单位代码：10293 密 级：

南京邮电大学

硕士学位论文

论文题目：基于结构与内容的社交网络水军团体识别

学 号 1013041231

姓 名 金礼仁

导 师 周国强

学科专业 计算机应用技术

研究方向 数据挖掘

申请学位类别 工学硕士

论文提交日期 二〇一六年三月

南京邮电大学学位论文原创性声明

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

本人学位论文及涉及相关资料若有不实，愿意承担一切相关的法律责任。

研究生签名：金礼仁 日期：2016.04.08

南京邮电大学学位论文使用授权声明

本人授权南京邮电大学可以保留并向国家有关部门或机构送交论文的复印件和电子文档；允许论文被查阅和借阅；可以将学位论文的全部或部分内容编入有关数据库进行检索；可以采用影印、缩印或扫描等复制手段保存、汇编本学位论文。本文电子文档的内容和纸质论文的内容相一致。论文的公布（包括刊登）授权南京邮电大学研究生院办理。

涉密学位论文在解密后适用本授权书。

研究生签名：金礼仁 导师签名：[illegible] 日期：2016.04.08

Structure And Content-Based Spammer Detection In Social Networks

Thesis Submitted to Nanjing University of Posts and Telecommunications for the Degree of

Master of Engineering

By

Liren Jin

Supervisor: Prof. Guoqiang Zhou

March 2016

摘要

随着在线社交网络的不断发展，基于社交网络的信息传播也越来越深入和广泛。然而近年来，有组织的网络水军的出现，导致社交网络上谣言信息盛行、欺诈活动猖獗，造成巨大的社会、经济损失，严重动摇了社交网络的安全基础，最终会影响社交网络的发展前景。所以进行网络水军识别研究是一项迫在眉睫的工作。在庞大的社交网络中，传统的水军识别工作，主要是基于单个特征进行的，没有把水军团体作为识别的目标，这类方法不能全面评价一个水军的特征，识别准确率和效率有提高的空间。因此，如何检测出社交网络中的水军团体，并提高社交网络水军检测的效率和准确率是一项重要的研究课题。

网络水军作为一个带有一定任务的团体，在他们的团体成员结构中会呈现出一种异常特征。基于这一思想，本文提出了一种基于结构与内容的社交网络水军团体识别方法。可以通过挖掘网络水军在社交网络中的结构特征，对社交网络中的水军团体进行识别；并结合节点本身所传播信息的内容特征，对社交网络中的水军团体进行综合分析，从而确认网络水军的身份。本文的具体工作：

（1）挖掘社交网络水军的网络结构特征。根据垃圾信息出现的时间，结合社交网络用户的转发记录构建社交网络中的转发关系网络，寻找其中传播信息能力强的重叠社区结构，初步识别网络水军团体。

（2）用户传播内容的特征挖掘。分析用户所发送内容的特征和垃圾信息的特征，通过度量它们之间的相似度，来判断一个用户是否传播过垃圾信息。

（3）综合用户的结构特征与发布内容的特征识别网络水军团体。在已识别的重叠结构的基础上，度量重叠社区内节点的内容与垃圾信息的相似度，寻找多次传播过垃圾信息的重叠社区节点，确定为网络水军。

本文基于网络水军整体结构为基点而得到的网络水军识别模式，具有全局性特征。在新浪微博数据集上，通过对比实验，验证了本文提出方法的有效性和可行性。相关成果可以为净化网络环境提供支持，因而具有一定的应用前景。

关键词：社交网络、水军识别、结构特征、内容特征、水军团体

Abstract

With the continuous development of the Internet, information dissemination based on social network is becoming far more in-depth and extensive. However, organized network spammers appearing recent years lead to the prevalence of rumors and the serious rise of cheating, which make huge damage to social and economic environment and influence the fundamental of social network, will finally harm the development of social network. As a result, detection of network spammers is indeed an imminent task. Traditional ways of detecting spammers, in the huge social network, are based on a single feature, which leave room for improvement. How to recognize spammer groups in social network and improve its efficiency and precision is therefore a significant research point.

Network spammer groups, as organizations with certain tasks, possess abnormal characteristics among their membership structure. For this reason, we propose our method to recognize network spammers based on combined structure and content features. We can detect spammer groups by figuring out their structure features in the social network. After considering content features they spread, we can are able to identify spammers by comprehensive analysis. Our work is as follows:

First, we distinguish structure features of spammers in social network. By recognizing temporal characters when garbage information appears and constructing forwarding relation networks using forwarding records among users to find overlapping community structures that are of strong propagating ability, we can make preliminary identification of network spammers.

Second, we extract features among contents that users spread. By measuring content characters users send and those of certain garbage information, we are able to make judgment whether garbage information a user has ever sent.

Finally, we recognize spammer groups considering both structure and content features. Based on overlapping communities that have been recognized, we measure similarity between contents in these communities and certain garbage information, we are capable of recognizing communities that have spread garbage information more than a certain times, and we confirm them as network spammers.

Our method has global characteristic in that the detection method it uses is based on the overall structure of network spammers. After comparison with other methods, experimental results on Sina Weibo data set have proven our work with higher efficiency and feasibility. Related technical

achievements may help purify the environment of social network and will have huge application prospect.

Keywords: social network, spammer recognition, structure feature, content feature, spammer group

目录

第一章 绪论 1
1.1 研究背景 1
1.2 网络水军识别现状研究 2
1.3 课题研究内容 4
1.4 论文的组织结构 5
1.5 本章小结 5
第二章 相关研究 6
2.1 社交网络概述 6
2.2 网络水军识别研究 8
2.2.1 基于内容特征的方法 8
2.2.2 基于行为特征的方法 9
2.2.3 基于网络特征的方法 10
2.2.4 基于影响力的方法 11
2.2.5 基于综合特征的方法 11
2.2.6 目前的研究难点和热点 12
2.3 网络水军识别研究总结 13
2.4 本章小结 14
第三章 基于网络结构特征的水军识别 15
3.1 社交网络水军的网络结构特征分析 15
3.1.1 对节点和边的度量 16
3.1.2 重叠社区结构的引入 18
3.2 构建社交网络中的转发关系网络 19
3.3 重叠社区结构发现算法 22
3.3.1 重叠社区结构发现研究 22
3.3.2 改进的重叠社区发现算法 24
3.3.3 算法伪代码 25
3.3.4 时间复杂度 28
3.4 本章小结 28
第四章 基于结构与内容的水军团体识别 29
4.1 基于内容特征识别水军问题分析 29
4.2 总体方案 29
4.3 内容特征的提取 30
4.3.1 主题模型简介 31
4.3.2 相似度计算 33
4.4 最终水军团体的确定 35
4.5 本章小结 35
第五章 实验 36
5.1 实验准备 36
5.1.1 实验数据集 36
5.1.2 数据集的处理 38
5.1.3 实验环境 38
5.2 重叠社区的发现 39
5.2.1 建立转发关系网络 39
5.2.2 评价指标 39

5.2.3 实验结果及分析 39
5.3 LDA 参数的确定 40
5.4 水军识别实验与分析 41
5.4.1 实验评估标准 41
5.4.2 实验及结果分析 41
5.5 本章小结 43
第六章 研究工作总结与展望 44
6.1 研究工作总结 44
6.2 未来的研究内容展望 44
参考文献 46
附录 1 攻读硕士学位期间撰写的论文 49
致谢 50

第一章 绪论

1.1 研究背景

当前，以 Twitter、Facebook、微信、百度贴吧、人人网等为代表的在线社交网络在新闻传播、网民互助、品牌营销、知识信息传播等方面表现出积极作用。基于社交网络的信息传播也越来越深入和广泛。研究表明，社交网络中的信息传播与传统媒体中的信息传播相比，呈现出规模大、实时、快速等特点。其对公民日常生活、国家经济和公共安全的影响越来越深入。然而近年来，社交网络上出现谣言信息盛行、欺诈活动猖獗等负面因素，特别是有组织的网络水军的出现，更加放大了这些负面因素，从而动摇了社交网络的安全基础，最终会影响社交网络的发展前景。2013 年 4 月 “雅安地震”，微博一方面成为最有力的信息传播媒体，各种“大 V”、政务微博、平民账号等充分利用微博的信息扩散能力，帮助救灾。但另一方面，也有不法分子利用微博传播谣言，欺骗公众，造成社会的不稳定和民众恐慌，带来极坏的后果。如何快速有效的识别出社交网络中的网络水军，遏制垃圾信息的传播，对于维护公民生活的正常秩序和国家公共安全具有重要意义。

网络水军识别主要运用 Web 信息挖掘技术，定义高区分度特征及行为模式发现隐藏的网络水军[1]。在早期的水军识别研究中，水军主要存在于邮件系统中，同时网络水军产生的内容具有明显的可识别特征，因此这个阶段的网络水军识别大部分是居于内容特征进行的。然而，随着网络环境的复杂化和用户辨别能力的增强，网络水军逐渐衍生出多样的欺骗手段，基于内容特征的识别方法对这类水军不再有效。社交网络服务的兴起，使丰富的用户信息在网络上不断累积，这也同时为水军提供了庞大的目标平台，刺激了水军的大量加入。网络水军的识别工作根据研究方法的不同，又出现了基于行为特征、基于网络特征和基于影响力等方面。

Web2.0 时代的社交网络中的网络水军检测，较早期网络水军检测的差异很大，表 1.1 列出了主要的几个方面。互联网环境的复杂化和社交网络用户的增加，使新型水军行为变得更加隐秘，且日益趋向于正常用户，使得传统的网络水军识别方法已无法满足社交网络中网络水军识别的需求。同时，仅从单个特征（内容特征、网络特征、行为特征、影响力等）入手进行网络水军检测，无法全面地分析社交网络中的网络水军特征。

表 1.1 社交网络水军识别和早期网络水军识别的差异

	早期网络水军识别	社交网络水军识别
水军的目标	电子邮件用户	社交网络用户
水军的规模	正常	极其庞大
水军行为的复杂性	不复杂	极为复杂接近正常用户
识别难度	一般	较大

所以，本文提出了一种基于结构与内容的社交网络水军团体识别方法。首先，以垃圾信息出现后的时间区间内传播过信息的用户作为种子节点，构建转发关系网络；然后，在转发关系网络上，挖掘此网络中传播信息能力强的结构（重叠社区结构），并结合此结构中节点的内容特征，检测出嫌疑网络水军团体；最后分析统计疑似网络水军团体中，多次传播过垃圾信息的成员，得到最终的网络水军团体。

1.2 网络水军识别现状研究

网络水军检测最初起源于邮件服务系统，当时的网络水军通过制造具有明显商业特点的垃圾邮件引起用户注意，引导用户点击商业广告站点。随着互联网的快速发展，社交网络中的网络水军识别已成为主流，陆续产生了大量的识别方法：

Almeida、Yamakami 等人[2]基于贝叶斯分类的方法，分析垃圾邮件的内容，关注垃圾邮件本身的特征。该方法效果较好，对垃圾邮件识别度高。

刘鸿宇、赵妍妍等人[3]对网络评论进行了对象抽取和倾向性判断，从而发现区别于正常用户的、由网络水军发布的虚假评论。

Niu、Chen 等人[4]从论坛站点、用户浏览和论坛水军等几个角度分析发现，网络论坛水军的主要目的是提高其发布垃圾内容的搜索引擎排名，并基于内容特征识别水军制造的垃圾内容。

以上这些方法都属于基于内容特征进行的网络水军识别，通过挖掘网络水军所发布垃圾信息内容与正常用户所发布信息内容的不同，识别网络水军。但随着网络环境的复杂化、用户对水军警惕性提高和对垃圾内容的反感增强，导致水军策略不断变化。这种方式的水军识别，已不能有效作用于新型网络水军。

因此，大量学者开始了对基于内容特征的水军识别方法之外的探索，通过对用户的行为特征、所处网络的特征和用户影响力等方面，进行网络水军识别。具体研究如下：

Sawaya、Kubota 等人[5]首次发现移动服务商骨干网络中的网络水军，有严格的时间序列特征和发送模式：突发性、周期性、持续性，并利用这些特征对其进行聚类分析。

Lim、Nguyen 等人[6]通过对 Amazon 的网络水军行为分析发现，为了减轻工作量获得最

大的利益，他们经常会复制已有的评论。可以通过挖掘内容相似的评论，识别网络水军。

Lee、Eoff 等人[7]利用“诱捕器”收集 Twitter 上的网络水军行为数据，通过对这些数据的分析，发现了网络水军在粉丝关系和用户链路上的特征，提高了水军的识别准确率。

上述基于行为特征的网络水军识别研究，通过分析已识别的网络水军的行为，定义其特征，再采用传统监督分类方法，识别其他用户是否为网络水军。但随着网络水军行为逐渐趋向正常用户，仅从用户行为特征这单一方面入手，进行社交网络中的水军识别已越来越困难。

在互联网中，用户的交互行为使他们逐渐形成一个关系网络。与普通用户不同，水军具有独特的网络特征，例如：当网络水军大规模制造垃圾邮件时，用户之间的关系会发生改变[8]，通过对此类征的识别来检测网络水军。

Bouguessa 等人[9]提出了一种无监督的自动化检测社交网络中水军的方法。作者首先通过分析邮件系统中用户邮件记录中的社交网络，在得到其中每个用户的合法分数后，采用混合 β 分布对合法性进行建模。

Murmann 等人[10]通过 Twitter 中用户的好友以及粉丝的特征，通过特征矩阵的方式，得到用户的信任度，判断该用户是否为网络水军。

Las-Casas、Guedes 等人[11]在网络环境层级对网络水军进行检测。他们通过网络服务商的数据集，发现水军为减轻工作负担获得最大利益，会在一段时间内集中发布大量垃圾信息，造成网络负载的突然增加。

由于网络水军在网络级别很难隐藏自己的行为，所以基于网络特征的水军识别具有很好的效果。同时，利用图论的方法并综合水军独特的特征不仅能够识别个体网络水军，也能发现网络水军团体，从而在源头上遏制网络水军的行为。但采用的不同的网络特征结构的识别效率和准确率也不尽相同，如何找到一个合理的网络特征结构是目前的一个难点。

另外，对网络水军识别的研究，还可以从节点影响力分析这个角度进行网络水军检测识别[12-15]，计算节点影响力排名，认为影响力高的节点，更有可能是社交网络中的网络水军。通过控制和监控影响力高的节点，来削弱网络水军的消息扩散能力。但把高影响力用户等同于网络水军看待，不符合现实情况。

从以上的网络水军识别研究中可以看出两个方面的问题：首先，根据单个特征的网络水军识别不能全面评价一个真实的网络水军，因此造成识别准确率不高；其次，现有的网络水军识别工作，都是以检测单个网络水军为主，这种“抓单”检测效率不高，也不符合现实情况。社交网络中的水军通常都是有组织的，通过对网络水军团体进行识别可以从源头遏制水军的发展。由于基于网络结构特征的水军检测，利用图结构理论可以初步检测出网络水军团

体，所以本文借鉴这个思想探索对社交网络中网络水军团体进行识别的方法。

综合上述内容，本文综合考虑水军的结构特征和所发信息的内容特征，避免单个特征识别水军的不足；并以网络水军团体为目标，直接对社交网络中的水军团体进行识别。

1.3 课题研究内容

随着社交网络在世界范围内的流行，人们花在微博、论坛、人人网上的时间也越来越多。网络水军的出现，导致网络环境遭到污染、网络秩序混乱，阻碍了网络的健康发展。因此，网络水军识别研究得到了广泛关注和参与。然而，目前的社交网络水军识别研究存在着大量缺陷，准确性和实用性不高。本课题的主要研究内容如下：

（1）建立社交网络中的转发关系网络

因为，网络水军只存在于垃圾信息出现后的一段时间内，发送过信息的用户群中。首先，从垃圾信息（比如谣言）入手，搜集相应时间内发布消息（微博、微信、论坛帖等）的用户节点信息；然后，寻找与这些用户节点（种子节点）有过转发记录的其他用户节点；最后，建立由这些社交网络用户节点组成的转发关系网络。

（2）初步识别社交网络中的水军团体

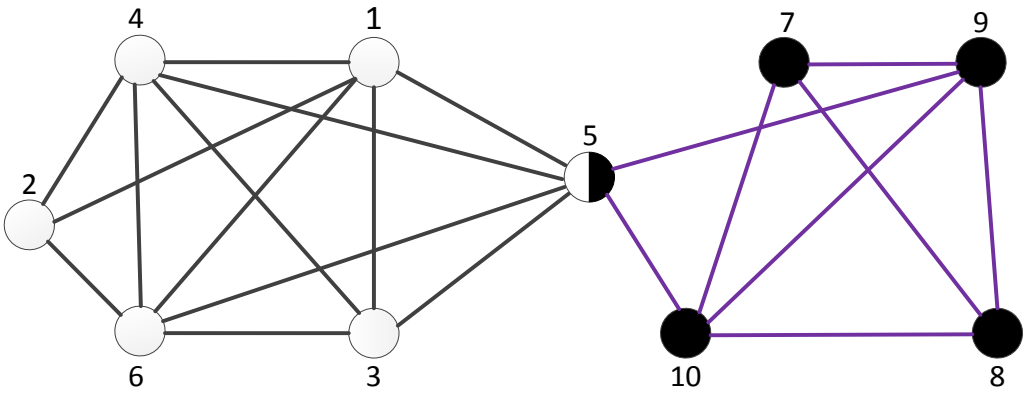


图 1.1 重叠社区结构

社交网络用户利用网络拓扑结构进行信息传播和影响力的放大，而网络水军所在网络的结构特殊性，使得他们传播信息的能力很强。通过文献[16]对节点传播影响力的分析可知，某节点的传播能力与和其直接相连的社区的数目有关，越多则传播速度越快、传播范围越广。进一步分析，若某节点隶属于多个社区，则它的传播能力应该更强，因为它不仅连接了多个社区，还直接与这些社区里的成员相连接，重叠社区结构（图 1.1 中的 5 号节点）就是具有这样特性的结构。通过挖掘转发关系网络中的重叠社区结构，得到传播信息能力强的结构，此结构中很有可能包含社交网络水军。

（3）基于结构与内容的社交网络水军团体识别

首先，采用经典的语义主题生成模型 LDA，对垃圾信息和重叠社区内的用户节点所发布的内容信息进行内容特征提取；如果重叠社区内的用户所发布内容与垃圾信息相似，则将其归入疑似网络水军团体；如果疑似网络水军团体中的用户所发布内容与多条其它的垃圾信息相似，则认为它是最终的网络水军团体成员。通过实验分析：本文提出的方法比以往的社交网络水军识别方法效果更好。

1.4 论文的组织结构

基于结构与内容的社交网络水军团体识别共分为六章，主要研究内容包括以下几个方面：

第一章：“绪论”，重点介绍了本课题的立题依据、课题的研究方案和研究的内容。

第二章：介绍了社交网络的特性，并对网络水军识别工作进行了相关研究，提出本文的研究问题。

第三章：基于网络结构特征的初步水军团体识别。本章中，首先分析了社交网络中的结构特征度量方法，分析选择基于重叠社区结构识别网络水军的原因；接着简单介绍了已有的重叠社区发现算法；最后提出改进的重叠社区发现算法。

第四章：计算重叠社区中用户与垃圾信息的相似度，以判断是否发送过垃圾信息。若发送过垃圾信息，则将其加入疑似社交网络水军团体。如果重叠社区内的用户节点多次发送过垃圾信息，则认为它是最终的社交网络水军团体成员。

第五章：实验设计和结果分析，并根据实验评价标准分析了实验结果。

第六章：总结与展望，对本论文进行全面总结，并指出需要进一步研究的问题和对所研究的课题进行展望。

1.5 本章小结

本章首先介绍了本文的研究背景，并对现有网络水军识别研究的现状进行了总结，然后叙述了本课题的研究方案和内容以及论文的组织结构。

第二章 相关研究

本章主要对社交网络特性、网络水军识别的相关研究进行了介绍。

2.1 社交网络概述

社交网络（Social Networks，SNS）[12]是指社会个体之间通过社会关系连结而成的复杂网络体系，它由社会中的个体以及个体间的关系组成。如图 2.1 是一个社交网络示意图，它向我们展示了 Facebook 用户和非 Facebook 用户通过各种社交关系被连结在一起，形成一个社交网络。通常我们将社交网络抽象为$G(V,E)$，其中V 代表社交网络中的用户节点集合，E 代表用户关系集合。以图 2.1 为例来说明，V 即是图中的小人集合，E 代表图中的线集合。社交网络中的边可以是无向的，例如 Facebook 中的朋友关系；也可以有向的，例如微博中的关注关系；还可以是有权值的，例如作家合著网中两个节点之间的边权重，代表两个作者之间合著的作品数目。

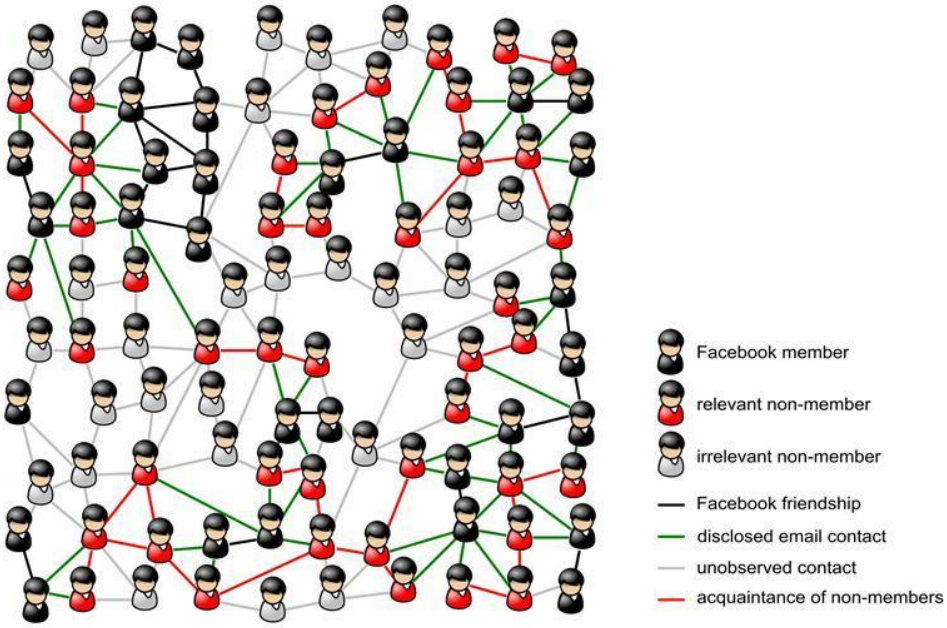


图 2.1 社交网络示意图

为了理解社交网络中水军的网络结构特征与传播信息能力之间的关系，就需要对真实的社交网络特性有比较透彻的了解。其中最著名的要属斯坦利.米格那姆于 1967 年提出的六度分隔理论（Six Degrees of Separation），也称为小世界理论。该理论认为：任何两个陌生人，通过他们认识的其他人，必然可以产生一定联系。更具体的说法是，两个陌生人之间的间隔

不超过六个人。其示意图如下所示：

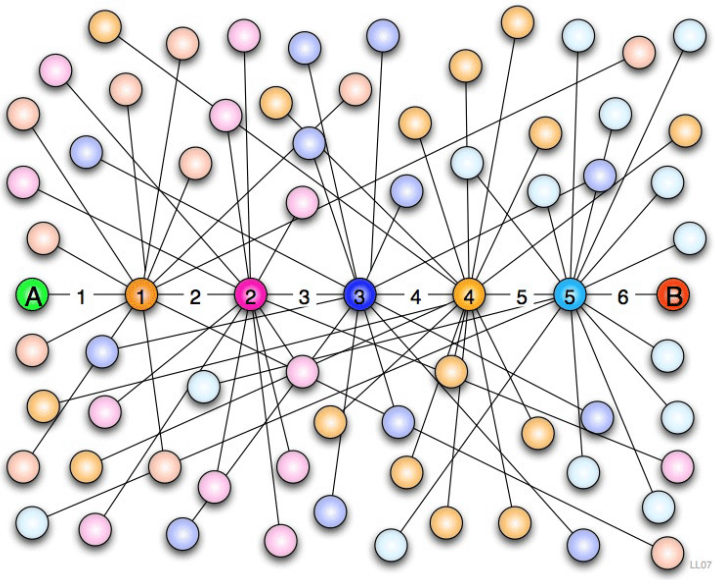


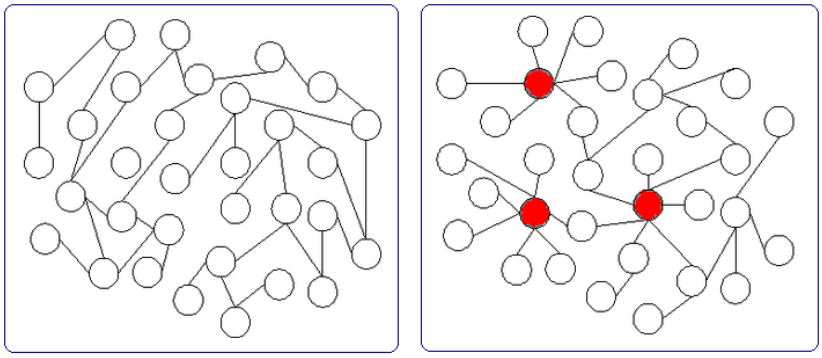
图 2.2 六度分隔理论示意图

在此基础之上，1998 年斯蒂文.斯特罗加茨和邓肯.瓦茨引进小世界网络，提出一种新的网络模型 WS 模型，它的主要特征为高聚集系数和低平均路径长度。其中聚集系数指的是，与一个节点相连的节点中相互连接的点对数与总点对数比值，可以用来描述“抱团”现象，也就是“朋友之间的相互认识的程度”。实际上，高聚集系数保证了较低的平均路径长度。

除了小世界理论外，社交网络还存在无尺度（无标度）特性。它指的是网络的度分布（随机从网络中抽取一个节点，与这个节点相连的节点数 d 的概率分布）满足幂律分布。也就是说 $d=\mathrm{k}$ 的概率正比于 k 的某个幂次（通常为负）：

$$P(d=k)\propto k^{-\alpha} \tag{2.1}$$

与随机网络的度分布（正态分布）不同，无尺度网络的度分布呈集散分布：大部分的节点只有比较少连接，少数节点却拥有大量连接，如图 2.3 所示。



(a) 随机网络 (b) 无尺度网络

图 2.3 随机网络与无尺度网络的对比

通过以上对社交网络特性的分析，我们可以看出：社交网络的小世界现象以及它的级联

传播方式[17]具有很强的催化放大作用，极大地加速了信息的传播和演化[15]，因此社交网络已成为最有力的信息传播媒体。社交网络在方便我们获得信息的同时，也给网络水军制造的垃圾信息的传播创造了优越的条件。社交网络的无尺度特性，显示出在社交网络中，存在着一些有“影响力”的节点。

2.2 网络水军识别研究

网络水军是指那些为达到如扰乱网络环境、影响网络民意等不正当目的，通过操纵软件机器人或水军账号，在互联网中制造、传播垃圾信息的总称[1]。Web2.0 时代，网络环境越来越复杂同时水军危害的增加，对其识别难度也逐渐加大。为了有效及时地发现社交网络中的网络水军团体，现有研究工作包括基于内容特征的方法、基于行为特征的方法 、基于网络特征的方法、基于影响力的方法以及基于综合特征的方法。本节将根据该分类对各个网络水军识别方法进行研究，并讨论各自的优缺点。

2.2.1 基于内容特征的方法

由于在早期的网络环境中，水军产生的垃圾内容具有很明显的可识别特征（比如，垃圾邮件和商业广告），所以在早期的水军识别工作中，主要使用了基于内容特征分析的方法[18,19]。该方法通过使用自然语言的处理手段，识别网络水军产生内容的显著特征（比如，URL 的模式特征和特殊的关键词），挖掘网络水军。对含有观点的文本处理，包括文本的分类[20]、倾向性分析[21]等方面。基于垃圾邮件内容的识别可以有效地检测垃圾邮件和邮件水军。但随着用户的警惕性提高以及邮件系统加入黑名单机制，以往简单通过制造大量垃圾邮件引诱用户点击商业广告的策略已发生很大改变，邮件水军制造垃圾信息不再具有很明显的可识别特征，通过内容特征的识别方法已不再有效。

Web2.0 时代，基于内容特征的网络水军识别工作主要在网络论坛领域。Chen 等人[22]着重研究了论坛水军所发的第一个帖子的特征，发现水军的第一个论坛帖：为了吸引正常用户，通常文字详细且配有图片；更加关注可以提升自己论坛等级的那些主题；并且喜欢在工作时间发帖。这些发现，为之后的论坛水军识别工作提供了参考。

总之，仅基于内容特征的网络水军检测，已经无法满足现今的社交网络中的水军识别要求，甚至连邮件网络系统中的水军识别要求都无法满足。目前，基于内容特征的网络水军识别，主要用于与其他特征结合，进行水军综合的识别。

2.2.2 基于行为特征的方法

用户的行为有时能极大地反映出一些用户的信息。比如，水军的突发性行为与正常的社交网络用户有着显著的区别。基于行为特征的网络水军识别研究，通过分析网络水军的行为，定义其特征，再采用相应的分类方法，识别其他用户是否为网络水军。

在传统的基于行为特征的网络水军识别中，文献[23]在分析了大量邮件网络水军的行为基础上，采用 URL 的利用率和邮件的发送时间这两大行为特征构建邮件分类器。此文献发现邮件网络水军通常利用一般人群的非工作时间段发送垃圾邮件，因为这样可以避免网络带宽的限制。Phithakkitnukoon 等人[24]从水军机器人制造的垃圾邮件的内容篇幅、类型和发送频率等特征入手，识别邮件机器人。并将大量邮件水军中相似度高的一群水军，聚类成邮件水军团体，通过分析此水军团体的共有行为，发现高度互助的水军团体。

在社交网络中，通过对网络水军行为特征的分析并进行建模[25-29]，来判断未知用户是否为网络水军，其中 Rodrigues 等人[25]分析了视频网站 Youtube 中的网络水军行为数据，并通过人工标记的方法，来标明数据集中的网络水军。之后分析并定义网络水军的行为特征，最后利用特征选择算法选取网络水军行为特征分辨力较强的，判断其他用户的身份。此方法，为社交网络水军识别中的标准方法。但由于数据集过于庞大，用人工先判别出一些网络水军再进行分析，效率不高。同时，网络水军的行为策略在不断发生变化[26]，给人工判别网络水军带来了很大困难。Stringhini、Kruegel 等人[27]在几个主流社交网络平台中，使用“诱捕器”收集实验数据，分析这些平台的网络水军行为特征。但其识别的方法过于简单，无法挖掘出日益趋向于正常用户的网络水军。文献[28]在 Twitter 中提出基于邻居节点特征发现网络水军的方法，并详细分析了 Twitter 中网络水军的隐藏策略。但从网络水军的邻居节点入手，会将问题复杂化，因为网络水军通常具有的庞大邻居节点。

传统的基于用户行为特征的邮件网络水军识别，已无法适应复杂的社交网络环境中网络水军识别的要求。但其中的思想，仍然值得借鉴。比如，上面提到的文献[24]，将相似度高的邮件水军聚类成水军团体的做法。基于这个思想，本文中，我们将具有相同网络结构的用户聚为一类，挖掘网络水军团体。在 Web2.0 时代的社交网络中，基于用户行为特征的网络水军检测已取得了大量成果，但随着网络水军行为的复杂多样化，仅仅从用户的行为特征入手，已很难准确发现社交网络中的网络水军。

2.2.3 基于网络特征的方法

基于网络特征的方法，主要分成两个方面：一个是基于网络结构特征的方法，另一个是基于网络环境层级特征的方法。

在传统基于网络结构特征进行网络水军识别的工作中，研究人员根据邮件往来记录构建用户关系网络。文献[30]通过邮件水军形成的网络拓扑结构特征，来提高 URL 过滤工具的准确性，并发现邮件网络中的水军可以很容易获得大量的网络资源来发布垃圾邮件。

社交网络中，由于网络结构具有一定稳定性，其拓扑特征不易被用户行为所影响[31]，网络水军不能掩饰他们在网络结构上的特征，所以基于网络结构特征的水军识别研究得到了广大研究人员的重视。文献[10]认为，合法用户很少是网络水军的粉丝，因此水军的粉丝较少，利用用户的粉丝数计算信任值。但这种做法过于简单，同时现实生活中网络水军往往具有大量粉丝。例如，微博网络上的网络水军，他们会先发布一段时间的吸引人注意的非垃圾内容（比如笑话、健康小知识等）来吸引其他用户关注，成为他们的粉丝。待粉丝达到一定数量后，再开始发布垃圾信息。Song、Lee 等人[31]解决了根据账户特征发现网络水军时，水军篡改账户信息和目标账户信息难以收集的问题。文中认为水军和目标用户间的距离为 3 至 4 个节点，同时出度越大，流通性越好，越有可能是网络水军。但此方法要在全网中进行搜索，计算量太大；并且文中只对距离为 3 至 4 个节点的用户对进行分析，准确率不高。

虽然隐蔽的网络水军对用户展现出的特征越来越趋向正常用户，但是在网络环境层级其异常行为无法被修饰[11,32-33]。例如，文献[11]基于网络水军产生时的网络环境特征进行水军识别。他们发现网络水军活动时，网络负载会突然增加，流量也会在某些链路中集中。在网络层级进行水军识别的准确率较高，但此类方法通常需要相应的网络服务提供商（ISP）数据集，因此其可推广率较低。

基于网络特征的方法目前已经成为水军识别研究的主要途径之一，其中基于网络结构特征的网络水军检测，不仅可以解决其他方法无法处理的日益趋于合法用户的水军问题；还可以发现网络水军形成的水军团体，从而遏制水军的发展。但在识别水军过程中，采用不同的网络结构特征，其识别效率也大为不同。同时，仅从这一单一特征识别社交网络中的网络水军，可能会出现虽然某些用户节点符合所采用的网络结构特征，但本身却没有传播过垃圾信息却被判断为是网络水军的情况。

2.2.4 基于影响力的方法

由 2.1 节中介绍的社交网络的无尺度特性可知，存在少数节点拥有大量连接，它们对与之相连的节点具有较强的影响力。一方面它们吸引更多节点连向它们；另一方面它们传播的信息有较大的接收群体。基于影响力分析的网络水军识别，通常先寻找社交网络中的意见领袖，即在相应环境下对其他人产生影响的个体，也就是有影响力的人[34]，再判断他是否为网络水军[12-15]。

其中文献[13]基于 PageRank 的思想，对网络中的用户进行评级。认为用户的评级越高，则传播虚假信息过程中起到的作用越大。通过对高评级的用户进行控制，来缩小虚假信息传播的覆盖面。但即使这个方法可以准确检测出有影响力的节点，也不能说明这些节点就是网络水军，它们的传播信息能力强。例如某明星，由于“明星效应”，他具有很大的影响力，但如果他长时间都不传播一条信息，那他的信息传播能力并不强，也就不可能是一个网络水军。文献[15]把对特定成员影响力构成，起到重要作用的网络群体称为支撑结构。通过寻找支撑结构，来间接检测网络水军。但在寻找支撑结构的过程中，对节点影响力的度量计算复杂，不利于实际运用。而且，即使找到了某些节点的支撑结构，若此支撑结构与其支撑的节点与其他社区孤立，它也无法将信息传播得更远。

综上所述，基于影响力分析的网络水军识别方法中，将有影响力的人默认为网络水军并不合理。但前人的研究也给我们带来了新的思路，寻找更本质的“影响力”特征——传播信息能力，本文通过它来初步识别网络水军。

2.2.5 基于综合特征的方法

由上分析可知，基于单一类型的网络水军识别工作，无法全面度量网络水军的特征，进而使其检测准确性存在瓶颈。在此基础上，基于综合特征的识别方法较基于单一特征的识别方法，在准确率上有较高的提升。

在邮件领域中，结合网络水军的显著特征和其所发送邮件的内容特征，可以进一步提高邮件水军的检测准确率。例如，文献[35]分析垃圾邮件的内容时发现，这些邮件中含有相同的 URL 并且由同一个团体的水军制造传播。从该特点出发，寻找属于多个水军团体的网络水军，该水军危害极高。

基于综合特征的社交网络水军识别方法捕捉水军在网络中的行为、关系、所发内容等特征。文献[36]首先根据图论初步寻找 Twitter 中的网络水军，这些网络水军通常花费大量时间

等待目标用户关注自己或关注目标用户，再结合用户的影响力特征进一步提高识别的准确率。文献[37]结合 Twitter 用户所发 Tweet 和其行为特征判断其是否为网络水军。但该方法的实验数据仅使用 3 个最热门主题的参与用户，对水军的识别准确率不到 70%。Chen、Wu 等人[38]基于网络水军的行为特征（发布帖子的数量、论坛活跃程度、回复帖子间隔等），综合其制造的垃圾内容的特点，发现国内门户网站论坛中潜藏的网络水军。

基于综合特征的方法，可以更全面的分析网络水军的特征，较单一特征的识别方法准确性更高。但选取哪些特征，才能更有效的识别社交网络中的网络水军，是本文要解决的关键问题。

2.2.6 目前的研究难点和热点

由于社交网络的飞速发展和大数据处理能力的加强，网络水军识别研究取得了一定成果，然而这个领域中，可以深入研究并可能取得一定成果的领域还有不少。图 2.4 所示，为其中主要的三个方面。

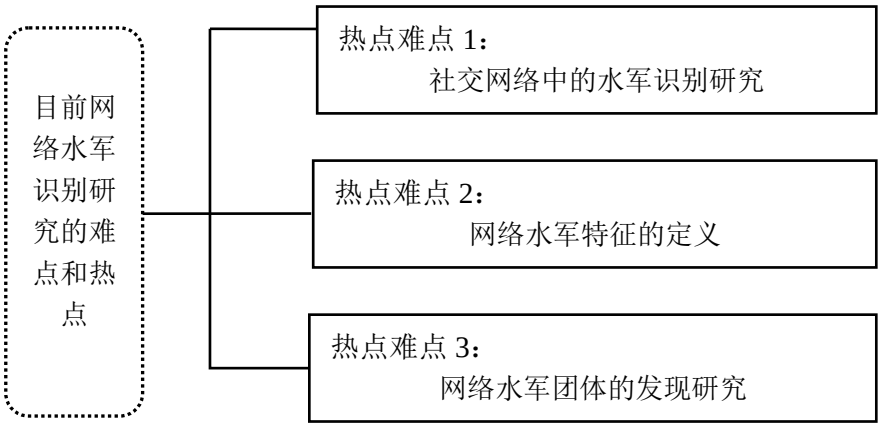


图 2.4 水军识别的热点和难点

（1）社交网络的水军识别研究

因为社交网络中用户群体庞大，潜藏着巨大的利益，所以社交网络领域已成为近年来网络水军活动最频繁的领域。通过对社交网络中水军的识别和防范，能够营造良好的网络环境，促进社交网络朝着更好更快的方向发展。

（2）网络水军特征的定义

准确定义网络水军的特征，是识别网络水军的关键，它将影响网络水军识别的最终表现。以往的工作通常从网络水军所发内容的特征、网络水军行为特征以及水军所在网络特征等方面入手进行检测。但不同目标领域的水军的行为不同，会表现出不同的特征。定义分辨率高

的水军特征，是网络水军识别研究的难度之一。

（3）网络水军团体的发现研究

社交网络中的网络水军要想对整个网络产生影响，必须形成一个水军团体，团体的影响力非单个水军能比拟。因此，网络水军团体的存在，才是对互联网最大的威胁。如何从庞大的社交网络中识别出这些具有极高危害的网络水军团体，将是网络水军识别研究中的重点[39]。

2.3 网络水军识别研究总结

从目前的研究现状来看，针对社交网络中的水军识别研究已取得了一定的进展，但还存在一些不足，原因是多方面的：

首先，目前的识别研究主要针对单个水军进行，这种一个个检测水军个体的方式效率不高，也不符合现实情况。社交网络中的水军通常都是抱团行动的，网络水军团体的“破坏力”不是单个水军可以比拟的。若能识别社交网络中的水军团体，我们就能更高效地识别出网络中的水军，减小他们对社交网络的影响力。

其次，基于单一特征的网络水军检测准确率不高。例如：在检测网络水军时，如果只考虑行为特征，可能出现检测出的用户虽然符合网络水军的行为特征，但其并没有发布过垃圾信息；其次，即使检测出用户符合水军的内容特征，但如果只是一次偶然评价，并不能确认他的水军身份。社交网络水军识别工作，应该避免把误传播垃圾信息的用户当成网络水军的情况。

针对以上问题，本课题从两方面进行研究。首先，通过对基于网络结构特征方法的分析，我们发现利用图模型理论可以发现网络水军中的水军团体。在本文中，我们通过分析社交网络中的转发关系网络，挖掘其中的重叠社区结构，来初步识别出可能的网络水军团体；然后，结合重叠社区结构与节点内容特征，对社交网络中的网络水军团体进行识别，避免单一特征识别时，无法全面分析网络水军的问题。

本文提出结合结构与内容特征，来对社交网络中的水军团体进行识别，先寻找有能力使传播的信息造成影响的重叠社区结构（“有能力造成危害”），再检测是否多次传播过垃圾信息（“有过多次水军行为”）。

2.4 本章小结

本章主要介绍了网络水军识别的相关研究、社交网络简介，并总结了以往水军识别工作的缺陷，提出本文的研究问题。主要包括以下几点：

（1）网络水军识别的背景知识介绍以及对各类网络水军识别方法进行了基本研究。

（2）社交网络概念的基本介绍和特性的研究。

（3）总结前人的工作，提出本文的研究问题。

第三章 基于网络结构特征的水军识别

本章对社交网络水军的网络结构特征进行了分析，找出了社交网络水军的网络结构特征——重叠社区结构。并通过构建社交网络中的转发关系网络，对庞大的社交网络进行预处理，提高识别效率。挖掘此网络中的重叠社区结构，从而初步发现社交网络中的水军团体，为下一章的工作准备。

3.1 社交网络水军的网络结构特征分析

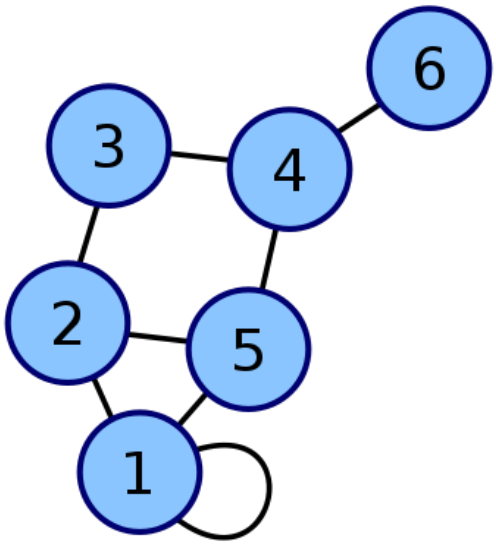
社交网络拓扑结构是用户在社交活动中遗留的“痕迹”，社交网络中的用户无法去掩盖。从网络拓扑结构的角度，可以体现用户的信息传播能力和影响力，并且社交网络结构的获取较为容易。网络结构中的节点和边分别代表用户和他们之间建立的关系，在分析传播信息能力和影响力的时候，它们都起着关键作用。

社交网络用户利用网络拓扑结构进行信息传播和影响力的放大，而网络水军所在网络的结构特殊性，使得他们传播信息的能力很强。本文所提到的信息传播能力，指的是一个节点有能力影响周围的用户并通过周围用户将信息传播出去，并且有多次传播信息的记录。我们通过分析社交网络结构特征，找出网络结构与影响力大小的联系，间接分析网络结构与信息传播能力大小的联系。

下文中，社交网络的结构用图$G(V,E)$表示，其中节点数$n=|V|$、v_i代表节点 i、e_{ij}代表节点 i 和节点 j 之间的边（连接）；$A_{n\times n}$代表图的邻接矩阵，a_{ij}是其中的元素，图G的邻接矩阵具有如下性质：

$$a_{ij}=\begin{cases}1 & \text{若}(v_i,v_j)\text{是}E(G)\text{中的边}\\ 0 & \text{若}(v_i,v_j)\text{不是}E(G)\text{中的边}\end{cases} \quad (3.1)$$

对于无向图（本文中涉及到的图结构，均为无向图）来说，它的邻接矩阵是对称的。图 3.1 为一个由 6 个节点，8 条边组成的网络图，它的邻接矩阵为A。



$$A=\begin{pmatrix}1&1&0&0&1&0\\1&0&1&0&1&0\\0&1&0&1&0&0\\0&0&1&0&1&1\\1&1&0&1&0&0\\0&0&0&1&0&0\end{pmatrix}$$

图 3.1 邻接矩阵例图

3.1.1 对节点和边的度量

（1）节点的中心度[40]

社交网络中节点的中心度主要用来度量网络中节点对其邻居的平均影响力，通常利用下面的公式来计算：

$$C^{DEG}(v_i)=\frac{\deg(v_i)}{n-1} \tag{3.1}$$

（2）节点的出度和入度

社交网络中与用户影响力相关的指标（如粉丝数、点赞数、回帖数等），可以用节点的出度（3.2）与入度（3.3）来衡量。

$$\deg^{in}(v_i)=\sum_j a_{j,i} \tag{3.2}$$

$$\deg^{out}(v_i)=\sum_j a_{i,j} \tag{3.3}$$

（3）紧密中心度[41]

节点的紧密中心度主要来计算当前节点对其他节点的间接影响力，或者该节点发送的信息传播到其他节点的距离。值越大，表明当前用户和其他用户的距离越近，影响其他用户的速度越快。紧密中心度通过下面的公式计算：

$$C^{CLO}(v_i)=\frac{1}{\sum_{v_j\in V\setminus v_i} g_{ij}'} \quad (3.4)$$

（4）介数中心度[42]

这个值用于度量节点在网络结构中位置的重要性，表明信息经过该节点的信息量。值越大该节点在信息传播中的影响力越大。如下式：

$$\Pr(v_i)=\frac{1-d}{n}+d\sum_{v_j\in L^{in}(v_i)}\frac{\Pr(v_j)}{\left|L^{out}(v_j)\right|} \quad (3.5)$$

（5）基于随机游走特征的度量

常用的基于随机游走特征度量影响力的指标有：特征向量中心度[43]、Katz 中心度[44]等。其中特征向量中心度值就是当前节点可达其他节点的权重线性和，该节点的影响力随其相连的其他节点的影响力增大而增大。Katz 中心度通过游走路径来计算两个节点之间的影响力，游走路径上距离 v_i 越远的节点，对 v_i 节点的 Katz 中心度影响越小。这两种度量方法的计算公式见表 3.1。

表 3.1 基于随机游走特征的度量

度量方法	计算公式
特征向量中心度	$\lambda x_i=\sum_{j=1}^{n}a_{i,j}x_j, i=1,2,...,n$
Katz 中心度	$C^{Katz}(v_i)=\sum_{k=1}^{\infty}\sum_{j=1}^{n}\alpha^k\left(A^k\right)_{ij}$

（6）局部聚集系数

社交网络中的用户由于联系十分紧密，因此有较强的趋势形成社团，这种趋势可以用聚集系数来度量。它表示存在一个节点 v_i，它的任意两个邻居节点 v_j 和 v_k 之间产生联系的可能性。例如，C 与 D、E 是朋友关系，那么聚集系数可以计算 D 和 E 是朋友关系的概率。计算方法如 3.6 所示：

$$C^{CLU}(v_i)=\frac{\left|\{e_{jk}:v_j,v_k\in Ng_i,e_{jk}\in E\}\right|}{\left|Ng_i\right|\left(\left|Ng_i\right|-1\right)/2} \quad (3.6)$$

基于节点度的方法表达的意义明确、计算简单，可以表示节点与邻居节点之间的关联程度。但此类方法仅在度量用户的局部影响力时较为有效，无法很好地计算用户在全网范围的影响力。

（7）Jaccard 相似度与边介数

对边的影响力计算，就是对边的两个节点（用户）之间影响程度的计算。常用的有 Jaccard 相似度和边介数[45]。其中 Jaccard 相似度，就是下面 3.3 节 LINK 算法中介绍的（3.8）。边介数类似与节点的介数中心度，用于衡量边在网络中的重要程度，它统计网络中经过某边的最短路径的总数量，表明信息经过该边的信息量。边介数的计算方式如下：

$$E^{BET}\left(e_{ij}\right)=\sum_{s<t}\left|g_{st}^{ij}\right| \tag{3.7}$$

其中，$\left|g_{st}^{ij}\right|$代表节点 *s* 和节点 *t* 之间的最短路径同时经过边e_{ij}的个数。

综上所述，基于网络拓扑结构的影响力计算，不考虑社交网络上海量的交互信息，仅抽取网络结构对用户影响力进行分析、建模，具有模型简单、易于应用扩展的特点，在社交网络这类大规模网络中具有较大优势。

3.1.2 重叠社区结构的引入

从上一小节社交网络拓扑结构与影响力关联的分析可知，计算社交网络中的影响力都是通过遍历拓扑图中的节点或边来进行的。当网络节点数量足够多时，基于节点或边影响力算法的运行时间将趋近于无穷大；然而，社交网络的节点数量都是亿量级的，如 Twitter 的用户数为 2.84 亿、新浪微博用户数 2.12 亿、Facebook 用户数则高达 22 亿之多。利用上述方法对整个网络进行影响力计算，效率太低、几乎不可行。本文通过选取种子节点，深度遍历与之有过转发关系的节点，构建网络结构；这种做法大大减小了网络的规模，排除了大量的无关节点，下一小节将详细介绍这种做法。同时，本文注意到，即使前人对影响力的度量是科学有效的，影响力仍然不能直接作为判断网络水军的指标。因为，有些节点的影响力是天生的，例如影视明星，在新浪微博中，即使他们不发送任何一条微博（或者极少发送微博），他们仍然有很大的影响力。所以，本文提出信息传播能力的概念，通过信息传播能力判断网络水军，更加合理、准确，它既体现了节点的影响力，又体现了节点有较强的传播信息意愿。

社交网络中，在一条信息传播之前，我们规定网络中的节点只有两个状态：激活态和非激活态，并且每个节点的状态只能由非激活态变成激活态，而不能由激活态再变为非激活态，也就是说，我们的激活过程是不可逆的，只是单一的一个方向变化。两个节点之间可能不相邻，但是可以通过其他节点相互联系。当一个节点与越多的其他节点直接或者间接的相邻，则它传播的信息就会有更大的可能被更多的节点接受（激活），而继续向其他节点传播。通过文献[16]对节点传播影响力的分析可知，某节点的传播能力与和其直接相连的社区的数目有

关，越多则传播速度越快，且传播范围越广。进一步分析，若某节点隶属于多个社区，则它的传播能力应该更强，因为它不仅连接了多个社区，还直接与这些社区里的成员有直接的连接，重叠社区中的节点就具有这样特性。重叠社区结构作为多个社区的共有部分，可以将所要推广的信息传递给各个社区里的成员，由各个社区里的成员继续推广下去。在现实中，这种现象很常见，推销人员常常注册很多 QQ 账号，加入几百个 QQ 群，向每个 QQ 群发送推广信息，以达到他们推广信息的目的。通过挖掘转发关系网络中的重叠社区结构，初步识别社交网络中的水军团体。

本文采用重叠社区结构作为检测网络水军的网络结构特征，具有以下意义：

（1）重叠社区中的节点传播信息能力强的，它更符合网络水军的网络结构特征。以往的网络水军识别工作，很少利用重叠社区结构进行水军检测。

（2）避免了从节点和边入手进行传播信息能力分析低效率的情况，挖掘重叠社区结构，可以一次性识别出社交网络中传播信息能力强的节点集，为识别网络水军团体提供条件。

（3）从社交网络中的转发关系网络中进行重叠社区结构的挖掘，体现了重叠社区结构的形成，是基于节点间的转发信息历史。它说明重叠社区中的节点，不仅影响力大，且有较强的传播信息意愿，因此它们的传播信息能力强。

3.2 构建社交网络中的转发关系网络

以往的水军识别研究中，都是在整个网络中识别网络水军，这种做法在对社交网络中水军进行检测时，可能会有以下两点缺陷：

（1）由于社交网络存在庞大的数据，节点以及边的数量都是千万级别的，使得在整个网络上检测网络水军容易产生计算风暴，无法运用到实际的社交网络水军识别工作中。

（2）在以往的检测水军过程中，网络的构成方式通常多种多样，其中通过粉丝关注关系形成的网络，也有通过合著关系形成的合著网络，还有评论记录形成的网络。其中根据关注关系和合著关系形成的网络无法体现信息的传播情况；由评论记录形成的网络，是对其他节点所传播信息内容的评论，但这种评论由于不被其他节点所见，所以对信息的传播产生不了帮助，即对相邻节点产生不了激活作用（如图 3.2 和图 3.3，在新浪微博中，对某条微博进行评论，但别人却看不到这条评论）。只有可以产生激活效果的行为（转发、回复等），才能达到网络水军将制作的垃圾信息传播给更多人的目的。因此，在这些网络上通过挖掘出重叠社区结构识别水军，准确率较低。

图 3.2 社交网络中的评论行为

图 3.3 被评论的微博没有出现在评论者的空间中

综上所述，社交网络中的很多社交行为（合著、关注、评论等）并不能直接体现网络水军的活动。而转发行为，则可以帮助信息的传播；转发就是将别人的所发的帖子或微博等社交网络信息转发到自己的主页上，让自己的好友也可以看到（如图 3.4 所示，转发小米公司的微博）。由于网络水军要起到传播特定内容的目的，在平时就需要与其他用户保持一定的“联

系”，所以社交网络的拓扑结构在一定时间内是结构稳定的[1]，其中的重叠社区结构基本保持不变。由文献[46]中的数据分析可知，转发行为一般发生在 10 小时以内，因此本文的转发关系网络，利用垃圾信息出现时间点后 10 小时内，发布过信息的用户作为种子节点，这其中就包括了传播过垃圾信息的节点；根据社交网络中的以往这些种子节点的转发记录，通过深度遍历来构建社交网络中的转发关系网络，由于六度分隔理论，本文中的遍历深度不超过 3 层。垃圾信息出现的时间，就是网络水军们活动的时间，通过这种做法，既减少了数据的规模，避免在全网络进行识别工作；又排除了很多与此垃圾信息无关的数据，提高水军识别的效率和准确率。

图 3.4 社交网络中的转发行为

表 3.2 为构建社交网络中的转发关系网络的伪代码（种子节点集已提前从社交网络数据集中获取）：

表 3.2 转发关系网络构建的伪代码

输入：种子节点集 *SEED*（垃圾信息出现时间点后 10 小时内，传播过信息的用户）
输出：转发关系网络 $G_{retweeted}$
WHERE（节点集 *SEED* 非空）
IF（$\forall n \in SEED \&\& n$有过转发行为）
THEN 以 *n* 为起点，根据节点转发记录进行深度遍历（深度不超过 3 层，遍历过程中将

两有转发历史的两个节点连边）；

ELSE *SEED-n*；//将节点 *n* 从 *SEED* 中删除

SEED-n；

RETURN $G_{retweeted}$；

3.3 重叠社区结构发现算法

在得到社交网络中的转发关系网络之后，本节中将对以往的重叠社区发现算法进行简要的介绍，并利用改进后的算法在转发关系网络中挖掘重叠社区结构，初步识别社交网络中的水军团体。

3.3.1 重叠社区结构发现研究

传统的非重叠社区发现方法，通常将网络划分成几个互不相连的社区，每个节点只能属于一个的社区。其中，代表算法包括层次聚类算法[47-48]、谱聚类算法[49-50]、模块度优化算法[51-52]等。但这不符合大多数现实的网络结构，因为各个社区之间通常都是存在联系的、相互重叠且彼此交叉的。在现实情况中，有一部分节点不只隶属于一个社区中，它们同时属于多个社区。例如，在社交网络中，每个人根据爱好的不同可以属于多个不同的社区（如读书、旅游、电影等）；在作家协作网络里，一个作家很可能会与多个作家有合作。因此，社交网络中重叠社区结构发现，更加具有现实的意义。

重叠社区发现是目前社区发现研究中的一个热点，根据现有的算法中采用标准及对象的不同，可以分为两大类：以网络中的节点为研究的对象，通过对节点进行社区划分、社区聚类、社区共有节点发现等方法寻找重叠节点；以网络中的边为研究的对象，按照边划分社区，由于一个节点可能与多个边相连，但一条边只能隶属于一个社区，间接得出属于多个社区的节点，发现重叠社区结构。

（1）派系过滤算法

2005 年 Palla 等人首次提出发现重叠社区的派系过滤算法 CPM（Clique Percolation Method）[53]。此方法以网络中的节点作为研究的对象，认为社区是具有共享节点的全连通子图集合。这些全连通的子图称为 k 派系，是由 k 个节点构成的全连通图。其中，两个 k 派系相邻指的是两个 k 派系存在 $k-1$ 个共有节点；而两个 k 派系连通指的是一个 k 派系能通过几个相邻的 k 派系到达另一个 k 派系。CPM 算法社区划分的过程，就是寻找全部连通 k 派系组成的

极大子图的过程。

如果存在一些节点，它们同时隶属于多个k派系，但这些k派系不相邻（k派系之间公共的节点不足$k-1$）。那么这些节点所隶属的多个k派系就不连通，即这些k派系不属于同一个k派系社区，所以这些节点属于多个不同的社区，即为几个社区的重叠结构。例如图 3.5 的 3 派系网络（三条黑粗线组成的三角形表示网络中的 3 派系，两个实心点表示 3 派系社区中的几个社区的共有节点）。但此算法对派系间的相邻关系定义较严格，所以一般只适合于全连通子图较多的网络。

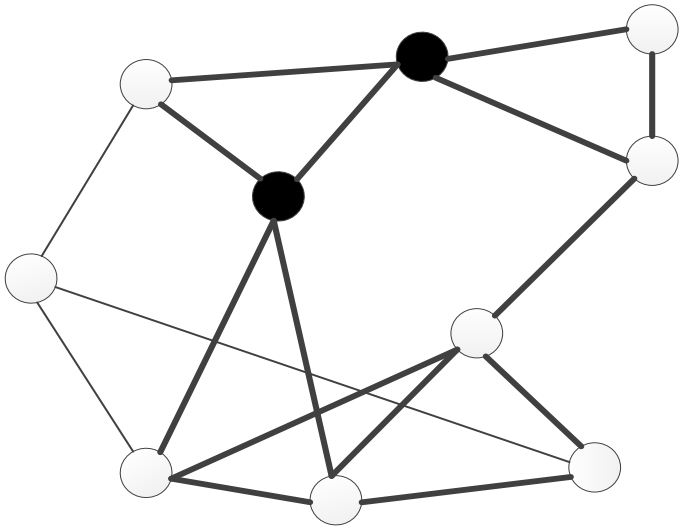


图 3.5 3 派系社区实例

（2）基于链接划分的算法

基于链接划分的算法[54-56]，首先对网络中的边进行划分，构造链接社区结构，并将链接社区结构转化为节点社区结构；最后，若连接一个节点的两条边分别属于不同的链接社区，则该节点为重叠节点。

Ahn、Bagrow 等人[54] 提出的基于层次聚类的链接划分算法 LINK，该算法中定义了网络中两条边的相似程度：给定一对具有共同节点u的边$e=(u,v)$和边$l=(u,w)$，它们之间的相似度为:

$$S(e,l)=\frac{|\Gamma(v)\cap\Gamma(w)|}{|\Gamma(v)\cup\Gamma(w)|} \tag{3.8}$$

上式中$\Gamma(v)$为节点v本身和它的所有邻居节点构成的点集，即$\Gamma(v)=\{x|(v,x)\in E\}\cup\{v\}$。

LINK 算法的过程是：首先，将各条边看作一个独立的链接社区，合并最相似的两个链接社区，直到预定义的目标函数达到最大为止；然后，将此结果转化为最终的社区节点集合，若存在两条边既隶属于不同链接社区又拥有同一个公共节点，则该节点为这两个社区的重叠

节点。如图 3.6 所示，节点 5 为社区 $\{1,2,3,4,5,6\}$ 和社区 $\{5,7,8,9,10\}$ 的重叠节点。

LINK 算法比较直观、时间效率高，但它将所有边都划到特定的链接社区中，可能会出现社区的“过度重叠”现象。在图 3.6 中，LINK 算法同样将边 $(1,7)$ 看作一个单独的链接社区，那么节点 1 和节点 7 就会被看作是重叠节点，这样的结果与真实情况不符。文献[55,57]对 LINK 算法进行了改进，其中 Kim、Jeong 等人[55]利用期望最大化的方法构造链接社区；Ball、Karrer 等人[57]把信息论的方法应用到链接聚类中。但他们只是提高了链接社区发现的准确率，不能有效阻止“过度重叠”的发生。

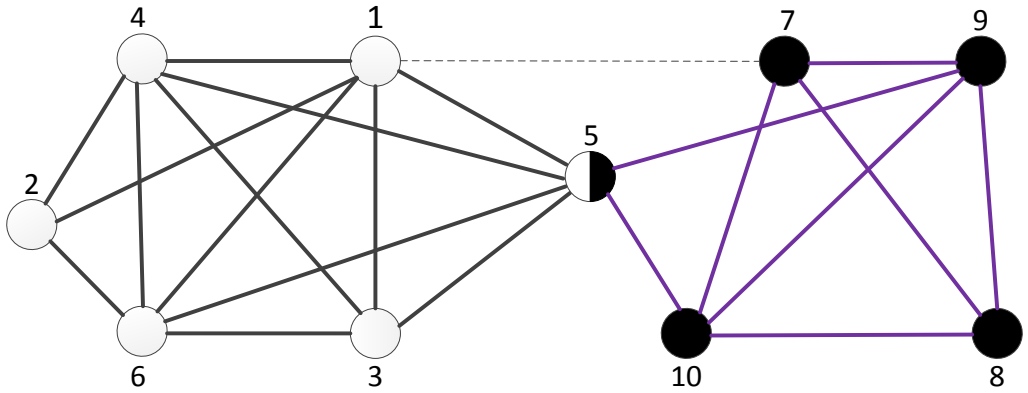


图 3.6 “孤立边”（1,7）

3.3.2 改进的重叠社区发现算法

由上一小节所述可知，基于链接划分的重叠社区发现方法中，LINK 算法比较直观、容易理解、时间效率高。但易出现“过度重叠”的情况，并不满足本文的水军识别的要求。因为节点 1 和节点 7 连接的节点较节点 5 少，因此它的传播能力比真正的重叠社区结构（5 号节点）弱，是网络水军的可能性小。通过定义核心边的方式可以排出“孤立边”（图 3.6 中的边（1,7））, 避免“过度重叠”现象的出现，提高网络水军识别准确率。

本文中改进的重叠社区发现算法的主要思想为：第一步，基于核心边的聚类算法将边集划分为若干互不相连的社区集，既可以发现高质量的链接社区 *LC*，又可以寻找核心边；第二步，将链接社区转化为节点社区 *C*，若存在两条边既隶属于不同链接社区又拥有同一个公共节点，则该节点自然成为这两个社区的重叠社区结构 *OS*（Overlapping Structure）中的一员。相关定义如下：

定义 3.1 边$e=(u,v)$的邻居边$N(e)$，指的是分别与节点u和v相连且不包括e的边集合，即：

$$N(e)=\{l=(v,i)\in E\mid i\in N(v)\}\cup\{l=(u,j)\in E\mid i\in N(u)\}-\{e\} \tag{3.9}$$

其中$N(v)$、$N(u)$代表节点v、节点u的邻居节点集合。

定义 3.2 边e的ε领域$N_{\varepsilon}(e)$是指在边e的邻居$N(e)$中，与边e的相似度大于或者等于ε的边集合，即：

$$N_{\varepsilon}(e)=\{l\in N(e)\mid s(l,e)\geq\varepsilon\} \tag{3.10}$$

定义 3.3 若一条边e的ε领域中的边数$|N_{\varepsilon}(e)|\geq\mu$，则称$e$为核心边。

定义 3.4 若边e为一个核心边且边l属于e的ε领域时，称边l直接密度可达于e，即：

$$Dir\,\mathrm{Re}ach(e,l)\Leftrightarrow e.Core=true\wedge l\in N_{\varepsilon}(e) \tag{3.11}$$

根据定义 3.1~3.4，算法的第一步是在本文的转发关系网络中，找出所有的链接社区；若某条边不属于任何社区，则它为孤立边，被自然排出在外。第二步，再根据下面的定义 3.5 和定义 3.6，将链接社区集合转化为相应节点社区集合，最终找出重叠社区结构。

定义 3.5 节点社区 C 即为链接社区 LC 中所对应的所有边的连接节点：

$$\forall e=(u,v)\in LC\Rightarrow u\in C\wedge v\in C \tag{3.12}$$

定义 3.6 如果存在两条边被同一节点 v 连接，同时这两条边分别属于不同的链接社区 LC 时，节点 v 即本文所寻找的重叠社区结构 OS 中的一员：

$$\begin{aligned}&OS(v)\Leftrightarrow\exists u_1,u_2,LC_1,LC_2(u_1\neq u_2\wedge LC_1\neq LC_2):\\&e=(v,u_1)\in LC_1\wedge e=(v,u_2)\in LC_2\end{aligned} \tag{3.13}$$

3.3.3 算法伪代码

算法 1 为本文中重叠社区发现算法的伪代码。算法包括两个步骤：第一步是聚类边集，构造链接社区的集合 LCS。在这一阶段中，算法先将每条边都设置为未被分类（行②）；遍历所有未被分类的边 e（行③），计算边 e 是否是一个核心边，若是则找出所有与之直接密度可达的边，构造链接社区 LC，加入链接社区集合 LCS 中（行③~⑦）。第二步，算法将遍历 LCS 中的每一个链接社区 LC，建立对应的节点社区 C，得到节点社区集合 CS（行⑧~⑫）；

最后寻找同时加入多个节点社区的节点，即重叠社区中的一员（行⑬~㉑）。

表 3.3 算法 1

输入：转发关系网络$G_{retweeted}$，ε，μ

输出：重叠社区结构 *OS*

/*对边集进行聚类*/

① $LCS \leftarrow \varnothing$， $CS \leftarrow \varnothing$；

② $\forall e \in E : e.IsClassified \leftarrow$ FALSE；

③ FOR EACH $e \in E$ DO

④ IF $!e.IsClassified$ THEN

⑤ IF *e.IsCore* THEN

⑥ $e.IsClassified \leftarrow$ TRUE；

⑦ $LCS = LCS \cup FindLinkComm(e)$；

/*将得到的链接社区转化为节点社区，并发现重叠社区结构*/

⑧ FOR EACH $LC \in LCS$ DO

⑨ $C \leftarrow \varnothing$；

⑩ FOR EACH $e = (u,v)$ IN LC_i DO

⑪ $C \leftarrow C \cup \{u,v\}$；

⑫ $CS = C \cup C_i$；

⑬ $\forall e \in LCS : e.IsClassified \leftarrow$ FALSE

⑭ FOR EACH $e \in LCS(e \in LC_x, e = (v,u) \in C_i)$ DO

⑮ IF $!e.IsClassified$ THEN

⑰ $e.IsClassified \leftarrow$ TRUE；

⑱ IF $(v \in C_j \vee u \in C_j) \wedge C_i \neq C_j$ THEN

⑲ IF $v \in C_j$ THEN

⑳ $OS \leftarrow OS \cup \{v\}$；

㉑ ELSE $OS \leftarrow OS \cup \{u\}$；

㉒ RETURN *OS*；

算法 2 为上面算法中 *FindLinkComs* 函数的伪代码，作用是找出所有与核心边 *e* 直接密度可达的边，构造链接社区 *LC*。先将边 *e* 添加至队列 *Q* 中（行②），接着按下面 4 个步骤迭代运行：

①从队列 *Q* 中取出队首元素 *x*（行⑤）；

②寻找与 *x* 直接密度可达的边集合 *R*（行⑥）；

③遍历 *R* 中的每一条边，若 *l* 没有被分类，则将 *l* 加入链接社区 *LC*；若 *l* 同时为核心边，将 *l* 加入队列 *Q* 中（行⑦~⑬）；

④删除队列 *Q* 的队首元素 *x*（行⑭）。

表 3.4 算法 2

输入：边 e，ε，μ；

输出：链接社区 *LC*

① $LC \leftarrow \varnothing$；

② $Q.Append(e)$；

③ $e.IsClassified \leftarrow$ TRUE；

④ WHILE $Q.IsEmpty()$ =FALSE DO

⑤ $x \leftarrow Q.Serve()$；

⑥ $R \leftarrow \{y \mid (E \mid Dir\,\mathrm{Re}ach(x, y, \varepsilon, \mu))\}$；

⑦ FOR EACH *l* IN R DO

⑧ IF $l.IsClassified \leftarrow$ FALSE THEN；

⑨ $l.IsClassified \leftarrow$ TRUE；

⑩ IF $l.IsCore$ THEN

⑪ $LC \leftarrow LC\{l\}$；

⑫ $Q.Append(l)$；

⑬ ELSE $LC \leftarrow LC\{L\}$；

⑭ *Q.Dequeue*（）;

⑮ RETURN *LC*;

3.3.4 时间复杂度

假设社交网络中的转发关系网络G包括n个节点、m条边，本文中的重叠社区发现算法的第一步遍历了每条边，并找出与该边直接密度可达的边集，其实只需要检查每一条边的邻居集合，时间复杂度为：$O(\sum_i d(e_i))$。其中$d(e_i)$表示边e_i的邻居个数，设$e_i=(u_i,v_i)$，那么$d(e_i)=k(u_i)+k(v_i)$，$k(u_i)$、$k(v_i)$为节点u_i和节点v_i的度。所以，若转发关系网络中节点的平均度为k，算法在第一步中的时间复杂度为$O(km)$.

改进的重叠社区发现算法的第二步是对转发关系网络中的每条边做处理，将每条边连接的两个节点都分配到相应的节点社区 C 中，其时间复杂度为$O(m)$。但实际的社交网络中，$k \ll m$， m和n呈线性关系，所以本文中的重叠社区发现算法的总体时间复杂度为$O(m)$或$O(n)$。

3.4 本章小结

本章主要是通过分析社交网络中的网络特征引出重叠社区结构特征，来初步识别社交网络水军团体，为下一步结合内容特征识别网络水军打下基础。主要内容包括：

（1）分析了社交网络中，网络结构的不同度量方法与信息传播能力的关系，找到了传播信息能力强的重叠社区结构；

（2）分析了以往在全网范围内进行网络水军检测的缺陷，提出构建转发关系网络，并在此网络上发现重叠社区结构，提高社交网络中水军识别的效率和准确率；

（3）为避免以往重叠社区发现算法中“过度重叠”的现象影响本文的水军识别工作，提出了改进的重叠社区发现算法，并介绍了原理和步骤，最终分析了此算法的时间复杂度。

第四章 基于结构与内容的水军团体识别

本章在第三章得到的社交网络中转发关系网络的重叠社区结构的基础上，进一步结合重叠社区内部节点的内容特征，识别社交网络中存在的水军团体。

4.1 基于内容特征识别水军问题分析

社交网络中的水军，平时看起来只是一个网络中具有较强传播信息能力的普通节点，它们存在于重叠社区结构中。但当他们有“任务”时，他们会“全体”出动，共同传播垃圾信息，使其成为热点。这也体现了网络水军团体在传播垃圾内容方面的特性——特定时间内（有“任务”时），他们所发送的内容与垃圾信息有高度语义相似性。通过对重叠社区内的节点，在特定时间内所传播的内容进行分析，来判断其是否发送过垃圾信息，进一步将重叠社区中的无关节点排除出去，得到水军团体。

本文认为处在重叠结构中的节点，如果存在转发、复制、重写垃圾信息等操作，则视为疑似网络水军，归入嫌疑水军团体中。本章中，根据上一章得到重叠社区结构，将其中节点转发、复制、重写内容的特征与当时的垃圾信息内容的特征进行相似性比较，确定疑似水军。为了避免出现，一次无意传播一条垃圾信息就被认定为网络水军的情况。本文中，将重叠社区结构中，多次传播过垃圾信息的节点认定为网络水军。即，如果疑似水军对多个垃圾信息进行了转发、复制、重写操作，那么就可以认为该疑似水军为网络水军，最终得到网络水军团体。

4.2 总体方案

图 4.1 展示了本文基于结合社交网络的结构特征与社交网络所承载内容的特征识别水军团体的总体方案。主要分为三个阶段：

（1）根据垃圾内容的相关信息，对社交网络数据集进行处理，构建转发关系网络，并发掘此网络中的重叠社区结构，得到水军团体的结构特征。

（2）分析重叠社区内节点包含内容的特征与垃圾信息的内容特征。

（3）结合第一阶段和第二阶段的成果，识别出网络水军团体。

其中，第一阶段已经在第三章中详细介绍并完成，剩下的两个阶段将在本章中全部完成。

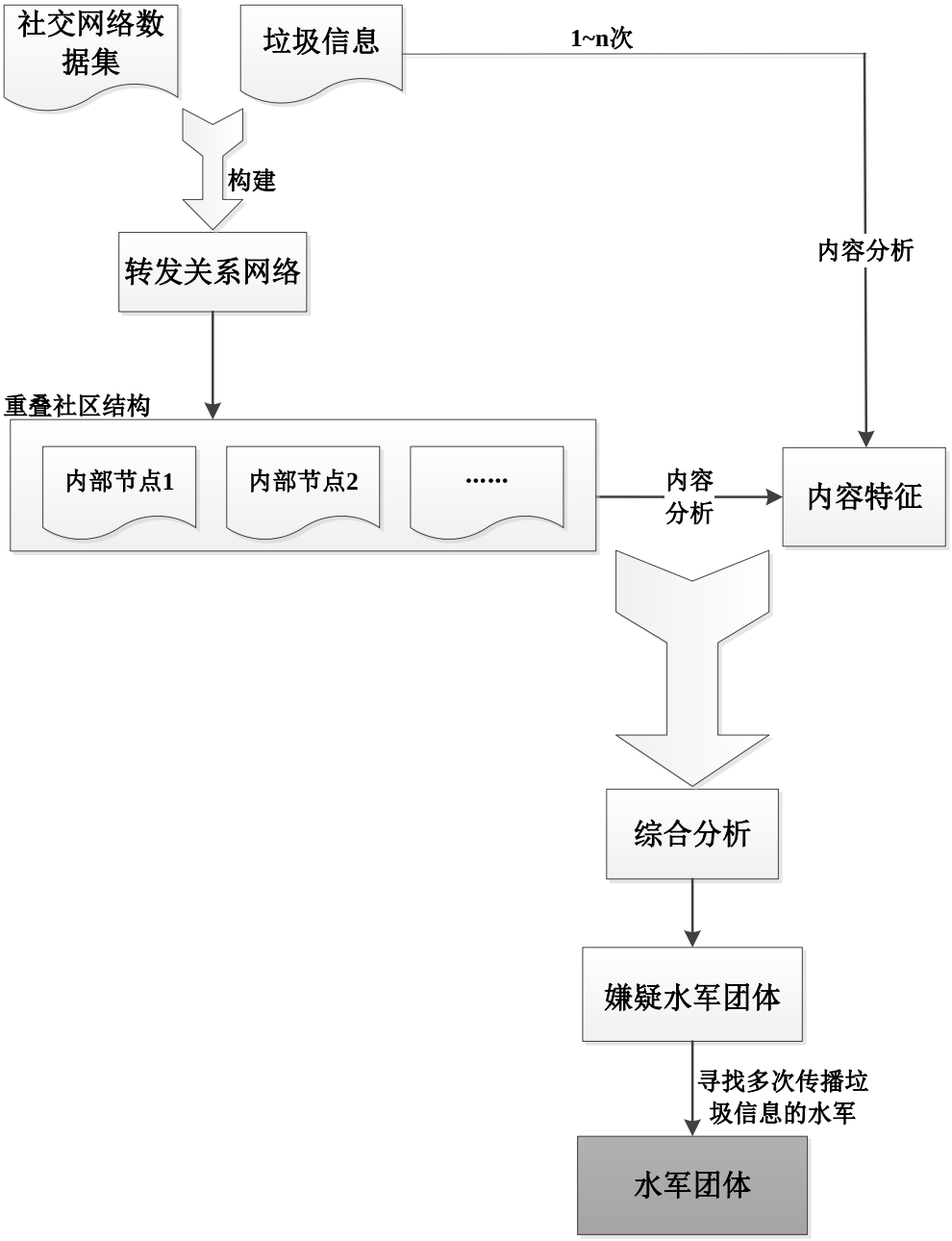


图 4.1 总体方案

4.3 内容特征的提取

本文通过计算重叠社区内部节点所传播过的信息与特定垃圾信息的相似性，来判断重叠社区内的节点是否传播过垃圾信息。传统文本相似性判断，通过统计某词或者短语共同出现的频率，如 TF-IDF[58,59]方法。但此类方法只考虑共有词语的多少，没有考虑到文字的语义特征。例如下面两个句子：

（1）“乔布斯刚刚去世了。”

（2）“苹果价格将会下降。”

尽管这两句话没有出现共同的词语，但人们可以很容易看出，这两句话表达的含义一致。在现实中，很多时候文本的相关程度取决于背后的语义联系，而非表面的词语重复。特别是狡猾的网络水军，很可能将其他水军所发布的垃圾信息改写，使其很少或不再出现相同的词语。因此，传统的通过“文本间重复词语的多少”来判断文本相似性的理论，已不能满足社交网络水军识别时的文本相似性检测的需求。而文本的语义，通常使用主题模型（topic model）来挖掘，下面将对它进行详细介绍。

4.3.1 主题模型简介

主题模型，实际上就是用来对文档中隐含的主题进行建模的一种方法。再来回看上面的例子，“乔布斯”这个词通常在“苹果公司”这个主题中出现，而“苹果”既可能属于“水果”主题，也可能属于“苹果公司”这个主题。当对这两个文档进行比较时，它们都体现了“苹果公司”这个主题，所以两个文档存在相似性。这里的“主题”，通常表示一个概念、一个方面。我们将一个主题，比作一个“词桶”，桶里装了很多与这个主题相关的一系列词。而一段文本往往包含不只一个主题，与这些主题相关联的词语，共同组成了这段文本。如图 4.2 中的四个桶，展示了四个主题以及它们混合组成的一段文本。

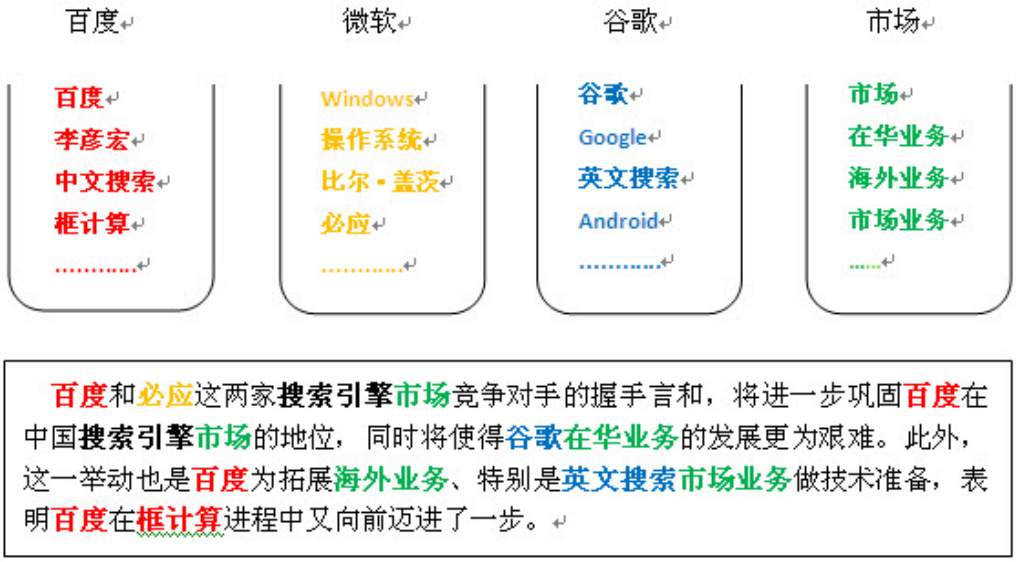


图 4.2 词桶和文本

以上文本中的段落包括四个主题，它主要表达的是百度和市场发展。因为，“百度”主题的词桶和“市场”主题的词桶中的词占整个文本的比例较大。可以看出其两个主题也有一定的体现，但不是主要的语义。主题模型认为，一篇文档中通过两个步骤生成：第一步，以一定概率选择某个主题；第二步，再从这个主题中以一定概率选择某个词语。在生成一篇文档的过程中，文档中每个词出现的概率：

$$p(\text{词语}|\text{文档}) = \sum_{\text{主题}} p(\text{词语}|\text{主题}) \times p(\text{主题}|\text{文档}) \quad (4.1)$$

式 4.1 还可以用矩阵乘法来表示，如下图所示：

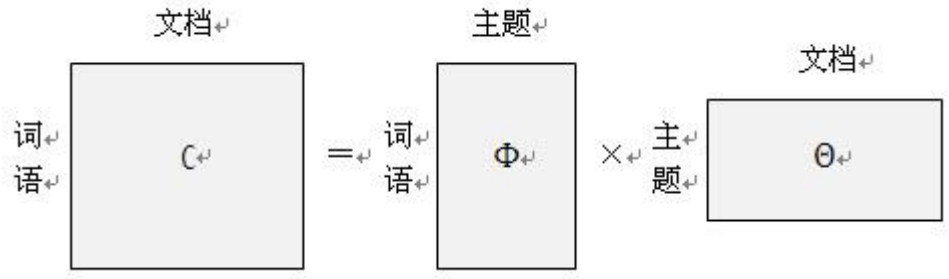


图 4.3 矩阵乘法表示

上图中，左边的C矩阵表示一篇文档中每个词语出现的概率 p(词语|文档)，这个是已知的；Φ矩阵代表每个主题中每个词语出现的概率 p(词语|主题)，也就是每个“词桶”；右边的Θ矩阵代表的是每篇文档中各个主题出现的概率 p(主题|文档)。例如，有大量文档，先分别对每个文档进行分词，得到每个文档的词汇列表，最终每篇文档被表示成为一个词语的集合。上图中的C矩阵，用每个词语在文档中出现的次数除以文档中词语的总数目就是 p(词语|文档)；对于剩下的两个右边矩阵，可以通过对大量已知的“词语—文档”C矩阵 （ p(词语|文档)）进行一系列的训练，推算出右边的“词语—主题”矩阵 Φ和“主题-文档”矩阵 Θ。本文的重点，就是得到一个文档中出现的主题分布 p(主题|文档)，即社交网络中的信息所表达的主要语义。

主题模型具有识别庞大文本集中潜藏的主题信息的能力，这非常适合处理社交网络中的海量文本数据，将大大地简化问题的复杂性。同时，社交网络中的水军，在进行垃圾信息传播的过程中无论是以何种形式（转发、复制、重写）进行的，垃圾信息的主题都不会改变，并且社交网络节点发布的信息通常都涉及一个或多个主题，这些特征正好和主题模型匹配。

主题模型有两种：pLSA（Probabilistic Latent Semantic Analysis）[60]和 LDA（Latent Dirichlet Allocation）[61]，但 pLSA 在计算文档对应主题概率时，没有统一的概率模型，如果参数过多，可能会导致过拟合现象。因此，LDA 引入了“文档-主题-词语”3 层贝叶斯模型（拓扑结构如图 4.4 所示），来避免上述情况的出现。

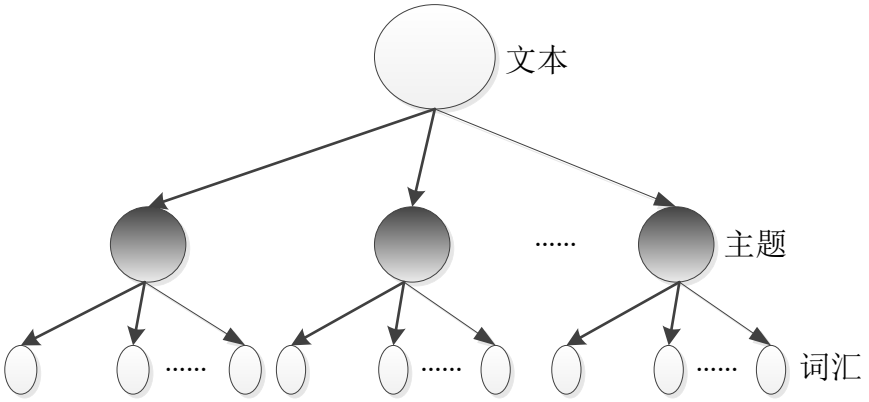


图 4.4 LDA 三层拓扑模型

LDA 模型首先选择一个主题分布向量θ，确定每个主题被选择的概率α。在生成每个词语时，从主题分布向量θ中选择一个主题z，按主题z的词语概率分布β生成一个词语。LDA 中，M份包含N个词语的文档生成过程如下：

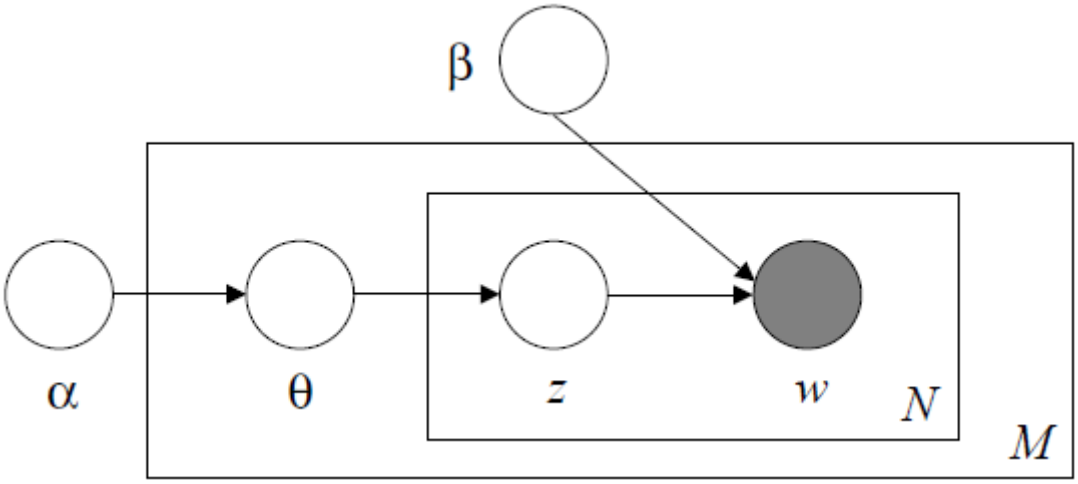


图 4.5 LDA 模型图

LDA 模型图中关键词w是可观察的变量，其他变量都为隐藏。箭头方向代表条件概率方向，大的矩形框表示从 Dirichlet 分布中为M份文本反复抽取主题分布，小矩形框表示从主题分布中反复抽取产生文本的N个词。

4.3.2 相似度计算

本文采用 LDA 主题模型来推断重叠社区内部节点所发信息和垃圾信息的主题分布。相关定义以及计算过程中用到的变量（表 4.1）如下：

定义 4.1　文本内容D的主题分布：令主题集合$C=\{C_1,C_2,...,C_T\}$，则$p(C_i|D)$表示D属于主题C_i的后验概率，由这T个后验概率组成的向量$\left(p(C_1|D),p(C_2|D),...,p(C_T|D)\right)$被称为文本内容$D$的主题分布。

定义 4.2 令集合 $D=\{D_1,D_2,...,D_d\}$ 表示重叠社区某个用户在垃圾信息发布后 1h 内，用户发布的文本信息的集合，D_t 表示已知垃圾信息。

定义 4.3 某用户发布文本内容的主题分布：用一个 T 维向量表示成 $(t_1,t_2,...,t_T)$，其中 $t_i=\frac{1}{d}\sum_{j=1}^{d}p(C_i|D_j)$。

表 4.1 推断过程中的变量列表

变量	含义
T	主题个数
V	词语个数
α	Dirichlet 分布的参数，用于生成一个主题 θ 向量
β	每个主题对应的词语概率分布
Z_{w_i}	词语 w_i 的主题
$Z_{D,\neg i}$	除了第 i 个词语的主题外，在文本 D 中其他所有词语的主题集合
$n(j,D)$	文本内容 D 中属于第 j 个主题的词语数
$n(D)$	文本内容 D 中的词语数
$n(j,v)$	主题 j 得到某一词语 v 的频数
$n(j,.)$	属于主题 j 的所有词语数

采用 Gibbs 抽样算法间接计算模型参数 $\theta_{D,j}$（文本内容 D 属于第 j 个主题的概率）。假设 D 由 n 个词语组成，记为 $\{w_1,w_2,...,w_n\}$，对文本内容 D 中的词语 w_i，$Z_{w_i}=j$（词语 w_i 属于第 j 个主题）的概率计算如下：

$$P(Z_{w_i}=j|Z_{D,\neg i},D,\beta,\alpha)\propto\frac{P(Z_{w_i}=j,Z_{D,\neg i},D|\beta,\alpha)}{P(Z_{D,\neg i},D|\beta,\alpha)}=\frac{n(j,D)+\alpha-1}{n(D)+T\alpha-1}\times\frac{n(j,v)+\beta-1}{n(j,.)+V\beta-1} \quad (4.2)$$

对式（4.2）反复迭代，并对所有主题进行抽样，最终达到抽样结果稳定，最终 $\theta_{D,j}$ 可以估计为：

$$\hat{\theta}_{D,j}=\frac{n(j,D)+\alpha}{n(D)+T\alpha} \quad (4.3)$$

文本内容 D 的主题分布为 $\hat{\theta}_D=(\hat{\theta}_{D,1},\hat{\theta}_{D,2},...,\hat{\theta}_{D,T})$。

本文通过衡量重叠社区内的用户所发信息与垃圾信息的主题相似度，来判别用户是否转发、复制或重写了垃圾信息以帮助其传播，并将相似度较低的用户，在此次检测中排除其是网络水军的可能。通过之前的推导，可以直接得到垃圾信息 D_t 的主题分布。根据本小节的定义 4.3，计算出重叠社区内所有用户的在 1h 内发布信息的主题分布 $(t_1,t_2,...,t_T)$。

由于网络用户所发信息（微博、论坛帖等）一般涉及的主题较少，即使是 1h 内的信息，网络水军也只会将重点放在要传播的垃圾消息上，所以主题分布很稀疏。针对这种情况，本文采用余弦相似性的方法，来度量每个重叠社区内的用户所发信息与垃圾信息的相似程度：

$$S_{\theta_{D_t},\theta_{D_i}} = \frac{\theta_{D_t} \bullet \theta_{D_i}}{\left\|\theta_{D_t}\right\| \left\|\theta_{D_i}\right\|} \tag{4.4}$$

求得所有的相似度结果后，设定阈值σ，若$S_{\theta_{D_t},\theta_{D_i}} \geq \sigma$，则将第 i 个用户加入嫌疑网络水军的团体中，得到由这条垃圾信息检测出的嫌疑网络水军团体。

若重叠社区内的某用户在指定的 1h 内未发布任何信息时，但在确定垃圾信息发送时间点的前后两个月的时间段内，有别的垃圾信息出现，则继续保持此用户的“嫌疑身份”，直接用其他的垃圾信息进行内容特征挖掘；反之在这条垃圾信息检测中不认定其为网络水军。

4.4 最终水军团体的确定

为了进一步提高检测的准确率，本文对从不同垃圾信息出发，检测得到的多个嫌疑网络水军团体，进行求交筛选操作：

$$\begin{aligned} I_{CO} &= CO_1 \\ I_{CO} &= I_{CO} \cap \sum_{i=2}^{n} CO_i \end{aligned} \tag{4.5}$$

其中，I_{CO} 为所有嫌疑网络水军团体的交集，初始值为CO_1。

最终得到所有团体的交集中，节点都是多次传播垃圾信息的社交网络节点，这种做法避免了一次误传播就被归类为网络水军的情况，这些节点是网络水军的可能性较之前更高。

4.5 本章小结

本章结合第三章得到的重叠社区结构，对社区内部节点所发信息和垃圾信息进行了主题特征挖掘，并通过两者的相似性，判断重叠社区内节点是否传播过垃圾信息。本章完成了基于结构与内容的社交网络水军团体识别的整个检测流程，并给出了相应的实现步骤。

第五章 实验

5.1 实验准备

5.1.1 实验数据集

本文在新浪微博平台上验证本文方法的有效性，实验所用数据集为新浪微博 2012 年 7 月至 2013 年 3 月之间，7000 多万新浪用户的个人信息和每个用户平均 1000 条的微博内容数据，此数据集在数据堂站点上购买。

图 5.1 数据集介绍

该数据集中的数据由 JSON 格式描述，其中的字段含义如下图：

字段名	字段类型	字段说明
created_at	string	微博创建时间
id	int64	微博ID
mid	int64	微博MID
idstr	string	字符串型的微博ID
text	string	微博信息内容
source	string	微博来源
favorited	boolean	是否已收藏，true：是，false：否
truncated	boolean	是否被截断，true：是，false：否
in_reply_to_status_id	string	（暂未支持）回复ID
in_reply_to_user_id	string	（暂未支持）回复人UID
in_reply_to_screen_name	string	（暂未支持）回复人昵称
thumbnail_pic	string	缩略图片地址，没有时不返回此字段
bmiddle_pic	string	中等尺寸图片地址，没有时不返回此字段
original_pic	string	原始图片地址，没有时不返回此字段
geo	object	地理信息字段 详细
user	object	微博作者的用户信息字段 详细
retweeted_status	object	被转发的原微博信息字段，当该微博为转发微博时返回 详细
reposts_count	int	转发数
comments_count	int	评论数
attitudes_count	int	表态数

图 5.2 数据集中字段介绍

图 5.3 表示的是 JSON 格式描述的用户个人信息和其在 2012 年 12 月 25 日 17 点 46 分发送的一条微博信息：


```
"user": {
    "id": 1404376560,
    "screen_name": "zaku",
    "name": "zaku",
    "province": "11",
    "city": "5",
    "location": "北京 朝阳区",
    "description": "人生五十年，乃如梦如幻；有生斯有死，壮士复何憾。",
    "url": "http://blog.sina.com.cn/zaku",
    "profile_image_url": "http://tp1.sinaimg.cn/1404376560/50/0/1",
    "domain": "zaku",
    "gender": "m",
    "followers_count": 1204,
    "friends_count": 447,
    "statuses_count": 2908,
    "favourites_count": 0,
    "created_at": "Fri Aug 28 00:00:00 +0800 2009",
    "following": false,
    "allow_all_act_msg": false,
    "remark": "",
    "geo_enabled": true,
    "verified": false,
    "allow_all_comment": true,
    "avatar_large": "http://tp1.sinaimg.cn/1404376560/180/0/1",
    "verified_reason": "",
    "follow_me": false,
    "online_status": 0,
    "bi_followers_count": 215
}
"created_at": "Tue Dec 25 17:46:55 +0800 2012",
"id": 11488051271,
"text": "好像开宝马的人，负面新闻特别多。 //@老沉:心理承受力差，没敢看视频",
"source": "<a href="http://weibo.com" rel="nofollow">新浪微博</a>",
"favorited": false,
"truncated": false,
"in_reply_to_status_id": "",
"in_reply_to_user_id": "",
"in_reply_to_screen_name": "",
"geo": null,
"mid": "5612814510546515491",
"reposts_count": 8,
"comments_count": 9,
```


图 5.3 JSON 格式的用户信息和微博信息

5.1.2 数据集的处理

由于以 JSON 格式提供的数据集中，存在很多与本文水军识别无关的属性。所以根据本文方法的数据需求，最终我们选择用户信息的 6 个属性和微博信息的 8 个属性，作为本文的实验数据。最终，将用户信息和微博信息数据存储于数据库 MySQL 中，如图 5.4、5.5 所示。

id	screen_name	decription	followers_count	friends_count	create_at
31709710	小小yu米	漂洋过海来看你 【此图胖里所有本人所画漫画可以转出，但不得去LOGO或二	1854	310	Sep 18 20:24:30 2009
35847663	PerFeeKid	Fight for love. Fight for life.	291	85	Apr 25 00:34:40 2012
38906982	大迪迪	子非鱼焉知鱼之乐...我欣然自得...自得其乐....penny.peng@vip.163.com	3005	386	Jan 09 01:52:50 2010
39231619	晋格格	留存珍惜，记录感动，审视生活。	502	217	Jun 04 18:23:10 2012
41275817	毛主席的戰士	http://blog.sina.com.cn/rjy369	146	77	Dec 16 22:02:50 2012
43077721	geji_geji	浪漫的爱情，美满的生活，传奇的人生	1195	1667	Mar 09 23:46:00 2010
43582170	佛法无处不在		26	114	Jul 10 15:41:20 2013
43989061	天河之情	w	183	696	Apr 08 00:04:40 2010

图 5.4 用户信息表

id	create_at	text	reposts_count	comments_count	attitudes_count	userId	retweeted_status
2498626480	Sep 10 21:16:00 2012	//@点评团_北京:【免费赢影票！】[太开心] 买好了还不过瘾？本周日（9月	0	1	0	1033337703	0
2692660108	Jul 16 23:40:10 2012	因为掏心掏肺的爱，所以没心没肺的放弃。	0	0	0	214909663	0
2770940953	Nov 21 04:51:40 2012	三点四十三 六近半 漂亮宝宝	0	1	0	1002647721	0
6093276241	Jan 05 22:49:20 2013	都是游戏惹的祸，有些人不是坏人呢，但是就是控制不了自己	0	0	0	1033337703	0
6093346717	Feb 05 22:51:30 2013	#2013我要#幸福、快乐每一天。参加"「可口可乐」新愿欢享中国年"活动即	0	0	0	1033337703	0
6291531509	Jul 12 10:22:10 2012	转发微博。	0	0	0	1040241737	0
6380363287	Feb 14 18:21:40 2013	西单大悦城	0	0	0	1040241737	0

图 5.5 微博信息表

同时，本文的数据集，在建立社交网络中的转发关系网络和基于 LDA 的内容主题特征挖掘阶段，要分别进行不同的预处理。具体如下：

（1）如 3.2 节的图 3.3，微博系统规定转发一条微博后，会在“//@”列出被转发人。通过这个规定，我们可以利用“//@”的位置，记录之后的被转发人，建立微博用户的转发记录，为建立社交网络中的转发关系网络做准备。

（2）对新浪微博内容进行 LDA 主题挖掘时，需要将文本转换成机器理解的格式。主要分为两步：第一，去除不含有任何意义主题词的文本，比如“@”符号、用户昵称、表情符号等；第二，去除停用词（如“非常”，“特别”，“十分”等）、不代表主题意义的单个汉字（如“这”、“那”等）等。

5.1.3 实验环境

实验代码用 Java 语言实现，实验环境为 Intel Core i5-4200M 2.5GHz 的 CPU，8GB 的内存，1TB 硬盘的 PC 机，操作系统为 Windows 8.1 中文版。实验结果图，使用 Microsoft Excel 进行处理。

5.2 重叠社区的发现

5.2.1 建立转发关系网络

本章选取了 2012 年 9 月 13 日、10 月 9 日和 12 月 17 日的三条谣言（垃圾信息），分别通过它们出现的时间抓取种子节点，再向下深度遍历得到包括种子节点在内的一共 79746 个微博用户和其中 208132 次转发回复关系；构建 R1、R2、R3 三个转发关系网络。在构建转发关系网络的过程中，已排除了用户过少的独立图结构（与其他图结构无联系），且两节点间的边只建立一次。具体信息如下表：

表 5.1 三个转发关系网络的信息

ID	节点数	边数
R1	30680	84316
R2	17133	54425
R3	15991	49602

5.2.2 评价指标

本文利用文献[62]中提出的模块度 EQ 来评价重叠社区发现算法，EQ 值越接近 1，表示网络划分出的社区结构的强度越强，划分质量越好，并通过将本文中的重叠社区发现算法（记为 LINK2）与具有代表性的重叠社区发现算法 CFinder（CPM 的实现）[63]、LINK 进行比较。

$$EQ=\frac{1}{X}\sum_{t}\sum_{i\in C_t, j\in C_t}\frac{1}{O_i O_j}[A_{i,j}-\frac{R_i R_j}{X}] \quad (5.1)$$

其中，节点d_i的度数为R_i，网络节点的总度数为X，A为网络的邻接矩阵，O_i为节点d_i所隶属社区的个数。

5.2.3 实验结果及分析

我们在转发关系网络 R1 进行重叠社区发现的对比实验，结果如表 5.2，可以看到本文中的算法（记为 LINK2），由于过滤掉“孤立边”，虽然在时间效率上比未改进的 LINK 算法略差，但它的 EQ 值明显高于 LINK 算法。同时，相对于 CFinder 算法 LINK2 也取得较高的 EQ 值。而 CFinder 算法，由于要寻找全部连通k派系组成的极大子图，所以十分耗时。

表 5.2 重叠社区发现算法比较

	LINK2	LINK	CFinder
EQ	**0.729**	0.351	0.68
耗时（s）	**65.5**	42.8	603.7

5.3 LDA 参数的确定

由于新浪微博较短，所以实验采用 LDA 分析重叠节点 1h 内的所有内容与谣言（垃圾信息）的主题分布。参照其他文献对超参数的选取，本文中的超参数的选择为：$\alpha = 50/T$，$\beta = 0.1$。

对于潜在主题数量 T 的取值，本文采用困惑度的方法进行选取，计算公式为：

$$Perplexity(M) = \exp\left\{-\frac{\sum_{m=1}^{M}\ln p(r_m)}{\sum_{m=1}^{M} N_m}\right\} \quad (5.2)$$

其中 M 为测试集，N_m 表示第 m 篇微博 r_m 的长度。$p(r_m)$ 表示 LDA 模型产生微博 r_m 的概率，公式如下：

$$p(r_m) = \prod_{i=1}^{n}\sum_{j=1}^{T} p(w_i \mid Z_{w_i} = j)\, p(Z_{w_i} = j \mid r_m) \quad (5.3)$$

在选取中，困惑度越低意味着模型产生文档的能力越高，主题数 T 的取值越接近真实情况。在实验中，对 T 取值为$10, 20, 30, 40, 50, 60, 70$的情况进行的测试，变化情况如图 5.6：

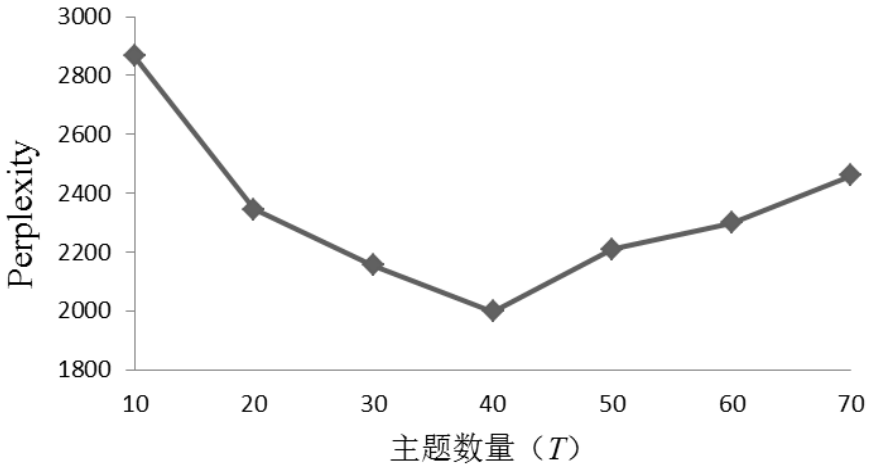


图 5.6 不同 T 值下的 Perplexity 值

从上图中我们可以看出，主题数在 10 至 40 这段时，困惑度在不断的减小；在 40 至 70 这段时，困惑度不断增大；主题数为 40 时，困惑度达到最小值。因此选取 T=40 作为主题的数目，此时模型的性能最好。

5.4 水军识别实验与分析

5.4.1 实验评估标准

目前为止，还没有一套大家公认的、统一的网络水军识别评估方法，这也是水军识别研究的难点之一。由于社交网络水军识别研究中，缺少公开可用的数据集，所以通常通过人工评价的方法，对算法的准确性进行判断。采用如下指标：

$$Accuracy = \frac{n}{N} \tag{5.4}$$

其中，N 为实验识别出的社交网络水军数目，n 为两名工作人员根据用户所有的微博内容、微博转发数以及用户本身的粉丝数三个指标，从检测出的网络水军中，判定出的真正社交网络水军数目。

5.4.2 实验及结果分析

为了得到更准确的实验结果，我们对比了以下 3 种方法：CAT、IM 和本文提出的水军识别方法（记为 NM）。其中 CAT 是 Amleshwaram 等人[64]综合用户的行为、内容、用户间关系等特征，实现对网络水军的识别。IM 是 Wang 等人[65]针对收集的评论数超过 7000 的微博及其用户间联系，提出的影响力在线算法检测网络水军。分别在之前建立的三个转发关系网络上进行对比实验，我们先观察下本文算法从单个谣言（垃圾信息）出发，即嫌疑网络水军团体的检测效果。结果如下：

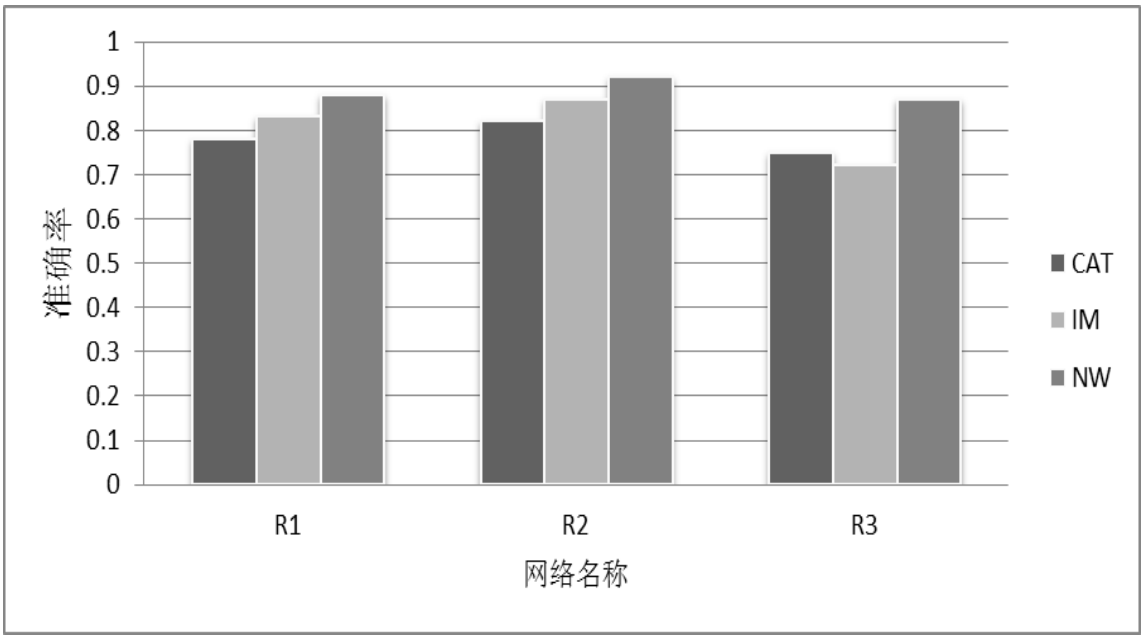


图 5.7 三种方法在三个网络上的对比结果

表 5.3 三种方法的准确率对比

Method	R1	R2	R3
CAT	0.78	0.80	0.75
IM	0.83	0.87	0.72
NW	**0.88**	**0.92**	**0.87**

从上面的图表中可以看出，由于 CAT 方法是根据行为、内容和用户间特征来进行水军识别工作的，但水军行为逐渐趋于正常用户同时他们所发送的内容不再具有显著的可以识别特征，这造成了这种方法的识别准确率在当前的社交网络中较低。而 IM 方法，通过根据评论数量和用户间的联系，计算用户的影响力，这种办法较 CAT 更科学；评论行为和其数量在一定程度上体现了一个用户的信息传播能力，但此方法并未考察用户所发信息的内容特征。较以上的两种方法，本文提出的水军识别方法，在嫌疑水军识别阶段的准确率已经最高。

对本文的方法进一步处理后，将本文提出方法在 R1、R2 和 R3 获得的嫌疑网络水军团体进行求交集运算，得到多次传播谣言（垃圾信息）的水军团体。并对比上一步实验中 CAT 和 IM 的最高准确率，结果如下：

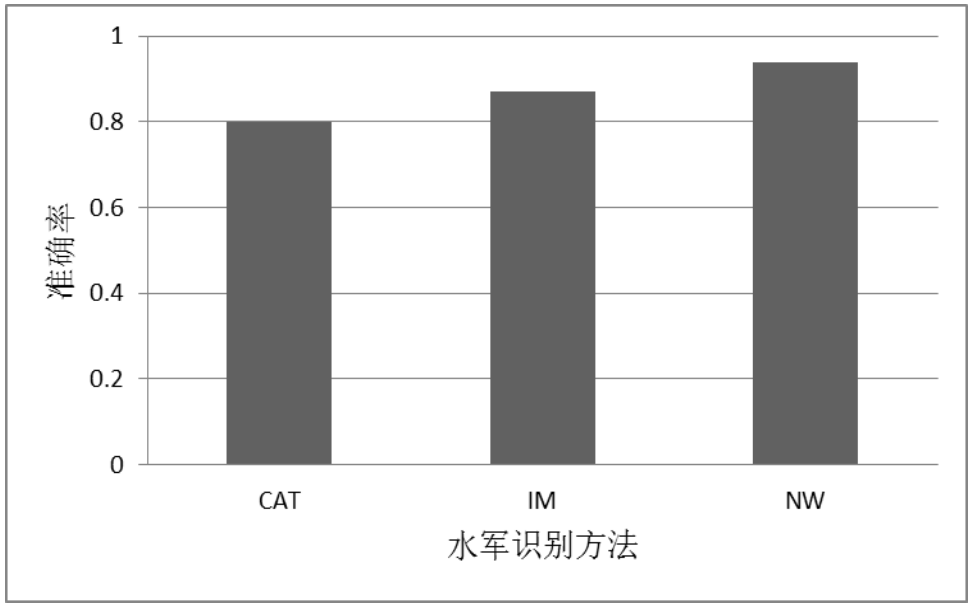


图 5.8 NW 方法最终的准确率对比

表 5.4 最终的准确率对比

Method	CAT	IM	**NW**
准确率	0.80	0.87	**0.94**

通过结果可以发现，对多个嫌疑网络水军团体进行求交运算，得到的最终网络水军团体，更加提高了识别的准确率；因为它排除了一些误传播垃圾信息的非水军用户。可以看出，本文提出的方法的合理性。较其他方法，本文的方法可以更准确的检测出社交网络中的网络水军团体。

5.5 本章小结

本章首先建立了由三个垃圾信息出发构建的转发关系网络，并对数据集进行了预处理；接着确定了 LDA 主题模型的参数；最终，通过与各个水军识别算法的比较，验证本文提出方法的效果。

第六章 研究工作总结与展望

6.1 研究工作总结

本文主要通过综合社交网络水军的网络结构特征与他们所发信息的内容特征，识别社交网络中的网络水军团体，同时对论文所涉及的相关领域进行了深入研究。其中包括：网络水军识别、重叠社区发现、主题模型等方面的相关内容及研究成果的论述；之后利用改进的重叠社区发现算法挖掘水军的网络结构，并结合 LDA 文档主题生成模型计算用户所发信息的主题分布；最后实现了基于结构与内容的社交网络水军团体识别方法。

本文提出的方法主要有如下贡献和创新：

（1）数据集的预处理。垃圾信息出现后，网络水军只存在于该时间片内发送信息的用户群中。从该时间点出发建立转发关系网络，排除了大量与这次垃圾信息传播无关的节点，避免了计算风暴，更具可行性。

（2）基于重叠社区的水军识别。网络水军必然是一群传播信息能力很强的节点，而重叠社区结构作为多个社区的公共部分非常有利于信息的传播。本文将重叠社区结构作为网络水军的网络结构特征，不仅符合网络水军本身的特征，还可以发现网络水军形成的水军团体。

（3）综合结构与内容特征的水军识别方法。本文综合考虑网络水军的网络结构特征和所发信息的内容特征，可以避免仅从单一特征检测水军的不足。同时分析了重叠结构中多次传播垃圾信息的节点，比一次传播垃圾信息就确定为网络水军的方法，得到的结论更加准确。

（4）实验。在新浪微博数据集上进行仿真，对数据集进行预处理并建立社交网络中的转发关系网络，发现其中的重叠社区结构并结合 LDA 生成的主题分布识别水军团体，最后用本文提出的水军识别方法与其他方法进行对比实验和分析。

综上所述，本文针对以往网络水军研究工作中的不足，提出了一种基于结构与内容的社交网络水军团体识别方法，综合考虑了社交网络中的水军团体所在结构的特性和节点本身的内容特性。此方法不仅能提高网络水军检测的准确率，还能找出网络水军形成的水军团体。利用新浪微博数据集，我们验证了本文提出方法的合理性和准确性。

6.2 未来的研究内容展望

近年来，国内网络水军识别研究被逐渐重视起来，许多机构和个人都展开了深入的研究。

但是对社交网络中的水军进行识别，在国内仍然处于起步发展阶段，与国际水平相比较，还需要进一步研究。同时，以往的水军识别方法在社交网络环境中，不能有效发现新型水军。本文对网络水军识别研究进行了深入的调查与研究，发现可以在以下几个方面进一步探索和研究：

（1）社交网络节点十分庞大，如何对社交网络庞大的数据集进行预处理，提高水军识别算法的处理效率，将是水军识别研究中的一个重点。

（2）由于水军的活动历史反映在网络结构上，他们无法进行隐藏，同时网络结构具有稳定性，所以基于网络结构特征的水军识别方法，可以很好的避免水军的行为日益趋于正常用户而难以进行识别的问题。本文就是利用这种方法，进行社交网络中水军的初步识别的。所以针对水军的结构特征进行识别工作，仍然是我们需要不断研究的方向。

（3）如何进行社交网络水军的实时检测，第一时间发现水军团体的活动，进行干扰和限制，为网络安全提供技术保障，将是未来网络水军识别工作的难点。

（4）网络水军的多种特征中，除了结合网络结构特征和内容特征，还可以通过哪些途径提高社交网络水军的检测效率和准确率。

（5）需要公开可用的社交网络数据集，使得水军识别研究能够有效的进行评估。

参考文献

[1] 莫倩, 杨珂. 网络水军识别研究[J]. Journal of Software, 2014, 25（7）.

[2] Almeida T A, Yamakami A, Almeida J. Probabilistic anti-spam filtering with dimensionality reduction[C]//Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010: 1802-1806.

[3] 刘鸿宇, 赵妍妍, 秦兵, 等. 评价对象抽取及其倾向性分析[J]. 中文信息学报, 2010 （1）: 84-88.

[4] Niu Y, Chen H, Hsu F, et al. A Quantitative Study of Forum Spamming Using Context-based Analysis[C]//NDSS. 2007.

[5] Sawaya Y, Kubota A, Yamada A. Understanding the time-series behavioral characteristics of evolutionally advanced email spammers[C]//Proceedings of the 5th ACM workshop on Security and artificial intelligence. ACM, 2012: 71-80.

[6] Lim E P, Nguyen V A, Jindal N, et al. Detecting product review spammers using rating behaviors[C]//Proceedings of the 19th ACM international conference on Information and knowledge management. ACM, 2010: 939-948.

[7] Lee K, Eoff B D, Caverlee J. Seven Months with the Devils: A Long-Term Study of Content Polluters on Twitter[C]//ICWSM. 2011.

[8] Brendel R, Krawczyk H. Application of social relation graphs for early detection of transient spammers[J]. WSEAS Transactions on Information Science and Applications, 2008, 5（3）: 267-276.

[9] Bouguessa M. An unsupervised approach for identifying spammers in social networks[C]//Tools with Artificial Intelligence （ICTAI）, 2011 23rd IEEE International Conference on. IEEE, 2011: 832-840.

[10] Moh T S, Murmann A J. Can you judge a man by his friends?-enhancing spammer detection on the twitter microblogging platform using friends and followers[M]//Information Systems, Technology and Management. Springer Berlin Heidelberg, 2010: 210-220.

[11] Las-Casas P H B, Guedes D, Almeida J M, et al. SpaDeS: Detecting spammers at the source network[J]. Computer Networks, 2013, 57（2）: 526-539.

[12] 周东浩, 韩文报. DiffRank: 一种新型社会网络信息传播检测算法[J]. 计算机学报, 2014, 4: 014.

[13] 王永刚, 蔡飞志, 胡建斌, 等. 一种社交网络虚假信息传播控制方法[J]. 计算机研究与发展, 2012（S2）: 131-137.

[14] 张玥, 张宏莉, 张伟哲, 等. 识别网络论坛中有影响力用户[J]. 计算机研究与发展, 2015, 50（10）: 2195-2205.

[15] 韩毅, 许进, 方滨兴, 等. 社交网络的结构支撑理论[J]. 计算机学报, 2014, 37（4）: 905-914.

[16] 赵之滢, 于海, 朱志良, 等. 基于网络社团结构的节点传播影响力分析[J]. 计算机学报, 2014, 37（4）: 753-766.

[17] 陈浩, 王轶彤. 基于阈值的社交网络影响力最大化算法[J]. 计算机研究与发展, 2015, 49（10）: 2181-2188.

[18] Almeida T A, Yamakami A, Almeida J. Probabilistic anti-spam filtering with dimensionality reduction[C]//Proceedings of the 2010 ACM Symposium on Applied Computing. ACM, 2010: 1802-1806.

[19] Almeida T, Yamakami A. Content-based spam filtering[C]//Neural Networks （IJCNN）, The 2010 International Joint Conference on. IEEE, 2010: 1-7.

[20] Sriram B, Fuhry D, Demir E, et al. Short text classification in twitter to improve information filtering[C]//Proceedings of the 33rd international ACM SIGIR conference on Research and development in information retrieval. ACM, 2010: 841-842.

[21] Liu B. Sentiment analysis and subjectivity[J]. Handbook of natural language processing, 2010, 2: 627-666.

[22] Chen Y R, Chen H H. Opinion spam detection in web forum: a real case study[C]//Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015: 173-183.

[23] Bhat V H, Malkani V R, Shenoy P D, et al. Classification of email using BeaKS: Behavior and keyword stemming[C]//TENCON 2011-2011 IEEE Region 10 Conference. IEEE, 2011: 1139-1143.

[24] Husna H, Phithakkitnukoon S, Palla S, et al. Behavior analysis of spam botnets[C]//Communication Systems Software and Middleware and Workshops, 2008. COMSWARE 2008. 3rd International Conference on. IEEE, 2008: 246-253.

[25] Benevenuto F, Rodrigues T, Almeida V, et al. Identifying video spammers in online social networks[C]//Proceedings of the 4th international workshop on Adversarial information retrieval on the web. ACM, 2008: 45-52.

[26] Gammereldin S E A, Musa M E M. Analyzing and revealing spammer accounts in email social network: SUST Mail case[C]//Proc. of the 2nd Int'l Conf. on Information and Communication Technology. Khartoum: Ministry of Communications and Information Technology. 2011: 3-11.

[27] Stringhini G, Kruegel C, Vigna G. Detecting spammers on social networks[C]//Proceedings of the 26th Annual Computer Security Applications Conference. ACM, 2010: 1-9.

[28] Yang C, Harkreader R, Gu G. Empirical evaluation and new design for fighting evolving Twitter spammers[J]. Information Forensics and Security, IEEE Transactions on, 2013, 8（8）: 1280-1293.

[29] Zhu Y, Wang X, Zhong E, et al. Discovering Spammers in Social Networks[C]//AAAI. 2012.

[30] Sadan Z, Schwartz D G. Social network analysis of web links to eliminate false positives in collaborative anti-spam systems[J]. Journal of Network and Computer Applications, 2011, 34（5）: 1717-1723.

[31] Song J, Lee S, Kim J. Spam filtering in twitter using sender-receiver relationship[C]//Recent Advances in Intrusion Detection. Springer Berlin Heidelberg, 2011: 301-317.

[32] Ramachandran A, Feamster N. Understanding the network-level behavior of spammers[J]. ACM SIGCOMM Computer Communication Review, 2006, 36（4）: 291-302.

[33] Schatzmann D, Burkhart M, Spyropoulos T. Inferring spammers in the network core[M]//Passive and Active Network Measurement. Springer Berlin Heidelberg, 2009: 229-238.

[34] 吴信东, 李毅, 李磊. 在线社交网络影响力分析[J]. 计算机学报, 2014, 37（4）: 735-752.

[35] Li F, Hsieh M H. An Empirical Study of Clustering Behavior of Spammers and Group-based Anti-Spam Strategies[C]//CEAS. 2006.

[36] Gayo Avello D, Brenes Mart ńez D J. Overcoming spammers in Twitter--A tale of five algorithms[J]. 2010.

[37] Benevenuto F, Magno G, Rodrigues T, et al. Detecting spammers on twitter[C]//Collaboration, electronic messaging, anti-abuse and spam conference （CEAS）. 2010, 6: 12.

[38] Chen C, Wu K, Srinivasan V, et al. Battling the internet water army: Detection of hidden paid posters[C]//Proceedings of the 2013 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. ACM, 2013: 116-120.

[39] Mukherjee A, Liu B, Glance N. Spotting fake reviewer groups in consumer reviews[C]//Proceedings of the 21st international conference on World Wide Web. ACM, 2012: 191-200.

[40] Freeman L C. Centrality in social networks conceptual clarification[J]. Social networks, 1979, 1（3）: 215-239.

[41] Sabidussi G. The centrality index of a graph[J]. Psychometrika, 1966, 31（4）: 581-603.

[42] Newman M E J. The structure and function of complex networks[J]. SIAM review, 2003, 45（2）: 167-256.

[43] Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2（1）: 113-120.

[44] Katz L. A new status index derived from sociometric analysis[J]. Psychometrika, 1953, 18（1）: 39-43.

[45] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proc. Natl. Acad. Sci.

USA, 2001, 99（cond-mat/0112110）: 8271-8276.

[46] 丁兆云, 周斌, 贾焰,等. 微博中基于多关系网络的话题层次影响力分析[J]. 计算机研究与发展, 2013, 50（10）:2155-2175.

[47] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the national academy of sciences, 2002, 99（12）: 7821-7826.

[48] Blondel V D, Guillaume J L, Lambiotte R, et al. Fast unfolding of communities in large networks[J]. Journal of Statistical Mechanics: Theory and Experiment, 2008, 2008（10）: P10008.

[49] Shen H W, Cheng X Q. Spectral methods for the detection of network community structure: a comparative analysis[J]. Journal of Statistical Mechanics: Theory and Experiment, 2010, 2010（10）: P10020.

[50] Jiang J Q, Dress A W M, Yang G. A spectral clustering-based framework for detecting community structures in complex networks[J]. Applied Mathematics Letters, 2009, 22（9）: 1479-1482.

[51] Newman M E J, Girvan M. Finding and evaluating community structure in networks[J]. Physical review E, 2004, 69（2）: 026113.

[52] Shang R, Bai J, Jiao L, et al. Community detection based on modularity and an improved genetic algorithm[J]. Physica A: Statistical Mechanics and its Applications, 2013, 392（5）: 1215-1231.

[53] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. Nature, 2005, 435（7043）: 814-818.

[54] Ahn Y Y, Bagrow J P, Lehmann S. Link communities reveal multiscale complexity in networks[J]. Nature, 2010, 466（7307）: 761-764.

[55] Kim Y, Jeong H. Map equation for link communities[J]. Physical Review E, 2011, 84（2）: 026110.

[56] 潘磊, 金杰, 王崇骏, 等. 社会网络中基于局部信息的边社区挖掘[J]. 电子学报, 2012, 40（11）: 2255-2263.

[57] Ball B, Karrer B, Newman M E J. Efficient and principled method for detecting communities in networks[J]. Physical Review E, 2011, 84（3）: 036103.

[58] Wang D, Zhang H. Inverse-Category-Frequency Based Supervised Term Weighting Schemes for Text Categorization[J]. Journal of Information Science & Engineering, 2013, 29（2）:209-225.

[59] 裴颂文, 吴百锋. 动态自适应特征权重的多类文本分类算法研究[J]. 计算机应用研究, 2011, 28（11）:4092-4096.

[60] Blei D M, Ng A Y, Jordan M I. Latent dirichlet allocation[J]. Journal of Machine Learning Research, 2003, 3:993-1022.

[61] Hofmann T. Probabilistic Latent Semantic Indexing[C]// In Proc Sigir. 2004:56-73.

[62] Shen H, Cheng X, Cai K, et al. Detect overlapping and hierarchical community structure in networks[J]. Physica A-statistical Mechanics & Its Applications, 2009, 388（8）:1706-1712.

[63] Adamcsek B, Palla G, Farkas I I, et al. CFinder: locating cliques and overlapping modules in biological networks[J]. Bioinformatics, 2006, 22（8）:1021-1023.

[64] Amleshwaram A A, Reddy N, Yadav S, et al. CATS: Characterizing automation of Twitter spammers[C]// Communication Systems and Networks （COMSNETS）, 2013 Fifth International Conference on. IEEE, 2013:1-10.

[65] Wang K, Xiao Y, Xiao Z. Detection of Internet water army in social network[C]//Proc. of the 2014 Int'l Conf. on Computer, Communications and Information Technology （CCIT 2014）. Amsterdam: Atlantis Press. 2014: 189-192.

附录 1 攻读硕士学位期间撰写的论文

（1）第二作者，计算机应用，已录用

（2）第二作者，SCIENCE CHINA Information Sciences，已投稿

致谢

时光飞逝，岁月荏苒，两年多的硕士研究生生涯即将结束。回想起整个研究生阶段，南京邮电大学给我留下了无数美好的回忆，我在这里学习到了先进的科学知识、认识了更多的朋友，也掌握了可以报效祖国的技能。在离别之际，情不自禁的流露出对校园生活的无限留恋。

经过两年多的科研工作和几个月来的论文写作与修改，我的毕业论文终于要完成了。在这里，我首先要深深的感谢我的导师周国强教授，是他一步一步指引我走进科学研究的世界，他的指导让我少走了很多的弯路；周老师对我的研究方向提出了很多宝贵的意见，我们之间的讨论让我学习到了很多先进的思路和研究方法；最后的毕业论文以及研究生阶段小论文的完成，都离不开周国强老师的帮助。周老师为人和蔼与我们是很好的朋友，对待学术研究十分严谨认真，在生活和学习上都给予了我很大的帮助，非常荣幸可以遇到这样的导师。同时我也要感谢大师门的张迎周、张卫丰、王子元老师，他们都曾在我的研究生阶段中提供了很多指导和照顾。从他们身上我学习到了宝贵的科学经验和坚忍不拔的科研态度，非常感谢！

另外，我还要感谢我的师兄刘旭、张文聪、林鹏的帮助，他们不仅在科研工作上给了我无数的科学经验和指导，还在生活上给我提供了很多无私的帮助。感谢仉雪玲、汪矿、陆洋、刘鸿舫、周洪飞、赵晨、崔丽娟等师门的兄弟姐妹，正是因为你们的存在，我两年多的研究生生活才能如此的丰富多彩。感谢那些在我生命中不顺时出现的朋友们，如果没有你们就不会有现在的我，朋友们谢谢！

我还要特别感谢我的家人，你们辛苦养育我 20 多年，在背后默默支持着我，不求回报为我付出了很多，非常感激你们，在这里说一声：你们辛苦了，我不会让你们失望的。

最后，大论文终于完成，研究生生涯也即将结束，但学习是永不会停止的，我会更加努力不断提升自己。感谢各位专家能在百忙之中抽出时间来对我的文章进行审阅和指导，谢谢！