

学校编码: 10384

分类号\_\_\_\_\_密级\_\_\_\_\_

学号: X2012230236

UDC\_\_\_\_\_

厦门大学

工程硕士学位论文

微博网络消息传播仿真系统的设计与实现

Design and Implementation of Weibo Message Propagation  
Simulation System

赵清

指导教师: 王备战教授

专业名称: 软件工程

论文提交日期: 2014年10月

论文答辩日期: 2014年11月

学位授予日期: 年 月

指导教师: \_\_\_\_\_

答辩委员会主席: \_\_\_\_\_

2014年 月

# 厦门大学学位论文原创性声明

本人呈交的学位论文是本人在导师指导下，独立完成的研究成果。本人在论文写作中参考其他个人或集体已经发表的研究成果，均在文中以适当方式明确标明，并符合法律规范和《厦门大学研究生学术活动规范（试行）》。

另外，该学位论文为（ ）课题（组）的研究成果，获得（ ）课题（组）经费或实验室的资助，在（ ）实验室完成。（请在以上括号内填写课题或课题组负责人或实验室名称，未有此项声明内容的，可以不作特别声明。）

声明人（签名）：

年 月 日

# 厦门大学学位论文著作权使用声明

本人同意厦门大学根据《中华人民共和国学位条例暂行实施办法》等规定保留和使用此学位论文，并向主管部门或其指定机构送交学位论文（包括纸质版和电子版），允许学位论文进入厦门大学图书馆及其数据库被查阅、借阅。本人同意厦门大学将学位论文加入全国博士、硕士学位论文共建单位数据库进行检索，将学位论文的标题和摘要汇编出版，采用影印、缩印或者其它方式合理复制学位论文。

本学位论文属于：

1.经厦门大学保密委员会审查核定的保密学位论文，  
于 年 月 日解密，解密后适用上述授权。

2.不保密，适用上述授权。

（请在以上相应括号内打“√”或填上相应内容。保密学位论文应是已经厦门大学保密委员会审定过的学位论文，未经厦门大学保密委员会审定的学位论文均为公开学位论文。此声明栏不填写的，默认为公开学位论文，均适用上述授权。）

声明人（签名）：

年 月 日

## 摘要

近年来，以微博为代表的社交网络得到了迅速的发展，例如国内的新浪微博和国外的 Twitter。热点问题在很短的时间内就得到广泛传播。对于微博传播行为的分析和预测就可以探测热点话题，以及可以监测和防止垃圾信息。

本文以微博用户间形成的复杂网络系统为研究对象，以现有的复杂网络的研究成果为基础，着重研究微博信息的传播行为。有趣的是：微博信息的传播和传染病病毒的传播行为是很相似的。病毒模型是医学领域的研究成果，基于这些模型可以预测未来的微博转发趋势。所以本文具体会研究三大基本传染病病毒模型 SI (Susceptible Infections), SIS (Susceptible Infections Susceptible), SIR (Susceptible Infections Recovery)。同时，对于不同的微博来说，每个博主的影响力是不同的，这个因素也会影响到微博的传播行为。文中对于两大影响力模型 IC (Independent Cascade) 和 LT (Linear Threshold) 模型也进行了一些研究。论文的主要工作表现在以下几个方面：

- (1) 根据现有的五大模型，实现算法；
- (2) 对于影响力模型中的影响因素，对于影响概率计算提出自己的计算公式；
- (3) 对复杂网络中的一些参数的计算；
- (4) 对微博信息的传播过程进行可视化。

本文将仿真微博传播行为，可视化出微博的传播路径，可以更清晰的看到在不同的模型下，微博的传播行为的异同。

**关键词：**微博；消息转发；仿真系统

## Abstract

Recently, the social networks develop fast as a representative of Micro Blog, like Sina micro blog in China, and Twitter in foreign countries. Hot topics currently being propagated can quickly become popular within a very short time. Analyzing and predicting the propagation behavior can detect hot topics, monitor and prevent spam message.

This dissertation studies up on the complex network system which is consist of the users of Sina micro blog. We base on the existing research results of complex network. We concentrate on studying the propagation behavior of micro blog. Interestingly, the propagation of message bears some similarity to the spread of infections diseases. The epidemic is the results of medical field. On the basis of these models, future propagating trends can be predicted. In this paper, we will study three epidemic models, SI (Susceptible Infections), SIS (Susceptible Infections Susceptible), SIR (Susceptible Infections Recovery). At the same time, to different micro blog, every blog user has different influence to others. This factor can also influence the propagation behavior of micro blog. So, we also study the two influence models namely IC (Independent Cascade) and LT (Linear Threshold). The main achievement we made as follows:

1. Achieve the placement algorithm basing on the five models.
2. Puts the Formula of influence factor in influence models.
3. Compute the parameters of the complex network system.
4. Visualize the process of propagation behavior.

This dissertation will simulate the process of propagation behavior. The propagate path can be visual. We can clearly observe the different propagation behaviors in different models.

**Key Words:** Micro Blog; Message Forwarding; Simulation System

## 目 录

<b>第一章 绪论</b> .....	<b>1</b>
<b>1.1 研究的意义与背景</b> .....	<b>1</b>
<b>1.2 国内外研究现状</b> .....	<b>2</b>
1.2.1 复杂网络研究现状 .....	2
1.2.2 消息传播模型研究现状 .....	4
<b>1.3 论文的组织结构</b> .....	<b>5</b>
<b>第二章 系统需求分析</b> .....	<b>6</b>
<b>2.1 传染病模型</b> .....	<b>6</b>
2.1.1 SI 模型 .....	6
2.1.2 SIS 模型 .....	7
2.1.3 SIR 模型 .....	9
<b>2.2 影响力模型</b> .....	<b>10</b>
2.2.1 独立级联模型 (IC 模型) .....	11
2.2.2 线性阈值模型 (LT 模型) .....	12
<b>2.3 系统功能分析</b> .....	<b>14</b>
2.3.1 系统仿真参数定义功能 .....	15
2.3.2 核心功能 .....	16
2.3.3 图形可视化功能 .....	17
<b>2.4 功能模块划分</b> .....	<b>17</b>
2.4.1 微博网络初始化模块 .....	19
2.4.2 微博信息发送引擎 .....	20
2.4.3 修改概率 .....	21
2.4.4 保存微博传播图 .....	21
<b>2.5 本章小结</b> .....	<b>22</b>
<b>第三章 系统设计</b> .....	<b>23</b>

<b>3.1 功能模块设计</b> .....	<b>23</b>
3.1.1 初始化微博网络.....	23
3.1.2 发送微博信息模块.....	24
3.1.3 动态显示微博传播路径模块.....	27
3.1.4 保存微博传播图模块.....	28
3.1.5 修改概率模块.....	29
3.1.6 其他功能模块.....	30
<b>3.2 流程图</b> .....	<b>30</b>
3.2.1 总流程图.....	30
3.2.2 初始化微博网络流程图.....	31
3.2.3 发送微博信息流程图.....	32
3.2.4 修改概率.....	34
3.2.5 动态显示图形.....	35
3.2.6 显示节点信息.....	37
3.2.7 增加节点及编辑节点信息.....	38
3.2.8 重新加载新图文件.....	39
3.2.9 保存图信息.....	40
<b>3.3 本章小结</b> .....	<b>41</b>
<b>第四章 系统实现</b> .....	<b>42</b>
<b>4.1 系统开发环境</b> .....	<b>42</b>
<b>4.2 系统算法实现</b> .....	<b>42</b>
4.2.1 传染病模型和影响力模型算法.....	42
4.2.2 微博节点消息转发算法.....	44
4.2.3 微博传播路径动态显示算法.....	48
<b>4.3 系统测试</b> .....	<b>51</b>
4.3.1 基本功能测试.....	52
4.3.2 不同模型结果对比分析结论.....	60
<b>4.4 本章小结</b> .....	<b>64</b>
<b>第五章 总结与展望</b> .....	<b>65</b>

5.1 总结.....	65
5.2 展望.....	66
参考文献.....	67
致谢.....	68

厦门大学博硕士论文摘要库



---

## Contents

<b>Chapter 1 Introduction .....</b>	<b>1</b>
<b>1.1 Research Background and Significance.....</b>	<b>1</b>
<b>1.2 Research Status .....</b>	<b>2</b>
1.2.1 Complex Networks Research Status .....	2
1.2.2 Complex Networks Research Status .....	4
<b>1.3 Thesis Structure.....</b>	<b>5</b>
<b>Chapter 2 System Requirement Analysis .....</b>	<b>6</b>
<b>2.1 Infectious Disease Model .....</b>	<b>6</b>
2.1.1 SI Model.....	6
2.1.2 SIS Model .....	7
2.1.3 SIR Model.....	9
<b>2.2 Influence Model.....</b>	<b>10</b>
2.2.1 IC Model .....	11
2.2.2 LT Model.....	12
<b>2.3 System Function Analysis.....</b>	<b>14</b>
2.3.1 System Parameter Settings.....	15
2.3.2 Key Function.....	16
2.3.3 Visualization Function .....	17
<b>2.4 System Modules Division.....</b>	<b>17</b>
2.4.1 Weibo Networks Initialization .....	19
2.4.2 Weibo Message Forwarding Engine .....	20
2.4.3 Modify Possibility.....	21
2.4.4 Saving Message Forwarding Picture .....	21
<b>2.5 Summary.....</b>	<b>22</b>
<b>Chapter 3 System Design .....</b>	<b>23</b>

---

<b>3.1 Function Modules Design .....</b>	<b>23</b>
3.1.1 Weibo Networks Initialization .....	23
3.1.2 Weibo Message Forwarding.....	24
3.1.3 Message Path Visualization .....	27
3.1.4 Saving Messag Forwarding Picture .....	28
3.1.5 Modifying Possibility.....	29
3.1.6 Miscellaneous Modules .....	30
<b>3.2 Flow Process Diagram .....</b>	<b>30</b>
3.2.1 Basic Flow Process Diagram .....	30
3.2.2 Weibo Networks Initialization Process .....	31
3.2.3 Weibo Message Forwarding Process .....	32
3.2.4 Modifying Possibility.....	34
3.2.5 Dynamic Visualization .....	35
3.2.6 Display Node Information .....	37
3.2.7 Add/Edit Node Information .....	38
3.2.8 Reload Image Files .....	39
3.2.9 Saving Image Files.....	40
<b>3.3 Summary.....</b>	<b>41</b>
<b>Chapter 4 System System Algorithms .....</b>	<b>42</b>
<b>4.1 Development Environment.....</b>	<b>42</b>
<b>4.2 System Algorithms System Algorithms .....</b>	<b>42</b>
4.2.1 Infectious Disease and Influence Algorithms .....	42
4.2.2 Weibo Node Message Forwarding Algorithms .....	44
4.2.3 Weibo Message Path Display Algorithms.....	48
<b>4.3 System Test .....</b>	<b>51</b>
4.3.1 Basic Function Test .....	52
4.3.2 Simulation Results and Analysis.....	60
<b>4.4 Summary.....</b>	<b>64</b>
<b>Chapter 5 Conclusions and Prospect .....</b>	<b>65</b>

<b>5.1 Conclusions</b> .....	<b>65</b>
<b>5.2 Prospect</b> .....	<b>66</b>
<b>References</b> .....	<b>67</b>
<b>Acknowledgements</b> .....	<b>68</b>

厦门大学博硕士论文摘要库

## 第一章 绪论

### 1.1 研究的意义与背景

微博，又叫微博客，英文名 Micro Blog，是 Web 2.0 时代的产物。微博可以看做是另一种形式的博客，微博与博客的区别在于微博的文本内容往往被限制在一定范围内（通常为 140 个汉字）。微博允许用户通过融合包括手机、博客、网页、即时通讯、论坛和 SNS 社区等多种渠道，来发布文字、图片、视频和音频等形式的富媒体信息。微博与传统博客相比，微博具有短小、灵敏和快速的特点。微博用户之间存在两种关系，包括粉丝关系和好友关系。例如，当用户张三关注了用户王五，那么张三便成为了王五的粉丝，满足是王五的粉丝关系；而当王五同样关注张三后，王五便成为了张三的好友，二者间便满足了好友关系。基于用户之间的粉丝和好友关系，微博用户间构成了一个复杂的社交网络。

自 2006 年 3 月，博客创始人埃文·威廉姆斯（Evan Williams）在全球首次推出微博服务“Twitter”后，微博在全球各地如雨后春笋般流行开来。2007 年 5 月，王兴成立了中国首个微博类网站“饭否”。2007 年 8 月 13 日，腾讯推出了自己的微博产品——腾讯滔滔。2009 年 8 月，新浪推出新浪微博。得益于其名人微博的营销策略，新浪微博在短时间内迅速成为中国用户量最大，活跃度最高的微博。截止至 2012 年 12 月底，中国微博用户数量达到 3.09 亿。与 2011 年底相比，用户数量增长了近 20%。

伴随着微博在国内市场的日益火爆，其在社交、经济、公关和社会舆论方面的影响力也越来越重要。微博平台不仅仅成为了人们日常交流和沟通的平台，更是成为了一个网络内容的产生和分发平台。越来越多的企业、政府机构在微博网络上开辟了自己的微博账户，用以进行产品宣传、信息发布和挖掘潜在的价值客户。为了有效挖掘微博背后的蕴含的巨大商业和社会价值和意义，越来越多的科研机构和企业部门开始对微博的信息传播机制进行研究。通过尝试研究微博网络的拓扑结构，微博消息的传播规律和特征，来更好的实现利用微博完成信息宣传的目标。

本文尝试将新浪微博网络建模为一个统一的复杂网络，并以之为研究对象，

深入挖掘微博用户间传递消息的特征,并进行相应的理论建模和仿真实验分析,以最终掌握微博消息的传播规律。本研究对掌握微博消息在网络的传播路径、传播时效和传播特征,具有重要的指导作用。在微博对当前社会舆论影响力日益增大的今天,本研究对社会舆情分析、商业信息推广以及社交类媒体的有效管控,同样有着较强的现实意义和经济意义。

## 1.2 国内外研究现状

本节将对研究微博消息传递模型中所采用的关键技术,及其国内外研究现状进行介绍。首先对于微博网络,由于其满足小世界性、自相似性、吸引性和自组织性,本文将微博网络建模为一种复杂网络。在此基础上,本文研究具体用户消息在该复杂网络模型上的传播路径、传播速率等传播规律。为了有效分析微博消息的传播特征,本文引入了传染病模型和影响力模型对微博消息的传播特征进行建模分析。将疾病的传播过程和微博消息的传播过程进行拟合,基于现有的疾病传播模型对微博消息传播过程建模,并最终得出其传播规律。1.2.1节首先分析了复杂网络的研究现状;1.2.2节则对传染病模型和影响力模型的研究现状进行介绍。

### 1.2.1 复杂网络研究现状

复杂网络(Complex Network)是指具有小世界性、自相似性、自组织性、吸引性、无标度性中一部分或全部性质的网络。

复杂网络是对复杂系统的一种更高层次的抽象,复杂系统中的个体被抽象为网络中的节点,系统中个体间按照某种规则而形成或人为构造的关联关系则被抽象为节点之间的边。现实世界中的各种类型网络均可以被看作复杂网络模型,如图 1-1 所示的朋友关系网络及合作网络等为代表的社会网络,如图 1-2 所示的以万维网以及电力网等为代表的技术网络。

近年来,复杂网络得到学术界学者的广泛研究。特别是,国际上有两项开创性工作掀起了一股不小的研究复杂网络的热潮:一是 1999 年 Barabas 和 Albert 在《Science》杂志上发表文章,他们发现从连接度分布的角度分析,许多实际复杂网络满足幂律形式的特点。由于幂律分布缺乏明显的特征长度,该类网络

又被称作无标度 (Scale-Free) 网络；二是 Watts 和 Strogatz 于 1998 年在《Nature》杂志上首次提出了小世界 (Small-World) 网络模型的概念，以解释完全规则网络和完全随机网络之间的转变规则。小世界网络拥有与规则网络相类似的聚类特点，同时具有与随机网络相类似的较小的平均路径长度。

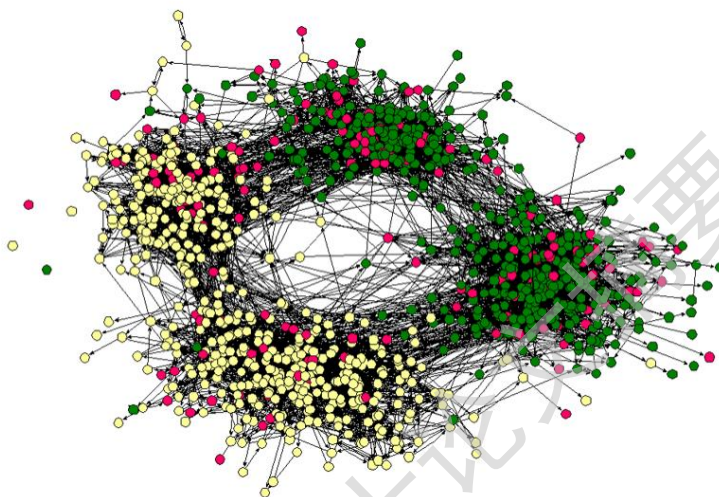


图 1-1 朋友网

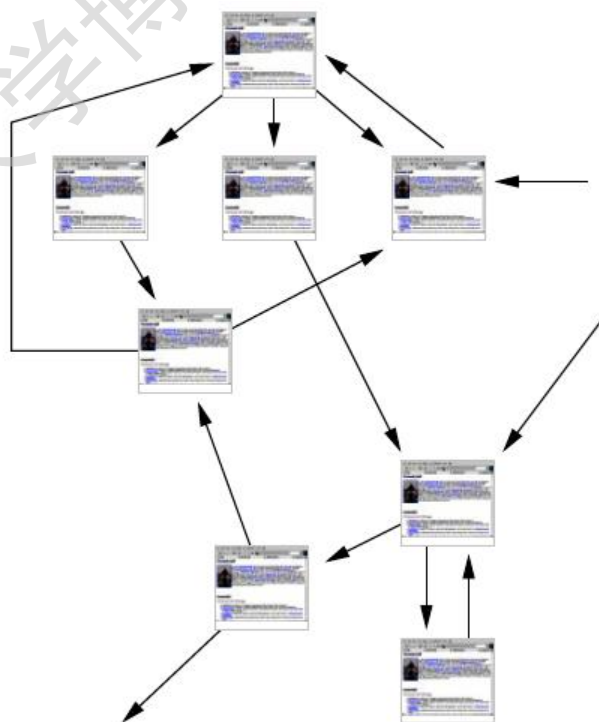


图 1-2 万维网

### 1.2.2 消息传播模型研究现状

由于微博信息在人群中的传播过程与传染病在人群中的传播过程类似, 均具有好友传播和集中扩散的特点, 很多关于微博信息转发模型的研究往往基于传染病模型和影响力模型两类。其中传染病模型包括 SI 模型, SIS 模型, SIR 模型三类基本模型, 影响力模型包括 IC 和 LT 模型两类基本模型。

自从 20 世纪以来, 对于病毒传播模型的研究逐步得到了广泛关注。1906 年, Hamer 为构建了针对麻疹疾病的离散时间模型。1911 年, Ross 首次提出一种不同的模型区, 用以研究疟疾在蚊子和人群中的传播方式。他最终证明, 如果将蚊子数量限制在一个阈值下, 疟疾是可以得到有效控制。这项研究成果使得它获得诺贝尔医学奖。1927 年, Kermack 和 Mckendrick 提出了著名的 SIR 模型。在《The Spread of Epidemic Disease on Networks》<sup>[1]</sup>中, M.E.J .Newman 对 SIR 模型进行了详细的分析论证。其研究成果使得 SIR 模型被应用于描述复杂网络上微博信息的传播过程成为可能。清华大学 WANG Hao 等人在论文《ReTweeting Analysis and Prediction in Microblogs--An Epidemic Inspired Approach》<sup>[2]</sup>中提出两种改进的传染病模型: 为 SISE 和 SISE+模型。针对基本的传染病模型中采用不变人口数和固定传播概率的这一假设缺陷, 他们在论文中采用了动态的传播概率, 使得分析预测结果更接近真实结果, 错误率更低。实验的结果表明, 如果传播概率是动态变化的, 当转发数量增大时, SISE+模型比 SISE 模型预测效果更好, 但是当转发量减少时, SISE 模型比 SISE+模型预测效果更好。但从整体结果来说, 上述两种模型的实验结果均比原有基本模型的预测结果更加准确。张发等《传染病传播模型综述》<sup>[3]</sup>中对各种传染病模型的特点进行总结比较。对于每种模型适用于那些领域, 以及需要解决那些问题, 怎样建模都进行了分析。丁玲芬的《几类传染病模型的定性分析》<sup>[4]</sup>中讨论一种传染病模型的变种 SEIR 模型, 一类具有垂直传染非自治的 SEIR 传染病模型, 以及一类具有垂直传染总人口数变化非自治的 SEIR 传染病模型。Luke Dickens 等在《Learning stochastic models of information flow》<sup>[5]</sup>中提出了一个可扩展的方法 用于一个基于独立级联模型的网络中的信息传播学习模型和预测信息传播的可能路径。参考文献 [6] 使用 MACD ( Moving Average Convergence-Divergence) 去预测趋势。文献[7]在 Twitter 上使用一种探测主题

的方法，通过时间，关键字和关系等因素来建立模型。

当然还有其他的一些模型例如：南京大学 DeChun Liu 发表的论文《Rumor Propagation in Online Social Networks Like Twitter -- A Simulation Study》<sup>[8]</sup>提出的 Rumor Propagation。对于网络中消息的传播也进行了分析与预测。并且取得了良好的结果。

### 1.3 论文的组织结构

第一章：绪论。本章主要介绍复杂网络的研究现状，同时介绍微博传播的一些理论模型。

第二章：系统需求分析。本章介绍常见的几种模型。这些模型包括，传染病模型，SI, SIS, SIR, 影响力模型，IC 和 LT 模型，介绍其基本理论，模型效果已经应用大微博传播的过程中的情况。然后对系统进行了功能需求分析，并进行了功能模块划分。

第三章：系统设计。本章主要是介绍本仿真系统的功能分析和设计思路，包括系统的功能模块设计、实现思路简介流程图等。

第四章：系统实现。本章主要介绍本系统的开发环境和具体实现过程，并对系统进行了测试。

第五章：总结与展望。本章对全文工作进行了总结，并对下一步工作进行展望。



Degree papers are in the "[Xiamen University Electronic Theses and Dissertations Database](#)". Full texts are available in the following ways:

1. If your library is a CALIS member libraries, please log on <http://etd.calis.edu.cn/> and submit requests online, or consult the interlibrary loan department in your library.
2. For users of non-CALIS member libraries, please mail to [etd@xmu.edu.cn](mailto:etd@xmu.edu.cn) for delivery details.

厦门大学博硕士论文摘要库