**西南石油大学研究生学位论文**

**开题报告**

**论文题目** 新浪微博数据消息传播机制的研究

**院 系** 计算机科学学院

**专 业** 计算机科学与技术

**年 级** 2013

**姓 名** 赵培

**导 师**  李平

填表时间 2015 年 6 月 29 日

|  |
| --- |
| 1. 目的意义和国内外研究概况 2. 目的意义   微博作为一种网络分享和交流平台，从2009年兴起并迅速风靡。如今各种社会媒体、政府机关、公众人物纷纷开通微博，微博已走入更多普通大众的生活，影响力越来越大。其中蕴含了海量的信息值得挖掘，越来越多的学界和业界的研究者开始把目光投向微博。  微博的传播机制是其中一个值得研究的课题。微博的传播机制，就是微博信息传播的制度与其各要素职能和联系的建立与运作，是包括传播者、传播过程、传播模式以及受众等所有构成的统一体，具有其独特性。微博的传播机制下，传播的过程具有五个要素：传播者、信息、媒介、受众、反馈，以它们为研究点可以对微博的传播过程进行系统地梳理。使用传播模式的方法可以来说明它们在传播过程中起到的作用以及相互之间的联系。  研究微博的信息传播机制对于链路预测、舆情监测、网络控制、社团划分等有重要意义。   1. 国内外研究概况   微博是一种在线社会网络（Online Social Networks，简称 SNS），国外比较著名的有 Fackbook、LinkedIn、Twitter 等，国内有新浪微博、腾讯微博、微信朋友圈、QQ空间等。SNS 的巨大影响力，其已经引起了众多国内外学者的关注。  哈佛大学心理学教授Milgram（1967）通过邮件传递实验得到著名的“六度分隔”理论。在2011年5月，研究人员发现全球Fackbook上两个用户之间的平均距离仅为4.74，比“六度分离”还要小。Golder 等人（2007）对 Facebook 数以百万的数据进行分析，研究好友和信息流动网络，发现了许多有价值的规律。Damon（2010）研究了用户行为在 SNS 中的扩散现象，发现网络结构对用户的行为有很大影响。  国外对微博的研究，以对 Twitter 的研究为主。在计算机信息技术领域，对微博的研究成果，主要包括信息处理、数据挖掘。Fiaidhi 等人（2012）研究了通过一系列算法和用户的 RSS 订阅识别 Twitter 中流行主题的方法，识别的流行主题要比 Twitter 推荐的更为准确。Kwak（2010）统计了 Twitter 中 4262 个流行主题、1 亿多条信息，发现 85%的流行主题是重要新闻事件，一条信息一旦被评论，就将快速的在网络中扩散。Kristina（2010）等研究了新闻事件在 Twitter 和 Digg 中的传播，发现社会网络在信息的传播中，起到关键角色，网络结构会影响信息传播。Tonkin等人（2012）通过案例研究了 Twitter 在医疗领域的应用，通过分析 Twitter 信息预测癫痫发作，虽然预测准确率较低，但却拓展了微博应用的领域，很有实践意义。  在中国知网核心期刊中，以微博为主题进行搜索可以得到189,155 篇文献（截止 2015 年 7 月）。而且对于微博的研究出现了逐年稳健增长的趋势。国内对微博的研究，多数是从新闻传播的角度出发，定性分析微博对信息传播的影响。邢立双（2012）研究了突发事件中微博信息的传播与管理，认为在可控制的情况下，让微博中意见领袖的思想和方法广泛传播，可以提高整个社会的智商，加快中国现代化的进程。王嘉（2012）研究了微博中话语权的问题，分析了微博中话语权的内涵、外延和分配问题，并对微博的发展趋势做了进一步的构想。许晓东等人（2011）研究了谣言在微博中传播的问题，从宏观角度分析了谣言在网络中的传播机理，构建模型并进行了仿真。平亮等人（2011）利用社会网络理论中的中心度和中心性等指标，研究了微博中意见领袖、权力等问题。此外，计算机领域的学者，从微博中信息数据分析、热点话题挖掘方面，也进行了一定的研究。 |
| 1. 论文的理论依据、研究方法、研究内容 2. 论文的理论依据   社会网络分析（Social network analysis）: 社会网络分析是研究一组行动者的关系的研究方法。一组行动者可以是人、社区、群体、组织、国家等，他们的关系模式反映出的现象或数据是网络分析的焦点。从社会网络的角度出发，人在社会环境中的相互作用可以表达为基于关系的一种模式或规则，而基于这种关系的有规律模式反映了社会结构，这种结构的量化分析是社会网络分析的出发点。  统计分析（statistical analysis）：在相关科学理论的指导下，利用统计调查并整理所掌握的大量资料及相关资料，运用统计的方法，对客观存在的社会经济现象及与之相关联的现象，进行分析研究，探求其发展变化的规律，为社会经济活动的实际决策提供依据。   1. 研究方法 2. 文献研究方法   通过对国内外文献进行梳理，对复杂网络理论、信息  传播理论、传播心理学理论等有了较深刻的认知。   1. 网络分析方法   运用了复杂网络理论中的网络拓扑结构分析法，选取多个指标，对微博信息传播复杂网络的拓扑结构进行了分析。   1. 建模和仿真分析方法   结合微博自身特征，对网络演化和信息传播过程，分别进行建模，并利用软件对其进行仿真。   1. 可视分析方法   通过运用可视化技术，将数据从不同维度展示出来，以发掘其中蕴含的规律，为深入研究提供突破口。   1. 研究内容   本文的主要研究内容是对新浪微博数据进行分析，发掘其中的消息传播机制。通过研究消息的传播深度、广度、速度与消息类型、消息参与者的好友关系等之间的联系和影响，得到网络消息传播的一般规律，并建立模型，根据实际数据训练相关参数，达到预测消息传播的最终目的。并根据部分数据来预测即将可能发生的好友关系，并与实际数据相印证，揭示在线社交网络信息传播的规律性、可预测性。 |
| 1. 研究条件和可能存在的问题 2. 研究条件   本研究在智能与网络化系统研究中心（Center for Intelligent and Networked Systems）进行，该中心主要从事网络科学、机器学习、大数据下的网络挖掘、中文自然语言处理等交叉学科的应用基础研究。承担了包括国家自然科学基金在内的研究项目，在权威国际期刊上发表多篇SCI论文。中心拥有100平方的独立实验室，2台专用高性能服务器，以及多台PC机，Hadoop云计算平台，具有良好的科学研究环境。   1. 可能存在的问题   （1）数据量过大硬件不足问题  （2）微博消息语义理解问题  （3）预测效率问题 |
| 1. 预期的结果 2. 得到“新浪微博”消息传播的一般规律 3. 得到“新浪微博”消息传播的预测模型 |
| 1. 进度安排   2015.7 – 2015.9 根据课题的要求，文献收集、理论研究，分析微博数据，进行实验  2015.10 – 2015.12 结合理论知识与实验结果，撰写论文初稿  2016.1 – 2016.3 对论文进行修改、定稿  2016.4 – 2013.6 论文审查、答辩 |
| 1. 论证小组论证意见：     论证小组签名：组长  成员  秘书  年 月 日 |