网络的鲁棒性指标进行重新定义，提出使用基尼系数作为网络鲁棒性评价的标准。与其他几种网络鲁棒性评价指标相比，基尼系数不仅考虑网络自身信息，还结合了基尼系数的经济学中的特殊含义，对网络鲁棒性进行评价。基尼系数反映了网络中不同节点的重要性差异，见公式(2-35)：

$$gini(S) = 1 - \sum \left( \frac{s_i}{\sum s_k} \right)^2 \quad (2-35)$$

公式(2-35)中 $gini(S)$ 表示网络 $S$ 的整体基尼系数， $s_i$ 表示节点 $i$ 的重要性。

## 2.7 本章小结

本章主要工作是对复杂网络的基本特征和模型、图神经网络算法和增边优化网络鲁棒性的相关理论进行了介绍，提出重要节点挖掘方法和增边优化网络鲁棒性的评价指标，为后文的工作提供理论基础。

### 第三章 基于图神经网络算法的重要节点挖掘方法

本章的主要研究工作是，将图神经网络算法应用在重要节点挖掘上，提出了带权的图自编码方法（WGA 方法）挖掘重要节点。本章提出的 WGA 方法是用于挖掘无向带权图上的重要节点，使用节点特征向量间的欧氏距离之和评判节点的重要性，使用 PNN 指标评价多种节点挖掘算法得到的节点重要性排序序列。后文将对 WGA 方法进行详细的介绍以及实验和实验分析。

#### 3.1 WGA 方法对网络中重要节点的挖掘

本节提出 WGA 方法挖掘网络中的重要节点，其总体流程如图 3-1 所示。

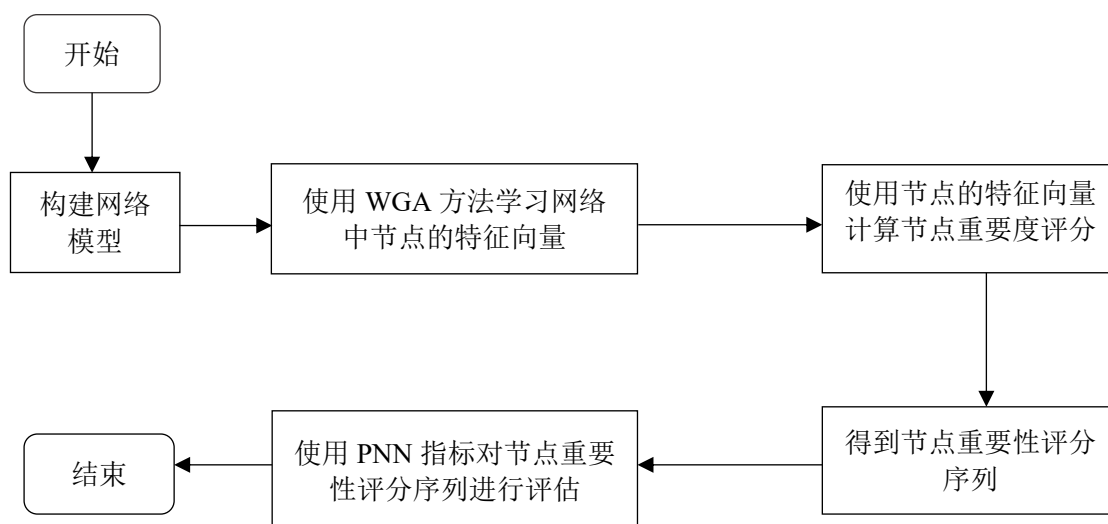


图 3-1 WGA 方法挖掘重要节点的总体流程图

##### 3.1.1 构建重要节点挖掘的网络模型

对于 WGA 方法，构建一个无向带权图的网络模型，对于这种一般的网络模型，网络中的某个节点的失效，将会导致该节点所关联的边全部失效。当我们从整个网络中删除一个节点 $i$ 时，就意味着与节点 $i$ 所关联的边全部失效。如图 3-2 所示，图中包含了 8 个节点，10 条边。当删除节点 1 时，节点 5 和节点 8 与节点 1 连接的边将全部删除。

例如，城市网络就是一种一般的无向带权图网络，当删除一个城市节点时，与该城市关联的所有航线都会失效。依据这样的定义，本文将城市抽象为节点，城市间的航线抽象为连接节点的边，航线上的航班的飞行次数作为边的权重，构建城市网络。

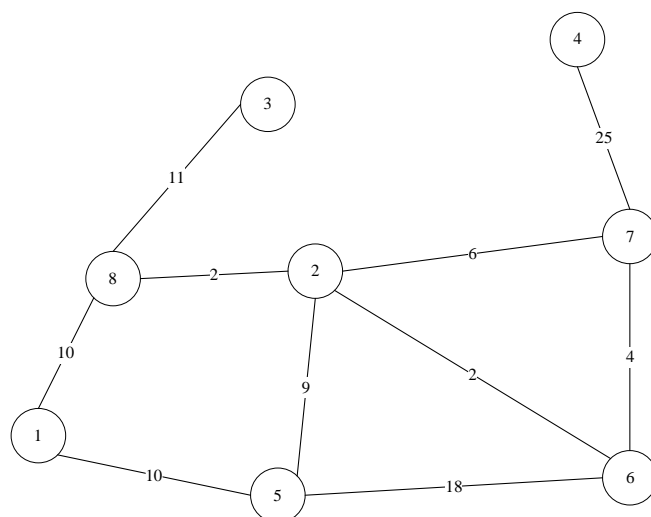


图 3-2 无向带权图网络

另外，还存在着另外一种网络的结构，如图 3-3 所示。当节点失效时，对它所关联的边有可能并不会造成巨大的影响。

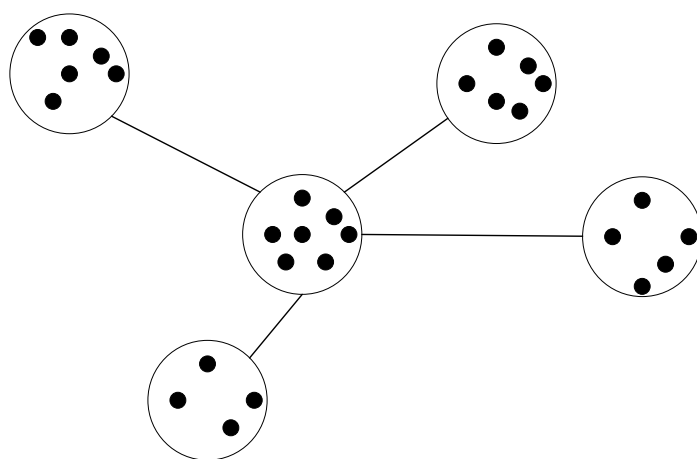


图 3-3 无向带权图网络

对于这种网络结构，节点的失效需要从节点之间所连接的边来考虑，但是有所不同。对于这条边来说，如果连接的某个节点在这条边上占据了一定的权值，这个节点的失效才会对网络产生影响；如果某节点在这条边上占据的权值过少，则该节点的失效不会对网络产生很大的任何影响。所以需要设定一个阈值来控制占比的数值。将连接这条边的所有节点在该边上贡献的权值之和作为边的总权值。

例如，在一个城市中，存在很多的航空公司，将航空公司抽象为节点，航线作为边。当某个航空公司失效时，它不一定会对航线产生很大的影响，除非它在这条航线上承担了一定比例的运能，航线的运能才会因为这个航空公司的失效而受到

很大的影响。所以，合并航空公司间的公共航线作为边，将航空公司对公共航线的贡献和，即两个航空公司在同一航线上的飞行次数之和，作为边的权重，构建航空网络。

### 3.1.2 使用 WGA 方法学习网络中节点的特征向量

在本节中，首先，将构建的网络模型称为网络  $G$ ， $A(G)$ 表示网络  $G$  的邻接矩阵， $W(G)$ 表示网络  $G$  的权重矩阵， $W(G)$ 的元素 $w_{ij}$ 对应每条边的权重，见式(3-1)；节点的权重矩阵  $D(G)$ ，见式(3-2)； $d_i$ 表示节点 $i$ 的权重，见式(3-3)。从而计算得出拉普拉斯矩阵  $L(G)$ ，表示为下式：

$$W(G) = \begin{bmatrix} 0 & w_{12} & \cdots & w_{1n} \\ w_{21} & 0 & \cdots & w_{2n} \\ \cdots & \cdots & 0 & \cdots \\ w_{n1} & w_{n2} & \cdots & 0 \end{bmatrix} \quad (3-1)$$

$$D(G) = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix} \quad (3-2)$$

$$d_i = \sum_{j=1}^n w_{ij} \quad (3-3)$$

计算拉普拉斯矩阵  $L(G)$ ：  $L(G)=D(G)-W(G)$ ，见式(3-4)。

$$L(G) = \begin{bmatrix} d_1 & -w_{12} & \cdots & -w_{1n} \\ -w_{21} & d_2 & \cdots & -w_{2n} \\ \cdots & \cdots & \ddots & \cdots \\ -w_{n1} & -w_{n2} & \cdots & d_n \end{bmatrix} \quad (3-4)$$

其次，使用自编码器（自编码器的作用是将输入的数据作为学习的目标，对输入的数据进行表征学习），将网络的权重矩阵作为输入，得到重构后的权重矩阵。再对权重矩阵的编码损失和权重矩阵的重构损失进行整体优化。

对于节点的权重矩阵的编码损失，其损失函数 $L_{1st}$ 定义如下，见式(3-5)：

$$L_{1st} = \sum_{i,j}^n w_{i,j} \|y_i - y_j\|_2^2 = 2tr(Y^T LY) \quad (3-5)$$

其中 $w_{i,j}$ 是权重矩阵中的元素， $Y$ 是得到的嵌入向量， $L$ 是网络对应的拉普拉斯矩阵， $tr(Y^T LY)$ 表示矩阵 $Y^T LY$ 的迹。。该损失函数 $L_{1st}$ 可以让网络中的相邻的两个节点对应的嵌入向量在隐藏空间接近。

对于权重矩阵的重构损失，其损失函数 $L_{2nd}$ 定义如下，见式(3-6)：

$$L_{2nd} = \sum_{i=1}^n \|w'_i - w_i\|_2^2 \quad (3-6)$$

$w'_i$ 表示重构后的权重矩阵元素， $w_i$ 表示输入的权重矩阵元素，使用网络的权

重矩阵 $W$ 作为输入的信息。在矩阵重构的过程中,可以让结构类似的节点得到相似的向量表示。

利用得到的权重矩阵的编码损失 $L_{1st}$ 和权重矩阵的重构损失 $L_{2nd}$ 进行整体优化,整体优化函数如下,见式(3-7):

$$L_{mix} = L_{2nd} + \alpha L_{1st} + \nu L_{reg} \quad (3-7)$$

$L_{reg}$ 为正则化项, $\alpha$ 为控制编码损失的参数(取值为 $10^{-5}$ ), $\nu$ 为控制正则化项的参数(取值为 $10^{-4}$ )。

最后,编码输出即为每个节点的向量表示(通常,特征向量维度都设为128维)。

学习节点特征向量整体架构如图3-4所示:

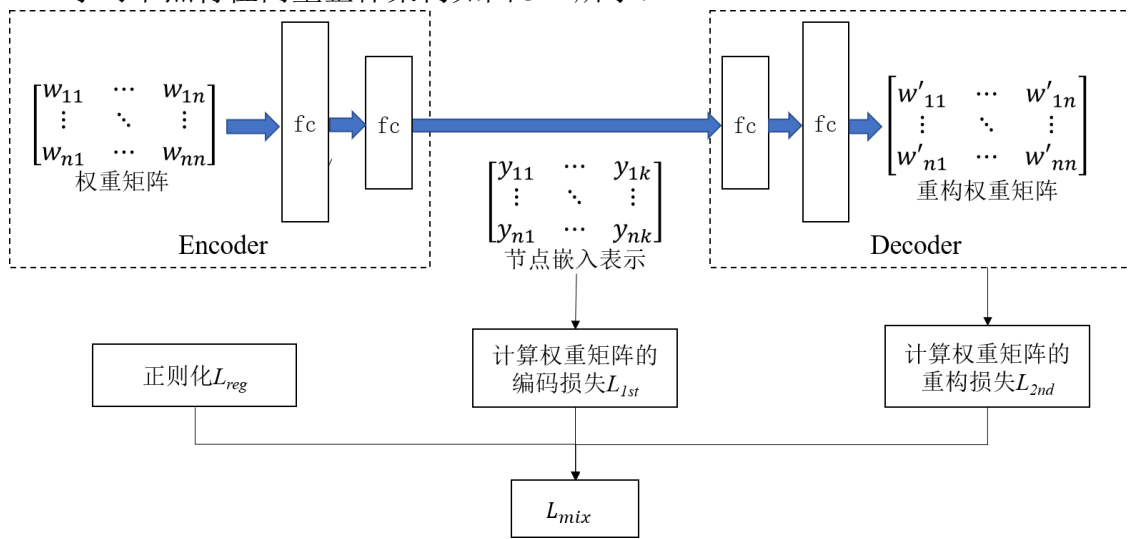


图 3-4 学习节点特征向量整体架构

图3-4中,fc表示全连接层(Fully Connected Layer),全连接层可以将学习得到的“分布式特征表示”转移到样本标记空间

### 3.1.3 对网络中的节点进行重要度评分

本节主要介绍的是,使用上一节通过WGA方法所学习的网络中节点的特征向量,对无向带权图的网络模型中的节点进行重要度评分,记为 $P_i$ 。

网络模型中节点 $i$ 的重要度评分 $P_i$ 的计算公式:  $P_i = \sum_{j \in N_i} dist(v_i, v_j)$ , 其中 $v_i$ 和 $v_j$ 分别表示节点 $i$ 和 $j$ 通过WGA算法学习到的特征向量, $N_i$ 表示除节点 $i$ 的所有节点集合, $dist(v_i, v_j)$ 表示特征向量 $v_i$ 和 $v_j$ 的欧氏距离。

对于节点 $i$ ,学习得到它的128维特征向量 $v_i=(v_{i1}, v_{i2}, \dots, v_{iG})$ ,  $1 \leq i \leq N, G=128$ ,  $N$ 表示节点数量。图3-2所构建的无向带权图的网络模型,使用WGA方法学习网络中8个节点的特征向量,如表3-1所示。

表 3-1 8 个节点的特征向量

节点	特征向量(128 维)
1	(0.121,...,0.231)
2	(0.241,...,0.352)
3	(0.103,...,0.464)
4	(0.171,...,0.353)
5	(0.203,...,0.321)
6	(0.211,...,0.324)
7	(0.193,...,0.254)
8	(0.217,...,0.297)

然后使用节点的特征向量的欧氏距离之和计算无向带权图的网络模型中节点重要度评分 $P_i$ 。对于图 3-2 所构建的无向带权图的网络模型，8 个节点的重要度评分 $P_i$ 如表 3-2 所示。

表 3-2 8 个节点的节点重要度评分 $P_i$ 

	1	2	3	4	5	6	7	8	$P_i$
1	0	0.249	4.169	4.378	4.679	4.576	4.460	4.400	26.911
2	0.249	0	4.104	4.320	4.668	4.562	4.459	4.397	26.759
3	4.169	4.104	0	0.322	4.411	4.276	4.334	4.283	25.899
4	4.378	4.320	0.322	0	4.524	4.397	4.594	4.546	27.081
5	4.679	4.668	4.411	4.524	0	0.181	4.711	4.680	27.852
6	4.576	4.562	4.276	4.397	0.181	0	4.560	4.528	27.079
7	4.460	4.459	4.334	4.594	4.711	4.560	0	0.095	27.212
8	4.400	4.397	4.283	4.546	4.680	4.528	0.095	0	26.930

对 $P_i$ 排序后，得到了一个节点重要性排序序列，按照节点重要度 $P_i$ 从高到低排序，得到节点的排序序列为：5、7、4、6、8、1、2、3。

### 3.1.4 评估节点排序序列的可靠性

针对 3.1.3 节所得到的节点重要度评分排序序列，我们需要某种指标来评判这个序列的可靠性。WGA 方法和几种其他的网络重要节点挖掘方法得到的节点排序

序列都是没有标签的，无法评判哪种方法在哪种网络模型下的节点排序序列是最优的。针对这个问题，在本节我们提出非关键节点比例指标（PNN 指标），来对比 WGA 方法与其他的重要节点挖掘方法的差异。

在本节中，我们采用 PNN 指标来评估网络中节点重要度计算方法的可靠性。PNN 表示删除节点后不会导致严重后果节点比，也就是说，当我们得到一个节点重要度排序序列，按照节点排序从高到低，从网络中依次删除节点，会得到一条对应的 PNN 值随删除节点变化的曲线。显然，如果 PNN 值随删除节点变化越快，则表示删除的节点在网络中越重要。

PNN 指标的定义，见式(3-8)和式(3-9)：

$$PNN(\theta) = \frac{1}{N} \sum_{i \in V} \delta(i) \quad (3-8)$$

$$\delta(i) = \begin{cases} 1, & \frac{N_{failed(i)}}{N} < \theta \\ 0, & otherwise \end{cases} \quad (3-9)$$

其中 $\theta$ 为容忍阈值， $N$ 表示节点总数， $N_{failed(i)}$ 表示删除节点 $i$ 导致的失效节点数。节点失效也要因情况而异。比如，删除一个节点，会导致整个网络结构或者局部网络结构破碎，使得网络的连通性大大降低，可以节点失效，又或者在带权网络中导致网络的运能大大降低，也可以称作节点失效。

对于一般的无向带权图的网络结构，当删除网络中一个节点时，该节点所连接的所有边都会失效，依据这一点，对网络的 PNN 指标中的 $N_{failed(i)}$ 和 $N$ 进行定义，如下：

$N_{failed(i)}$ 定义如下，见式(3-10)：

$$N_{failed(i)} = \sum_j w_j^i \quad (3-10)$$

$N$ 定义如下，见式(3-11)：

$$N = \sum_i \sum_j w_j^i \quad (3-11)$$

$w_j^i$ 表示与节点 $i$ 关联的第 $j$ 条边的权值， $N$ 表示与所有节点关联的边的权值之和。

对于另外一种网络结构，对 PNN 指标中的 $N_{failed(i)}$ 和 $N$ 进行另外的定义，如下：

$N_{failed(i)}$ 定义如下，见式(3-12)和式(3-13)：

$$N_{failed(i)} = \sum_k \mu\left(\frac{w_k^i}{\sum_j w_j^i}\right) \quad (3-12)$$

$$\mu(x) = \begin{cases} 1, & x > \alpha \\ 0, & x \leq \alpha \end{cases} \quad (3-13)$$



$w_k^i$ 表示与节点 $i$ 关联的第 $k$ 条边的权值； $\alpha$ 是预先设定的一个阈值，表示节点 $i$ 失效时，边的失效的程度； $N$ 表示全部边的数量。

## 3.2 WGA 方法的实验及实验分析

本节主要是使用 WGA 方法、Node2vec 算法、度中心性指标、介数中心性指标、PageRank 算法、DIL 方法和 NL 方法在多个数据集上进行实验，每种方法得到对应的 PNN 曲线，并进行分析。根据 PNN 的定义，PNN 曲线变化得越快，说明方法越有效。

### 3.2.1 实验数据及实验环境

本文的实验使用的数据包括美国航空飞行数据集、USAir 网络、UCI 网络、PB 网络和 Email 网络。这些数据集的网络属性，如表 3-3 所示。

表 3-3 实验数据集属性分布

	飞行数据集	USAir 网络	UCI 网络	PB 网络	Email 网络
节点数	809	332	1899	1222	1133
边数	6101	2126	13838	16714	5451
平均度	15.083	12.810	14.570	27.360	9.620
平均聚类系数	0.679	0.749	0.138	0.360	0.254
平均路径长度	3.147	2.738	3.055	2.741	3.611

绘制飞行数据集、USAir 网络、UCI 网络、PB 网络和 Email 网络的度分布图，分别如图 3-5、图 3-6、图 3-7、图 3-8 和图 3-9 所示。

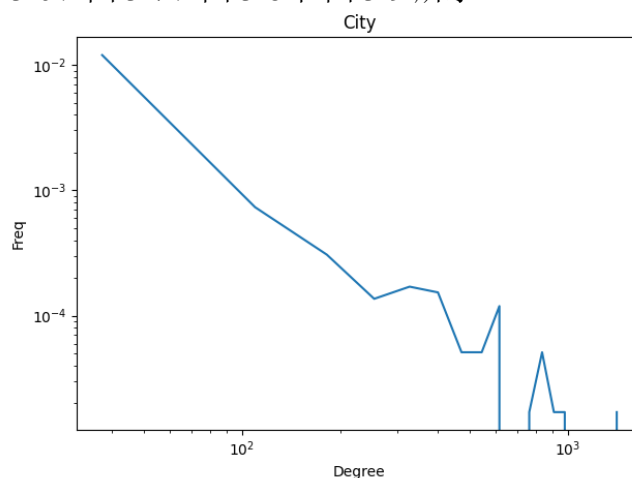


图 3-5 飞行数据集度分布图

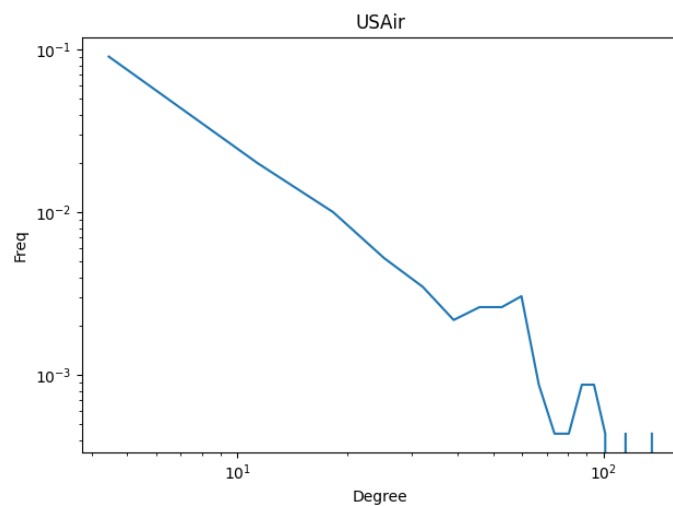


图 3-6 USAir 网络度分布图

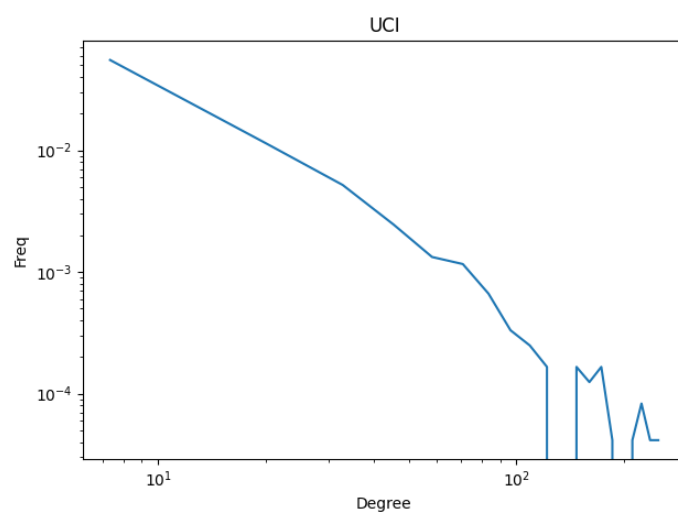


图 3-7 UCI 网络度分布图

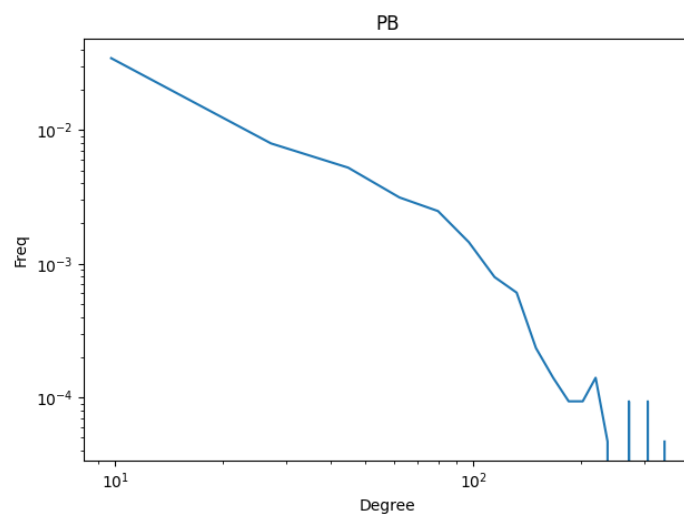


图 3-8 PB 网络度分布图

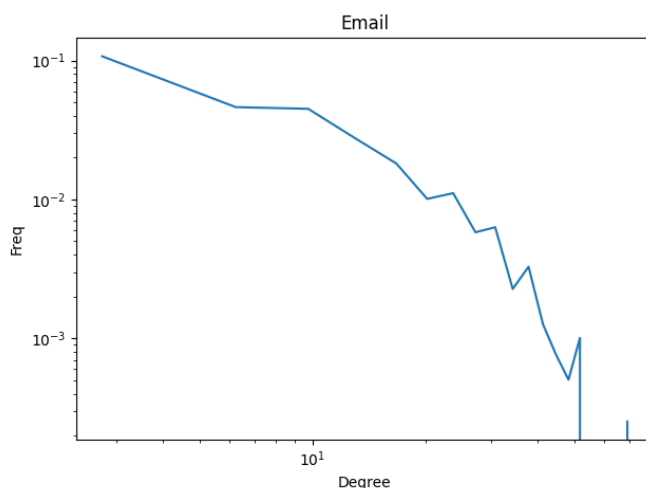


图 3-9 Email 网络度分布图

这五种数据集的度分布都大致服从幂律分布，表明大约 20% 的节点是大度节点，剩下的大部分是小度节点，节点度之间存在异配的关系。

本文有关的实验使用的是 python 语言，使用 Pycharm 进行编程，版本为 python3.6。实验配置见表 3-4。

表 3-4 实验配置

计算机属性	计算机属性值
处理器	Intel(R) Core(TM) i5-4210 CPU @1.70GHz 2.40GHz
显卡	AMD Radeon R7 M265
已安装内存	8.00 GB
操作系统	Windows 10 教育版
操作系统类型	64 位
实验环境	Python 3.6 Pycharm

## 3.2.2 重要节点挖掘的实验结果及分析

### 3.2.2.1 PNN 实验结果及分析

在图 3-10，图 3-11，图 3-12，图 3-13，图 3-14 和图 3-15 中，使用 WGA 方法、Node2vec 算法、度中心性指标、介数中心性指标、PageRank 算法、DIL 方法和 NL 方法分别在城市网络、航司网络、USAir 网络、UCI 网络、PB 网络和 Email 网络中的重要节点进行挖掘，然后对每种方法得出的对应的节点重要性序列进行评估，按照节点重要性从高到低从网络中移除节点，使用 PNN 作为评价指标，分析每种方法在不同网络下的实验结果。

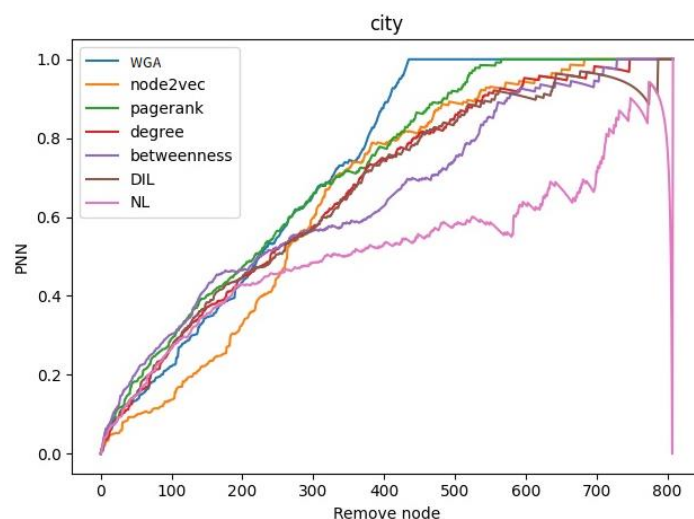


图 3-10 城市网络 PNN 曲线

在图 3-10 中，在移除前几十个左右节点的时候，WGA 方法的 PNN 曲线变化的最快，说明 WGA 方法对城市网络中重要节点的挖掘是最有效的。在移除节点达到 200 个左右的时候，除了 Node2vec 算法外，其余六种方法的 PNN 曲线变化相似，说明 WGA 算法与其余五种重要节点排序算法在该城市网络上效果差不多。而 NL 方法在最后几个移除节点上，其 PNN 曲线发生震荡的原因，很可能是由于，最后几个节点变成了孤立节点，PNN 值会迅速变为 0，当节点全部删除的时候，PNN 值跳变为 1。随着节点的不断删除，本方法也以最快的速度达到了 PNN 的最大值。证明了本方法在网络中重要节点挖掘上的可行性。

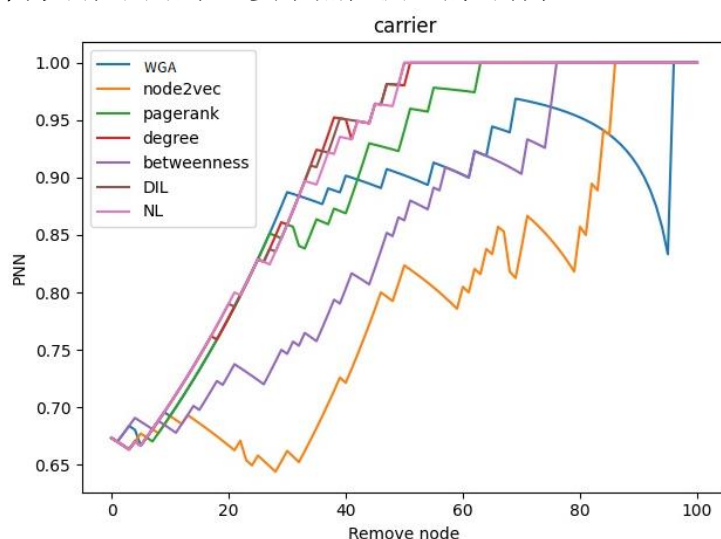


图 3-11 航司网络 PNN 曲线

在图 3-11 中，PNN 变化曲线是不平滑的，可能有几种原因：一是航司节点数目偏少（101 个），二是航司网络是一种特殊的网络结构，它的 PNN 定义与一般的

无向带权图的 PNN 定义有所不同。航司网络 PNN 的定义从实际出发，考虑了多方面的因素，因此可能会产生曲线不平滑的结果。而且，由于节点数目偏少，我们只需要分析前 20 个节点即可。在移除前几个节点时，我们可以看出本方法与 NL 方法是最优的，NL 方法是 2019 年提出来的比较新颖的节点重要性方法，说明本方法较于其他方法在该航司网络中也是具有优越性的。在移除节点达 20 个时，WGA 方法与多个方法效果一致，同样说明了 WGA 方法的可行性。

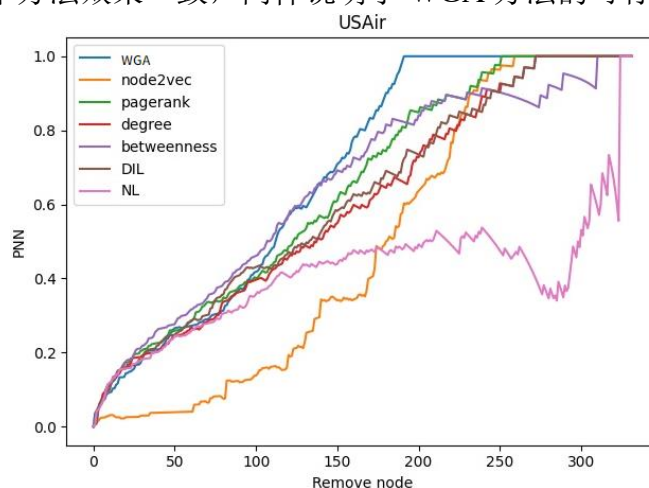


图 3-12 USAir 网络 PNN 曲线

在图 3-12 中，USAir 网络的 PNN 曲线中，在移除前 150 个节点时，Node2vec 算法在该网络中的效果相对较差，但是 WGA 方法在该网络中的效果非常不错。WGA 方法中 PNN 曲线变化比较快，在移除不到 200 个节点的时候，PNN 值就已经到达了最大值，而其中 PNN 值最快到达 1 的方法有 PageRank 和 Node2vec 算法，但是这两种方法移除节点数达到了 250，其余方法更加缓慢。这说明了，在 USAir 网络，WGA 方法的效果是最好的。

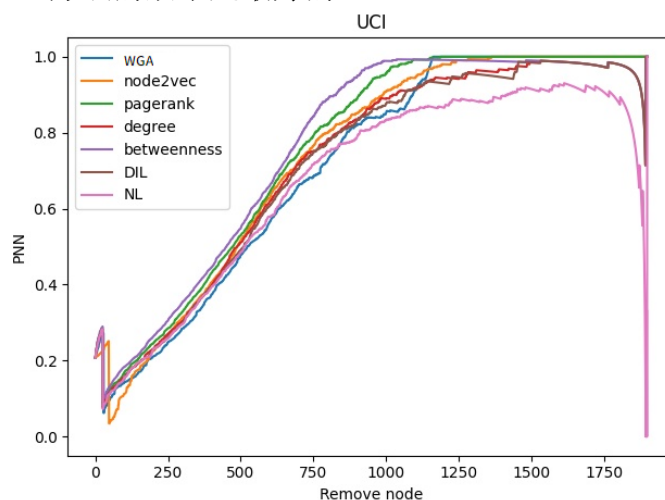


图 3-13 UCI 网络 PNN 曲线

在图 3-13 中, 在 UCI 网络中, 在移除前 750 个节点时, 每种方法表现都较好。WGA 方法与其他几种方法效果近似。NL 方法与 DIL 方法在最后几个节点处 PNN 值跳变为 0 的原因, 很有可能最后剩下的节点为孤立节点, 当最后移除全部节点时, PNN 值又跳变为 1。总而言之, WGA 方法在 UCI 网络上适用的。

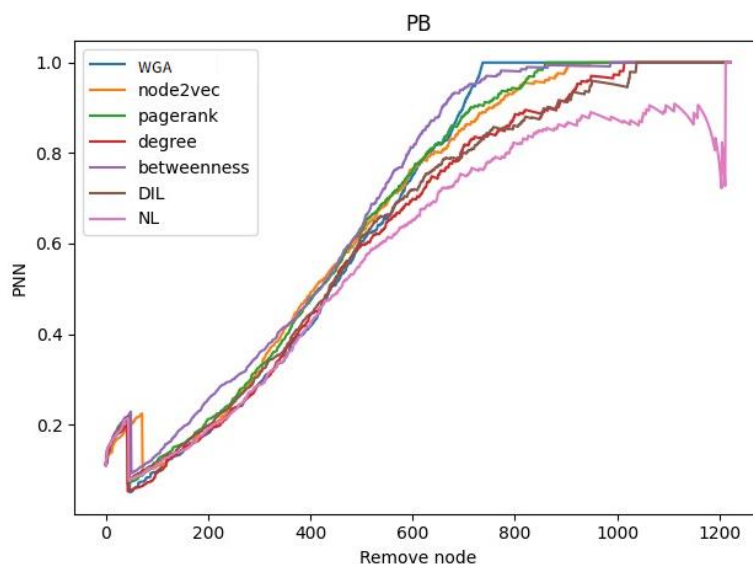


图 3-14 PB 网络 PNN 曲线

在图 3-14 中, PB 网络的 PNN 曲线与图 3-13 中 UCI 网络的 PNN 曲线较为类似, 在移除前 600 个节点时, 每种方法的效果也较好, 差距也不明显。但是 WGA 方法的 PNN 值是最快达到 1 的, 相比于其他方法, 少移除了近 200 个节点, 要知道 PB 网络总共也就 1200 多个节点, 因此, 在 PB 网络上, WGA 方法是优于其他方法的。

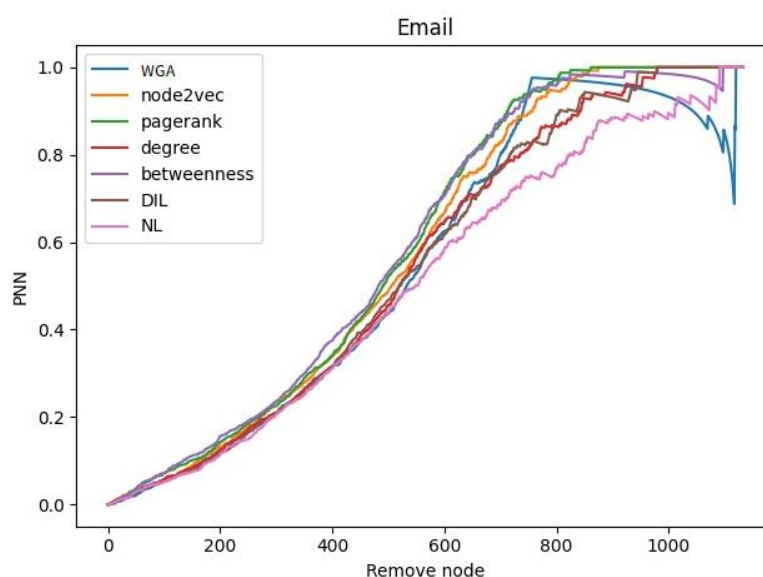


图 3-15 Email 网络 PNN 曲线

在图 3-15 中, Email 网络的 PNN 曲线上, 在移除前 400 个节点左右的时候, 每种方法在网络上的效果都较为类似, 当移除节点数在 700 个左右的时候, Node2vec 算法与 WGA 方法效果良好。由于 Email 网络只有一千多个节点, 只考虑前几百个节点就可以说明方法的可行性。所以, 在 Email 网络上, WGA 方法在重要节点的挖掘上是可行的。

### 3.2.2.2 WGA 方法挖掘重要节点的可靠性分析

为了分析 WGA 方法在真实情境下的挖掘重要节点的可靠性, 我们以城市网络和航司网络为例, 在使用 WGA 方法对网络中的重要节点进行挖掘, 在真实的美国地图上展示所挖掘到的重要城市节点。

如图 3-16 所示, 表示的是通过 WGA 方法对城市网络中的节点进行重要性排序的结果, 节点的大小反映了城市节点的重要程度。

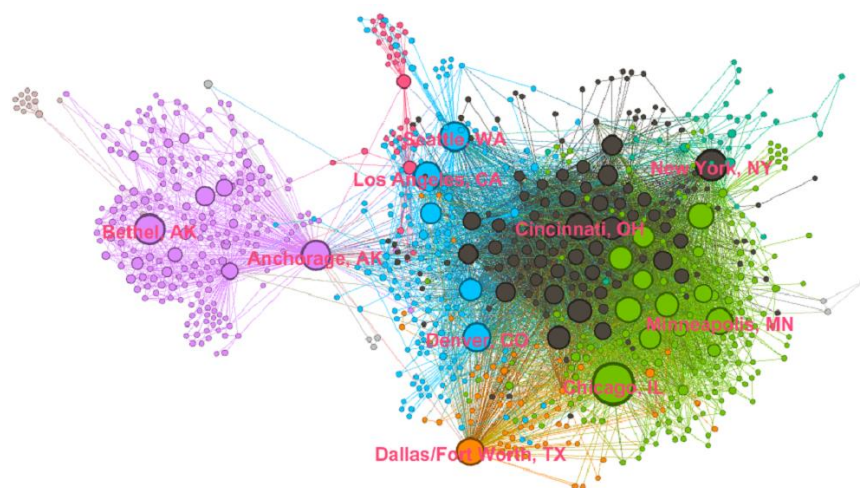


图 3-16 WGA 方法对城市网络节点的挖掘

图 3-16 中, 显示了节点排名前十的城市节点, 并且可以清楚地看出城市节点 Chicago,IL 最大, 表明城市 Chicago,IL 在该城市网络中的重要性最高。同时, 城市 Anchorage,AK 也是非常重要的节点, 从图 3-16 中可以看出, 城市 Anchorage,AK 起到了桥梁和中转节点的作用, 分布在城市 Anchorage,AK 两边的城市节点, 大多数都需要经过它来连接。

接下来, 在真实的美国地图上, 将图 3-16 中的排名前十的城市标注出来, 分析 WGA 算法在挖掘重要城市节点上的可靠性。如图 3-17 所示。

在图 3-17 中, 清晰地显示了 Chicago, IL、New York, NY 等著名城市在美国地图上的地理位置。根据我们的常识, Chicago, IL、New York, NY、Seattle, WA、Los Angeles, CA 等在图 3-15 中的十个城市都是在美国航空网络中占据重要地位的城



市。而且，从地理位置上看，城市 Anchorage,AK 位于美国阿拉斯加州的边缘，与美国其余州隔洋相望，Anchorage,AK 也就成了阿拉斯加州与其余州交流的连接点。在本文所构建的城市网络中，得出 Chicago, IL 和 Anchorage,AK 等是很重要的城市，这也是非常符合实际的。因此，WGA 算法挖掘城市网络中的重要节点是符合我们日常认知的。

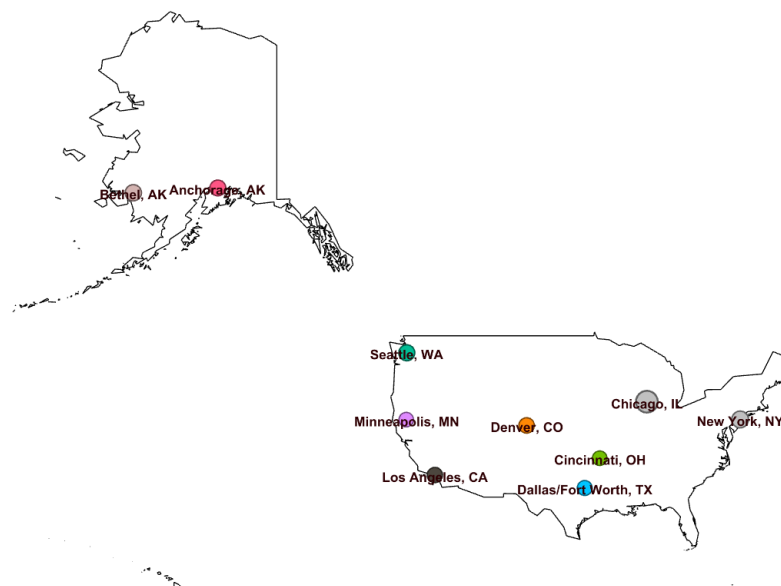


图 3-17 重要性前十的城市（美国地图）

在图 3-18 中，表示的是通过 WGA 方法对航司网络中的节点进行重要性排序的结果，节点的大小反映了航司节点的重要程度。

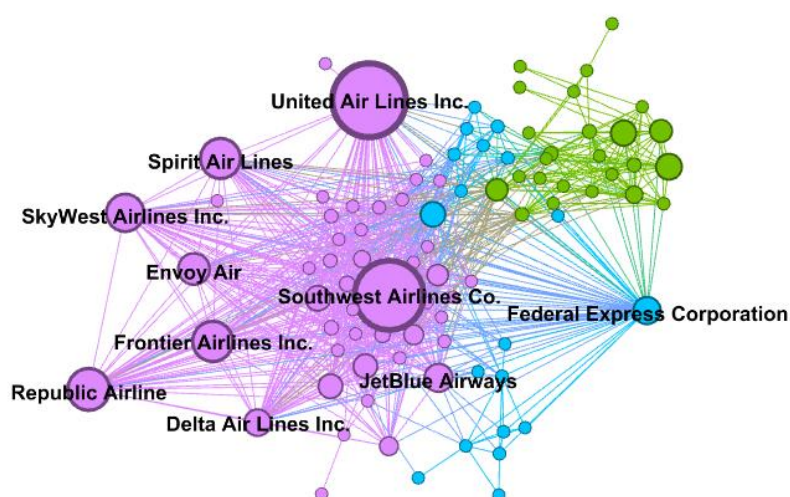


图 3-18 WGA 方法对航司网络节点的挖掘

图 3-18 中，显示了节点排名前十的航司节点，并且可以清楚地看出航司节点



United AirLines Inc.最大, 表明航空公司 United AirLines Inc.在该航司网络中的重要性最高。

接下来, 在真实的美国地图上, 将图 3-18 中的排名前十的航空公司标注出来, 分析 WGA 方法在挖掘重要航司节点上的可靠性。如图 3-19 所示。

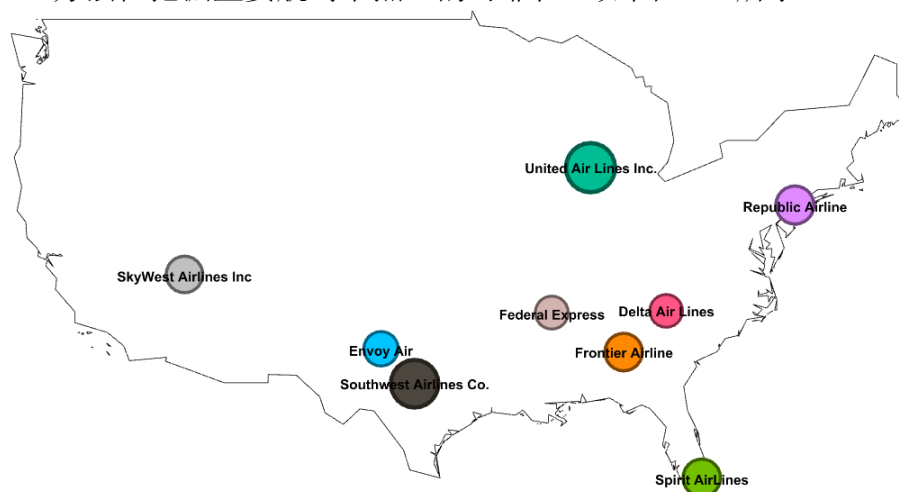


图 3-19 重要性前十的航空公司 (美国地图)

在图 3-19 中, 航空公司的地理位置使用的是它所在的城市的坐标, 由于 Republic Airline 和 JetBlue Airways 都位于 New York, NY, 为防止节点名字重合, 故只显示了九个航空公司的节点标签。图中显示了 United AirLines Inc.、Southwest Airlines Co.等重要航空公司在美国地图上的地理位置。航空公司 United AirLines Inc.位于城市 Chicago, IL, Republic Airline 位于 New York, NY 都印证了所挖掘到的航空公司的重要性。在本文所构建的航司网络中, 得出 United AirLines Inc.是最重要的航空公司, 这也是非常符合实际的。因此, WGA 方法挖掘航司网络中的重要节点也是合理的。

在上面的多个实验中, PNN 曲线在随着节点的删除时, 在某些网络中, 并没有单调递增甚至出现了震荡, 经过分析知, 当随着节点不断地删除, 有些网络可能变成了孤立网络, 或者一个个孤立的社团结构, 所以当删除这些节点时, 出现这些问题也是可能的。但是并不影响实验的结果, 对于每种网络, 都是删除最后少量不重要节点的时候, 才会出现这样的结果, 删除前面大量的重要节点时, PNN 曲线还是单调递增的。

总而言之, 本文所提出的 WGA 方法在网络上挖掘重要节点的效果上是可行的, 并且在真实情境下得到了验证。虽然 WGA 方法挖掘重要节点在有些实验中效果不是最好的, 但是 WGA 方法是首次将图神经网络运用在节点挖掘方向上, 并具有一定的可靠性。

### 3.3 本章小结

总而言之，本文所提出的 WGA 方法所提出的针对重要节点挖掘的方法，在网络上挖掘重要节点的效果上是可行的，并且在真实情境下得到了验证。最重要的是，本文从另外一种角度思考了重要节点的挖掘，拓宽了节点挖掘的思路。

## 第四章 网络鲁棒性优化方法

在第三章中,我们将图神经网络算法运用在了网络中重要节点的挖掘上,提出了 WGA 方法,在本章我们将要讨论如何优化网络的鲁棒性。首先,本章基于节点权值的方法对节点进行排序,然后提出修改基尼系数的定义作为网络鲁棒性的评价指标,在最重要的节点的邻居节点构成的局部网络上进行。使用邻居间节点增边的策略优化网络的鲁棒性,提出在增边条数和增边容量限制下的贪心增边策略和动态规划增边策略,并对比了在度值和最大的节点间增边的方法以及在介数值最大的节点间增边的方法。最后在多个数据集下进行了验证。

### 4.1 网络鲁棒性的优化

对于网络鲁棒性的优化,传统的网络鲁棒性指标都是针对图本身进行定义的,例如连通度,连通子图相对大小和网络效率。本章对网络鲁棒性指标进行重新定义,结合现实,提出基尼系数作为网络鲁棒性评价指标,不仅从网络结构本身,而且考虑了基尼系数本身的含义,从新的角度分析网络的鲁棒性。

#### 4.1.1 构建网络鲁棒性研究的模型

构建鲁棒性研究的网络模型可表示为:  $G(V, E, W)$ , 其中  $V = \{v_i | i = 1, 2, \dots, n\}$  表示网络中所有的节点,  $E = \{e_{ij} | i, j = 1, 2, \dots, n\}$  表示网络中所有的边,  $W = \{w_{ij} | i, j = 1, 2, \dots, n\}$  表示网络中所有边的权值。所构建的是一个无向带权图的网络模型。

例如构建一个含有 10 个节点和 15 条边的网络模型,边上的数值表示边的权值,如图 4-1 所示。

网络鲁棒性的优化,首先要基于某种重要节点排序方法。

所以,为了计算的方便和直观,本文对网络模型中的节点的重要性计算采用的基于节点权值的方法。节点连接的所有边的权值之和作为节点的重要性,边的权值之和越大,节点就越重要。

节点的重要度  $s_i$  的计算公式如下,见公式(4-1):

$$s_i = \sum_{j \in N(i)} w_{ij} \quad (4-1)$$

其中  $N(i)$  表示节点  $v_i$  的邻居。所有节点的重要度表示为  $S = \{s_i | i = 1, 2, \dots, n\}$ 。 $s_i$  越大,表示节点越重要。

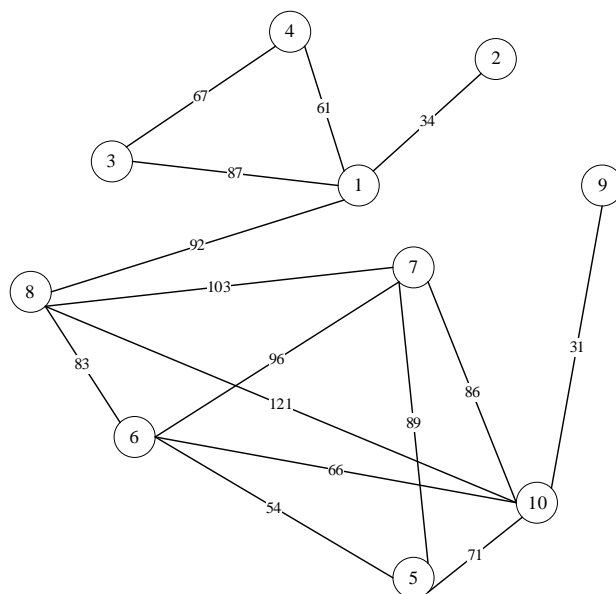


图 4-1 无向带权图的网络模型

对于图 4-1 所构建的无向带权图的网络模型, 分别对网络中的节点进行重要度计算, 计算结果, 见表 4-1。

表 4-1 网络中节点重要度计算表

节点	重要度 $s_i$
1	$s_1 = 92+87+61+34=274$
2	$s_2 = 34$
3	$s_3 = 67+87=154$
4	$s_4 = 67+61=128$
5	$s_5 = 54+89+71=214$
6	$s_6 = 83+96+66+54=299$
7	$s_7 = 103+86+89+96=374$
8	$s_8 = 92+103+121+83=399$
9	$s_9 = 31$
10	$s_{10} = 86+31+71+66+121=375$

基于节点权值的节点排序方法, 在无向带权的网络中, 如果这个网络的节点和边的数量非常巨大的话, 当网络受到蓄意攻击的时候, 往往重要度评分高的节点比重要度评分低的节点更容易受到攻击, 因此重要度评分高的节点在网络中更应该受到关注。同时, 如果不考虑增加网络边的数量, 对一个庞大的无向带权网络的优化, 计算量也是相当大的。所以, 针对一个庞大的无向带权网络, 可以选择特定的优化范围, 以减少所增加的边的数量。

因此,考虑从最重要节点的邻居节点集合 $N$ 中选出一些节点构成局部网络。而节点集合 $N$ 中的节点重要性不一定都较高,所以从节点集合 $N$ 中选择重要性排名前20%的节点集合 $H$ 构成优化的局部网络节点。在确定局部网络的节点集合 $H$ 后,随机选择 $H$ 中节点构成边的30%作为候选边集合 $E'$ , $E'$ 作为实验的测试数据集,以模拟工程中提供的外部候选边集资源。最终,在原有的网络 $G$ 中移除候选边集 $E'$ ,形成子网络 $G'$ ,网络 $G'$ 为真实的网络。可以表示为如下步骤:

1. 选择最重要节点的邻居节点集合 $N$ 。
2. 选择 $N$ 中重要性排名前20%的节点集合 $H$ 。
3. 随机选择 $H$ 中节点构成边的30%作为候选边集 $E'$ 。
4. 在原网络中移除候选边集 $E'$ ,形成子网络 $G'$ , $G'$ 就是需要进行鲁棒性优化的网络。

#### 4.1.2 网络鲁棒性指标的分析

而在本章中,结合所构建的网络模型,对网络的鲁棒性指标进行重新定义。首先提出基尼系数的概念,基尼系数可以反映网络中不同节点的重要性差异,结合基尼系数的含义,使用基尼系数作为网络鲁棒性的指标,见公式(4-2):

$$gini(S) = 1 - \sum \left( \frac{s_i}{\sum s_k} \right)^2 \quad (4-2)$$

公式(4-2)中 $S = \{s_i | i = 1, 2, \dots, n\}$ 表示节点 $v_i$ 的重要性。

所以,对于图4-1中所构建的网络模型,结合基尼系数对网络中的节点进行网络鲁棒性值的计算,计算结果见表4-2。

表 4-2 网络鲁棒性指标计算表

节点	重要度 $s_i$	$(s_i / \sum s_k)^2$
1	$s_1 = 274$	0.0144
2	$s_2 = 34$	0.0002
3	$s_3 = 154$	0.0046
4	$s_4 = 128$	0.0031
5	$s_5 = 214$	0.0088
6	$s_6 = 299$	0.0171
7	$s_7 = 374$	0.0269
8	$s_8 = 399$	0.0306
9	$s_9 = 31$	0.0002
10	$s_{10} = 375$	0.0270

计算得出，图 4-1 所构建的网络的基尼系数为

$$gini(S) = 1 - \sum \left( \frac{s_i}{\sum s_k} \right)^2 = 0.8671 \quad (4-3)$$

对于公式(4-2)所提出的基尼系数指标，可以这样理解，在一个无向带权图的网络中，基于某种节点排序方法，使用基尼系数作为网络鲁棒性评价的指标评价整个网络的鲁棒性。

当对网络进行增边时，选择边使得网络整体的基尼系数增加的越快，说明选择的边是合适的。因此，后文将对选择的边进行详细的分析。本节着重对基尼系数进行分析。

举个简单的例子，如图 4-2 所示的包含 4 个节点和 3 条边的带权图 $N$ ，每条边上的值表示边的权值。

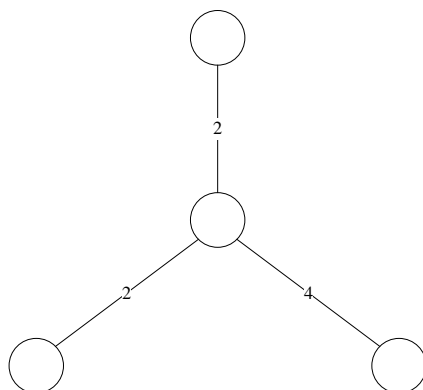


图 4-2 星型网络 $N$

对于图 4-2 所示的网络 $N$ ，计算网络 $N$ 的基尼系数，为：

$$gini(N) = 1 - \left( \left( \frac{2}{16} \right)^2 + \left( \frac{2}{16} \right)^2 + \left( \frac{4}{16} \right)^2 + \left( \frac{8}{16} \right)^2 \right) = \frac{11}{32} \quad (4-4)$$

如果我们对网络 $N$ 加权值为 2 一条边，变成网络 $N'$ ，如图 4-3 所示。

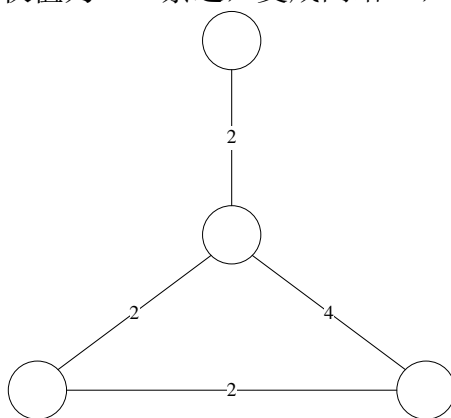


图 4-3 网络 $N'$

对于图 4-3 所示的加边网络 $N'$ ，计算网络 $N'$ 的基尼系数，为：

$$gini(N') = 1 - \left( \left( \frac{2}{20} \right)^2 + \left( \frac{4}{20} \right)^2 + \left( \frac{6}{20} \right)^2 + \left( \frac{8}{20} \right)^2 \right) = \frac{7}{10} \quad (4-5)$$

而对于一个完全图 $N''$ 来说，如图 4-4 所示。

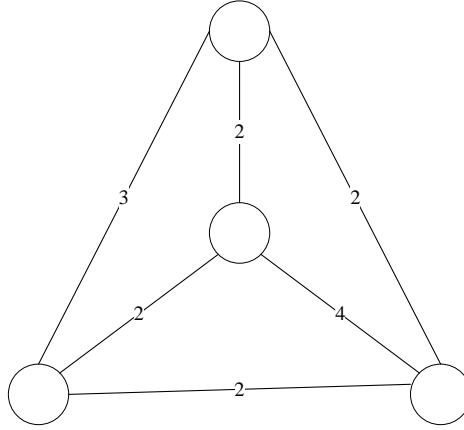


图 4-4 完全图网络 $N''$

对于图 4-4 所示的完全图网络 $N''$ ，计算网络 $N''$ 的基尼系数，为：

$$gini(N'') = 1 - \left( \left( \frac{7}{30} \right)^2 + \left( \frac{8}{30} \right)^2 + \left( \frac{8}{30} \right)^2 + \left( \frac{7}{30} \right)^2 \right) = \frac{23}{30} \quad (4-6)$$

经过比较，发现 $gini(N'') > gini(N') > gini(N)$ 。这也是符合常理的。这主要是通过增边的方法，降低了网络中核心节点的重要性，使得网络中的核心节点受到蓄意攻击的时候，网络还能保持一定的结构完整性和连通性。

接下来，对公式 $gini(S) = 1 - \sum \left( \frac{s_i}{\sum s_k} \right)^2$ 进行分析。

设 $S$ 表示当前节点评分集合， $S'$ 表示在节点 $i$ 和节点 $j$ 之间增加权重为 $w_{ij}$ 的边 $(i, j)$ 后的评分集合，有：

$$gini(S) = 1 - \sum \left( \frac{s_i}{\sum s_k} \right)^2 = 1 - \frac{s_1^2 + \dots + s_n^2}{(s_1 + \dots + s_n)^2} \quad (4-7)$$

$$\begin{aligned} gini(S') &= 1 - \frac{s_1^2 + \dots + (s_i + w_{ij})^2 + \dots + (s_j + w_{ij})^2 + \dots + s_n^2}{[s_1 + \dots + (s_i + w_{ij}) + \dots + (s_j + w_{ij}) + \dots + s_n]^2} \\ &= 1 - \frac{s_1^2 + \dots + s_n^2 + 2w_{ij}^2 + 2s_i w_{ij} + 2s_j w_{ij}}{(s_1 + \dots + s_n + 2w_{ij})^2} \end{aligned} \quad (4-8)$$

令 $p = s_1^2 + \dots + s_n^2$ ， $q = s_1 + \dots + s_n$ ， $p$ 、 $q$ 均为常数。

$$gini(S) = 1 - \frac{p}{q^2}, \quad gini(S') = 1 - \frac{p + 2w_{ij}^2 + 2s_i w_{ij} + 2s_j w_{ij}}{(q + 2w_{ij})^2} \quad (4-9)$$

在节点 $i$ 和节点 $j$ 之间增加权重为 $w_{ij}$ 的边 $(i, j)$ 后，基尼系数的增量为：

$$\Delta gini = gini(S') - gini(S) = \frac{p}{q^2} - \frac{p+2w_{ij}^2+2s_iw_{ij}+2s_jw_{ij}}{(q+2w_{ij})^2} \quad (4-10)$$

当增加的权重为 $w_{ij}$ 的边 $(i, j)$ 使得 $\frac{p+2w_{ij}^2+2s_iw_{ij}+2s_jw_{ij}}{(q+2w_{ij})^2}$ 最小时,  $\Delta gini$ 最大, 即当基尼系数增量最大。

接下来, 对基尼系数和基尼增量进行收敛性分析。

$$(s_1 + \dots + s_n)^2 \leq n(s_1^2 + \dots + s_n^2) \quad (4-11)$$

$$\frac{p}{q^2} = \frac{s_1^2 + \dots + s_n^2}{(s_1 + \dots + s_n)^2} \geq \frac{1}{n} \quad (4-12)$$

对于公式 (4-12), 当且仅当 $s_1 = s_2 = \dots = s_n$ 时,  $\frac{p}{q^2} = \frac{1}{n}$ 。

$$\lim_{n \rightarrow \infty} gini(S) = \lim_{n \rightarrow \infty} 1 - \frac{p}{q^2} \leq \lim_{n \rightarrow \infty} 1 - \frac{1}{n} = 1 \quad (4-13)$$

对于公式 (4-13), 若  $n$  足够大, 当且仅当 $s_1 = s_2 = \dots = s_n$ 时,  $gini(S) \rightarrow 1$ , 因此基尼系数指标是收敛的。当 $s_1 = s_2 = \dots = s_n$ 时,  $\lim_{n \rightarrow \infty} gini(S) = 1 - \frac{1}{n} \rightarrow 1$ , 表明此时网络最稳定, 这是符合常理的。

对于 $\Delta gini$ , 所增加的边的权值 $w_{ij}$ 通常满足 $w_{ij} \leq \min(s_i, s_j)$ 这个条件。

$$\lim_{w_{ij} \rightarrow 0} \Delta gini = 0 \quad (4-14)$$

当  $n$  足够大时, 不论 $w_{ij}$ 的值多大, 有:

$$\lim_{n \rightarrow \infty} \Delta gini = 0 \quad (4-15)$$

由公式 (4-14) 和公式 (4-15) 可知, 基尼增量是收敛的。

所以, 我们要找出某种在最重要节点的邻居间的增边方法, 使得网络的基尼系数增量尽可能大。

总而言之, 本节提出了一种在无向带权网络中的使用基尼指数作为网络鲁棒性指标的方法。通过增边的方法, 降低重要节点在网络中的重要性, 从而保护重要节点, 优化网络鲁棒性, 是接下来要探讨的问题。

### 4.1.3 基于贪心策略的网络增边优化方法

在节点 $i$ 和节点 $j$ 之间增加权重为 $w_{ij}$ 的边 $(i, j)$ 后得到网络的 $gini$ 值, 见公式(4-8)。本节提出三种贪心增边策略 Greedy\_A、Greedy\_B、Greedy\_C, 具体过程如下:

(1) 为了使基尼系数增加的越快,  $gini(S)$  对 $s_i$ 求偏导, 见公式 (4-16), 使得偏导数 $\frac{\partial(gini(S))}{\partial(s_i)}$ 最大。

$$\frac{\partial(gini(S))}{\partial(s_i)} = \frac{2 \sum_{k \neq i} s_k^2 - 2 s_i \sum_{k \neq i} s_k}{(\sum_{k \neq i} s_k + s_i)^3} \quad (4-16)$$

因此贪心策略 Greedy\_A 的具体的增边步骤如图 4-5 所示:



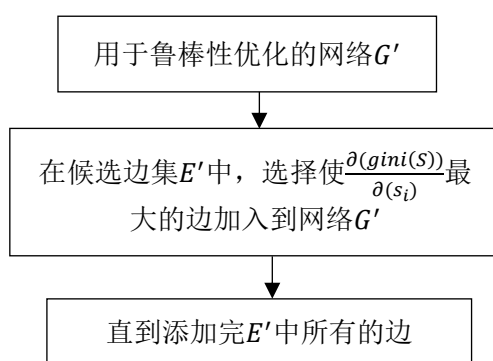


图 4-5 贪心策略 Greedy\_A 增边过程

(2) 贪心策略 Greedy\_B 考虑的是增加边后前后基尼系数的差值, 将  $\frac{p+2w_{ij}^2+2s_iw_{ij}+2s_jw_{ij}}{(q+2w_{ij})^2}$  记为  $e_i$ 。每次在网络中增加一条边, 重新计算  $e_i$  的值。贪心策略 Greedy\_B 具体的增边步骤如图 4-6 所示:

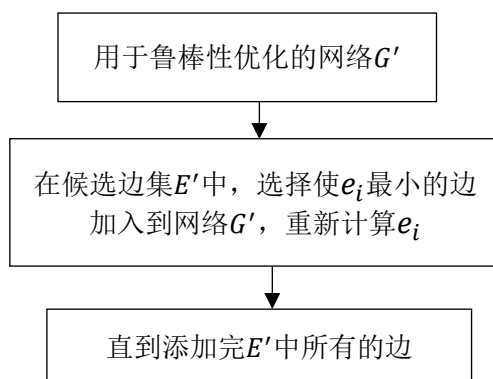


图 4-6 贪心策略 Greedy\_B 增边过程

(3) 在上述方法中, 每次加入边时需要计算  $\frac{p+2w_{ij}^2+2s_iw_{ij}+2s_jw_{ij}}{(q+2w_{ij})^2}$ , 记为  $e_i$ , 计算过程非常耗时。为了减少计算资源, 设计了贪心策略增边优化的近似方法 Greedy\_C。首先将候选边集中所有边的  $e_i$  值计算出来, 再进行增边。贪心策略 Greedy\_B 增边的近似贪心策略 Greedy\_C 的实验步骤如图 4-7 所示:

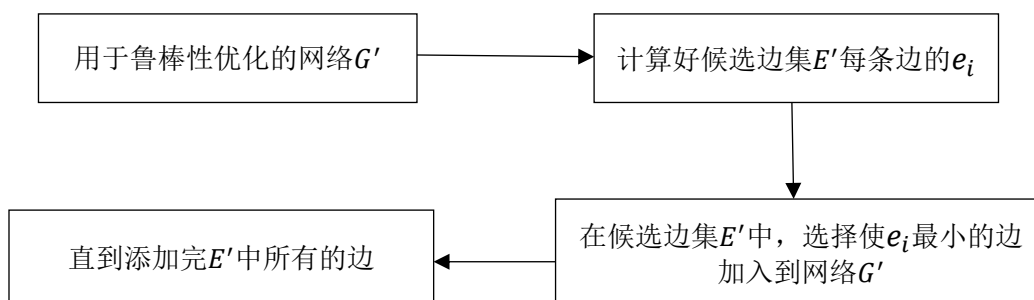


图 4-7 贪心策略 Greedy\_C 增边过程

#### 4.1.4 基于动态规划的网络增边优化方法

在网络增边容量和条数存在限制下，其增边的容量限制： $\sum w_i \leq C$ ，其增边条数限制 $\sum l_i \leq L$ 。此时问题为：从候选边集中选择适当的边，在增加边的总容量不超过 $C$ 、条数不超过 $L$ 的情况下，使增边后的 $gini$ 增量最大。采用的增边方法如图 4-8 所示：

- 1) 计算候选边集 $E'$ 中每条边的 $gini$ 增量 $\frac{p}{q^2} - \frac{p+2w_{ij}^2+2s_iw_{ij}+2s_jw_{ij}}{(q+2w_{ij})^2}$ ，记为 $v_i$ 。
- 2) 记候选边集中每条边的权重为 $w_i$ ，每条边的条数 $l_i = 1$ 。
- 3) 近似使用动态规划求解，定义 $dp(i, j, k)$ 表示前 $i$ 条边中，增边总容量不超过 $j$ ，增边条数不超过 $k$ 的最大 $gini$ 增量。
- 4) 状态转移方程：

$$dp = \begin{cases} 0, & i = 0 \\ dp(i-1, j, k), & j < w_i \text{ or } k < l_i \\ \max\{dp(i-1, j, k), v_i + dp(i-1, j-w_i, k-l_i)\}, & j \geq w_i \text{ and } k \geq l_i \end{cases} \quad (4-17)$$

最终，在满足限制条件下，使 $gini$ 增量最大的边为需要添加的边。

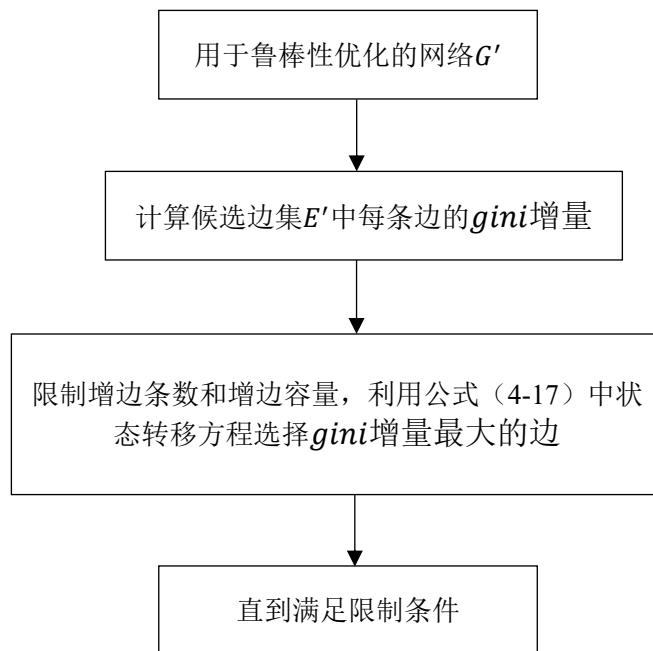


图 4-8 动态规划增边过程

#### 4.1.5 基于节点的度和介数值的网络增边优化方法

(1) 计算每个节点的度值的大小，按照节点的度的大小顺序，在两个度值和最大的节点间进行增边，在增边的条数或增边的容量的条件下，计算增边的基尼系

数曲线。

(2) 计算每个节点的介数值的大小, 按照节点的介数值的大小顺序, 在两个介数值和最大的节点间进行增边, 在增边的条数或增边的容量的条件下, 计算增边的基尼系数曲线。

## 4.2 增边优化网络鲁棒性的实验及实验分析

本节的实验与第三章所使用的数据集和实验环境是一样的。下面对实验进行分析。

### 4.2.1 动态规划增边策略与贪心增边策略的对比

(1) 基于贪心策略 Greedy\_A, 贪心策略 Greedy\_B, 贪心策略 Greedy\_C 以及随机加边策略实验结果如图 4-9 所示。

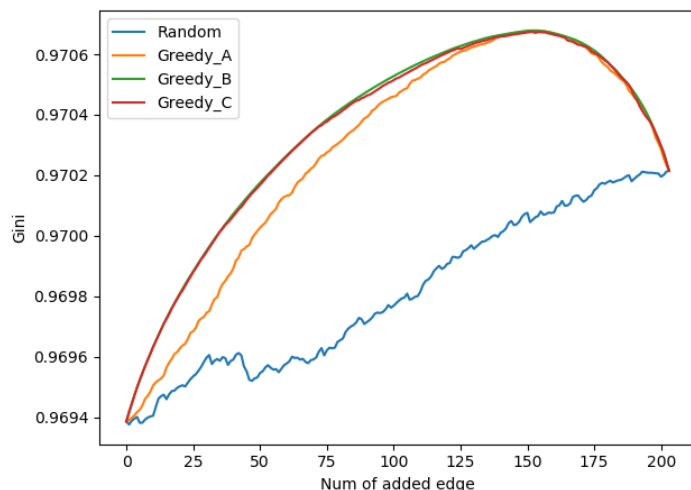


图 4-9 贪心策略实验结果

在图 4-9 的结果中, 贪心策略 Greedy\_B 的效果也一直在贪心策略 Greedy\_A 的之上。根据基尼系数增量的分析, 由贪心的思想选择使当前基尼增加最大的边。结果显示贪心方法选择的边使基尼系数先增加后下降, 并且只需加少量的边就能达到较高的基尼系数值, 明显优于随机选择。说明贪心算法选择使基尼增量最大的边对网络优化有明确的可行性。经过分析可知, 直接对基尼系数作差相比于求导更加的精确和直观。但二者都比随机加边策略更加的有效。

但是由于贪心策略 Greedy\_B 每次加边后需要重新计算每条边的基尼增量, 非常耗时且耗费大量的计算资源, 而贪心策略 Greedy\_C 使用第一次的网络近似替换每次更新后的网络, 发现差别不大, 结果显示近似处理与不近似处理结果非常相近, 节省了计算的时间。因此, 说明可以采用近似贪心方法提高计算效率, 既节省了时

间，也不影响实验效果。

因此，在后文的多个实验中，都是使用 Greedy\_C 进行实验的。

在网络增边，存在增边容量限制下，限制容量 $C=50$ ，采用动态规划的增边策略与贪心策略进行对比，计算集合 $H$ 中节点重要性评分的基尼系数，如图 4-10 所示。

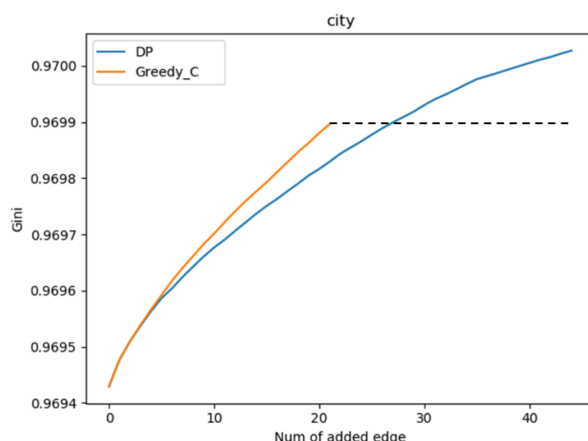


图 4-10 存在容量限制的动态规划策略与贪心策略的实验结果

在图 4-10 的实验结果中，加入了容量限制进行优化，此时采用了动态规划近似处理成背包问题进行求解，同时和贪心算法进行对比。在增边容量不超过 20 的情况下，贪心策略加边条数有限，在加入 20 多条边的时候，基尼系数值达到最大，但不是最优的。而动态规划的方法，可以基尼系数指随着加边条数的不断增加而上升，且最大值也是大于贪心增边策略的。根据图 4-10，基于动态规划的增边效果是明显优于贪心增边策略的。

(2) 在网络增边，同时存在运能容量限制和增边条数限制下，图 4-11，图 4-12 和图 4-13 分别表示在容量限制 $C=150$ 、 $C=200$ 、 $C=250$ ，条数限制 $L=30$ 的设置下的实验结果。

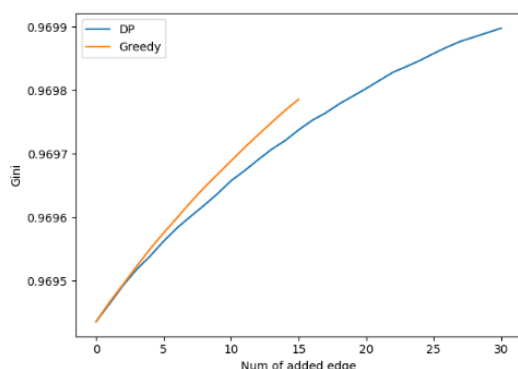


图 4-11  $C=150$ ,  $L=30$

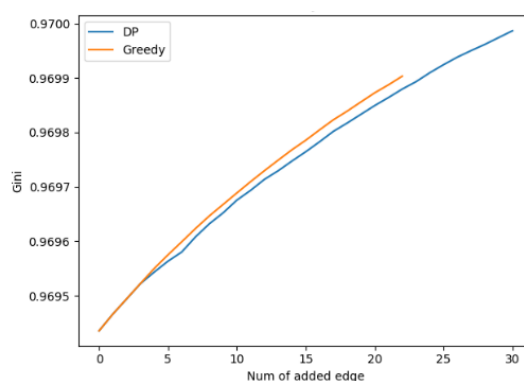


图 4-12  $C=200$ ,  $L=30$

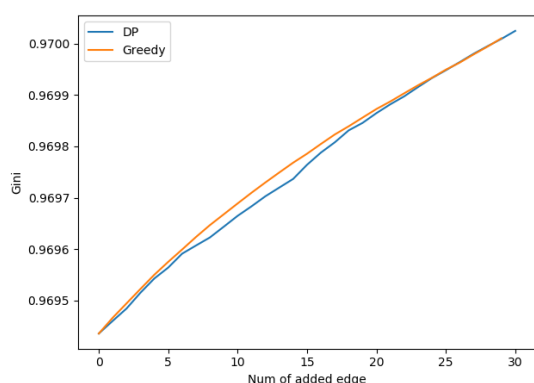


图 4-13  $C=250$ ,  $L=30$

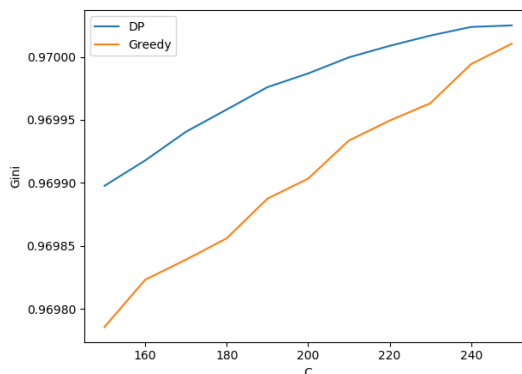


图 4-14  $L=30$ ,  $C$  不断变化

根据图 4-11, 图 4-12 和图 4-13 所示, 当增边条数限制固定为 30 条时, 运能容量限制等于 150, 200 时, 动态规划增边策略是明显优于贪心增边策略的。但是, 运能容量限制更大时, 贪心增边策略的基尼系数增量更大, 这也是符合常识的。当运能容量限制达到 250 时, 贪心增边策略效果几乎等同于动态规划策略。所以, 存在这样一个问题, 固定增边条数不变, 随着运能容量限制的增大, 贪心增边策略的效果是否会超越动态规划增边策略。因此, 本节补充了一个实验, 固定增边条数限制  $L=30$ , 运能容量限制不断变化的情况下的实验结果如图 4-14 所示。

根据图 4-14 的实验结果, 在增边条数固定, 容量不断变化的情况下, 动态规划增边策略的基尼系数值一直大于贪心增边策略的基尼系数, 这表明动态规划增边策略是更优的方法。但是随着运能容量的不断增大, 贪心增边策略的效果也是不断逼近动态规划增边策略的, 但不会超过动态规划的。总而言之, 基于动态规划的增边策略的效果是优于贪心策略的。

(3) 在网络增边, 同时存在运能容量限制和增边条数限制下, 图 4-15, 图 4-16 和图 4-17 分别表示在容量限制  $C=200$ , 条数限制  $L=10$ ,  $L=30$ ,  $L=50$  的设置下的实验结果。

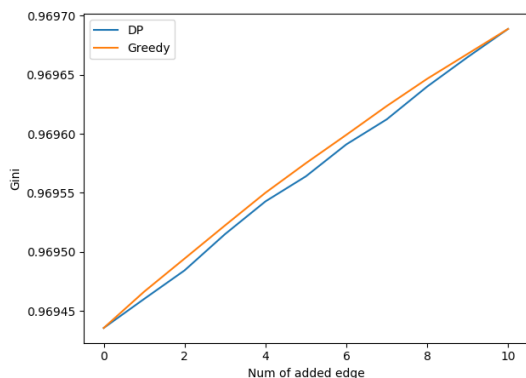


图 4-15  $C=200$ ,  $L=10$

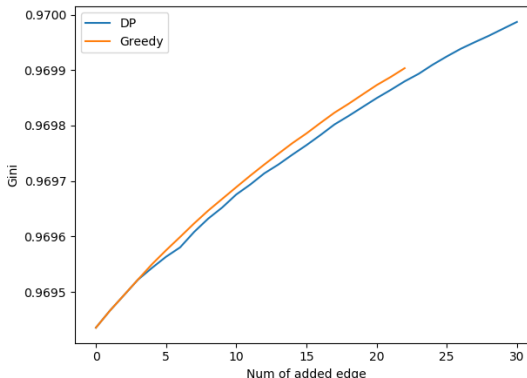
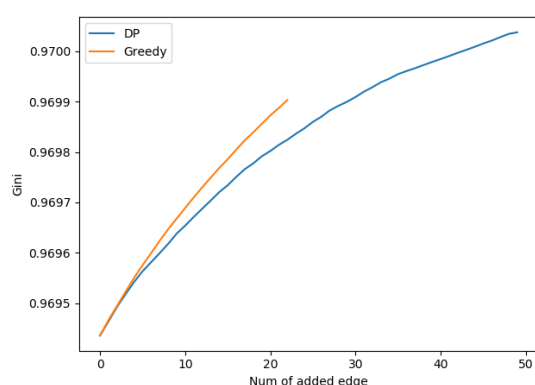
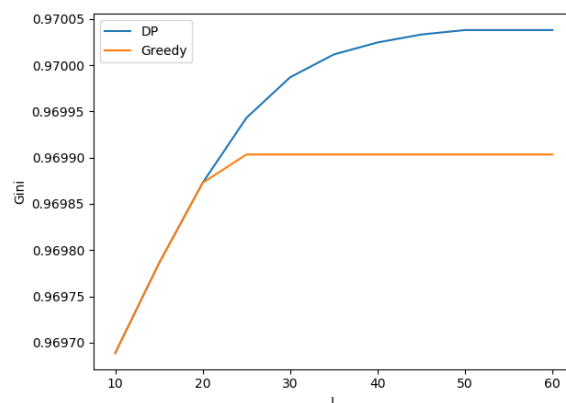


图 4-16  $C=200$ ,  $L=30$

图 4-17  $C=200$ ,  $L=50$ 图 4-18  $C=200$ ,  $L$  不断变化

根据图 4-15, 图 4-16 和图 4-17 所示, 当运能容量限制固定为 200 时, 增边条数限制等于 30, 50 时, 动态规划增边策略是明显优于贪心增边策略的。贪心增边策略受到增边的条数的限制, 导致了基尼系数不能无限的增大。而在增边条数比较少的时候, 像图 4-12 中所示, 增边条数限制等于 10 的时候, 动态规划增边效果甚至低于贪心增边的效果这也是正常的现象。但是, 当运能容量限制固定时, 随着增边的条数的增加, 是否会导致动态规划增边策略的效果远远优于贪心增边策略呢? 因此, 本节增加了一个实验, 固定增边容量限制  $C=200$ , 增边条数不断变化, 实验结果, 如图 4-18 所示。

根据图 4-18 的实验结果, 在容量限制固定, 增边条数不断变化的情况下, 两者的基尼系数值都是不断上升的。在容量限制比较少少的情况下, 图中显示 20 条以内的时候, 贪心增边策略是与动态规划增边策略效果一致的。这也是符合客观知识的。增边条数太少, 会限制动态规划方法的效果。但是, 随着增边条数的不断增多, 在增边条数不到 30 条时, 贪心增边策略的基尼系数值就达到了它的最顶峰, 而, 动态规划增边策略的基尼系数值还在不断攀升, 并且明显大于贪心增边策略的基尼系数值。随着增边条数达到比较大的数值的时候, 两者增边策略都不会使得基尼系数增加, 说明网络已经非常稳定了。

(4) 接下来 USAir 网络、UCI 网络、PB 网络和 Email 网络中在固定增边容量限制或者增边条数的限制下对贪心增边策略和动态规划增边策略进行实验对比。

图 4-19 和图 4-20 分别表示在 USAir 网络下两种贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

根据图 4-19 和图 4-20 的实验结果, 在固定增边容量时, 增边条数在 30 条左右的时候, 贪心增边策略的基尼系数值就已经达到了最大, 基尼系数值不会再随着增边条数而增加了。而动态规划增边策略的基尼系数值还在不断增长, 且涨幅明显, 在增边条数到 50 条时, 基尼系数才到达最大, 且基尼系数值的最大值也大于贪心

增边策略的最大值。而当固定增边条数时，当不断扩大增边容量时，动态规划增边策略也是优于贪心增边策略的。所以，在 USAir 网络中，动态规划增边策略是优于贪心增边策略的。

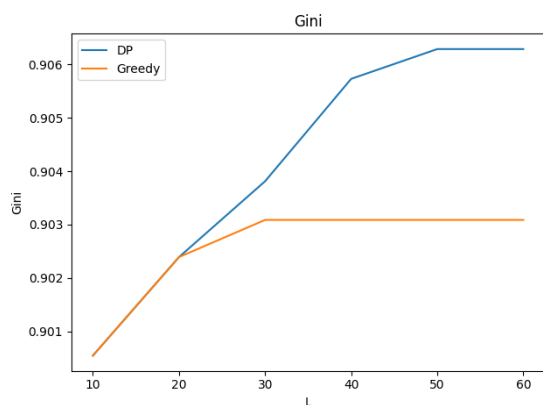


图 4-19  $C=100$ ,  $L$  不断变化

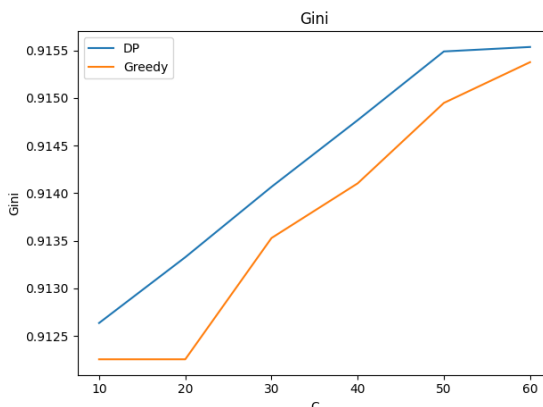


图 4-20  $L=30$ ,  $C$  不断变化

图 4-21 和图 4-22 分别表示在 UCI 网络下两种贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

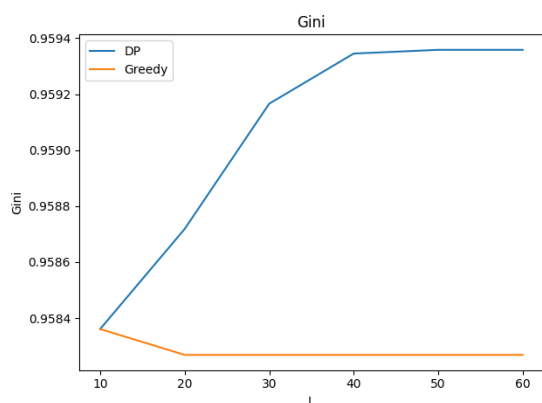


图 4-21  $C=100$ ,  $L$  不断变化

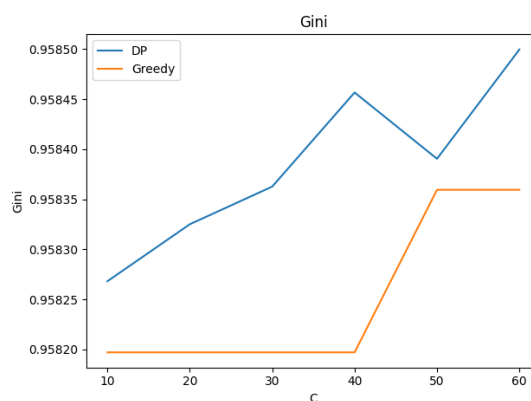
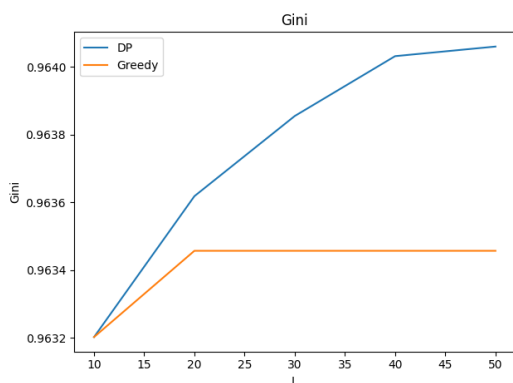
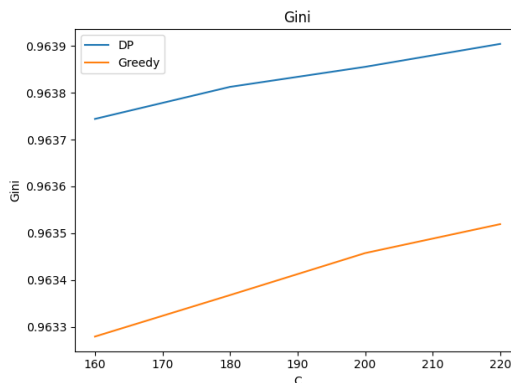


图 4-22  $L=30$ ,  $C$  不断变化

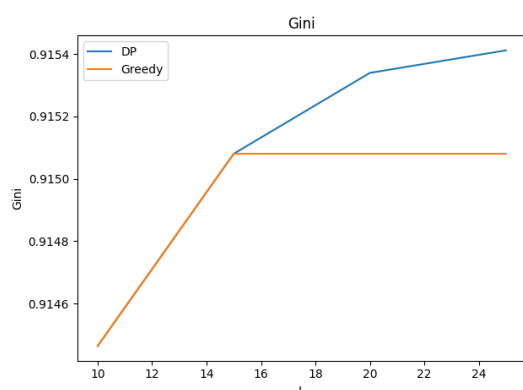
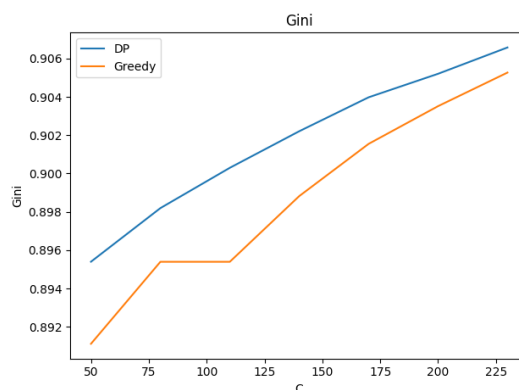
根据图 4-21 和图 4-22 的实验结果，在固定增边容量时，增边条数在 40 条左右的时候，动态规划增边策略的基尼系数值达到最大，且呈现上升趋势。而在本实验中，贪心增边策略的效果不太好，但也不影响动态规划增边策略的优越性。当固定增边条数时，当不断扩大增边容量时，动态规划增边策略的基尼系数值也是大于贪心增边策略的，虽然有所波动，但是动态规划增边策略也是优于贪心增边策略的。所以，在 UCI 网络中，动态规划增边策略是优于贪心增边策略的。

图 4-23 和图 4-24 分别表示在 PB 网络下两种贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

图 4-23  $C=100$ ,  $L$  不断变化图 4-24  $L=30$ ,  $C$  不断变化

根据图 4-23 和图 4-24 的实验结果, 在固定增边容量时, 增边条数在 20 条左右的时候, 贪心增边策略的基尼系数值达到最大, 而动态规划增边策略的基尼系数值还在不断上升, 表明动态规划增边策略的效果更好。当固定增边条数时, 当不断扩大增边容量时, 动态规划增边策略的基尼系数值也一直是大于贪心增边策略的。所以, 在 PB 网络中, 动态规划增边策略是优于贪心增边策略的。

图 4-25 和图 4-26 分别表示在 Email 网络下两种贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

图 4-25  $C=100$ ,  $L$  不断变化图 4-26  $L=30$ ,  $C$  不断变化

根据图 4-25 和图 4-26 的实验结果, 在固定增边容量时, 增边条数在 16 条左右的时候, 贪心增边策略的基尼系数值达到最大, 而动态规划增边策略的基尼系数值还在不断上升, 表明动态规划增边策略的效果更好。当固定增边条数时, 当不断扩大增边容量时, 动态规划增边策略的基尼系数值也一直是大于贪心增边策略的。所以, 在 Email 网络中, 动态规划增边策略是优于贪心增边策略的。

综合本节的多个实验, 比较了动态规划增边策略和贪心增边策略在多个数据集下的实验结果, 表明, 动态规划贪心增边策略的优越性。接下来, 在下一节使用



动态规划增边策略与几种常见的增边策略做对比。

#### 4.2.2 动态规划增边策略与其他的增边策略对比

在本文中，适合本文的增边策略有，按照节点的度数从高到低加边（Degree）和按照节点的介数值从高到低加边（BS）。

根据这个思想，选择最重要节点的邻居节点构成的部分网络，在该网络中选择两个节点度值和最大的节点进行增边，或者两个节点介数值和最大的节点进行增边。比较动态规划增边策略的实验结果。

对于每个不同的数据集，需要对增边容量  $C$  和增边条数  $L$  的参数值进行调参。因为每个数据集的节点的个数是不一样的，在节点数量比较少的数据集中，实验效果是肯定不如节点数量比较多的数据集的。这与节点的邻居的部分网络的构成有关。还需要保证网络是一个无向带权的网络。

图 4-27 和图 4-28 分别表示在飞行数据集下两种贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

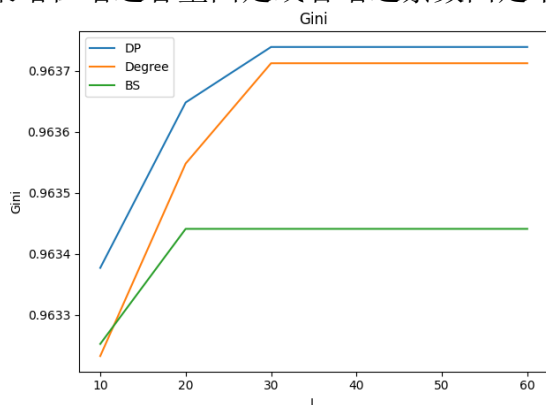


图 4-27  $C=100$ ， $L$  不断变化

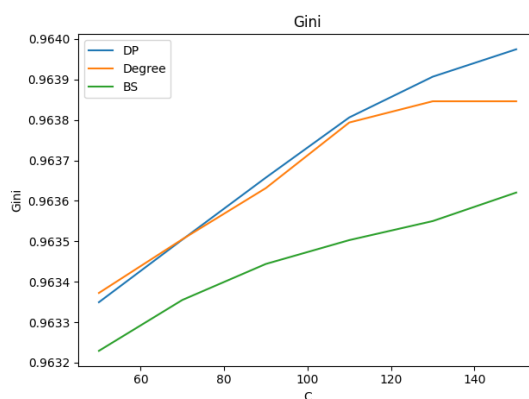


图 4-28  $L=30$ ， $C$  不断变化

根据图 4-27 和图 4-28 的实验结果，飞行数据集中，在容量限制固定，增边条数不断变化的情况下，三种方法的基尼系数值都是不断上升的。但是，基于动态规划增边策略的效果是明显优于另外两种方法的。在增边条数固定下，动态规划增边策略的基尼系数值也是大于另外两种方法的。所以，在飞行数据集下，动态规划增边策略是最优的。

图 4-29 和图 4-30 别表示在 USAir 网络中，贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

根据图 4-29 和图 4-30 实验结果，USAir 网络中，增边容量固定，增边条数不断变化的情况下，动态规划增边策略的基尼系数值是最大的，所以动态规划增边策略的效果是明显优于另外两种方法的。在增边条数固定下，增边容量不断变化的情

况下, 动态规划增边策略的基尼系数值也是大于另外两种方法的。所以在 USAir 网络中, 动态规划增边策略是优于另外两种方法的。

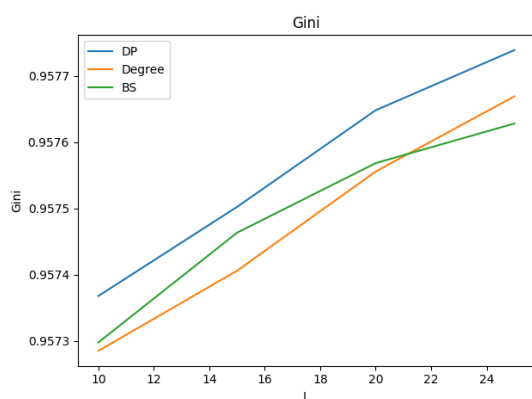


图 4-29  $C=100$ ,  $L$  不断变化

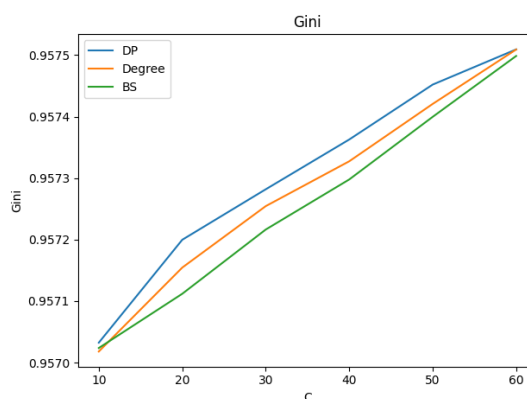


图 4-30  $L=30$ ,  $C$  不断变化

图 4-31 和图 4-32 分别表示在 UCI 网络中, 贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

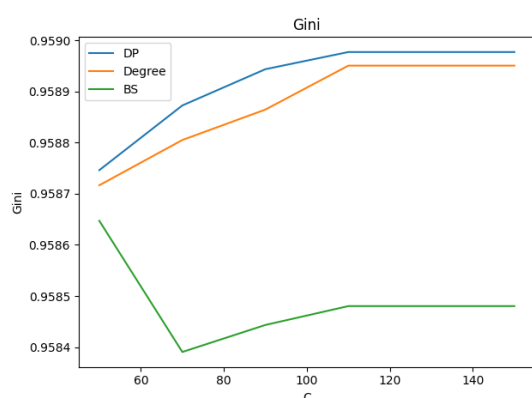


图 4-31  $C=100$ ,  $L$  不断变化

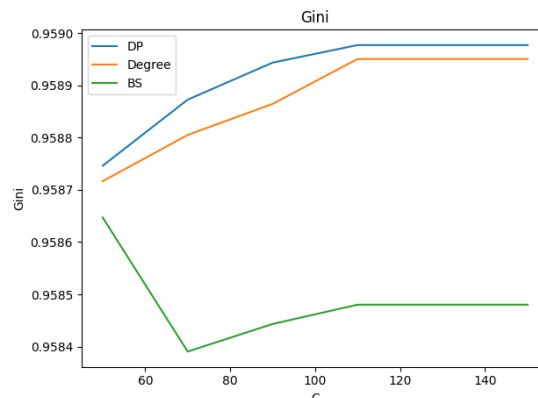


图 4-32  $L=30$ ,  $C$  不断变化

根据图 4-31 和图 4-32 的实验结果, UCI 网络中, 增边容量固定, 增边条数不断变化的情况下, 动态规划增边策略的基尼系数值也是最大的, 所以动态规划增边策略的效果是优于另外两种方法的。在增边条数固定下, 增边容量不断变化的情况下, 动态规划增边策略的基尼系数值也是大于另外两种方法的。所以在 UCI 网络中, 动态规划增边策略是优于另外两种方法的。

图 4-33 和图 4-34 分别表示在 PB 网络中, 贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

根据图 4-33 和图 4-34 的实验结果, PB 网络中, 增边容量固定, 增边条数不断变化的情况下, 动态规划增边策略的基尼系数值也是最大的, 虽然在增边条数 20 到 30 条时, 基尼系数值有所下降, 这与网络的度分布有关。但是动态规划增边策

略的效果还是优于另外两种方法的。在增边条数固定下，增边容量不断变化的情况下，动态规划增边策略的基尼系数值也是大于另外两种方法的。所以在 PB 网络中，动态规划增边策略是优于另外两种方法的。

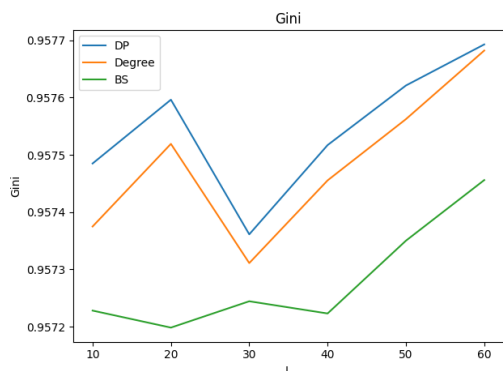


图 4-33  $C=100$ ,  $L$  不断变化

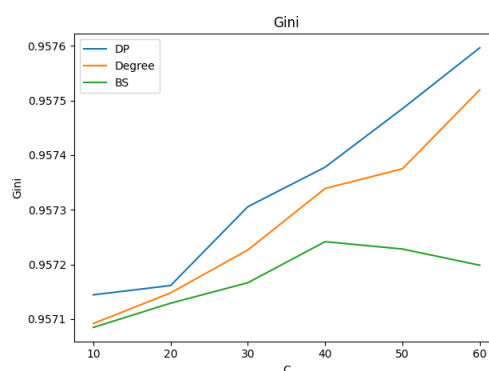


图 4-34  $L=30$ ,  $C$  不断变化

图 4-35 和图 4-36 分别表示在 Email 网络中，贪心增边策略和动态规划增边策略在增边容量固定或者增边条数固定下进行的实验。

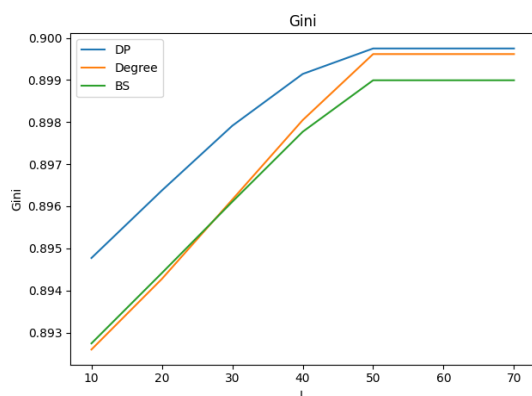


图 4-35  $C=100$ ,  $L$  不断变化

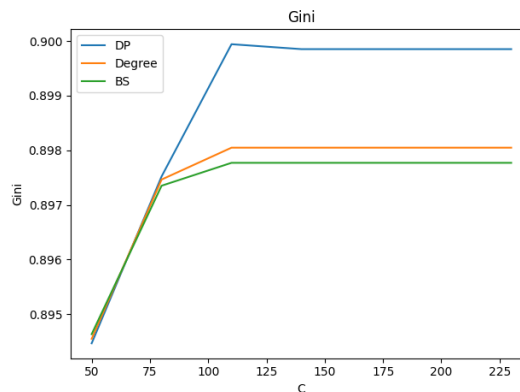


图 4-36  $L=30$ ,  $C$  不断变化

根据图 4-35 和图 4-36 实验结果，Email 网络中，增边容量固定，增边条数不断变化的情况下，动态规划增边策略的基尼系数值也是最大的，所以动态规划增边策略的效果是优于另外两种方法的。在增边条数固定下，增边容量不断变化的情况下，动态规划增边策略的基尼系数值的最大值也是最高的。所以在 Email 网络中，动态规划增边策略是优于另外两种方法的。

综合本节的多个实验，我们可以得出结论，基于动态规划的增边方法，无论是固定增边容量还是固定增边条数，与贪心策略增边方法、在度值最大两个节点间增边和在介数值最大两个节点间增边相比，基尼系数都是最大的，效果最好。总而言之，动态规划增边策略是最优的增边方法。

### 4.3 本章小结

本章详细地介绍了网络鲁棒性的定义，并改进基尼系数含义来定义了新的网络鲁棒性，此外重点提出了贪心增边策略和动态规划增边策略的网络鲁棒性优化方法，对比了多种增边策略，得出动态规划增边方法是最好的网络鲁棒性优化方法。

## 第五章 全文总结与展望

近年来,复杂网络的发展是越来越快,现实中的复杂系统几乎都可以表示成复杂网络的结构进行研究。而且,图神经网络领域的蓬勃发展,也使得复杂网络可以结合图神经网络中的算法模型,对网络中的重要节点进行挖掘。同时,网络鲁棒性优化在现实中的很多领域,例如航空物流领域,城市交通领域等,都受到了极大地关注。

### 5.1 全文工作总结

在第一章中主要介绍了复杂网络的研究背景和意义,重点介绍了复杂网络研究的重要节点挖掘方向和网络鲁棒性优化的国内外研究现状。首次提出将图神经网络算法运用在了网络重要节点的挖掘上,将图神经网络与复杂网络相结合,提出 WGA 方法对网络中的重要节点进行挖掘。此外,提出网络增边的方法来保护重要节点,从而对网络进行鲁棒性优化。

第二章中首先介绍了复杂网络的几种基本特征和基本网络模型,然后介绍网络中的重要节点挖掘方法,其次介绍了图神经网络的研究现状和两种算法模型,再介绍网络鲁棒性优化的邻居节点间的加边方法,分别采用贪心策略加边和动态规划方法进行加边,最后介绍了节点挖掘方法和网络鲁棒性优化方法的评价指标

第三章详细的介绍了基于图神经网络算法的重要节点挖掘方法——WGA 方法,提出新的节点重要性指标,对于每种方法得出的节点重要性排序序列,提出 PNN 指标作为每种方法的评判标准,以及 WGA 方法所挖掘的重要节点在真实情境下的分析,最后进行了详细的实验以及分析。

第四章详细的介绍了网络鲁棒性优化的方法,提出重新定义的基尼系数作为评价网络鲁棒性指标,在重要节点的邻居节点构成的网络上使用贪心增边策略和动态规划增边策略优化网络鲁棒性,最后进行了详细的实验。

第五章对本文的工作进行了总结,介绍了本文的主要贡献和未来的工作展望。

### 5.2 本文主要贡献

- (1) 首次将图神经网络算法运用在节点挖掘领域,并提出 WGA 方法;
- (2) 提出新的节点重要性定义:利用 WGA 方法所学习到的节点的特征向量之间的欧氏距离之和作为节点重要性评价的指标;
- (3) 结合不同的网络模型,提出 PNN 指标对节点重要性排序序列进行评估;

(4) 在真实情境下, 分析了所挖掘到的重要节点的可靠性, 为节点挖掘在提供了现实依据。

(5) 提出新的网络鲁棒性优化策略: 使用贪心增边策略和动态规划增边策略对网络的鲁棒性进行优化, 并对比了几种其他的网络增边策略。

(6) 提出新的网络鲁棒性评价指标: 结合基尼系数, 对实验中的网络鲁棒性评价指标进行了重新定义, 并取得很好的实验效果。

(7) 本文的几种方法, 都是基于真实网络进行的, 具有很强的现实意义。可以在真实的场景下, 节约时间和经济成本, 具有不错的经济效益。

### 5.3 未来工作展望

(1) 在网络中重要节点挖掘的问题上, 图神经网络算法取得了与多种重要节点挖掘方法相近的实验效果。但是, 在有些实验中的效果不是最优的, 所以我们需要考虑结合的图神经网络算法。其次对节点重要性进行评分时, 使用的指标是节点的特征向量之间的欧氏距离之和作为节点的重要性评分, 是否有更准确的节点重要性评分的定义, 这也是值得思索的问题。

(2) 在网络鲁棒性优化的问题上, 结合贪心策略和动态规划策略的增边方法, 取得了不错的效果。采用重新定义的基尼系数作为网络鲁棒性评价指标, 是具有一定现实价值的。未来还可以结合机器学习或者深度学习中的具体概念知识, 对网络鲁棒性指标进行探索。

## 致 谢

时光荏苒，七年的成电生涯就要画上句号了。在这个美丽的校园里，我认识了许多品学兼优的同学，他们带领我学习成长。

在这里，感谢本科 421 宿舍的两个室友，我们三个又一起继续在科大学习生活了三年。嘉健带我打羽毛球，叶芊带我出去玩，丰富了我的日常生活。但是，在这里我不得不说，叶芊你还是少喝点酒，读了研之后你真的是变成了一个酒鬼，哈哈。另外，感谢实验室的诸多同学，老黎、顺顺、远远、佳成、跃华、璐姐等，我们是一个 team，虽然我研二以后和你们不在一个实验室了，但我还是非常喜欢去你们那里串门的，喜欢和你们一起出去耍，你们一个个那么的有才华，说话还那么好听，和你们一起真的有意思。另外，感谢我自己的师弟，亮光和文波。我的论文和研究生成果离不开两位学弟的帮忙。

最应该感谢的还是刘震老师，论文的选题，实验的进行都是在刘震老师悉心的指导下完成的。而且，论文的撰写，一遍又一遍地修改，都是在刘震老师的监督下完成的。最终论文答辩也取得了不错的成绩。另外，刘震老师还指导我进行两项专利的撰写工作，并成功申请了发明专利。而且，刘震老师为我提供了前往华为实习的机会，让我收获良多。在这里，我再次对刘震老师表示诚挚的感谢。

最后，感叹一句，时间过得真快啊！再过一个多月就要踏上社会了，也不知道还要过多久才能和这些良师益友再次相聚了。谢谢你们！

## 参考文献

- [1] Marx V. The big challenges of big data[J]. Nature, 2013, 498(7453): 255-260
- [2] 周涛.大数据:商业革命和科学革命[J].半月谈:时事资料手册, 2013, 98(4): 14-17
- [3] 朱涵,王欣然,朱建阳.网络“建筑学”[J].物理, 2003, 32(6): 364-369
- [4] 周涛,张子柯,陈关荣,等.复杂网络研究的机遇与挑战[J].电子科技大学学报, 2014, 43(1): 1-5
- [5] 何宇,赵洪利,杨海涛,等.复杂网络演化研究综述[J].装备指挥技术学院学报, 2011, 22(1): 120-125
- [6] 张盛.基于复杂网络理论的供应链管理研究综述及其展望[J].中国管理信息化, 2018, 21(14): 61-63
- [7] 舒睿.组合星型网络的同步动力学研究[D].北京:北京邮电大学, 2019, 1-5
- [8] 王洋,崔红波. WS 小世界网络统计特性研究[J].价值工程, 2016, 35(4): 1-5
- [9] Watts DJ, Strogatz SH. Collective dynamics of “small world” networks[J]. Nature, 1998, 393(6684): 440-442
- [10] Newman MEJ, Watts DJ. Renormalization group analysis of the small-world network model[J]. Physica A, 1999, 263(4-6): 341-346
- [11] Barabási AL, Albert R, Jeong H. Mean-field theory for scale-free random networks[J]. Physica A, 1999, 272(1-2): 173-187
- [12] Pimm SL. Food web[D]. Chicago: University of Chicago Press, 2002, 1-10
- [13] Jeong H, Mason S, Barabasi AL, et al. Lethality and centrality in protein networks[J]. Nature, 2001, 411(6833): 41-42
- [14] Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. The Journal of Mathematical Sociology, 1972, 2(1): 113-120
- [15] Chen D, Lü L, Shang MS, et al. Identifying influential nodes in complex networks[J]. Physica A, 2012, 391(4): 1777-1787
- [16] 任卓明,邵凤,刘建国,郭强,汪秉宏.复杂网络中节点重要性排序的研究进展[J].物理学报, 2013, 62(17): 178901
- [17] South T, Roughan M, Mitchell L. Popularity and centrality in Spotify networks: critical transitions in eigenvector centrality [J]. Journal of Complex Networks, 2021, 8(6): 1-8
- [18] Berkhin P. Fabrication and Properties of Solution-Cast Polyaniline@Carboxymethylchitin Blend Films[J]. Interactive Mathematics, 2005, 2(3): 73-74



- [19] Lu L Y, Zhang Y C, Yeung C H, Zhou T. Challenges in CRISPR-CAS9 Delivery Potential Roles of Nonviral Vectors[J]. PLoS One, 2011, 6(2): 1-6
- [20] Kleinberg J M. Authoritative sources in a hyperlinked environment[J]. Journal of the ACM, 1999, 46(2): 604-605
- [21] Azim Nouran Ayman R, Gharib Tarek F, et al. Influence propagation: Interest groups and node ranking models[J]. Physica A, 2020, 10(1016): 12424
- [22] JiaQi L, XueRong L, JiChang D. A survey on network node ranking algorithms: Representative methods, extensions, and applications[J]. Science China Technological Sciences, 2020, 10(1007): 1-11
- [23] 王兴隆, 苗尚飞, 贺敏, 刘明学. 基于改进 K-shell 算法的空中交通信息物理系统节点排序[J]. 中国科技论文, 2020, 15(10): 1144-1149
- [24] 卢鹏丽, 于洲. 基于度与 H 指数扩展的复杂网络节点排序方法[J]. 兰州理工大学学报, 2020, 46(5): 100-106
- [25] Albert R, Jeong H, Barabasi A. Error and attack tolerance of complex networks[J]. Nature, 2000, 406(6794): 378-382
- [26] 朱炳, 包家立, 应磊. 生物鲁棒性的研究进展[J]. 生物物理学报, 2007, 23(5): 357-363
- [27] 章伟豪. 基于承载力预压装备式框架结构鲁棒性分析[D]. 西安: 西安科技大学, 2020, 1-5
- [28] 朱银. 基于对抗攻击的孪生网络跟踪算法鲁棒性研究[D]. 广州: 暨南大学, 2020, 1-10
- [29] 邓贵仕, 武佩剑, 田炜. 全球航运网络鲁棒性和脆弱性研究[J]. 大连理工大学学报, 2008, 48(5): 765-768
- [30] 黄小原, 晏妮娜. 供应链鲁棒性问题的研究进展[J]. 管理学报, 2007, 4(4): 3-8
- [31] 张家年. 国家安全保障视角下安全情报与战略抗逆力融合研究——伊朗核设施遭“震网病毒”攻击事件的启示[J]. 情报杂志, 2016, 37(2): 8-14
- [32] 安天研究院, 广东省电力系统网络安全企业重点实验室. 委内瑞拉大规模停电事件的初步分析与思考启示[J]. 信息安全与通信保密, 2019, 5(1): 28-39
- [33] Zhong N, Zeng G. Prevention and treatment of chronic respiratory diseases in China[J]. Chronic Disease and Translational Medicine, 2019, 5(4): 209-213
- [34] 李佳遥. 面向协同网络组织的风险级联效应研究[D]. 北京: 北京交通大学, 2020, 1-8
- [35] Albert R, Jeong H, Barabasi AL. Error and attack tolerance of complex network[J]. Nature, 2000, 406(6794): 678
- [36] 任广建, 朱金福, 卢朝阳. 基于相对熵理论的航路网络鲁棒性研究[J]. 武汉理工大学学报, 2019, 43(5): 820-824+831

- [37] 裴燕如,武英达,于强,张启斌,胡雅慧.荒漠绿洲区潜在生态网络增边优化鲁棒性分析[J].农业机械学报, 2020, 51(2): 172-179
- [38] 张振江,张玉召,王小荣.铁路快捷货运网络鲁棒性分析[J].中国安全科学学报, 2020, 30(3): 150-156
- [39] 刘迪洋,张震,张进.基于社区结构的复杂网络鲁棒性优化策略[J].计算机工程, 2020, 26(4): 162-165
- [40] 王哲,李建华,康东,冉淞丹.复杂网络鲁棒性增强策略研究综述[J].复杂系统与复杂性科学, 2020, 17(3): 1-26+46
- [41] United States Department of Transportation. Bureau of Transportation Statistics[EB/OL]. [https://transtats.bts.gov/Tables.asp?DB\\_ID=111&DB\\_Name=Air%20Carrier%20Statistics%20%28Form%2041%20Traffic%29-%20All%20Carriers&DB\\_Short\\_Name=Air%20Carriers](https://transtats.bts.gov/Tables.asp?DB_ID=111&DB_Name=Air%20Carrier%20Statistics%20%28Form%2041%20Traffic%29-%20All%20Carriers&DB_Short_Name=Air%20Carriers), February 11, 2019
- [42] Lü L, Zhou T, Zhang Q, et al. The h-index of a network node and its relation to degree and coreness[J]. Nature Communications, 2016, 7(1): 440-442
- [43] 王建伟,荣莉莉,郭天柱.一种基于局部特征的网络节点重要性度量方法[J].大连理工大学学报, 2010, 50(5): 822-826
- [44] Zhong L, Chao G, Zili Z, Ning S, Jiajin H. Identifying Influential Nodes in Complex Networks: A Multiple Attributes Fusion Method[J]. Physica A, 2012, 391(6): 11-22
- [45] 任卓明,邵凤,刘建国,郭强,汪秉宏.基于度与集聚系数的网络节点重要性度量方法研究[J].物理学报, 2013, 62(12): 522-526
- [46] Travencolo BAN, Costa LD. Node importance ranking of complex networks[J]. Physics Letters A, 2008, 373(8): 89
- [47] 李鹏翔,任玉晴,席西民.网络节点(集)重要性的一种度量指标[J].系统工程, 2004, 22(4): 13-20
- [48] 谭跃进,吴俊,邓宏钟.复杂网络中节点重要度评估的节点收缩方法[J].系统工程理论与实践, 2006, 1(11): 79-83+102
- [49] Bryan K, Leise T. The \$25,000,000,000 Eigenvector: The Linear Algebra behind Google[J]. SIAM Review, 2006, 48(3): 569-581
- [50] Schneider CM, Mihaljev T, Herrmann HJ. Inverse targeting-An effective immunization strategy[J]. EPL, 2012, 98(4): 46002
- [51] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks[J]. Nature Physics, 2011, 6(11): 888-893
- [52] 闫艳芳.基于信息融合的复杂网络节点重要性研究[D].徐州:中国矿业大学, 2020, 20-41

- [53] Wang JY, Hou XN, Li KZ, et al. A novel weight neighborhood centrality algorithm for identifying influential spreaders in complex networks[J]. Physica A, 2017, 475(3): 88-105
- [54] Liu J, Xiong QY, Shi WR, et al. Evaluating the importance of nodes in complex networks[J]. Physica A, 2016, 452(3): 209-219
- [55] 刘书磊,杜家乐,邵增珍.一种基于邻居节点和边的复杂网络节点排序方法——NL 中心性算法[J].山东科学, 2019, 32(2): 130-136
- [56] Farabet C, Couprie C, Najman, et al. Learning Hierarchical Features for Scene Labeling[J]. IEEE, 2013, 35(8): 1915-1929
- [57] 郑逸涵.针对小样本数据的图神经网络模型研究[D].广州:华南理工大学, 2020, 1-10
- [58] Grover Aditya, Leskovec Jure. node2vec: Scalable Feature Learning for Networks[C]. Acm Sigkdd International Conference on Knowledge Discovery & Mining, Palo Alto, 2016, 526-535
- [59] Wang D, Peng C, Zhu W. Structural Deep Network Embedding[C]. Acm Sigkdd International Conference on Knowledge Discovery & Mining, Beijing, 2016, 1225-1234
- [60] 刘震,游慧明,左文波.一种基于自编码神经网络的航空网络重要节点发现方法[P].中国,发明专利, 202010348635.6, 2020 年 4 月 28 日
- [61] 雷艳菊.基于遗传算法和截断凝聚 SGD 训练 WGAN[D].大连:大连理工大学, 2020, 1-7
- [62] Sen A, Mazumder A, Banerjee J, et al. Identification of K most vulnerable nodes in multi-layered network using a new model of interdependency[C]. IEEE, Toronto, 2014, 831-836
- [63] Gong M, Wang Y, Wang S, et al. Enhancing robustness of interdependent network under recovery based on a two-layered protection strategy[J]. Scientific Reports, 2017, 7(1): 12753
- [64] 王世锦,苏思雨,李海云,等.基于 PCNC 的航路网络拓扑结构鲁棒性优化[J].南京航空航天大学学报, 2019, 51(6): 756-762
- [65] 刘震,潘亮光,游慧明.一种航空网络鲁棒性优化方法[P].中国,发明专利, 202010310454.4, 2020 年 4 月 20 日
- [66] 刘爱霞.一些正则图的条件连通度和限制诊断度[D].太原:山西大学, 2020, 1-10
- [67] 丁涛,高双,花泽眩.基于复杂网络的人道物流网络鲁棒性分析[J].物流技术, 2016, 35(3): 82-86+108
- [68] 党亚茹,丁飞雅,史丽媛,李雪娇.基于航线失效模式下的航班流网络抗毁性分析[J].中国民大学学报, 2014, 32(1): 64-69

## 攻读硕士学位期间取得的成果

- [1] 物流采购大数据挖掘,横向项目,2019 年 10 月至 2020 年 10 月
- [2] 刘震,游慧明,左文波.一种基于自编码神经网络的航空网络重要节点发现方法[P].中国,发明专利, 202010348635.6, 2020 年 4 月 28 日
- [3] 刘震,潘亮光,游慧明.一种航空网络鲁棒性优化方法[P].中国,发明专利, 202010310454.4, 2020 年 4 月 20 日