**学号：2014200790**

**北京化工大学**

**硕士研究生学位论文开题报告**

**论文题目：复杂网络社区发现问题研究**

**学 院 名 称：信息科学与技术学院**

**专 业：计算机科学与技术**

**研究生姓名： 胡海霞**

**导师姓名：耿志强**

**开题日期：2015年9月25日**

|  |  |  |
| --- | --- | --- |
| **考核**  **成绩** |  | |
| **审核**  **小组**  **成员**  **以及**  **职称** | **姓 名** | **职 称** |
| **袁洪芳** |  |
| **李学斌** |  |
| **万洪杰** |  |
| **万静** |  |
| **尚颖** |  |

目录

**1 课题来源及项目名称 ........................................................................................................1**

**2 文献综述 ...........................................................................................................................1**

2.1 研究背景...........................................................................................................................1

2.2 社区发现研究情况............................................................................................................2

2.2.1 社区结构概述...........................................................................................................2

2.2.2 社区发现研究进展...................................................................................................3

2.3 社区发现方法...................................................................................................................4

2.3.1 基于图分割的社区发现.............................................................................................4

2.3.2 基于聚类思想的社区发现.........................................................................................4

2.3.3 基于优化指标的社区发现.........................................................................................5

2.3.4 重叠社区发现.............................................................................................................5

2.4 课题创新...........................................................................................................................5

2.5 参考文献...........................................................................................................................6

**3 研究计划 ...........................................................................................................................8**

3.1 选题的目的及意义...........................................................................................................8

3.2 主要研究内容...................................................................................................................8

3.3 研究方案...........................................................................................................................9

3.3.1 技术方案..................................................................................................................9

3.3.2 实施方案所需要的条件..........................................................................................9

3.3.3 拟解决的关键问题.................................................................................................10

3.4 课题难点分析.......................... .......................................................................................10

3.5 预期研究成果及创新点.................................................................................................10

3.6 工作计划进度.................................................................................................................10

1. **课题来源及项目名称**

课题来源：自选题目

项目名称：复杂网络社区发现问题研究

1. **文献综述**

**2.1 研究背景**

现实世界中无处不在的就是网络，而随着网络的规模不断增大，网络关系也不断增多，就会形成许多复杂而庞大的系统，我们称之为复杂系统。而我们对于复杂系统一般以网络或者图的形式来描述，展现出来的就是复杂网络。与我们生活息息相关的如交通运输网络、手机联系网络、Web主题搜索网络、朋友社交网络，还有生物领域的蛋白质交互网络[1]、新陈代谢网络[2]等都是属于复杂网络的研究范畴。网络中最基础的单元就是节点与边，一般来说，网络中的节点代表一个关系系统里的实体，如在朋友社交网络中的节点就代表了这个朋友社交网络中的人员；网络中的边即代表在特定网络中实体之间的相互关系，如在朋友社交网络中的边是代表人员之间的关系。从这种意义上来说，在我们周围的诸多系统我们都可以抽象为网络来进行研究。

复杂网络的研究最开始是由图论的兴起而发展，著名的七桥问题是数学中图论研究的起源，也可以说是针对复杂网络研究的图方法的一次有益创举。之后，著名的匈牙利数学家Erdos和Reny提出了随机图理论，社会心理学家stanley Milgraml做了小世界实验，也就是六度分离推断，还有随后提出的弱连接的强度，这些都是为复杂网络研究的兴起做了理论基础，而具有突破性意义的则是二十世纪关于小世界[3]与无标度性质[4]的两篇文章的发表。随着计算机硬件和Internet的迅速发展，人们可以通过各种方法收集并处理种类不同的各种实际网络数据，针对复杂网络的研究也慢慢受到人们的关注与追逐，并推动它成为了一个热点课题，而且在各方面已经颇有成果，如在传播动力学方面各模型的研究，文献[5]研究了复杂网络中传播现象与疾病在现实生中的传染有很大程度上的相似，运用传染病模型模拟网络中出现的传播现象，在结构上对基础网络模型即无标度模型进行基于优化理论的研究[6]以及对社交网络的结构理论研究[7]，在复杂网络的同步特性上也有许多研究，如文献[8]研究了针对移动网络社区的舆情传播特性。当然，作为一个与其他学科交叉融合的一个学科，复杂网络在生物学科、物理学科、社会学科等领域也都有相关的研究成果，这些都是因为复杂网络在表示真实复杂系统方面的优势，它的兴起为许多其他学科的研究提供了一个全新的思路与方法。

根据目前的关于复杂网络方面的研究，涉及的方面由具体到笼统，具体的有如节点的研究[9]、模型的研究，笼统的如网络演化[10]范畴的网络同步的本质研究[11,12]、博弈理论的研究，还有对于网络的性质如稳定性、可控性、抗毁性等的研究，从这些研究成果来看，本质上就是关于网络结构的研究，就如人由骨骼来支撑，一个网络的骨骼就是它的结构。所以对于一个网络而言，研究它的结构具有很大意义。

**2.2 社区发现研究情况**

**2.2.1 社区结构概述**

研究表明[13~16]，复杂网络具有社区结构特性，但是就目前的研究成果而言，还没有针对社区的统一定义，一般来讲，我们认为社区是一组具有相似属性的节点集合，基本特征是社区内部边稠密，社区之间边稀疏。从社区的这一特征来看，社区在复杂网络中的真实意义与其网络息息相关，例如在人际交往网络中，社区是代表着一个因为兴趣或某种特定关系（如家人、朋友等）而形成的群体，在Web搜索网络中，社区是代表着具有相同或相近主题的网络链接。

自Girvan和Newman提出复杂网络具有社区结构特性以来，研究者们根据各自所研究的网络性质给出了不同的定义。但是运用比较广泛的是Santo Fortunato总结的三种社区的定义方法，即基于局部的定义、基于全局的定义、基于节点相似性的定义，分别如下：

基于局部的定义： 社区是与网络中其他部分只存在少量联系的一部分网络。局部定义是着眼于社区与网络其他部分的关系，而非从全部意义上进行考虑。进一步来说，社区是具有自治系统的单独实体。局部定义主要关注的是可能包含直接相连的邻居在内的子图，但是忽略了图的剩余其他部分的对整个网络的作用。如CPM算法[17]的派系过滤的“clique”就是从局部的角度考虑社区是一个独立的子团，而且一般从局部定义的思想出发而得到的社区都是极大子图，直到往该子图中添加任何节点或边都会导致相应标准值的降低，从这个意义来说，局部定义更侧重于社区内部关系的紧密性。

基于全局的定义： 社区结构也可以从网络整体上进行定义。在全局定义中,社区结构是对网络的一个划分或者覆盖，并且这种划分或者覆盖使得特征函数具有最优解。如果采用一种全局的特征函数使得划分的结果得到一种最优解，结合系统中所有社区结构进行的一种全局性定义，全局特征函数的设计可以根据具体的应用需求进行变化。这样就对社区结构给出了一种量化的定义。Newman等人提出的模块函数Q就是一种全局的特征函数，通过比较当前网络与随机网络之间的社区内部边密度，来衡量社区结果的优劣。

基于节点相似性的定义: 社区结构可以定义为是网络中一些具有相似性的节点集合。在异构的网络中也往往体现出一些社区节点的相似性特征，这些节点组成一些小的类似的团体，类似与社会学研究的“组织”或“群体”的概念。相似度定义是根据社区内部节点的相似性进行定义，而不考虑这些节点之间的具体关系，许多研究者给出了不同的相似度衡量标准。这种社区的发现和节点的相似性度量通常采用层次聚类的方法来实现。

复杂网络中，社区可以包含多个含义，如社团、簇、类、群、组模块等。社区是宏观网络的一个缩影，它们是构成网络的局部单位。在我们周围存在的网络可以看作是由一些较小的社区组成，而这些社区又可以由更小的社区组成。因此，揭示网络中的社区结构对于研究整个网络的功能和分析网络的组成结构具有重要的意义。

**2.2.2 社区发现研究进展**

最早的关于社区结构的研究可以追溯到Weiss和Jacobson在1955年针对政府机构工作人员之间关系的研究。他们将连接了不同工作小组之间的关系移除，从而得到了各个独立的工作小组，这一思想成为后来社区发现算法的基础。而在2002年Girvan和Newman提出了复杂网络普遍具有社区结构特征的理论，也给出了社区发现的经典算法，即GN算法，它是基于分裂思想的一种社区发现方法，一直被认为是经典的社区发现算法，类似的，Newman提出了一种凝聚的社区发现算法——Newman快速算法，后来，又出现了改进的CNW算法。而随着Girvan和Newman提出了用于社区结构的模块度函数（Q函数）[15]后，引领了许多以Q函数为依据的基于模块度的社区发现的算法，如贪心优化算法、模拟退火算法、极值优化等，但是这类启发式算法会得到局部最优解，而且不断研究发现通过优化Q函数进行社区发现的算法会存在着分辨率极限[18]的问题，所以这类方法的理论基础还需进一步的研究与探索。

由于复杂网络的表现形式与图的很相似，所以基于图论的社区发现方法也是一个研究热点。基于图论的算法主要思想是将图中的节点划分为若干个子集，然后要求划分后的各子集之间连接的变数最少。最典型的就是谱方法和Kernighan-Lin算法，基于图论的还有如团渗理论的社区发现算法、最大流算法、多层次图分割算法等等，都有许多研究成果。传统的社区发现算法还有很大一部分比例是被聚类所占据，如基于层次聚类的方法、基于划分聚类的方法等。 由于现实生活中许多网络数据是多种类或者说是有多重标签的，所以针对社区发现问题的动态算法也是受到许多研究者们的关注，如基于自旋模型的社区发现算法、基于随机游走理论的社区发现算法。其他的还有关于动态网络社区发现[19~21]的研究也因为与实际网络更切合而得到许多关注，而发展更甚的则是关于重叠社区的社区发现研究，因为实际网络中一个节点由于自身的属性是可以属于多个社区的。关于重叠社区的研究[22]是对传统社区概念的一个拓展，目前在这方面的研究已有不少，如基于团渗理论的方法[23]、基于局部扩展的方法[24]等，但是对于社区重叠性的判断，对社区重叠性的度量标准，还有对于如何简单但又高效地发现重叠社区等这些方面的研究还需要更多的探讨与改进。对于社区发现的研究，还有许多研究是针对不同视角来进行的，如针对网络组成元素的，即基于节点[25,26]的社区发现研究、针对网络中标签属性[27,28]的研究，还有研究针对社区内部密度[29]的研究，值得关注的是，从网络中边的角度[30]来进行社区发现也成为一种新的研究途径。

**2.3 社区发现方法**

上一节介绍了社区发现大致的研究范围和情况，而关于社区发现的算法种类颇多，在这里介绍几类社区发现方法。

**2.3.1 基于图分割的社区发现方法**

图分割方法中典型的两类分别是谱方法和Kernighan-Lin方法。谱方法基本思想是将社区划分为二次型优化问题，通过计算特殊矩阵的特征向量来优化预定义的“截”函数，该方法是基于“截”函数的，它通过根据不同的问题定义不同的“截”函数，然后社区划分问题就转化为了最小化“截”函数的问题。但是这类方法存在着不能自动确定聚类数、划分多类需要迭代多次的缺点。Kernighan-Lin方法则是一种典型的基于贪心思想的社区二分法，基本原理是提出一种采用试探优化方法来划分网络结构社区，基本思想就是引入了一个效益函数Q，再将节点移动到其他社区或者交换不同社区的节点来使最初定义的效益函数Q最大。

**2.3.2 基于聚类思想的社区发现方法**

在聚类思想中，应用最广泛的是层次聚类。而层次聚类又分为分裂的层次聚类和凝聚的层次聚类。分裂的层次聚类思想就是在图中找到已连接的相似度最低的节点对，然后移除它们的边，重复这个过程，逐步把整个网络分成越来越小的各个部分，它是一个自顶往下的过程，经典的GN算法就属于分裂层次聚类算法；而凝聚的层次聚类思想是依据某种评价标准计算出节点之间的相似性，然后自底向上的逐个合并最有可能位于同一社区的两个节点，逐步构成越来越大的连通分支，Newman贪心算法就属于典型的凝聚型聚类方法。其他的聚类思想的方法有许多，如划分型聚类思想，将网络划分问题转化为传统向量空间样本点的聚类问题，K-means算法就是此类方法的经典算法。

**2.3.3 基于优化指标的社区发现方法**

这里的指标也包括别的指标，如适应度函数，但是一般来说都是指模块度函数（Q函数）。由于Girvan和Newman提出了用于社区结构的模块度函数，后来许多研究者以通过优化Q函数来发现社区结构作为一种新的思路。基于优化指标的方法一般假定更高的Q值意味着更好的社区结构，所以对应着具有最高Q值的一个划分即为最佳的社区划分，这就是基于优化指标的基础思想。这类算法又可以细分为贪心方法、模拟退火、极值优化、频谱优化等。

**2.3.4 重叠社区发现方法**

重叠社区发现的研究主要可以分为三类：基于核心的方法、基于节点分裂思想的方法、基于优化思想的方法。基于核心的方法如CFinder算法就是将派系看成是网络社区的核心来进行重叠社区的研究。CONGA算法是典型的利用节点分裂思想的方法，它是将GN算法进行扩展，通过分裂节点介数来实现的。而基于优化思想的方法主要是针对于各种指标，如LFM算法就是通过优化局部适应度函数来划分重叠社区并且能发现层次社区，另外对于优化Q函数的研究更是层出不穷。除了前面的三类，还有其他许多方法，如用标签传播方法来发现网络中的重叠社区、基于链接社区结构来进行重叠社区结构的发现等。

**2.4 课题创新**

社区发现一直是复杂网络研究中一个重要的研究领域。通过阅读文献，可以知道，社区发现与聚类分析有很大的相似之处，然而目前还没有相应的揭示两者之间联系的相关论述，所以这是需要探索与研究的。另在重叠社区发现方面，大部分关注都在于节点上，从边的角度来考虑进行重叠社区发现也是个很好的视角。所以课题第一部分工作是针对社区发现与聚类分析的联系，探索两者之间的转换关系，主要是如何高效地将社区发现问题转化为聚类分析问题，第二部分工作则是基于第一点，构造网络中边的相似度，然后设计相应的聚类算法来进行重叠社区发现。第三点是针对加权复杂网络，提出相应的基于变相似度的甲醛复杂网络社区发现算法。在数据方面，可以将算法也用于动态网络数据集中，但是会根据动态网络数据的特性改变相应的算法中涉及到的结构。

**2.5 参考文献**

[1] Wang Z, Zhang J. In search of the biological significance of modular structure in protein networks. PLOS Computational Biology, 2007, 3(6): e107.

[2] Ravasz E, Somera AL, Mongru DA. Hierarchical organization of modularity in metabolic networks. Science, 2002, 297(5586): 1551-1555

[3] Watts DJ, Strongatz SH. Collective dynamics of ‘small-word’ networks. Nature, 1998, 393(6684): 440-442.

[4] Barabsi AL, Albert R. Emergence of scaling in random networks. Science, 1999,286(5439): 509-512.

[5] Kitsak M, Callos L K, Havlin S. Identification of influential spreaders in complex networks[J]. Nature Physics, 2010,6(11):888-893.

[6] 吴泓润，覃俊，易云飞，李德毅，郑波尽. 基于优化理论的社区无标度网络模型. 计算机学报，2014年，第38卷（2），337-348.

[7] 韩毅，许进，方滨兴，周斌，贾焰. 社交网络的结构支撑理论. 计算机学报，2014年，第37卷（4），905-914.

[8] LIU H, CHEN J, LU J. Generalized synchronization in complex dynamical networks via adaptive couplings[J]. Physica A, 2010, 389: 1759-1770.

[9] 秦李，杨子龙，黄曙光. 复杂网络的节点重要性综合评价. 计算机科学，2015年，第42卷（2），60-64.

[10] YANG Zhongming, QIN Yong, HUANG Han. Research on the Evolution of Complex Networks and their Application in Internet Load Balancing[J].COMPUTER ENGINEERING & SCIENCE,2011,33(2):37-41.

[11] X. F. Wang, G. Chen. Synchronization in small-world dynamical networks [J]. Int. J. Bifurcation Chaos Appl. Sci. Eng., 2002, 12:187.

[12] X. F. Wang, G. R. Chen. Synchronization in scale-free dynamical networks: robustness and fragility [J], IEEE Trans. Circuits and System-I, 2002, 49:54-62.

[13] M. Girvan and M. Newman, Community structure in social and biological networks , PNAS, 2002, vol.99,no.12,7821-726.

[14] G. Palla, 1.Derenyi, 1.F arkas, and T. Viesek, Uncovering the overlapping community structure of complex networks in nature and society, Nature, 2005,vol.435,814-818.

[15] M.E.J. Newman and M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E, 2004.

[16] S.Boeealetti, V.Latora, Y.Moreno, M.Chavez, and D-U. Hwang, Complex networks: Strueture and dynamics, Physics Reports, 2006 ,175-308.

[17] Palla G, Derenyil, Farkas I, Vicsek T. Uncovering the overlapping community structures of complex networks in nature and society[J]. Nature, 2005, 435(7043): 814-818.

[18] Santo Fortunato, Marc Barthelemy, Resolution limit in community detection, PNAS, 2007,vol 104,36- 41.

[19] 王莉，程学琪. 在线社会网络的动态社区发现及演化. 计算机学报，2014年，第37卷.

[20] 孙怡帆，李赛. 基于相似度的微博社交网络的社区发现方法. 计算机研究与发展. 2014年，第51卷（12），2797-2807.

[21] Chonghui Guo, Jiajia Wang, Zhen Zhang. Evolutionary community structure discovery in dynamic weighted networks. Physica A, 2014, 565-576.

[22] Xu Zhou, Yanheng Liu, Jindong Zhang, Tuming Liu, Di Zhang. An ant colony based algorithm for overlapping community detection in complex networks. Physica A, 2015, 289-301.

[23] Evans T S. Clique graphs and overlapping communities[J/OL]. Journal of Statistical Mechanics: Theory and Experiment, 2010.

[24] Lee C, Reid F, McDaid A. Detecting highly overlapping community structure by greedy clique expansion. ACM, 2010, 33-42.

[25] Qiong Chen, Ting-Ting Wu, Ming Fang. Detecting local community structures in complex networks based on local degree central nodes. Physica A, 2013, 529-537.

[26] Chonghui Guo, Haipeng Zhao. Community structure discovery method based on the Gaussian kernel similarity matrix. Physica A, 2012, 2268-2278.

[27] Heli Sun, Jiao Liu, Jianbin Huang, Guangtao Wang, Zhou Yang. CenLP:A centrality-based label propagation algorithm for community detection in networks. Physica A, 2015, 767-780.

[28] Shenghong Lin, Hao Lou, Wen Jiang, Junhua Tang. Detecting community structure via synchronous label propagation. Neurocomputing , 2015, 1063-1075.

[29] Hong Jin, Shuliang Wang, Chenyang Li. Community detection in complex networks by density-based clustering. Physica A, 2013, 4606-4618.

[30] Yong Yeol Ahn, James P. Bagrow, Sune Lehmann. Link communities reveal multiscale complexity in networks. Nature, 2010, 761-765.

1. **研究计划**

**3.1 选题的目的及意义**

社区结构普遍存在于现实网络中，不同的社区代表着不同的意义，例如在社会网络中，属于同一个社区的个体一般是具有相同兴趣爱好或者是相似的文化背景；在Web网络中，同一社区内的页面通常是表达着相近主题的。在一定程度上，社区结构可以反映出真实系统的拓扑关系，社区表示了网络中的具体的功能实体，研究网络社区对于理解网络结构、分析网络特性、发现网络中的隐藏规律有很大帮助，在实际应用方面比如市场营销、病毒传播研究、社交网站个性推荐等都有很大的应用。总而言之，研究社区结构能帮助人们更加直观地认识复杂网络的结构和功能从而可以更好地理解和利用复杂网络。

也正是因为社区结构的研究对于人们在理解复杂网络的结构和功能方面的重要性，研究者们在社区发现方面投入了很大的关注度，也提出了许多成果。前文中对于社区发现领域的研究进展以及算法已经给出了概述，成果显然是很卓著的，但是这一领域仍然还有许多问题需要我们去研究与探索。从目前研究结果来看，聚类分析运用于社区发现的算法有很多，通过阅读文献，可以发现复杂网络中的聚类分析问题与社区发现问题二者有很大的联系，当然对应着也有区别，但是目前关于这方面还没有相关的论述，所以对于二者之间的关系研究可以进行探索。对于传统的社区发现算法，虽然经典，但是在网络划分准确度或者时间复杂度方面或多或少存在着缺陷，所以我们应该尽力将这些缺陷最小化。对于重叠社区发现的研究，已经有许多针对节点来研究的各种理论，但事实上如果以边作为考虑对象，对于发现重叠社区更有利，因为边社区之间可能会出现一些重叠的节点，所以它很自然地对应了节点的重叠社区结构。另外，现有的社区发现算法绝大部分都是针对于无向无权网络，但是实际网络中存在着许多有向网络、加权网络，所以在这方面的研究也是需要努力的方向。

**3.2 主要研究内容**

关于复杂网络中社区发现问题的研究，课题的研究内容主要从以下几点展开。

（1）揭示复杂网络中社区发现问题与聚类分析问题之间的关系。从社区发现问题和聚类分析问题的本质来看，社区发现问题是可以在一定条件下转化为聚类分析问题的，而聚类分析问题中将数据构造一个网络模型也可以转化为社区发现问题，通过实验结果分析验证两者之间的关系是本课题的第一个主要内容。

（2） 提出基于边相似度的LGN算法。GN算法是经典的网络社区发现算法，但是复杂度高所以不能用于大规模网络，而基于节点的相似度在一些领域并不适用，如生物领域中的节点很难被划分一个独立模块，所以研究定义边相似度替代边介数的计算，得到新的算法。

（3）针对加权网络提出基于权重的新的边相似度定义，并结合聚类分析提出基于边相似度的加权复杂网络社区方法。复杂网络存在着边的权重特性，对于加权网络，若忽略其权重特性则会丢失包含于权重里的有用信息，所以对于加权网络也是社区发现的重要问题。

**3.3 研究方案**

**3.3.1 技术方案**

针对课题研究内容，对应解决的技术方案如下：

（1） 针对第一点研究内容，其一，利用节点相似度作为媒介，采用Jaccard、 new-Jaccard、和signal相似度构造方法与AL算法、AP算法进行组合，在真实数据集（如Zachary 空手道俱乐部网络）和人工数据集上进行实验分析社区发现结果；其二，通过不同方式构造网络模型为媒介，采用T-neighborhood邻域连通图和K-nearest neighborhood的网络模型构造方法与GN算法、Kernighan-Lin算法进行组合，在真实数据集和人工数据集上进行实验分析聚类结果。根据对应的结果分析两者之间的关系。

（2）改进的算法LGN中关键的是边相似度定义，考虑到网络中两条边的相似性，主要在于共同拥有的节点的相关信息，拟定义边的相似度为：

S

将LGN算法运用于Zachary 空手道俱乐部网络数据集和新陈代谢网络数据集上，分析实验结果并与GN算法和传统基于节点相似的算法进行优劣比较，验证算法速度上的改善效果及在发现重叠节点上的有效性。

（3）在第二点的基础上研究针对加权网络的社区发现方法，关键是提出合适的基于权重的边相似度衡量标准，结合聚类分析算法应用于真实数据集（Karate Club网络、Dolphin网络）和人工数据集上，因为数据集都是知道其结构的，所以通过分析验证实验结果与真实情况比较验证算法的有效性和准确性。

**3.3.2 实施方案所需要的条件**

课题的研究需要大量的文献信息作为基础，因此前期的工作需要就是阅读课题相关的国内外文献，了解基础算法并对相关算法有清晰的分析，充分利用网络资源、学校图书馆各项资源，最后确定课题的详细方法；后期主要工作需要就是课题算法的具体编程以及收集数据进行实验验证分析，因此只需要可用于编程环境的计算机一台以及实验验证所需要的数据集。

**3.3.3 拟解决的关键问题**

课题的关键问题有三个，其一是网络模型的构造，包括如何设置网络模型构造参数、分析不同的网络模型对于实验结果的影响，还有分析不同的相似度方法对于实验结果的影响；其二，对于LGN算法，关键在于充分权衡网络图中的各个属性定义边相似度；其三，加入了权重的参考因子，并结合聚类方法，关键在于基于权重的相似度的定义和聚类的过程。

**3.4 课题难点分析**

对于课题研究内容的第一点，难点在于如何设置构造网络模型的参数。T-neighborhood邻域连通图的方法是通过设置阈值，连接距离小于某个阈值的两个节点从而建立网络；K-nearest neighborhood方法是通过连接每个节点与距离最近的K个节点从而建立起网络，而网络模型的好坏会直接影响聚类效果，所以如何选择适当的参数（如阈值、近邻数K）是一大难点。由于网络中具有的性质很多，如何权衡各个网络性征从而定义边相似度确定LGN算法是难点之二。针对于加权复杂网络，对于如何将权重因子考虑在边的相似度判断中，再结合合适的聚类分析方法提出有效的加权网络的算法也是课题的一大难点。

**3.5 预期研究成果及创新点**

预期的研究成果主要有两方面。第一，通过实验研究分析，可以表明社区发现问题和聚类分析问题是可以通过媒介相互转化的，通过将社区发现问题构造合适的相似度方法就可以完全将社区发现问题转化为聚类分析问题，对应的通过将数据构造成合适的网络模型，聚类分析问题就完全可以转化为社区发现问题；另一方面两者若是媒介选择的不好会产生哪些不利于问题的后果也要进行研究并给出相应的结论。改进的LGN算法比较传统的GN算法在计算复杂度上会得到大幅度改善但不失准确性，且相对于传统的基于节点相似度的算法在对于重叠节点的发现上更有效率和准确。针对于加权复杂网络上的研究，提出的算法在对于加权网络的社区划分上更合理。

**3.6 工作计划进度**

2015年09月 ~ 2015年10月 整理相关文献，明确课题

2015年11月 ~ 2015年12月 实验方法设计

2015年12月 ~ 2016年02月 实现算法并应用于数据集验证算法的有效性

2016年03月 ~ 2016年05月 对实验结果分析整合

2016年06月 ~ 2016年11月 根据实验结果进一步深入研究

2016年12月 ~ 2017年04月 撰写大论文

|  |
| --- |
| **指导教师意见：**  **指导教师签名：**  **年 月 日** |
| **审核小组意见：**  **审核小组组长签字：**  **年 月 日** |
| **研究生根据审核小组意见对开题报告的改进措施：**  **年 月 日** |
| **备注：** |