

本科毕业论文（设计）

**论文题目（中文）**  **基于库仑定律的标签传播社团检测算法**

**论文题目（英文）** An algorithm of label propagation community

detection based on coulomb’s law

学生姓名 蒋坤元

指导教师 赵志立

学 院 信息科学与工程学院

专 业 计算机科学与技术（数据科学方向）

年 级 2019级

兰州大学教务处

**诚信责任书**

本人郑重声明：本人所呈交的毕业论文（设计），是在导师的指导下独立进行研究所取得的成果。毕业论文（设计）中凡引用他人已经发表或未发表的成果、数据、观点等，均已明确注明出处。除文中已经注明引用的内容外，不包含任何其他个人、集体已经发表或未发表的论文。

本声明的法律责任由本人承担。

论文作者签名： 日 期：

**关于毕业论文（设计）使用授权的声明**

本人在导师指导下所完成的论文及相关的职务作品，知识产权归属兰州大学。本人完全了解兰州大学有关保存、使用毕业论文（设计）的规定，同意学校保存或向国家有关部门或机构送交论文的纸质版和电子版，允许论文被查阅和借阅；本人授权兰州大学可以将本毕业论文（设计）的全部或部分内容编入有关数据库进行检索，可以采用任何复制手段保存和汇编本毕业论文（设计）。本人离校后发表、使用毕业论文（设计）或与该毕业论文（设计）直接相关的学术论文或成果时，第一署名单位仍然为兰州大学。

本毕业论文（设计）研究内容：

□可以公开

□不宜公开，已在学位办公室办理保密申请，解密后适用本授权书。

（请在以上选项内选择其中一项打“√”）

论文作者签名： 导师签名：

日 期： 日 期：

**基于库仑定律的标签传播社团检测算法**

**中文摘要**

社团检测是复杂网络分析的一个重要部分，其主要目的是对网络中密切相关的子集进行分类。社团检测在现实生活中有着广泛的应用，例如在社会网络推荐系统、公共安全、嫌疑人追踪和城市规划中。基于标签传播的社团检测是比较经典的社团划分方法之一，但其结果具有较大的随机性。为此，本文主要研究基于标签传播的社团检测优化问题，提出一种基于库伦定律的标签传播社团检测算法LPACL。~~本文提出的社团检测算法优化的中心思想是类比库仑定律电荷间吸引力的特性尽量规避LPA算法随机性和不稳定性的缺陷，从而达到优化算法的目的。~~主要工作如下：

1. 提出了种基于库伦定律的标签传播社团检测算法LPAC。类比迁移库仑定律中描述的电子间作用力来避免LPA的缺点，将网络中的节点度数视为电子的电荷量，节点间的距离类比电子之间的距离，根据库仑定律计算所有两个节点之间的吸引力，得到节点吸引力矩阵。

2. 优化LPAC算法降低其复杂度。根据节点间吸引力矩阵进行社团检测迭代，为了避免大量冗余计算，提出了深度矩阵d的概念来限定矩阵计算。

3. LPAC方法性能分析。本文使用模块度、NMI和ARI三种经典的社团划分评价指标，对比分析LPAC算法与经典LPA算法结果。

在真实网络和人工网络上的实验结果显示，在绝大多数情况下，LPACL在模块度、NMI和ARI三种社团划分评价指标上都表现出了不同程度的提高，且社团划分后形成的社团数量更少，由此可以证明LPACL在社团检测上的有效性，也证明了LPACL相较于经典的LPA算法有更加准确和稳定的结果。

**关键词：**社团检测算法；LPA；库仑定律；节点距离；库伦矩阵

**An algorithm of label propagation community detection based on coulomb’s law**

**Abstract**

Community detection is an important part of complex network analysis, whose main purpose is to classify closely related subsets of the network. This has a wide range of real-life applications, such as in social network recommendation systems, public safety, suspect tracking and urban planning. The defect of random selectivity of label propagation does not contribute to the stability and accuracy of the algorithm. This paper focuses on the optimization problem of label propagation-based relevance detection algorithm and proposes an algorithm called LPACL. The main work of this paper is as follows:

1. design the algorithm to avoid the drawbacks of LPA by analogy with the inter-electron forces described in the migration Coulomb's law, consider the node degrees in the network as the charge of electrons, the distance between nodes analogous to the distance between electrons, calculate the attraction between all two nodes according to Coulomb's law, and get the node attraction matrix.

2. The community detection iteration is performed according to the inter-node attraction matrix, and the concept of depth matrix d is proposed to qualify the matrix calculation in order to avoid a large number of redundant calculations.

3. In this paper, the experimental results of three classical community classification evaluation metrics, namely, modularity, NMI and ARI, are compared with the results of the classical LPA algorithm to check their accuracy and performance advantages.

In summary, the central idea of the proposed community detection algorithm optimization is to avoid the defects of randomness and instability of the LPA algorithm as much as possible by analogy with the properties of the Coulomb's law charge-to-charge attraction, so as to achieve the purpose of optimizing the algorithm. According to the specific experimental results on real and artificial networks, in most cases, LPACL shows different degrees of improvement in three association division evaluation indexes, namely, modularity, NMI and ARI, and the number of associations formed after community division is smaller, which can prove the effectiveness of LPACL in community detection and also proves that LPACL compared with the classical LPA algorithm has more accurate and stable results.

**Keywords**: Community detection algorithm; LPA; Coulomb's law; nodal distance; Coulomb matrix

**目 录**

[第一章 绪论 1](#_Toc133422003)

[1.1 研究背景及意义 1](#_Toc133422004)

[1.2国内外研究现状 1](#_Toc133422005)

[1.3 本文主要研究内容 2](#_Toc133422006)

[1.4 论文组织结构 3](#_Toc133422007)

[第二章 相关理论知识与技术 4](#_Toc133422008)

[2.1 复杂网络概述 4](#_Toc133422009)

[2.1.1 复杂网络的表现形式 4](#_Toc133422010)

[2.1.2 复杂网络的主要研究内容 6](#_Toc133422011)

[2.2 社团检测算法的研究 6](#_Toc133422012)

[2.2.1 基于标签传播的社团检测算法 7](#_Toc133422013)

[2.2.2 基于模块度的社团检测算法 8](#_Toc133422014)

[2.2.3 基于层聚类的社团检测算法 8](#_Toc133422015)

[2.2.4 基于随机块模型的社团检测算法 9](#_Toc133422016)

[2.2.5 基于非负矩阵分解的社团检测算法 9](#_Toc133422017)

[2.3 相关符号及定义 10](#_Toc133422018)

[2.4 评价指标：模块度、NMI和ARI 10](#_Toc133422019)

[2.4.1 模块度 10](#_Toc133422020)

[2.4.2 NMI 10](#_Toc133422021)

[2.4.3 ARI 11](#_Toc133422022)

[2.5 本章小结 11](#_Toc133422023)

[第三章 基于库伦定律的社团检测算法 12](#_Toc133422024)

[3.1 库仑定律及库伦矩阵 12](#_Toc133422025)

[3.1.1 库仑定律 12](#_Toc133422026)

[3.1.2 定义节点间吸引力 12](#_Toc133422027)

[3.1.3 库伦矩阵 12](#_Toc133422028)

[3.2 基于库伦定律的社团检测算法 13](#_Toc133422029)

[3.2.1 算法设计 13](#_Toc133422030)

[3.2.2 算法实现 13](#_Toc133422031)

[3.3 算法流程 13](#_Toc133422032)

[3.4 时间复杂度分析 15](#_Toc133422033)

[3.5 本章小结 15](#_Toc133422034)

[第四章 实验及结果分析 16](#_Toc133422035)

[4.1 实验数据集 16](#_Toc133422036)

[4.2 关于参数的选取 16](#_Toc133422037)

[4.3 社团检测结果比较 17](#_Toc133422038)

[4.4 本章小结 18](#_Toc133422039)

[第五章 总结与展望 19](#_Toc133422040)

[5.1 论文工作总结 19](#_Toc133422041)

[5.2 未来展望 19](#_Toc133422042)

[参考文献 21](#_Toc133422043)

[致 谢 23](#_Toc133422044)

**第一章 绪论**

1.1 研究背景及意义

现实世界中的许多复杂系统都是以网络的形式存在的，例如，社会网络、生物和其他网络，以及Web网络都是复杂的网络。复杂网络使自然界和人类社会的不同方面被抽象为一个看似不相关的复杂系统。研究如何处理这些复杂网络的共性问题就可以更有效地促进发现隐藏在复杂系统中的信息线索。一般的，用于表示复杂系统的网络的价值是随着其规模的增加而呈指数级增长。随着现代的进步和变化，这些复杂系统的规模继续快速增长。这意味着这些复杂系统中的信息具有巨大的价值，并在未来将继续增长，使其成为数据分析和数据挖掘研究的沃土。

主要的复杂网络研究中存在的问题是社团检测（Community detection），这是一个非常重要的课题。社团检测是复杂网络分析领域的重要组成部分，其主要目的是划分网络中紧密关联的子集，在现实生活中有着非常广泛的应用场景，比如社交网络的推荐系统，公安追踪嫌疑人，城市规划等，综上所述，研究社团检测算法并提高精度软件应用体验、社会治安以及对于丰富和保障人类生活质量都有重大的意义。

在大数据时代，数据量不断增加，网络的规模也在快速增长。而随着网络规模的增长，网络维度的数量也开始呈指数级增长。所以，复杂网络中的图划分和社团检测是一个非确定性的多项式时间困难（Non-deterministic Polynomial-time Hardness）问题，当网络规模比较大时，没有有效的精确解决方案。一方面是由于网络规模庞大和社团结构的重叠性，另一方面则是因为复杂网络中的监督标注信息不足，不能单纯使用某种指标去量度算法的优劣，所以所有研究的目的旨在优化算法。

1.2国内外研究现状

钱学森对于复杂网络给出了一种严格的定义：具有自组织、自相似、吸引子、小世界、无标度中部分或全部性质的网络称之为复杂网络。抽象提取了许多真实系统的复杂网络已经成为科学研究中的一个流行方向。不同的关联感知方式导致了研究者对峰会集和边缘集的细分标准和方法不同，从而产生了不同风格的算法，从网络的动态角度来看，可以分为基于静态网络的社团检测算法和基于动态网络的社团检测算法这两类。

基于静态网络的社团检测算法可大致分为以下几种：

基于模块度的社团检测算法：模块度是常用一种的评价社团检测结果好坏的评价指标，目前针对未知真实 社团划分情况的网络进行社团检测时，模块度是最常用的评价指标。比如由 Clauset、Newman 和 Moore 在2004年提出的CNM算法[1]。

基于标签传播的社团检测算法：标签传播的基本思想是，网络中一个节点的标签取决于其邻居的最大标签数。这一思想首先被Raghavan应用到了社团检测的领域，他于2007年依据标签传播思想提出了LPA(Label Propagation Algorithm)算法[2]。标签传播的基本思想是，网络中节点的标签取决于其邻居节点的标签中数目最多的标签， LPA迭代使用的随机性选择可能会影响算法的稳定性，即不太重要的节点会影响更重要的节点，并且还可能导致振荡现象。此外，LPA仅使用了一跳邻域信息，而来自其邻居的其他连接被忽略，从而导致网络全局拓扑结构给出的重要信息的丢失。这些缺陷并不利于算法的稳定性和准确性，本文将类比迁移库仑定律中描述的电子间作用力来避免LPA的上述缺点。

基于层次聚类的社团检测算法：大多数常见的分层社团检测算法都是基于聚集或划分过程的，Girvan和Newman 2002年提出的GN算法[3]是一种基于划分过程的经典分层聚类算法。

基于非负矩阵分解的社团检测算法：非负矩阵分解一种机器学习算法，通过将给定的特征矩阵分解为两个非负的矩阵，实现了特征矩阵的非线性的维度降低，其中有 NMF[4]、NAGC[5]等都是近几年在非负矩阵分解框架下进行改进的社团检测算法。

这些社团检测算法基本上是为特定类型的数据集设计的。换句话说，没有一个单一的社团检测算法是可以为所有类型的网络提供良好的社团划分结果。在现实世界中有许多抽象的复杂网络，根据某些特征的不同，可以分为几种类型，如加权和非加权、有向和无向、动态和静态等等。并且，在现实网络中，一个节点可能属于多个关联，所以关联检测可以分为两类：非重叠关联检测和重叠关联检测。由于复杂网络的多样性，研究人员提出了大量的关联识别方法，关联识别的研究领域也越来越活跃和深入。

在动态网络中进行社区检测的算法主要分为两大类：基于演化聚类的社区检测算法以及基于增量聚类的社团检测算法。Javadi等人[6]提出了一种动态网络中社群结构的增量检测算法，以社团领袖的重要性和持久性为指标，能够有效发现真实和人工动态网络中的社群结构。Chakrabarti等人[7]提出了一种基于进化分组的社群检测算法，该算法考虑了动态时间对分组的影响，并利用一般分组框架将进化分组问题转化为一个双目标优化问题。

1.3 本文主要研究内容

本文主要研究基于标签传播的社团检测算法优化。标签传播的随机选择性的缺陷并不利于算法的稳定性和准确性，本文的主要工作如下：

1. 设计算法，类比迁移库仑定律中描述的电子间作用力来避免LPA的缺点，将网络中的节点度数视为电子的电荷量，节点间的距离类比电子之间的距离，根据库仑定律计算所有两个节点之间的吸引力，得到节点吸引力矩阵。

2. 根据节点间吸引力矩阵进行社团检测迭代，为了避免大量冗余计算，提出了深度矩阵d的概念来限定矩阵计算。

3. 本文会将实验结果与经典的LPA算法结果作比较以检验其精度和性能优势。

1.4 论文组织结构

本文将类比迁移库仑定律中描述的电子间作用力来避免LPA的随机性缺点，并会将实验结果与现有LPA算法的其他变体结果作比较以检验其精度和性能优势。具体章节安排如下：

**第一章 绪论。**本章主要介绍了在复杂网络中社团检测算法的研究背景、目的、研究现状和意义，并引出本文的研究设计思路和主要研究内容。

**第二章 相关理论知识与技术。**本章介绍了现阶段社团检测算法相关内容，主要包括：常用的主流社团检测算法及算法功能评价指标。

**第三章 根据库仑定律类比推移完成算法设计。**此节将详细叙述算法的设计和运行迭代流程。

**第四章 实验数据与结果展示。**描述实验数据，并将实验结果与其他经典社团检测算法对比在网络划分上的表现，以验证其准确性、稳定性和精度。

**第五章 总结。**分析不足，并提出展望。

**第二章 相关理论知识与技术**

2.1 复杂网络概述

在社会科学、生命科学和信息科学中，存在着具有非常复杂拓扑特性的网络结构。这些网络的形状既不完全是规则的，也不完全是随机的，学位分布有肥尾，集群系数高，边缘之间有相称性与非相称性，有社团结构以及分级结构（Hierarchy structure）等。在有向图网络中，还存在其他特征，如相互性和三角突出性。然而，在复杂网络的概念被引入之前，这种特征在网络的数学模型中是不存在的。

复杂网络的两个最著名和研究最充分的模型是小世界网络和无尺度网络。前者的特点是路径长度短，内聚力系数高，而后者的特点是根据幂律的顺序分布递减。此外，随着对复杂网络研究的深入和拓展，各种具有其他特征的复杂网络模型也引起了人们的关注。

2.1.1 复杂网络的表现形式

为了更直观地表示网络，可以将其抽象为一种图，用点和线表示，这是研究复杂网络的普遍共识。将网络表示为图，为理解现实世界网络的拓扑特性提供了一种更直观的方式。这些属性只与网络的大小和节点的连接边等情况有关。

具体的网络可以被分解成两部分：节点集V和边集E。这就形成了一个图G=(V,E)，节点数N=|V|，边数M=|E|，其中每条边E由V中的一对节点组成。

如果按照有向性和权重对图进行分类，可以分为四类：若图的一条边是无向和无权的，就说是无向和无权的。这里，无向意味着任何顶点对的边是（i, j）∈E和（j, i）∈E，代表相同的边，无权也意味着图的所有边都有相同的权重。若一个图是无定向的，但有带权重的边，它被称为无定向图。如果一个图的边是有方向的，而且没有权重，那么它就被称为有方向的无权图。这里，有向意味着图中任何一对顶点都代表两个不同的边（i, j）∈E和（j, i）∈E，其中节点i是边（i, j）∈E的起点，节点j是终点。如果图中的边是有方向的，那么图中的每条边都有一个相应的权重，表明成对的节点之间的接近程度根据它们的权重而不同。如果图中的一条边是有向和有权的，那么就说该图是有向有权图。



1. 无向无权图 (b) 无向有权图

(c) 有向无权图 (d) 有向有权图

图1. 图的分类

上述四种类型的图有一些区别，但可以通过某些变换转化为其他类型的图。如图2所示，有向图能通过对称化转换为无向图，通过阈值化转换为无重有向图，无向图也可以通过阈值化转换为无重无向图，无重有向图可以通过对称化转换为无重无向图。



图 2. 四种类型图之间的关系

为了利用计算机有效地分析网络特性，网络被抽象为两种一般的表示方法，即相邻表和相邻矩阵，这两种表示方法在实际科研中被广泛使用。如果图G(V，E)是一个包含n个节点的无定向和非加权图，那么相应的邻接矩阵是一个n×n矩阵，通常用A表示，定义如下。

|  |  |  |
| --- | --- | --- |
|  |  | (1) |

因此，图的邻接矩阵表示为：

|  |  |  |
| --- | --- | --- |
|  |  | (2) |

由于大型复杂网络通常不是很密集，使用邻接矩阵来进行存储会直接导致存储空间的浪费，这是因为邻接矩阵的大多数元素都是0。因此，为了表示稀疏的无向图，提出了一个邻接表的表示方法，其对于每个节点，创建一个包含邻接信息的单一链接表。用邻接表表示图形可以大大减少内存浪费，但由于每条边在无向无重的图形中出现两次，从邻接表中删除重复的边可以节省更多存储空间。

2.1.2 复杂网络的主要研究内容

研究复杂网络的主要方面包括网络的结构稳定性、网络上模型的特性、网络进化的统计规律、网络的几何特性、网络形成的机制、网络进化动态的机制等。在自然科学中，网络主要在节点、社区和图层面上进行研究。在节点层面，包括聚集及其分布特性、秩序相关性、最短距离及其分布特性、节点相互依赖及其分布特性、节点秩序及其分布特性、。社区层面主要包括社区的发现和发展。

在复杂网络分析中，有三个成熟的研究方向和领域：链接预测、影响最大化和关联检测。链接预测是一种预测在网络中尚未形成连接边缘的两个节点之间建立链接的概率的方法，其依据是已经知道的关于节点和网络结构的信息。影响力最大化是影响力传播领域中被广泛研究的一个基本问题。它涉及到在一个给定的网络传播模型中寻找一组节点来最大化一组节点的最终影响力。社团检测一般指的是搜索网络中密切相关的部分，称为社团。虽然每个研究方向都有不同的方面和重点，但它们都追求相同的研究目标：发现和描述网络结构的拓扑特性；找到这些拓扑特性的更好的衡量标准；通过研究网络中单个节点的特性和整个网络的结构，进行分析和预测网络行为；并且追求的是同样的研究目标：建立网络模型，以更好地理解产生网络统计特性的意义和机制；提出改善网络性能的有效方法。网络建模可以更好地理解产生网络统计数据的意义和机制，并提出改善网络性能的有效算法。

2.2 社团检测算法的研究

纽曼等人[3]在2002年第一次提出了社团结构的相关概念，概念一经提出，复杂网络分析中的社团检测问题正式流行了起来。社检测是一个研究领域，它利用网络中包含的信息，通过某些搜索方法来揭示网络中的社团结构。这是理解复杂网络的复杂结构的重要一步。

近几十年来，社团检测受到了研究人员的极大关注，并提出了许多社团检测算法。它们大致可分为基于模块化的社团检测算法、基于标签传播的社团检测算法、基于分层分组的关联识别算法、基于随机块模型的社团检测算法和基于非负矩阵分解的社团检测算法。在这篇文章中，将介绍具体的社团检测算法对各种社团的检测。

2.2.1 基于标签传播的社团检测算法

自2007年发表第一个用于社区检测的标签传播算法[2]以来，研究人员一直试图增强算法的稳定性并提高结果的准确性。LPA是一种具有许多缺点的非确定性算法。LPA在更新标签时只使用一跳邻域提供的信息，这忽略了高层结构提供的信息。另一个问题，所有的节点和边都是相等的，并且没有给出重要性函数。此外，算法使用的随机性导致其他人中的不同社区结构在同一网络上运行。因此，已经研究了各种策略来改进LPA算法。在下文中，我们将更详细地讨论它们。

LPAm：Barber和Clark [8]使用LPA来优化目标函数，其最大值对应于社区解。在这种方法中使用的模块化最大化遭受分辨率限制问题[9]，并且可以获得较差的局部最大值。

LPAm+：Liu和Murata [10]提出了一种先进的模块化标签传播算法，使用多步贪婪凝聚算法（MSG），可以同时加入多对社区。MSG被用于避开LPAm中模块度优化过程所涉及的局部极大值。LPAm和LPAm+两个算法的复杂度由于模块度函数的最大化而不是线性的。

LPA-CNP：为了达到提高LPA的准确性的目的，Lou，Li和Zhao在[11]中提出的算法使用加权相干邻域邻近度来计算一对顶点在同一社区内的概率。节点在LPA迭代过程中采用概率最大的标签。

LPAc：K.Zhang等人[12]使用边聚类系数来更新节点而不是随机邻居节点的标签。LPAc不考虑节点的重要性，并使用三角形计数，这在计算上是昂贵的[22]

NIBLPA：基于节点影响，Xing等人[13]试图通过更新具有最大k壳层值的节点的标签来进一步提高LPA的性能。

LPALC：X. -K。Zhang等人[14]通过选择具有局部循环的最近邻居取代选择随机邻居来更新标签，从而改进了LPA过程。

LPA\_NI：为了避免LPA迭代过程中的随机性，X. -K。Zhang等人在[15]中将节点的标签更新顺序固定为节点重要性的降序。当多个标签是候选时，它们计算每个标签的影响，并选择影响最大的标签来达成更新节点标签。

GA-LP：Francisquini，Rosset，and Nascimento [16]研究了基于LPA和遗传算法关键算子给出的局部信息的有向网络中的遗传算法。该算法在有向网络和无向网络中都表现出了良好的性能和较低的复杂度，但它更倾向于小规模社区的划分。

LPA-E：标准LPA只关注一跳的邻居，降低了算法的性能。Chen等人[17]提出的方法是基于直接邻居和间接邻居的互信息来计算节点对之间的关系。该方法可以检测不相交和重叠的社区。

LP-LPA：Berahmand和Bouyer [18]提出了一种基于链接影响的LPA算法。其主要思想是利用标签影响顺序进行节点选择，利用邻居的链接强度进行标签更新。

WLPA-LEB：Joghan，Bagheri，and Azad [19]将Girvan纽曼算法与LPA算法相结合，提出了基于局部边介数的加权标签传播算法，使其具有最佳的准确率和性能。其思想是从具有较低局部边介数的邻居节点选通标签。

许多最近的社区检测方法也以某种形式采用了LPA算法。为了避免随机更新顺序，作者在[20]中使用基于局部信息的节点相似性来传播标签。[21]的工作使用局部拓扑信息来测量节点之间的亲密度。在[22]中，定义了基于随机游走的相似性矩阵。

现有的研究并没有挖掘出网络的全部全局拓扑信息，而且大多局限于一阶或二阶邻域。所以本文提出了一种基于库伦定律的改进标签传播社团发现方法，该方法综合利用了网络的局部和全局拓扑信息。

2.2.2 基于模块度的社团检测算法

模块度是评价相社团检测结果的一个常见指标，也是在实际相关性分类未知的网络的相关性检测中最常用的指标。因此，人们开发了一套社团检测方法，目的是最大限度地提高模块化程度，但事实证明，模块化最大化问题是一个非确定性的多项式时间困难问题，许多基于模块的社团发现算法依赖于近似算法，如模拟退火、极值优化和光谱优化[23]。

CNM算法[24]是一种社团发现算法，它采用了Clauset、Newman和Moore提出的最大化模块化的贪婪算法的思想，因此缩写为CNM，是一种在计算和更新堆数据结构中的模块化，以 达到模块化的目的。在算法初始化时，假设每个节点都是独立的社团，并且模块化值为零，计算初始模块化增量矩阵，其中每一行的最大元素可以形成最大堆，从最大堆中选择最大的模块化增量矩阵，进行相应的融合关联，并且 更新模块化增量矩阵和最大堆；最后，循环上述步骤，直到所有节点都合并成一个单一的社团。上述过程中的CNM算法的模块化只有一个最大值，即如果模块化增加矩阵的最大元素从正值变为负值，则认为网络的相关性分割是当前网络的相关性结构。

2.2.3 基于层聚类的社团检测算法

在现实世界中，网络的实际划分经常是未知的，这种情况下，分区聚类等策略有可能并不符合标准，因为它们对给定图中聚类的数量和聚类的大小做出了不合理的假设。然而，网络的图形可能是分层的，有多层的节点集群。

常用的分层社团检测的算法大多是基于分裂过程或者聚集过程，Girvan 与 Newman 提出的 GN 算法[3]，它是基于分裂过程的一种经典层次聚类算法，其以网络边缘的中位数，这意味着网络中所有节点之间的最短路径经过一条边的次数是检测社团的标准，不同相关性的边之间的中值必须相对较高。中位数值大的边缘为 人们认为，通过删除它们可以实现网络的相关性结构，但GN算法需要对网络中的实际社团数量有一个事先的了解。

Shen等人[25]采用聚集框架提出了EAGLE(Agglomerative hierarchical clustering based on maximal clique)算法，通过处理一组最大集群来检测网络的相关结构和层次，并通过扩展模块化来评估算法的性能。

通过结合这两个过程，Toujani等人[26]提出了一种基于优化和混合分层分组的混合方法来探索网络中的相关结构。由于混合分层模型通常会陷入局部最优，他们还引入了一个基于模块化的定向函数，并使用遗传算法对目标函数进行优化，以确保这种方法不会陷入局部最优。

2.2.4 基于随机块模型的社团检测算法

随机块模型是一类普遍的随机图模型，一般用于生成图的结构。这个模型的思想是，对于网络中的一个特定节点，首先要确定该节点的相关性，然后确定该节点与其他与之相关的节点之间的关系。该模型很容易描述，但可以产生各种不同的网络结构。例如，将概率矩阵放在对角线上会产生一个具有不连接成分的网络，而从对角线上增加较小的元素则会产生一个传统的相关结构。概率矩阵的其他选项允许创建核心-外围、分层和多人结构。由于其通用性和易分析性，基于块的概率模型被用于许多情况，这些方法也受到社团检测领域的研究者的欢迎。

随机块模型除了上述的优点还具有一定的局限性，Karrer 等人[27] 认为由于大多数现有的块状模型忽略了网络中节点顺序的变化，因此不适用于顺序分布通常很广的真实网络，他们提出了一种方法，将节点的顺序结合到随机块状模型中，并使用随机块状模型来提高关联度，以此利用随即模块模型改进检测的目标函数。

2.2.5 基于非负矩阵分解的社团检测算法

非负矩阵分解为一种机器学习算法，可用于特征提取和减少数据大小。近年它被广泛地应用于社区检测问题，并取得了许多相关研究结果。非负矩阵分解法是由Lee和Seung[28]在1999年提出的。目前以非负矩阵分解框架为基础的社 团检测算法的变体有非常多种，例如，Zhang 等人[29] 通过用扩散核（即给定网络的拉普拉斯矩阵）取代输入特征矩阵，可以得到一种检测更多重叠关联的关联检测算法。这种方法可以检测出模糊的群体。这种方法的一个特点是，它提供了关于一个节点属于某个社团的程度的信息。Zarei等[30]提出了一种基于NMF的社区检测算法，该方法引入了节点与节点之间相关性矩阵作为新的特征矩阵。

2.3 相关符号及定义

复杂网络可以被建模为无向图G =（V, E），其中n =|V|节点，m =|E|节点之间的边数。N（i）= {j, vj ∈V和{ vi, vj }∈E}表示顶点vi（所有相邻节点）的邻域集合，deg（vi）=|N(vi)|是节点vi的度。节点vi和vj之间的测地距离是两个节点之间的最短路径中的跳数。

2.4 评价指标：模块度、NMI和ARI

2.4.1 模块度

Newman在[1]中提出的模块度度量了随机图中社区内的边减去其期望值的分数。它可以计算如下：

|  |  |  |
| --- | --- | --- |
|  |  | (3) |

其中，Q为模块度；m为网络的边数；A为网络的邻接矩阵。如果节点i和j相邻，则Aij = 0。如果不相邻，则Aij = 0；ci，cj为节点i和j的标号，如果节点i和j在同一社区中则δ（ci, cj）= 1，否则δ（ci, cj）= 0。

模块度定义在-1和1之间，通常在真实网络中，取值范围为[0.3, 0.7][10]。Q的值越高，指示高质量的社区检测结果。

2.4.2 NMI

许多研究使用NMI [31]来评估或比较不同的社区检测算法。NMI度量通常用于具有已知参考划分的基准网络或具有由领域专家注释的划分的真实世界网络。

考虑用于将n个顶点划分成社区组的两个划分X和Y，X和Y之间的归一化互信息（NMI）被定义为：

|  |  |  |
| --- | --- | --- |
|  |  | (4) |

使用顶点被分配到方案X中的社区x的概率P（X），将互信息I定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (5) |

H是熵，定义如下：

|  |  |  |
| --- | --- | --- |
|  |  | (6) |

NMI测量两种不同划分方案共有的信息量。它始终采用0和1之间的值。NMI的值越高意味着性能越好，当X和Y相同时，NMI等于1。

2.4.3 ARI

ARI取值范围为[-1, 1]，值越大越好，反映社团划分的重叠程度，使用该度量指标需要数据本身有类别标记。

用C表示实际的类别划分，K表示聚类结果。定义a 为在C中被划分为同一类，在K中被划分为同一簇的实例对数量。定义b为在C中被划分为不同类别，在K中被划分为不同簇的实例对数量。定义Rand Index（兰德系数）：

|  |  |  |
| --- | --- | --- |
|  |  | (7) |

其中，n表示实例总数，RI的取值范围为[0，1]，值越大说明聚类效果越好。

Rand Index无法保证随机划分的聚类结果的RI值接近0。于是，提出了Adjusted Rand index（调节的兰德系数）：

|  |  |  |
| --- | --- | --- |
|  |  | (8) |

2.5 本章小结

本章对社团检测与复杂网络的相关理论和技术做了介绍。首先，在2.1小节中对复杂网络的定义和表现形式做了详细介绍，并简述了在复杂网络中的经典研究问题。本章2.2小节介绍了适用于各种网络的社团检测算法，如基于随机块模型的社团检测算法、基于模块化的社团检测算法、基于分层分组的关联识别算法、基于标签传播的社团检测算法和基于非负矩阵分解的社团检测算法，并描述了各种算法的特点和发展。在本章2.3小节介绍了复杂网络和本文的相关符号和定义。最后，在本章的2.4小节介绍了社团检测的评价指标模块度Q，NMI以及ARI的定义和算式。

**第三章 基于库伦定律的社团检测算法**

3.1 库仑定律及库伦矩阵

3.1.1 库仑定律

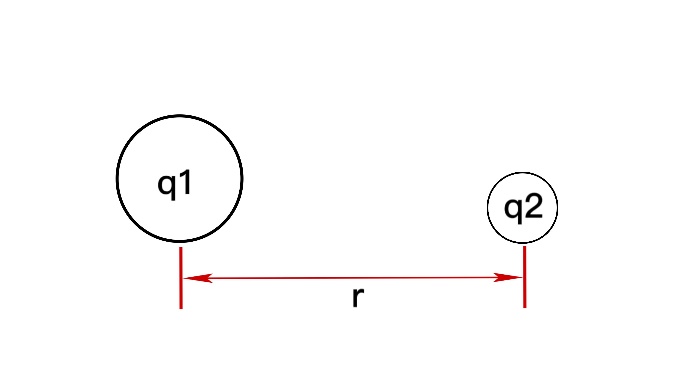


图 3. 距离为r的带电粒子q1和q2之间的电力

给定两点电荷q1和q2相隔距离r（图1），根据库仑定律（1785），它们之间的排斥力（电荷符号相同）或吸引力（电荷符号相反）的大小如下：

|  |  |  |
| --- | --- | --- |
|  | *，k为比例常数* | (9) |

3.1.2 定义节点间吸引力

与物理学领域类似，本文把网络想象成一组带电粒子。为了得到节点i和j之间的吸引力Fij，我们设k = 1，并令节点的度为它的电荷。根据库仑定律，其公式如下：

|  |  |  |
| --- | --- | --- |
|  |  | (10) |

很明显，简单网络中吸引力的定义是对称的，即Fij = Fji。

许多网络都有多个节点具有相同的度。然而，通过考虑测地距离，所定义的吸引力给出了网络中节点位置的更全局可视化。两个节点之间的吸引力与两个节点之间的测地线距离的平方成反比，因此，当两个节点之间的距离越小时，吸引力越大，并且力与节点的度成正比。

3.1.3 库伦矩阵

G =（V,E）的库仑矩阵是由V索引的n × n矩阵C，其（i,j）项定义为：

|  |  |  |
| --- | --- | --- |
|  |  | (11) |

一个节点所受的力之和表现了该节点在整个网络中的重要性。由于网络中所有最短路径的计算复杂度较高，本文定义了一个深度为d的矩阵来计算最大距离为的每对节点之间的库仑力。可以在区间[1，D]中确定d的值。D的值是指网络中一对顶点之间的最大距离（图的直径）。这样，当深度太高时，由于要计算的最短路径的数目最大，所以会导致算法计算量很大，但结果达到稳定状态。因此，d值越高，库仑矩阵越精确。

3.2 基于库伦定律的社团检测算法

3.2.1 算法设计

可以将网络中的节点度数视为电子的电荷量，节点间的距离类比电子之间的距离，根据库仑定律计算所有两个节点之间的吸引力，得到节点吸引力矩阵，为了避免冗余运算量过大，可以定义一个深度系数来限定某单位距离内的节点间的测地线距离（比如限定为3个单位）。两个节点之间的吸引力与两个节点之间的测地线距离的平方成反比，因此，当两个节点之间的距离越小时，吸引力就越高，并且力与节点的度数成正比

3.2.2 算法实现

为每个节点赋予一个唯一的标签，计算规定深度的库伦矩阵。计算每一个节点与其他节点的吸引力之和，并将所得值按升序排序，为每个节点按吸引力最大的邻居节点更新标签。每一次迭代遍历更新所有的节点标签，更新迭代至社团划分结果不再有变化或达到迭代次数阈值则结束社团划分。

更新标签时，如果有几个节点与主节点具有相同的吸引力，则选择最常见的邻居的标签，如果共同邻居的数量也相同，则随机选择标签，成为更新的节点标签。

3.3 算法流程

本文基于库伦定律的社团检测算法的主要步骤如下：

(1) 每个节点被认为是一个社区，并获得一个唯一的标签；

(2) 计算深度为d的库仑矩阵；

(3) 根据库仑力的和将网络节点按升序排列，并将结果存储在列表X中；

(4) 设置t=1；

(5) 对于每一个x∈X，根据库仑矩阵迭代更新节点标签，节点x从其吸引力最大的邻居处获得标签。当最大值与多个标签相关联时，我们选择具有最大公共邻居的标签。否则，选择是随机的；

(6) 如果每个节点的标签没有改变，停止迭代。否则，设置t +1，转步骤(5)；

(7) 具有相同标签的节点组成一个共同体。



图 4. 算法流程图

三次LPA迭代所采用的随机性选择可能会影响算法的稳定性，导致不太重要的节点影响比较重要的节点的逆流现象，也可能导致振荡现象[9]。此外，LPA只利用了一跳邻居信息，忽略了来自邻居的其他连接，导致网络全局拓扑结构所给出的重要信息丢失。

为了避免这些问题，该算法首先按照库仑力的升序对图中的节点进行排序，然后从具有较高库仑力的邻居节点中选择一个标签。当在具有相似标签的相邻节点之间存在多个节点时，测量相同标签的力的总和。然后，选择求和后具有最大值的标签。当最大值被分配给许多标签时，也避免了随机性，并且选择来自具有最大公共邻居的节点的标签。

3.4 时间复杂度分析

给定一个网络G（V,E），其中n=|V|表示节点的数量，m=|E|表示边的数量。就时间复杂度而言，d是图的直径，<k>是指网络中节点的平均程度。该算法的每个步骤的时间复杂度估计如下：

**步骤1**：该算法的计算时间实际上受构建库仑矩阵的时间限制。要做到这一点，应该遍历网络中的所有节点和所有节点的邻居，直到一个深度d，使1≤d≤D（最大图直径）。因此，计算时间复杂度为 计算时间复杂度为O(n<k>d)，随着d的增加而快速增长。

**步骤2**：采用快速排序算法，将节点按其库仑力的升序排列的时间复杂度为O(n log n)。的时间复杂度为O(n log n)。

**步骤3**：用唯一的标签初始化每个节点的时间复杂度是O(n)。标签传播过程的时间复杂度为t×O(m)（其中t为最大迭代次数）。在算法结束时将具有相同标签的节点分配给一个社区的时间复杂度为O(n)。

因此，整体时间复杂度大致为O(n<k>d)+t×O(m)+2×O(n)+O(n log n)，当计算所有距离，即无定向和无权重网络内的所有最短路径时，LPA的总计算复杂度被O(nm)所约束。

3.5 本章小结

本节根据库仑定律的定义对基于标签传播的社团检测算法进行了类比推移以达到改进效果，定义了节点之间的吸引力，并根据节点之间吸引力定义了库伦矩阵。本节展示了关于这种算法的详细设计，并给出了伪代码和算法流程，其核心思想为该算法首先按照库仑力的升序来对图中的节点进行排序，再从具有较高库仑力的邻居节点中选择一个标签，如果符合的有多个节点时，测量相同标签的力的总和，再选择求和后具有最大值的标签。这样可以避免LPA算法的随机性弊端。

**第四章 实验及结果分析**

4.1 实验数据集

本文使用两个真实社交网络的数据集评估了社团检测模型的性能，这些数据集包含了节点的拓扑结构和特征信息，并提供了社团数量和社团成员的分类信息。数据集中的每个节点都有一个独特的社团标签。关于实验数据集的详细的信息可以在表1中查找。

表1. 数据集信息

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据集 | 节点数 | 边数 | 节点特征数 | 节点类 | 孤立点数 | 描述 |
| cora | 2708 | 5278 | 1433 | 7 | 151 | 引文网络 |
| cornell | 195 | 286 | 1703 | 5 | 18 | WebKB数据集的子网络 |

Cora的引文网络[32]包含2708个出版物，分为七个类别。在节点特征信息中，每个节点由一个1/0值的向量来描述，表示它是否有相应的特征，用于描述出版物中是否出现1433个独特的词。Cornell是一个WebKB超链接网络[33]，由877个节点和1608条边组成，其网络的节点分为五大类：student、course、faculty、 project 和 staff。节点特征信息共包含1703个节点特征。这些数据集包含拓扑结构、节点特征和关联分类信息，直接适用于这里描述的社团检测方法。请注意，这两个网络都包含一些孤立的点，它们与网络中的其他节点没有联系，因此很难检测网络中的关系。两个数据集都可以从UCSC网站下载。

本文还使用了两个人工网络art\_network1和art\_network2，关于两个人工网络的详细信息可在表2中查找。

表2. 人工网络数据集

|  |  |  |
| --- | --- | --- |
| 数据集 | 节点数 | 边数 |
| art\_network1 | 256 | 602 |
| art\_network2 | 256 | 354 |

4.2 关于参数的选取

本文提出了一种改进的标签传播社区检测方法基于库仑定律，它使用的局部和全局拓扑信息。首先，该算法量化每一对节点之间的力量获得相似矩阵。相似矩阵建设期间，可以设置一个深度参数d限制相似性的计算节点在一个距离等于该参数。最后，算法执行使用一个标签传播方法的聚类分析。

由于在大型网络中库伦矩阵的计算量会很大，所以本文定义一个深度系数d，在区间[1, D]内确定深度系数的值。d的最大值D是指网络中一对顶点之间的最大距离（图的直径）。这样一来，当深度过高时，由于要计算的最短路径数量最多，算法就很复杂，结果就会进入一个稳定状态。因此，深度系数的值越高，库仑矩阵就越精确。

一般来说，当d取值为图的直径时，最远的邻居对计算节点重要性的影响很小，这导致了有序的节点集和社区检测结果的稳定状态，但大型网络会使库伦计算占用过多资源，建议根据网络的大小和稀疏程度来调整d。

4.3 社团检测结果比较

选取数据集cora， cornell， art\_network1， art\_network2，其中数据集cora和cornell是真实网络，art\_network1和 art\_network2是人工网络。下面展示LPA和LPACL在网络中的各类评价指标表现：

表3. LPA和LPACL在数据集cora上的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 模块度 | NMI | ARI | 社团数量 |
| LPA | 0.697598 | 0.381378 | 0.123527 | 298 |
| LPACL | 0.705052 | 0.415410 | 0.182873 | 134 |
| 提升 | 1.06% | 8.92% | 48.02% | - |

表 4. LPA和LPACL在数据集cornell上的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 模块度 | NMI | ARI | 社团数量 |
| LPA | 0.317192 | 0.155553 | 0.021326 | 20 |
| LPACL | 0.348533 | 0.117824 | 0.079056 | 10 |
| 提升 | 9.88% | - | 270.70% | - |

表5. LPA和LPACL在数据集art\_network1上的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 模块度 | NMI | ARI | 社团数量 |
| LPA | 0.354179 | 0.308941 | 0.161885 | 33 |
| LPACL | 0.461531 | 0.516519 | 0.335072 | 10 |
| 提升 | 30.31% | 67.18% | 106.98% | - |

表6. LPA和LPACL在数据集art\_network2上的实验结果

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 算法 | 模块度 | NMI | ARI | 社团数量 |
| LPA | 0.571978 | 0.313742 | 0.043335 | 90 |
| LPACL | 0.633004 | 0.376326 | 0.077498 | 60 |
| 提升 | 10.67% | 19.95% | 78.83% | - |

可以看出，在绝大多数情况下，LPACL在模块度，NMI和ARI三种社团划分评价指标上都表现出了不同程度的提高，且社团划分后形成的社团数量更少，由此可以证明LPACL在社团检测上的有效性，也证明了LPACL相较于经典的LPA算法有更加准确和稳定的结果。

4.4 本章小结

在本文中，介绍了一种基于库仑定律的改进的LPA算法LPACL。首先，该算法计算库仑矩阵，并将节点按其受力的升序排列。在标签传播过程中，LPACL使用许多策略来避免原始LPA的随机选择，使结果更加稳定。为了证明LPACL相较于经典的LPA算法的社团划分结果更优秀且有更加准确稳定的结果，本章展示了两种方法在社团检测的三种评价指标：模块度，NMI，ARI在数据集真实网络cora和cornell，人工网络art\_network1和art\_network2的具体实验结果，可以得出结论，LPACL更稳定，在获得更高的模块度值和NMI方面也更准确。考虑到大型网络会使库伦计算占用过多资源，本章讨论了深度系数d设置的必要性，并建议根据网络的大小和稀疏程度来调整d。

**第五章 总结与展望**

5.1 论文工作总结

社团检测是复杂网络分析的一个重要部分，其主要目的是对网络中密切相关的子集进行分类。这在现实生活中有着广泛的应用，例如在社会网络推荐系统、公共安全、嫌疑人追踪和城市规划中。随着大数据时代数据量的增加，网络的规模也在快速增长。因此，社团检测算法需要进行优化，以便在复杂的现实世界网络中获得稳定和准确的分类结果。因此，本文主要研究基于标签传播的相关性检测算法的优化问题，主要工作如下：

1. 设计算法：考虑到标签传播算法（LPA）迭代中用于检测社团的随机选择可能会损害算法的稳定性。此外，由于LPA只使用单跳邻域信息，网络的整体拓扑结构所提供的重要信息就会丢失。这些缺点无助于算法的稳定性和准确性。本文设计了LPACL算法，通过类比库仑定律描述的电子间作用力来避免LPA算法的上述缺点。具体设计如下：网络中节点的度数被认为是电子的电荷，节点之间的距离类似于电子之间的距离，根据库仑定律计算任意两个节点之间的吸引力，得到节点的吸引矩阵，两个节点之间的吸引力与两个节点之间的距离的平方成反比，这就是测地线的吸引力 因此，两个节点之间的距离越小，吸引力就越大，力与节点频率成正比。

2. 引入深度系数d：由于所考虑的大型复杂真实网络的规模，本文提出的LPACL算法对每对节点之间的力量进行量化，以获得相似性矩阵，这是一个操作密集的过程，对于计算库伦矩阵来说，资源消耗太大。为此，引入了一个深度参数d来约束相似性，计算和设置与该参数相同距离的节点，以最大限度地降低大型网络的时间复杂度，因为当d等于图的直径时，最远的邻居对计算节点的重要性影响不大、 得到一个有序的节点集，社区检测结果也很稳定。建议根据网络的大小和稀疏程度调整d。

3. ~~实验并得出结论~~：本章展示了两种方法在社团检测的三种评价指标：模块度，NMI，ARI在数据集真实网络cora和cornell，人工网络art\_network1和art\_network2的具体实验结果，在绝大多数情况下，LPACL在模块度，NMI和ARI三种社团划分评价指标上都表现出了不同程度的提高，且社团划分后形成的社团数量更少，由此可以证明LPACL在社团检测上的有效性，也证明了LPACL相较于经典的LPA算法有更加准确和稳定的结果。

5.2 未来展望

本文提出了一种基于库仑定律的标签传播社团检测算法，但是这种方法在网络类型应用和运算量上仍然有一定局限性，为了更好的提升LPACL的性能，本文的未来研究方向是继续研究如何改进库仑矩阵的计算，其重点是如何在不损失精确度和稳定性的前提下降低大型网络中的库伦矩阵运算。其次，本文未来将提出的方法扩展到有向加权网络，以解决更大范围的网络应用问题。

**参考文献**

[1] M.E.J. Newman, M. Girvan, Finding and evaluating community structure in networks, Phys. Rev. E 69 (2004) 026113, http://dx.doi.org/10.1103/PhysRevE.69.026113.

[2] U.N. Raghavan, R. Albert, S. Kumara, Near linear time algorithm to detect community structures in large-scale networks, Phys. Rev. E 76 (2007) 036106, http://dx.doi.org/10.1103/PhysRevE.76.036106

[3] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the National Academy of Sciences of the United States of America, 2002, 99, 7821-7826.

[4] Mankad S, Michailidis G. Structural and functional discovery in dynamic networks with non-negative matrix factorization[J]. Physical Review E, 2013, 88, 1-14.

[5] Maekawa S, Takeuch K, Onizuka M. Non-linear attributed graph clustering by symmetric NMF with pu learning[J]. https://arxiv.org/abs/1810.00946, 2018.

[6] Javadi S H S, Gharani P, Khadivi S. Detecting Community Structure in Dynamic Social Networks Using the Concept of Leadership[J]. Springer, 2018.

[7] Chakrabarti D, Kumar R, Tomkins A. Evolutionary Clustering[C]. Acm Sigkdd International Conferenceon Knowledge Discovery & Data Mining. 2006.

[8] M.J. Barber, J.W. Clark, Detecting network communities by propagating labels under constraints, Phys. Rev. E 80 (2009) 026129, <http://dx.doi.org/10.1103/PhysRevE.80.026129>.

[9] S. Fortunato, M. Barthelemy, Resolution limit in community detection, Proc. Natl. Acad. Sci. 104 (2007) 36–41, <http://dx.doi.org/10.1073/pnas.0605965104>.

[10] X. Liu, T. Murata, Advanced modularity-specialized label propagation algorithm for detecting communities in networks, Phys. Stat. Mech. Its Appl. 389 (2010) 1493–1500, <http://dx.doi.org/10.1016/j.physa.2009.12.019>.

[11] L. Tang, H. Liu, Community detection and mining in social media, Synth. Lect. Data Min. Knowl. Discov. 2 (50) (2010) <http://dx.doi.org/10.2200/S00298ED1V01Y201009DMK003>.

[12] X.-K. Zhang, X. Tian, Y.-N. Li, C. Song, Label propagation algorithm based on edge clustering coefficient for community detection in complex networks, Internat. J. Modern Phys. B 28 (2014) 1450216, <http://dx.doi.org/10.1142/S0217979214502166>.

[13] Y. Xing, F. Meng, Y. Zhou, et al., A node influence based label propagation algorithm for community detection in networks, Sci. World J. 2014 (2014) 1–13, <http://dx.doi.org/10.1155/2014/627581>.

[14] X.-K. Zhang, S. Fei, C. Song, et al., Label propagation algorithm based on local cycles for community detection, Internat. J. Modern Phys. B 29 (2015) 1550029, <http://dx.doi.org/10.1142/S0217979215500290>.

[15] X.-K. Zhang, J. Ren, C. Song, et al., Label propagation algorithm for community detection based on node importance and label influence, Phys. Lett. A 381 (2017) 2691–2698, <http://dx.doi.org/10.1016/j.physleta.2017.06.018>.

[16] R. Francisquini, V. Rosset, MCV. Nascimento, GA-LP: A genetic algorithm based on label propagation to detect communities in directed networks, Expert Syst. Appl. 74 (2017) 127–138, <http://dx.doi.org/10.1016/j.eswa.2016.12.039>.

[17] N. Chen, Y. Liu, H. Chen, J. Cheng, Detecting communities in social networks using label propagation with information entropy, Phys. Stat. Mech. Its Appl. 471 (2017) 788–798, <http://dx.doi.org/10.1016/j.physa.2016.12.047>.

[18] K. Berahm, A. Bouyer, LP-LPA: A link influence-based label propagation algorithm for discovering community structures in networks, Internat. J. Modern Phys. B 32 (2018) 1850062, http://dx.doi.org/10.1142/S0217979218500625.

[19] H.S. Joghan, A. Bagheri, M. Azad, Weighted Label Propagation Algorithm based on Local Edge Betweenness, Vol. 29, 2019.

[20] C. Song, G. Huang, B. Yin, et al., Label propagation algorithm based on node similarity driven by local information, Internat. J. Modern Phys. B 33 (2019) 1950363, http://dx.doi.org/10.1142/S0217979219503636.

[21] H. Kong, Q. Kang, C. Liu, et al., An improved label propagation algorithm based on node intimacy for community detection in networks, Internat. J. Modern Phys. B 32 (2018) 1850279, http://dx.doi.org/10.1142/S021797921850279X.

[22] X.-K. Zhang, C. Song, J. Jia, et al., An improved label propagation algorithm based on the similarity matrix using random walk, Internat. J. Modern Phys. B 30 (2016) 1650093, http://dx.doi.org/10.1142/S0217979216500934.

[23] Fortunato S. Community detection in graphs[J]. Physics Reports-Review Section of Physics Letters, 2010, 486, 75-174

[24] Clauset A, Newman M E J, Moore C. Finding community structure in very large networks[J]. Physical Review E, 2004, 70, 1-6.

[25] Shen H, Cheng X, Cai K, Hu M B. Detect overlapping and hierarchical community structure in networks[J]. Physica a-Statistical Mechanics and Its Applications, 2009, 388, 1706-1712.

[26] Toujani R , Akaichi J . GHHP: genetic hybrid hierarchical partitioning for community structure in social medias networks[C]//2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI). 2018, 1146-1153.

[27] Karrer B, Newman M E J. Stochastic blockmodels and community structure in networks[J]. Physical Review E, 2011, 83, 1-12.

[28] Lee D D, Seung H S. Learning the parts of objects by non-negative matrix factorization[J]. Nature, 1999, 401, 788-791.

[29] Zhang S, Wang R S, Zhang X S. Uncovering fuzzy community structure in complex networks[J]. Physical Review E, 2007, 76, 1-7.

[30] Zarei M, Izadi D, Samani K A. Detecting overlapping community structure of networks based on vertex-vertex correlations[J]. Journal of Statistical MechanicsTheory and Experiment, 2009, 1-17

[31] L. Tang, H. Liu, Community detection and mining in social media, Synth. Lect. Data Min. Knowl. Discov. 2 (50) (2010) <http://dx.doi.org/10.2200/S00298ED1V01Y201009DMK003>.

[32] Sen P, Namata G, Bilgic M, Getoor L, Gallagher B, Eliassi-Rad T. Collective classification in network data articles[J]. AI Magazine, 2008, 29, 93-106.

[33] Craven M. Learning to extract symbolic knowledge from the World Wide Web[C]//AAAI'98/IAAI'98: Proceedings of the fifteenth national/tenth conference on Artificial intelligence/Innovative applications of artificial intelligence. 1998, 509-516.

[34] B. Laassem, A. Idarrou, L. Boujlaleb, M. Iggane, Label propagation algorithm for community detection based on Coulomb’s law[J]. Physica A: Statistical Mechanics and its Applications, Volume 593, 2022. DOI: [10.1016/J.PHYSA.2022.126881](https://schlr.cnki.net/Detail/doi/GARJ2021_2/SJESA606D9083987A9282BCA9CF2874D3962)

致 谢

行文至此落笔，我的本科生活也即将在不久以后结束。时光荏苒，犹记2019年入校时对大学生活满怀憧憬的心情。在这四年里的所有美好时光仿佛历历在目。在毕业设计完成之际，我想对所有曾给予我帮助和鼓励的人表示最诚挚的感谢。

感恩我的父母对我二十余年的养育之恩，古语有云：“父母之爱子，非为报也，不可内解于心。”他们在对我的培养道路上从不吝啬时间和金钱以及心意的支出，总愿意把最好的都留给我，在我学习的道路上毫不犹豫的投资我，重视我的教育和德育，督促我一步一步变成更好的人，给我机会和支持去站在更高的地方看我想要的远方。父母恩情没齿难忘，今后我也会好好生活，不让他们失望和担心。

感谢我的论文指导老师赵志立老师，赵老师总是很有耐心的和我进行关于学习方面的谈话，也很关心我的校园生活。在我向赵老师提出出国留学的计划时，赵老师给我了莫大的指导和帮助。赵老师对我们做的好的地方并不吝啬夸奖，并时常秉承严谨治学一丝不苟的态度对待科研，同样也用更高的要求来指导我们，总是提出关键而宝贵的意见，使我的论文得以完善。同时也感谢曾给我指导的所有老师，我会带着各位老师的谆谆教诲继续求学，希望老师们工作顺利，事事顺心。

感谢大学四年我遇到的每一位同学，前辈和学弟学妹。我多数时候是一个个性有些粗糙的人，而我却被每一个朋友照顾的很好。谢谢每一次相遇，让我们有机会一起成长，互相支持。我也遇到了很多三观合得来，懂我脾气，开玩笑有度，温柔有教养的好朋友，是他们照亮了我的四年。也很感谢我一直保持联系的来自家乡的朋友，人生长河漫漫，他们最少陪我走过六年旅途，最多有相识二十年之久。大学生活难免有摩擦，在我每一次失落彷徨时他们都会在远方给我支持。最后也很感谢我的每趟旅途中的奇妙相遇和新朋友，让我知道只要我一直前行，远方总会有很多有趣有盼的事物等着我发掘，使我能一直不畏遗憾永远乐观。

感谢学校和学院，为我提供这样好的平台和机会去求学，尤其感谢兰州大学是北海道大学的合作友好学校，使我有机会在大三时参与北海道大学的暑期学校，为我今后的留学生活铺垫了一个好的开端。

最后我感谢努力的自己，在每一次逆境没有放弃，始终心怀梦想砥砺前行，今后我依旧会怀着初心，戒骄戒躁，脚踏实地做好自己。

聚散终有时，未来皆可期，希望所遇一切都值得！

毕业论文（设计）成绩表

|  |
| --- |
| 导师评语  **建议成绩 指导教师（签字）** |
| **答辩委员会意见**  **答辩委员会负责人（签字）** |
| **成绩 学院（盖章）**  **年 月 日** |