# 社交网络信息传播

# SOCIAL NETWORK INFORMATION DIFFUSION

张熙　编著

中国工信出版集团

電子工業出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
http://www.phei.com.cn

# 社交网络信息传播

张　熙　编著

電子工業出版社
Publishing House of Electronics Industry
北京 · BEIJING

## 内 容 简 介

社交网络信息传播是计算机科学、传播学、社会学、管理学等领域的重要研究问题，在舆情分析和网络营销领域具有广泛的应用。目前，同类著作更多地站在传播学或管理学角度介绍信息传播的模型、原理和应用。而本书主要从计算机科学角度出发，介绍了该领域的经典问题和最新成果，包括传播模型、话题检测、影响力最大化等问题。此外，本书面向实际应用场景，阐述了如何开发舆情分析和网络营销系统。

本书可供社交网络分析与数据挖掘研究领域的研究者了解该方向的前沿基础工作，也可供信息传播与网络舆情领域的工程实践人员作为系统构建的参考和指导。

**未经许可，不得以任何方式复制或抄袭本书之部分或全部内容。**
**版权所有，侵权必究。**

**图书在版编目（CIP）数据**

社交网络信息传播 / 张熙编著. —北京：电子工业出版社，2016.8

ISBN 978-7-121-29783-0

Ⅰ. ①社… Ⅱ. ①张… Ⅲ. ①互联网络－信息－传播－研究 Ⅳ. ①G206

中国版本图书馆 CIP 数据核字（2016）第 205185 号

责任编辑：徐蔷薇
特约编辑：赵海军　赵海红等
印　　刷：三河市鑫金马印装有限公司
装　　订：三河市鑫金马印装有限公司
出版发行：电子工业出版社
　　北京市海淀区万寿路 173 信箱　　邮编：100036
开　　本：720×1000　1/16　　印张：13.75　　字数：220 千字
版　　次：2016 年 8 月第 1 版
印　　次：2016 年 8 月第 1 次印刷
定　　价：59.00 元

凡所购买电子工业出版社图书有缺损问题，请向购买书店调换。若书店售缺，请与本社发行部联系，联系及邮购电话：（010）88254888，88258888。

质量投诉请发邮件至 zlts@phei.com.cn，盗版侵权举报请发邮件至 dbqq@phei.com.cn。

本书咨询联系方式：xuye@phei.com.cn。

# 前　言

随着互联网进入 Web 2.0 时代，以新浪微博、网络社区、Twitter 和 Facebook 为代表的社交网络得到飞速发展，信息的传播速度更快、影响范围更广，正在深刻改变着人们的思维方式、行为模式和社会形态。深入理解社交网络中的信息传播模式和规律具有重要的科学价值，如能将其合理利用，将带来巨大的经济价值和社会价值。

社交网络信息传播涉及计算机科学、传播学、社会学、管理学和心理学等多个学科领域。目前，同类著作更多地站在传播学或管理学角度介绍信息传播的模型、原理和应用，而本书主要从计算机科学角度出发，基于近些年在数据挖掘和社交网络分析领域的研究经历与相关成果，系统梳理了社交网络信息传播的经典问题和最新研究成果。另外，面向实际应用中的需求，介绍了如何实现对传播信息和网络舆情的监测、分析和处理。

本书分为上、下两篇共 7 章。上篇从理论研究出发，第 1 章传播模型，介绍了社交网络中信息的两种传播模型，分别解释模型和预测模型；第 2 章热门话题检测，介绍了几种话题检测的算法，并结合实例进行了分析和对比；第 3 章影响力最大化，总结分析了几种社交网络影响力最大化传播模型及其优化算法；第 4 章收益最大化，介绍并分析了营销模型及策略，描述了相关的算法。下篇从工程实践出发，介绍了作者团队近年来开发的网络舆情监测系统（第 5 章）、品牌推荐和保护系统（第 6 章），以及其中涉及的一项核心技术——网站验证码识别（第 7 章）。

本书可供社交网络分析与数据挖掘研究领域的研究者了解该方向的前沿基础工作，也可供信息传播与网络舆情领域的工程实践人员作为系统构建的参考和指导。

感谢参与本书内容讨论、资料收集、内容编纂、成果贡献和审查校对的北京邮电大学可信分布式计算与服务教育部重点实验室的老师和同学：吴旭老师和颉夏青老师，博士生苏援和许晋，硕士生侯玉锋、李金兰和曲思宇，以及北京邮电大学国际学院的高炘、麦艺琼、吕浩然和郭鲲鹏同学。感谢“973”项目“社交网络分析与网络信息传播的基础研究”对本书的支持。

由于作者水平有限，书中难免有错误和疏漏之处，恳请读者批评指正。



此为试读，需要完整PDF请访问：www.ertongbook.com

# 目　录

## 上篇　理论研究

此为试读，需要完整PDF请访问：www.ertongbook.com

# 上篇

## 理论研究

# 第 1 章　传播模型

## 1.1　引言

近年来，随着社交网络的发展，对于信息传播模型的研究一直很活跃。本章介绍一些基本的传播模型，并描述这些模型如何推断出底层传播级联机制或预测消息传播过程。

在流行病学领域，对复杂系统中传染病传播过程的研究已经持续了几个世纪，例如，在某些条件下病毒增殖传播的预测。在线社交网络信息传播领域的研究也广泛地借鉴了流行病学的研究方法，但是过程更加复杂。我们不能直接套用传染病模型，一方面原因是在线社交网络规模非常大，网络更新和传播速度也更快，而很多原有模型的效率太低，难以实际应用；另一方面是用户类型和消息类

型更加多样，各个网络平台的传播规则也不尽相同，需要设计新的模型。

信息传播的预测具有广阔的应用场景，如市场营销、安全监测和网络搜索等。例如，对于市场营销来说，如果我们知道哪些特征主导传播过程，就可以更好地宣传产品或者保护其不受到网络攻击。同时，市场营销也可以通过合理选择初始投放广告节点来使得收益最大化，或者通过确定营销行动之间的时延来获利。另外，在安全维护的场景下，刑事调查员通常需要了解特定成员之间的信息流，以提取关于一个人或一群人是否有犯罪嫌疑的线索[16]。最后，对于 Web 搜索，一个传播预测模型可以帮助用户根据某话题热度的预期来增长订阅最热门的话题。这些都反映了传播预测模型的广泛作用。

传播过程的特征由两方面描述：第一个方面为结构，用一幅网络图描述出哪些节点间可以相互影响，网络的拓扑结构是怎样的；第二个方面为动态变化，如传播速率的演变，即在一段时间内接收某条消息的节点数量。

描述信息传播过程的基本方法是考虑网络中的一个节点是否可以被信息激活。因此，传播过程可以看作节点连续激活的序列，具体查看定义 1.1。

定义 1.1：（激活序列）网络中的一组有序节点连续接收某条消息，这组节点序列被称为激活序列。

在通常情况下，在线社交网络（Online Social Network，OSN）背景下的模型都只假设用户只接受相互连接邻居的影响。也就是说，一个 OSN 是一个封闭的世界，并且假设信息级联导致了信息的传播。这也是为什么网络中一条消息的路径通常被称为传播级联，如定义 1.2 所示。

定义 1.2：（传播级联）有向树的根可以作为激活序列的第一个节点，这棵树表示了节点之间的影响关系（有向边显示了信息传播方向），并且以激活序列依次展开。

激活序列如图 1-1 所示，黑色节点表示参与某一话题传播的活跃节点。但我们并不清楚这个消息如何传播及为何传播。因此，有必要建立模型来描述传播过程的底层机理。传播模型可以分为两类：解释模型和预测模型。

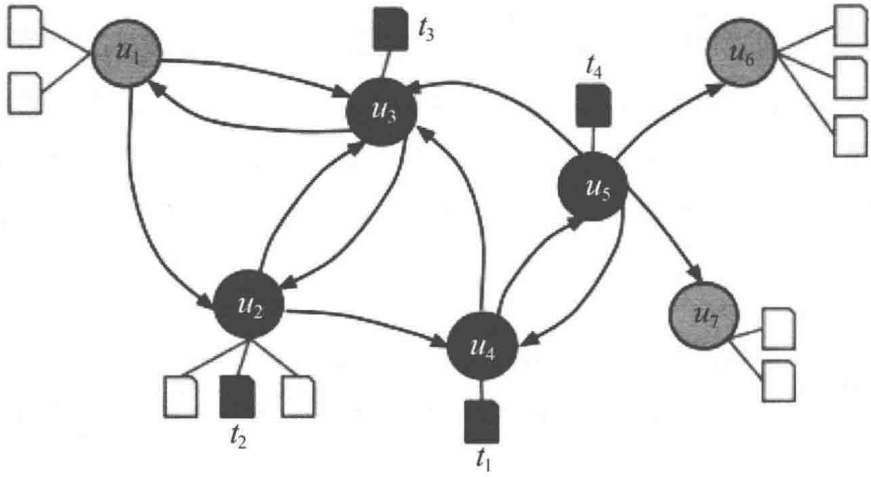


图 1-1　激活序列

本章主要内容安排如下：第二节介绍解释模型及相应算法；第三节介绍预测模型及代表性算法；第四节对两种模型进行总结。

## 1.2　解释模型

解释模型的目的是在给出完整的激活序列后，推断出底层传播级联机制。这些模型能够帮助我们了解消息是如何传播的。

### 1.2.1　问题描述

网络中的消息传播可以类比传染病扩散的过程。节点何时被感染往往可以直



此为试读，需要完整PDF请访问：www.ertongbook.com