



在线社会网络的 用户行为建模与分析

◎ 郭强 刘建国 著

Modeling and Analysis of Online-social-network
User Behaviors

 科学出版社

在线社会网络的用户行为 建模与分析

郭 强 刘建国 著

科学出版社

北 京

内 容 简 介

本书介绍在线社会网络及其研究进展,旨在反映在线社会网络系统下的用户行为规律。全书共分九章。第1章介绍在线社会网络的背景及基础知识。第2章引入超网络的概念与模型。第3章介绍在线社会网络的用户行为模式,并进行实证分析举例。第4章介绍网络中的节点重要性度量。第5章详细介绍推荐算法的基础知识。第6章介绍协同过滤推荐算法。第7章介绍基于网络结构的推荐算法。第8章介绍基于内容的推荐算法。第9章介绍混合推荐算法。每章都包含复杂系统科学研究中心在该领域的相关研究工作及发表在SCI期刊论文中的部分成果。

本书可供有志于探索在线社会网络的有关研究人员以及高等院校有关专业的研究生、本科生阅读,也可为从事智能电子商务、复杂性科学、科学知识图谱分析、知识管理、超网络模型构建与分析、推荐算法、传播动力学、时序行为模式分析以及大数据分析相关领域的教学、科研人员提供参考。

图书在版编目(CIP)数据

在线社会网络的用户行为建模与分析/郭强,刘建国著. —北京:科学出版社, 2017.5

ISBN 978-7-03-052667-0

I. ①在… II. ①郭…②刘… III. ①互联网络—研究 IV. ①TP393.4

中国版本图书馆CIP数据核字(2017)第098588号

责任编辑:王腾飞/责任校对:张 胶

责任印制:张 伟/封面设计:许 瑞

科 学 出 版 社 出版

北京东黄城根北街16号

邮政编码:100717

<http://www.sciencep.com>

北京九州迅驰传媒文化有限公司印刷

科学出版社发行 各地新华书店经销

*

2017年5月第 一 版 开本:720×1000 1/16

2017年5月第一次印刷 印张:17 1/4

字数:348 000

定价:89.00 元

(如有印装质量问题,我社负责调换)

前 言

时至今日，科技的迅速发展极大地方便了人们的生活，甚至改变了人们的生活方式。自 20 世纪 90 年代到 21 世纪初，在线社会系统经历了诞生和发展，如今呈现一片繁华景象。忙碌的社会、快节奏的生活，让人们已分配不出足够的时间去维系传统的社会交际。在这样的情况下，在线社会系统应运而生，将现实中人与人的交互成功地转向虚拟的网络，不但节省了成本，更衍生出了诸如 Facebook、微博这样各具特色的交互平台。本书的目的在于向广大读者介绍并分析在该系统下人类的行为规律，希望能给广大读者在研究的道路上以抛砖引玉的作用。

网络已经成为描述形形色色复杂系统最重要的工具之一，被广泛地用于交通、通信、社交、医学和生物等诸多领域。在线社会系统亦是一个个复杂的网络，拥有自身的体系和规律，本书的任务就是寻找蛛丝马迹，归纳分析，透过现象挖掘本质。大家所熟悉的在线社会系统无非两大类。一类是像国外的 Facebook、Twitter 以及国内的人人网这样的社交网络，旨在交友互动以及分享轶事趣闻。二是如 Amazon、京东商城这样的电子商务在线系统，电子购物已成为时下最流行的购物方式之一。那么在这样的在线网络中，每个人最多能与多少人保持亲密的联系？怎样的电子商务系统能够最快最好地给顾客推荐想要的产品？这些都是本书即将要解答的问题。

本书共分九章。为了能让读者更清晰地理解在线社会系统，本书按照背景、建模、实证研究和算法这样的编排顺序来分配章节。第 1 章，不仅介绍基本概念，更希望让读者对在线社会系统有一个宏观的认识，主要介绍在线社会系统自发展以来的背景知识。包括在线社会网络和个性化推荐系统两大领域。第 2 章，向大家介绍超网络的概念，并建立模型。超网络还只是一个概念，对超网络的边界也没有明确的规定。大多数超网络都可以看做一类特殊的复杂网络。这种复杂性主要表现在属性上，而非表现在规模上。对超网络的研究首先要考虑的问题，是探索并发现实际生活中的一些网络由于哪些特征才需要而且可以把它们看做超网络问题，以及怎样建立概念模型、结构模型乃至数学模型。而在该章，构建的是基于知识传播的科研合作超网络模型，并介绍其应用。第 3 章，以实证分析为主，根据采集来的数据，分析包括电子商务网站中用户的兴趣，社交网络上用户的行为规律以及结构特征和社会影响对用户选择行为的影响。第 4 章，向大家介绍的是节点重要性的度量。在一个网络中，识别重要节点具有特殊意义。如社交网络

中，它往往代表的是具有强影响力的人；而在个性化推荐系统中，它往往代表着用户最感兴趣的产品。当然不同的度量方式会得到不同的结果，也适用于不同的网络。该章先总结前人已经研究出的度量方法，而后又根据我们的实验分析得到几种新方法，进一步完善节点重要性度量的理论结构。第5~9章，详细介绍个性化推荐系统的相关知识。第5章先为大家普及基本概念，包括常用数据集和评价指标等，并主要介绍推荐算法中常用到的二部分网络的知识。个性化推荐算法可分为四大类：协同过滤推荐算法、基于网络结构的推荐算法、基于内容的推荐算法和混合推荐算法。因此，第6~9章，每一章介绍一类推荐算法。值得一提的是，这四类算法关系密切，很多时候并没有很明确的界限区分它们，这里只是按照作者的理解划分，以期帮助读者更深入地理解。

本书不是一本教科书，预期的目的是给相关领域或者对此领域感兴趣的研究者一个系统的梳理和总结。或许读起来枯燥乏味，但作者依然希望其中的某个章节能够带给您新的灵感和启发，为整个体系的完善贡献出自己的一份力量。

这里需要感谢上海理工大学复杂系统研究中心的各位老师同学，本书所有提及的研究成果皆出自研究中心的实验小组。尤其要提出的是，第4章和第6章分别出自杨光勇和任卓明的研究，他们的独立研究，自成一个章节，为本书的完整性作出了巨大贡献。此外还要感谢冷瑞、石珂瑞、李洋、周继平、胡兆龙、邵凤、李旭东、侯磊、张一璐和宋文君，正是他们辛勤的努力和刻苦的研究，才有了构思本书的想法，并顺利出版成书。研究的脚步不会停止，在线社会系统这个领域也期待着更多的有志之士贡献力量，并一步步完善。

虽然已经尽力做到字字斟酌句句推敲，但由于作者水平有限，书中难免存在不妥之处，恳请广大读者批评指正。

目 录

前言

第 1 章 在线社会系统	1
1.1 在线社会网络	1
1.1.1 社交网络中的基本概念	2
1.1.2 社交网络的理论基础	3
1.1.3 社交网络的国内外发展状况	6
1.1.4 社交网络的优势和劣势	7
1.2 个性化推荐系统的蓬勃发展	8
1.2.1 产生背景	8
1.2.2 个性化推荐系统的应用发展	10
1.2.3 研究意义	19
参考文献	22
第 2 章 超网络模型的构建及其应用	25
2.1 超网络相关研究	25
2.1.1 超网络的基本概念	25
2.1.2 超网络的研究概述	25
2.1.3 超网络研究存在的问题	27
2.2 知识传播相关研究	28
2.2.1 知识的基本概念	28
2.2.2 知识传播的研究概述	31
2.2.3 知识传播研究存在的问题	32
2.3 科研合作超网络模型的建立与分析	32
2.3.1 已有的两种超网络演化模型	32
2.3.2 LWH 超网络模型的建立	35
2.3.3 LWH 超网络拓扑特性的分析	36
2.4 科研合作超网络上的知识传播研究	44
2.4.1 知识传播模型	44
2.4.2 知识传播模型的参数设置及评价指标	47
2.4.3 结果分析	48

2.5	科研合作超网络上的知识创造研究	51
2.5.1	知识创造超网络模型的建立	52
2.5.2	数值模拟	55
2.6	小结	59
	参考文献	60
第3章	用户行为模式分析	65
3.1	用户行为在个性化推荐算法中的重要地位	65
3.2	用户行为模式分析	65
3.2.1	基于集聚系数的度量方法	66
3.2.2	基于信息熵度量用户兴趣的多样性	70
3.2.3	在线打分的记忆效应	75
3.3	微博中基于用户结构的信息传播分析	81
3.3.1	微博网络的相关机制	81
3.3.2	突发事件的信息传播分析	82
3.4	Facebook 中个人中心网络的统计特性分析	87
3.4.1	模型的建立	88
3.4.2	数据描述	90
3.4.3	实证统计	91
3.4.4	随机模型的运用	94
3.5	社会影响对用户选择行为的影响	96
3.5.1	社会影响与用户偏好网络模型建立及其结构特性	96
3.5.2	网络数据分析	103
3.5.3	数值模拟	107
	参考文献	110
第4章	网络中的节点重要性度量	113
4.1	网络中节点重要性排序的研究进展	113
4.1.1	基于网络结构的节点重要性排序方法	114
4.1.2	基于传播动力学的节点重要性排序方法	124
4.2	复杂网络中最小 k -核节点的传播能力分析	125
4.2.1	理论基础与方法	126
4.2.2	数值仿真与结果分析	128
4.3	基于 k -核与距离的节点传播影响力排序方法研究	131
4.3.1	基于 k -核与距离的节点传播影响力排序度量方法	132
4.3.2	实验数据及相关参数	132

4.3.3 数值仿真与结果分析	133
4.4 基于度与集聚系数的网络节点重要性度量方法研究	137
4.4.1 理论基础与方法	138
4.4.2 实例验证	141
参考文献	143
第 5 章 个性化推荐系统的相关理论概念	148
5.1 二部分网络	148
5.2 个性化推荐算法	152
5.2.1 基于协同过滤算法的推荐系统	152
5.2.2 基于内容的推荐系统	152
5.2.3 基于网络结构的推荐系统	153
5.2.4 基于混合推荐算法的推荐系统	153
5.2.5 其他推荐算法	153
5.3 常用数据集	154
5.3.1 MovieLens 数据集	154
5.3.2 Netflix 数据集	155
5.3.3 Delicious 数据集	155
5.3.4 Amazon 数据集	155
5.4 评价指标	157
5.4.1 推荐的准确度	158
5.4.2 被推荐产品的流行性	159
5.4.3 推荐产品的多样性	159
5.4.4 分类准确度、准确率与召回率	160
5.4.5 F 度量	160
5.4.6 新颖性	161
5.5 相似性	161
5.5.1 基于打分的相似性	161
5.5.2 结构相似性	162
5.6 小结	165
参考文献	167
第 6 章 协同过滤推荐系统的算法研究	170
6.1 协同过滤推荐算法	170
6.1.1 基于用户的协同过滤推荐算法	170
6.1.2 基于产品的协同过滤推荐算法	172

6.2	用户关联网络对协同过滤推荐算法的影响研究	174
6.2.1	用户关联网络简介	174
6.2.2	用户关联网络统计属性	175
6.2.3	基于用户关联网络的协同过滤推荐算法	178
6.3	考虑负相关性信息的协同过滤推荐算法研究	180
6.3.1	算法介绍	181
6.3.2	实验结果分析	182
6.4	集聚系数对协同过滤推荐算法的影响研究	185
6.4.1	产品集聚系数对协同过滤推荐算法的影响研究	185
6.4.2	用户集聚系数对协同过滤推荐算法的影响研究	186
6.4.3	数值结果分析	187
6.5	基于 Sigmoid 权重相似度的协同过滤推荐算法	190
6.5.1	基于 Sigmoid 权重相似度的协同过滤推荐算法	190
6.5.2	实验过程及结果分析	193
	参考文献	195
第 7 章	基于网络结构的推荐算法研究	197
7.1	基于热传导的推荐算法	197
7.2	二部分图中局部信息对热传导推荐算法的影响研究	198
7.2.1	HC 数值模拟结果	199
7.2.2	改进的 HC 数值模拟结果	200
7.3	基于物质扩散过程的推荐算法	202
7.4	基于物质扩散过程的协同过滤推荐算法	204
7.4.1	基于物质扩散过程的二阶协同过滤推荐算法	205
7.4.2	算法的数值实验结果	205
7.5	考虑用户喜好的物质扩散推荐算法	207
7.6	产品之间的高阶相关性对基于网络结构推荐算法的影响	209
7.6.1	基于网络结构的推荐算法	210
7.6.2	通过去除重复性的改进的算法	211
7.6.3	实验数据结果	213
7.7	有向相似性对协同过滤推荐系统的影响	215
7.7.1	用户相似性的方向性对 CF 算法的影响	216
7.7.2	基于最大相似性的 CF 算法	218
7.7.3	数值结果分析	219
7.8	二阶有向相似性对协同过滤推荐算法的影响	225

7.8.1 改进的算法	228
7.8.2 实验结果分析	229
7.9 时间窗口对热传导推荐模型的影响研究	233
7.9.1 基于局部信息的用户相似性指标	233
7.9.2 实证结果分析	234
7.10 考虑负面评价的个性化推荐算法研究	239
7.10.1 基于物质扩散模型	239
7.10.2 基于热传导模型	245
7.11 一种改进的混合推荐算法研究	250
7.11.1 模型与方法	250
7.11.2 实证结果分析	251
参考文献	254
第 8 章 基于内容的推荐算法研究	256
参考文献	259
第 9 章 混合推荐算法研究	260
参考文献	261

第 1 章 在线社会系统

1.1 在线社会网络

在线社会网络 (online social network, OSN), 是一个可以在一定程度上延伸现实生活关系的平台, 人们可以在这个平台上分享兴趣爱好、活动。它最早出现在 2003 年的美国, 进而风行全世界。市场研究机构 eMarketer 在 2013 年 11 月公布的数据显示, 全世界约有 16.1 亿人每个月至少使用一次在线社会网络。其中, 全球最大的社交网站 Facebook 在全球活跃用户的普及率可以达到 51%。预计到 2017 年, 各社交网站 (Twitter、Facebook、MySpace、Friendster、Google+ 等) 的总用户数将超过 23.33 亿。此组数据显示, 2013 年全世界有 22.8% 的人每月至少使用一次社交网络。而到 2017 年, 全球人口总数预计会超过 74 亿, 届时全球将会有超过 30% 的人使用在线社会网络。

随着 Web 2.0 的迅速发展, 在线社会网络越来越受到人们的广泛关注, 人们的生活、交流以及获取信息的方式也发生了巨大的变化。尤其是近几年微博、微信等社交工具的迅速崛起, 即时通信工具、社交网站等在线社会网络已经成为人们不可或缺的沟通工具。随着网络化进程的加快, 人们越来越倾向于把日常生活转移到网络上。与现实世界相比, 在线社会网络的沟通不受时间和空间的限制, 用户与他们的家人、朋友、同事, 甚至陌生人都可以随时随地保持联系。如今信息的更新越来越及时, 也越来越高效和便捷, 这使得人们可以随时随地将自己的所见所闻发布到网上, 虚拟社交与现实世界之间的交叉性越来越强^[1]。

近年来, 微博和博客成为人们发布新闻、获取信息、分享感受的主要平台之一。在线社会网络在信息传达上的及时、高效和便捷等特性, 使其在社会生活、政治、疾病预防等方面发挥了极大的作用。例如, 2011 年 8 月, 美国弗吉尼亚州发生地震后, 在传统媒体报道这一突发事件前, 大量消息已经在 Facebook、Twitter 等社交网站上广为传播, 这显示了新型媒体在信息传播速度方面的优势。与此同时, 社交网络的发展也引起了一系列的疑问, 如新型社交网站的整体结构如何? 这些网站在提供给用户更便捷、更低成本的交互服务的同时, 是否提高了人们的社交能力, 是否能够超越人类大脑皮层的能力限制? 抑或是与计算器相类似, 仅仅提高了简单计算的速度, 并没有提高人类对于数学的认知能力, 当今的数字网络是否也仅仅是降低了人们互相沟通与联系的成本, 并没有提高社交能力的生理限制?

互联网的不断发展使网络上的服务越来越贴近人们的生活, 人们的生活也变

得更便捷和低成本。与传统媒介相比，在线社会网络信息更新的及时、高效和便捷等特性，使其在信息发布、意见收集、信息交流、突发事件预测、预防疾病传播、事件定位、寻找新兴话题和广告营销等诸多方面起着举足轻重的作用。但与此同时，网络上也存在着流言肆虐、隐私泄露等问题，这严重影响人们的正常生活。如何正确利用在线社会网络的优势来规避风险成为在线社会网络研究的重要目标之一。而对社交网络的拓扑结构和用户的行为模式的研究可以让我们更深入地认识用户的行为与心理状态。此外，对人类朋友圈上限值的研究也具有重要的理论价值和现实意义。

作为在线社会网络最重要的两种形式，微博和 Facebook 等社交网络服务(SNS)网站对于人们的生活、工作尤其重要。而由于其固有的特性，微博和 Facebook 侧重于人们生活的不同方面，微博注重信息的传播，Facebook 则注重朋友间的交互，所以可以将微博看做获取信息的平台，而将 Facebook 看做交友的平台。

1.1.1 社交网络中的基本概念

社交网络是由一个或多个行动者和他们之间的一种或多种关系组成。而在线社会网络，是指人和人之间通过朋友、血缘、兴趣、爱好等关系建立起来的社交网络平台，包括 Facebook、Twitter、MySpace、人人网、开心网等。根据哈佛大学心理学教授 Milgram 在 1967 年创立的六度分隔理论，即“你和任何一个陌生人之间所间隔的人不会超过六个”，也就是说，最多通过六个人你就能够认识任何一个陌生人。按照六度分隔理论，以认识朋友的朋友为基础，可以不断扩大自己的交际圈，最终整个社会形成一个巨大的网络。这种基于社会网络关系思想的网站就是在线社会网络。这种网络平台致力于其用户关系的建立和维护。在网络中，用户可将个人信息展示在“个人主页”上，内容包括文字、图片以及视频等，其他用户可浏览该用户的信息，同时该用户也可以浏览其他用户的信息，用户之间还可以通过评论、提及、转发等功能实现交流互动以及信息共享。与其他大多数网站不同，在线社会网络的内容都是用户自己生成的。众多的在线社会网站只是为用户提供服务的一个平台。比较著名的某些社交网站如图 1-1 所示。

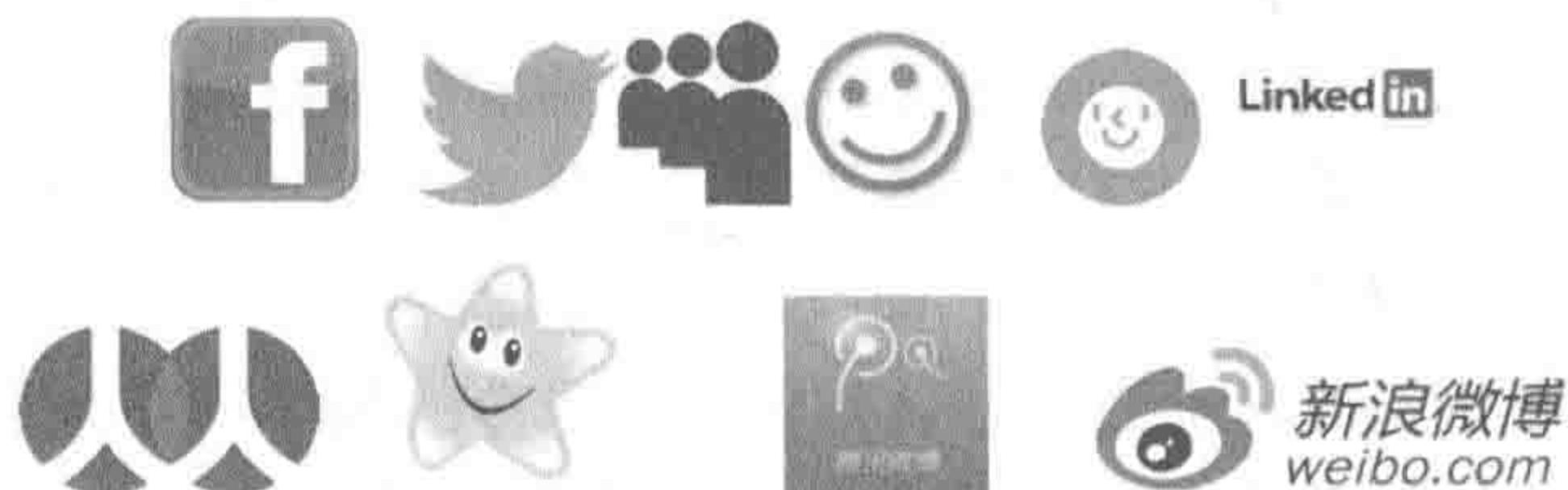


图 1-1 比较著名的社交网站

目前关于社交网络还没有统一的定义,本书在已有研究的基础上,尝试给出社交网络的描述性定义。社交网络是一个有边界的系统,其中:①系统的主体是公开或半公开个人信息的用户;②用户能创建和维护与其他用户之间的朋友关系及个人发布的内容信息,如日志、照片等;③用户通过链接可以浏览其他用户的主页和分享的信息,并进行转发和评论。

一般来讲,社交网络具有几个基础性的关键概念,包括行动者、关系、二元图、三元图、群等。

(1) 行动者。行动者是指组成社交网络的社会实体,包括团体中的个人、企业的部门、城市服务机构或世界体系中的民族、国家。在网络分析中,每个社会实体都可以被刻画成一个节点。

(2) 关系。关系是指行动者之间的联系。行动者通过社会联系彼此连接,联系的范围和类型非常广泛,而其特征就是建立了一对行动者之间的连接,如甲对乙进行评价、共同参加一项活动、上级对下级的领导等。在网络分析中,每对行动者之间的联系通常被刻画成两个节点之间的连边,而联系的强弱则通过边的权重来表示。

(3) 二元图。在最基础的层次,一个连接或关系建立了两个行动者之间的联系。这个联系在本质上是属于这两个行动者的,而不是仅仅从属于某一个行动者。二元是由一对行动者和他们之间的联系构成。二元分析关注成对关系的属性,注意联系是否互惠,多重关系的某些类型是否同时出现等。

(4) 三元图。许多社交网络的分析方法和分析模型关注的是三元图——三个行动者子集和他们之间的联系。其中最受大家关注的是三元图是否可传递(例如,若行动者 a 喜欢行动者 b , 行动者 b 喜欢行动者 c , 是否行动者 a 也喜欢行动者 c) 和是否是平衡的(例如,若行动者 a 和 b 彼此喜欢,则他们对第三个行动者 c 的态度是相似的,而若行动者 a 和 b 相互不喜欢,则他们对第三个行动者 c 的态度是不同的)。

(5) 群。二元图是成对的行动者及其联系,三元图是三个一组的行动者及其联系,则行动者群可定义为任意集合的行动者和他们之间的联系。

1.1.2 社交网络的理论基础

六度分隔理论,也称“小世界”理论,具有高集聚系数和短平均路径的特性。如前所述,社交网络理论来源于六度分隔理论,即任意两个人之间可以通过不多于六个中间人而建立联系。它要表达的观点为,任何两个素不相识的人都可以通过一定的渠道产生必然联系,完全没有关系的两个人是不存在的。

这个理论最早是由美国著名社会心理学家 Milgram 于 20 世纪 60 年代提出的。后来, Milgram 设计的信件传递实验证明了这一理论的正确性,随机发送的 160

个信件大多都在经过五、六个步骤后到达了目标人手里。

强连接和弱连接。六度分隔理论肯定了人们之间联系的普遍性，但没有区别这种联系的关系强弱。人们会在生活中认识数以百计的人，其中既有经常联系的亲密的亲人、同事、朋友，也有对于我们无足轻重的只是认识的人。人们和不同的人的联系方法和强度是不同的，而六度分隔理论却将这些联系看做无差别的，没有强弱之分。因而在实际运用中，对连接强度进行加权有时是很必要的。

贝肯数。贝肯数是由六度分隔理论演变而来的。贝肯只是一个普通演员，在电影中从来都不是主角，但是他在很多电影中与众多影视明星合作过，人们将当时贝肯要与其他影视明星之间产生连接所需要的中间人数量称为贝肯数。这一数字说明，一个人要成为网络的中心，他不一定非要成为一个大人物，一个经常出现的小人物也可以非常接近网络中心。

邓巴数。邓巴数 (Dunbar number)^[2]是英国牛津大学人类学家邓巴于 1992 年提出的，指的是可以和特定人物保持亲密关系的最多人数，一般范围为 100~230，通常人们使用 150。邓巴数也称“150 定律”，即一个社群能够保持稳定联系的规模大约为 150 人，超过这个数，人们相互间的互动和影响会迅速降低，这是由人的大脑皮层容量决定的。此处限定的人际关系是指某个人知道其他人是谁并且了解那些人之间的关系。支持者认为，人数多于邓巴数的团队，需要更加苛刻的法律、规章、制度来保持其稳定性和凝聚力。

动物学家通过对灵长类动物的研究发现，社群规模大小受到其大脑皮层容量的限制，即大脑皮层容量的大小限制了其能维持稳定联系的人数的上限。动物学家认为在一个特定物种与其社群规模大小之间存在一个系数，而该系数可由该物种的大脑皮层容量大小来计算。

1992，牛津大学人类学家邓巴根据从灵长类动物界观测到的相关系数来预测人类社群的大小，从而提出了邓巴数理论，即“150 定律”，意味着人类社群规模一般为 150 人左右，这是一个人类个体能够和其他人维持稳定联系的理论上的上限值。一个人类社群规模一旦超过这个数字，群组内成员就不能进行有效的沟通和协调，社群结构会变得松散。

1992 年，邓巴开始研究英国人寄圣诞卡的习惯。在邓巴作研究的那个年代，社交网络尚未诞生，他希望找到一个办法衡量人们的社交关系。邓巴不仅想知道研究对象认识的人数，更对每个个体真正在乎的人数感兴趣。他发现，可以通过研究圣诞贺卡来探寻这种情感关联。要送出贺卡，前提是你首先必须知道收贺卡人的邮寄地址，然后去买贺卡、买邮票，再写上几句祝福的话，最后寄出去。这一系列活动都属于一种投资，需要人花费时间、金钱来完成，而大多数人是不愿意为一个无关紧要的人这样费心费力的。

邓巴研究发现，人们把约 25% 的贺卡寄给了自己的亲人，近 67% 寄给了朋友，

约 8% 寄给了同事。不过，其中最重要的研究发现是这样的一个数字：对于一个人寄出的全部贺卡，收到贺卡的家庭的人数之和的平均数为 153.5 人，即 150 人左右，这一数字与邓巴数理论非常吻合。

在邓巴看来，其原因很简单，人不能超出生理条件的限制而实现无限的可能，例如，人不能在地球上像鸟儿一样飞翔，五秒内跳不了一百下，人耳听不到频率低于 20Hz 的次声波和超过 20000 Hz 的超声波。大多数人最多只能与 150 左右的人建立具有实际意义的联系，不同个体之间可能会有所不同，但不会比 150 多出太多。然而，这一规律也不完全绝对，因为既有不善交际、内向自闭的人，也有善于交际、开朗的公众人物。但总体来讲，一个群体的规模一旦超过 150 人，成员内部之间的关系就会开始淡化。尽管现在的文明程度较以前有了很大的提高，但人类的社交能力并没有获得大的提升。邓巴说道：“150 个人似乎是我们能够建立稳定社交关系人数的上限，而在这种社交关系中，我们不仅仅是了解他们是谁，而且也了解他们与我们自己之间的关系。”

根据邓巴的研究，人类的社会结构表现为图 1-2 所示的同心圆模型：5 人左右的亲密接触圈；12~15 人的同情圈，在此范围内若有人去世，其他人会非常悲伤；50 人左右组成群落，共同生产、活动；150 人左右称为氏族，他们可能具有相同的信仰或习俗；500 人左右组成部落，他们使用相同的语言（这里的“语言”只指一些经常交流的人之间约定俗成的词语和概念）；5000 人左右为大群落，他们通常具备共同文化。参照此模型，当交际范围超过 150 人时，个体之间的交流、影响就会明显降低，只能靠共同的语言来维系；而当人数上升到 5000 人左右时，维系此社会结构则只能依靠共同的文化。一个人能够维系的稳定人际关系处于 100 到 230 之间，通常人们认为是 150。这里的人际关系是指某个人知道其他人是谁并且了解那些人之间的关系。

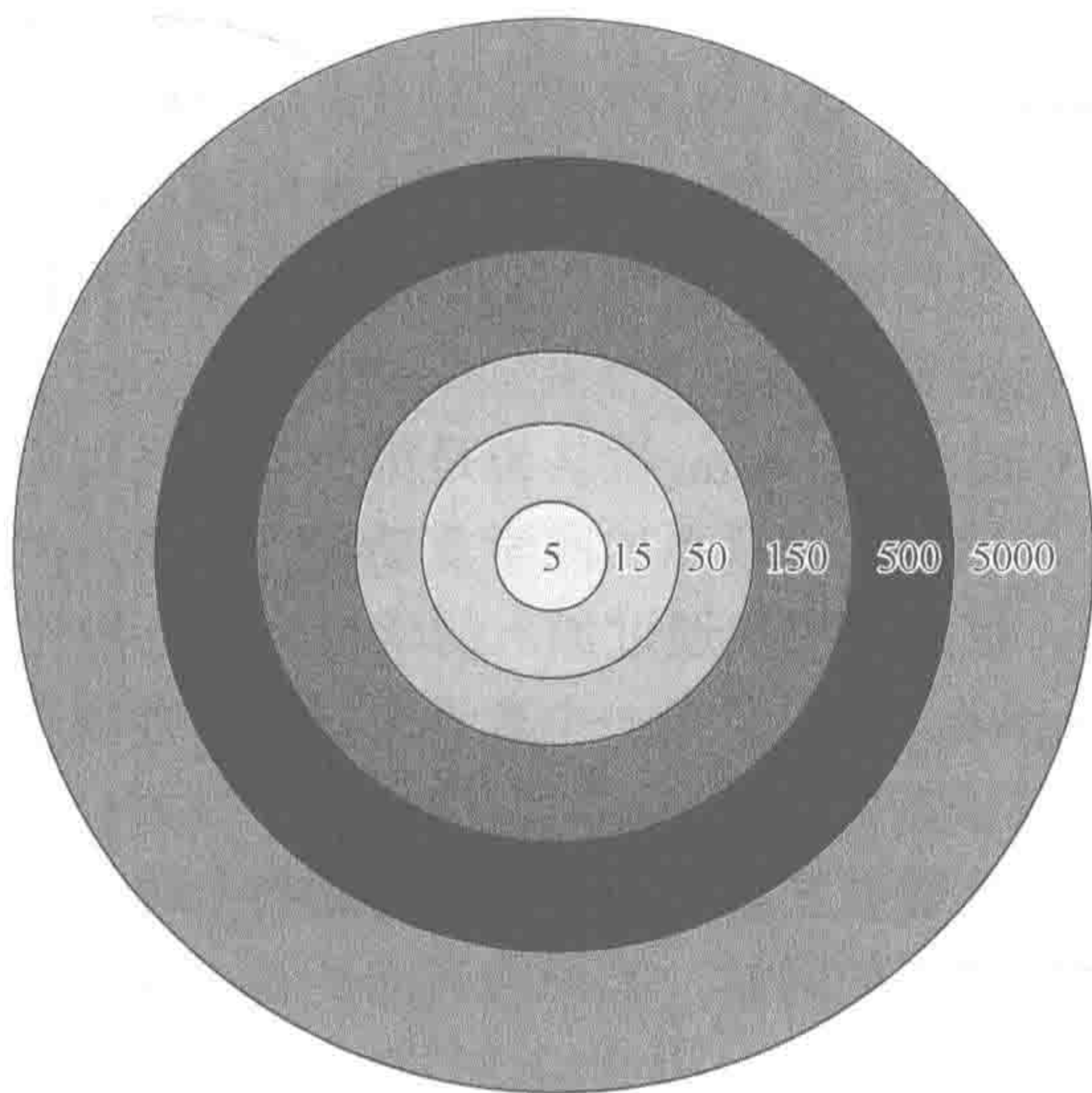


图 1-2 邓巴数理论同心圆模型

图论。对于一个网络，可以从多个方面对其进行观察，而最实用的一个方面就是将其看做由线和点组成的图。图论思想被广泛应用于社交网络分析中。图论为标记和表示社会结构提供了许多词汇，而且图论中的数学运算和观念的运用对于社交网络的许多属性的量化和测量起到了不可估量的作用。例如，度量用户网络结构的度、网络密度、连通分量等指标都来源于图论。

图论来源于著名的哥尼斯堡七桥问题。在普莱格尔河上，七座桥将河中的两个小岛与两边的河岸连接起来，问题要求从小岛或河岸出发，不重复地通过每座桥一次，最后再回到起点。很多人尝试解决这个问题，但都没有成功。而欧拉通过将小岛和陆地都抽象化成一个点，每座桥抽象化成一条边，从而得到一个“图”，如图 1-3 所示，将图 (a) 抽象成图 (b)。欧拉证明了这个图是无解的，并且提出了欧拉路径和欧拉回路问题，这就是第一个图论问题。

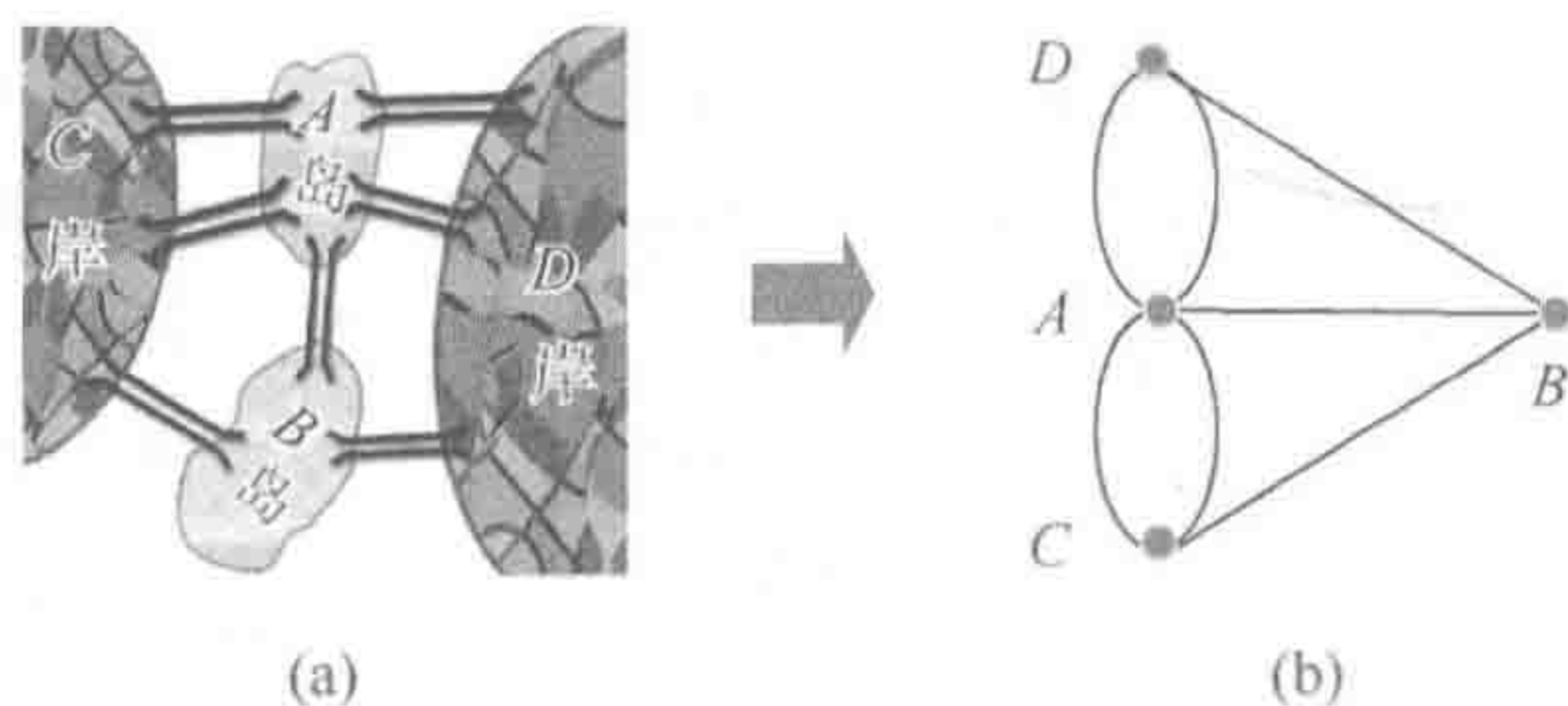


图 1-3 哥尼斯堡七桥问题

1.1.3 社交网络的国内外发展状况

社交网络源自网络社交，网络社交的起点是电子邮件，BBS 把网络社交推进了一步。随着网络社交的悄悄演进，便出现了社交网络。社交网络在人们的生活中扮演着重要的角色，它已成为人们生活的一部分，并对人们的信息获得、思考和生活产生不可低估的影响。

1995 年成立的 Classmates.com 被认为是第一个允许用户之间进行交流的网站，它可以帮助用户与已经失去联系的同学重新获得联系，然而，它不允许用户之间建立直接的联系，仅允许用户通过加入的学校来进行沟通。而后，1997 年建立的六度网站 (six degrees.com) 是第一个允许用户之间建立直接联系的真正的社交网站。

当越来越多的用户接触到互联网的时候，在线社会网络迅速流行起来。在 21 世纪初，出现了大量的交友的网站，其中最著名的是 Friender，还有同时期的其他相似网站，如 Cyworld 等。

2003 年，MySpace 成立，它允许用户订制自己的主页，受到了用户的欢迎，

迅速成长为最大的在线社会网络。然而，在2008年5月，2004年上线的Facebook取代MySpace，成为全球最大的在线社会网站。直到现在，Facebook仍然是全球最大的在线社会网站。

随着社交类网站的流行，许多其他类型的网站也开始具有社交性质，如视频分享网站Flickr、YouTube、Zoomr，博客类网站LiveJournal和Blogspot，专业网站LinkedIn和Ryze，新闻门户网站Digg和Reddit等。

据GlobalWebIndex 2013年第一季度的“社会化平台季度更新”报告显示：根据活跃用户数渗透率排名，Facebook为全球排名第一的社交网站，网民渗透率为51%；Google+排名第二，为26%；YouTube排名第三，为25%；Twitter排名第四，为22%，但新浪微博是增长速度最快的社交网站。具体社交网站排名如图1-4所示。

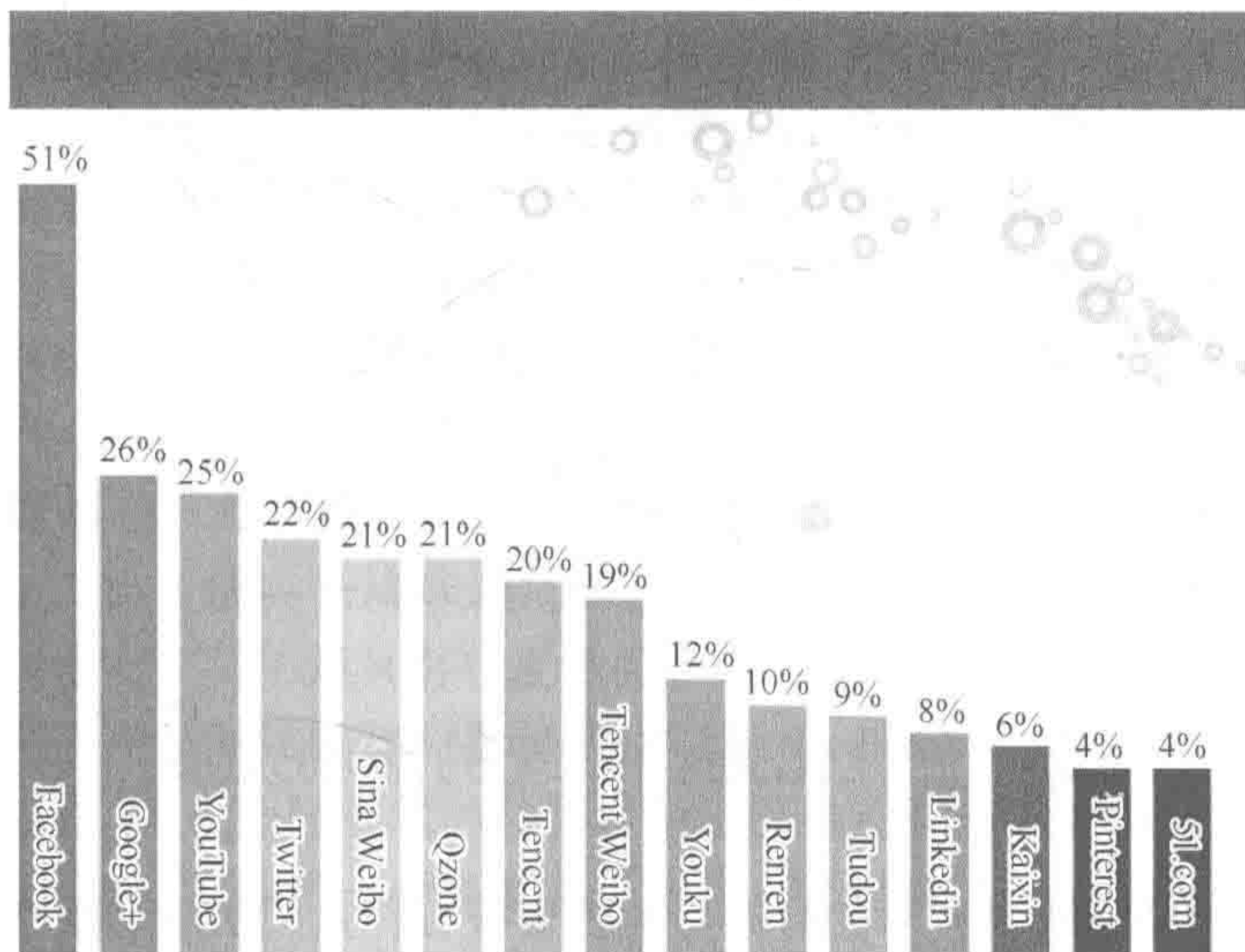


图 1-4 全球前 15 强社交媒体网站排名（根据活跃用户数渗透率排名）

1.1.4 社交网络的优势和劣势

在线社会网络的优势包括以下几点。

(1) 通过在线社会网络，人们可以与朋友保持更加直接紧密的联系，其搜索用户的功能更能帮助用户寻找到之前失去联系的朋友甚至素未谋面的陌生人，大大拓宽了人们的交际范围。

(2) 在线社会网络用户可随时随地将周边发生的事发表出来，通过朋友的转发可以迅速扩散开来，所以这对新闻的传播、流行话题的检测等多方面具有