财经类高等学校
“十三五”通识教育规划教材

科技进步与科学精神

大数据与社会网络

陈　媛◎编著

BIG DATA & SOCIAL NETWORK

上海财经大学出版社

财经类高等学校
“十三五”通识教育规划教材

科技进步与科学精神

大数据与社会网络

陈　媛◎编著

上海财经大学出版社

图书在版编目(CIP)数据

大数据与社会网络 / 陈媛编著. —上海：上海财经大学出版社，2017.5

财经类高等学校“十三五”通识教育规划教材·科技进步与科学精神

ISBN 978-7-5642-2713-5/F.2713

Ⅰ.①大… Ⅱ.①陈… Ⅲ.①数据管理—高等学校—教材②社会网络—高等学校—教材 Ⅳ.①TP274②C912.3

中国版本图书馆 CIP 数据核字(2017)第 090635 号

责任编辑：李志浩
封面设计：张克瑶

DASHUJU YU SHEHUIWANGLUO

大数据与社会网络

财经类高等学校“十三五”通识教育规划教材·科技进步与科学精神

作　　者：陈　媛　编著
出版发行：上海财经大学出版社有限公司
地　　址：上海市中山北一路 369 号(邮编 200083)
网　　址：http://www.sufep.com
经　　销：全国新华书店
印刷装订：上海华教印务有限公司
开　　本：710mm×1000mm　1/16
印　　张：15.5
字　　数：253 千字
版　　次：2017 年 5 月第 1 版
印　　次：2017 年 5 月第 1 次印刷
定　　价：40.00 元

内 容 简 介

人们之间的互联与共享是信息技术发展的原动力之一，从人与人之间交往形成的隐形网络，到计算机之间互相连接的网，再到大数据时代两种网络的交织，人们的社交方式从线下到线上。本书针对这种现象，从互联与共享角度出发，综合介绍互联网与大数据的基础知识、社会网络及在线社会网络的分析方法。

本书分为上、下篇，上篇为互联网与大数据，主要介绍互联网与大数据的相关基础知识，下篇为社会关系与社会网络，系统介绍社会关系、社会网络以及在线社会网络的相关思想和分析方法。本书是本科生的入门教材，同时也适合对该领域感兴趣的读者阅读。

前　言

我们生活在互联网时代，数据大爆炸的同时，人们的社交模式正在发生改变。网络将万物联接，让人与人、机器与机器、人与机器发生新的关系。人与人（或拟人）的关系通过网络中节点的交互显现出来，即构成了虚拟化的社会网络关系。大数据背景下，面对海量数据，如何根据网络运行数据描绘出人与人的互动关系，识别数据背后所隐藏的社交模式，进而分析相应的社会现象是计算机科学、管理学及社会学等多学科学者需要面对的共同问题。

本书源于我们开设的一门课程。十余年前，笔者开始接触社会网络分析理论与数据分析相关研究工作，随着互联网的大规模应用与渗透，人们对背后的数据与社会网络变化产生了浓厚的兴趣。基于此，笔者萌生了开设一门专门课程的想法，意图传播和探讨相关的理念与方法。2015 年初次开设此课程后，吸引了来自法学、会计学、人文、计算机、电子商务等多个专业的不同层面学生。课程几经修订，即构成本书的主体内容。书中的基本概念和方法来自于多个领域的研究文献，本书将这些思想和成果汇集与融合，使其可以适合多学科不同层次的学生群体学习与研究的需要。

全书分为上、下篇共 8 章。上篇侧重介绍互联网和大数据基础知识，其中，第一章回顾了互联网的产生与发展历程，介绍互联网相关技术。第二章介绍进入大数据时代，传统的数据收集、存储、处理及分析技术相应的发展和变化。第三章介绍大数据的处理与分析及大数据计算平台环境等方面技术。下篇重点介绍社会关系与社会网络相关理论与方法，其中，第四章介绍社会关系与社会网络的基本概念及相关理论。第五章分析社会网络理论的起源、发展历史和最新进

展，梳理相关研究成果。第六章分别介绍社会网络表示的矩阵和图方法。第七章介绍社会网络分析的基本方法，并结合一些领域内研究成果讨论社会网络的特征。第八章介绍在线社会网络分析和数据获取方法。

本书的工作受到国家自然科学基金资助(No. 71371115)，同时感谢参与本书编写工作的谢美萍、陈炤宇和马瑾，每周三下午的书稿讨论是令人难忘的美好时光！

由于编者水平限制，书中一定存在不少缺点或疏漏，恳请读者批评指正。

陈　媛

2017 年 2 月于上海财经大学

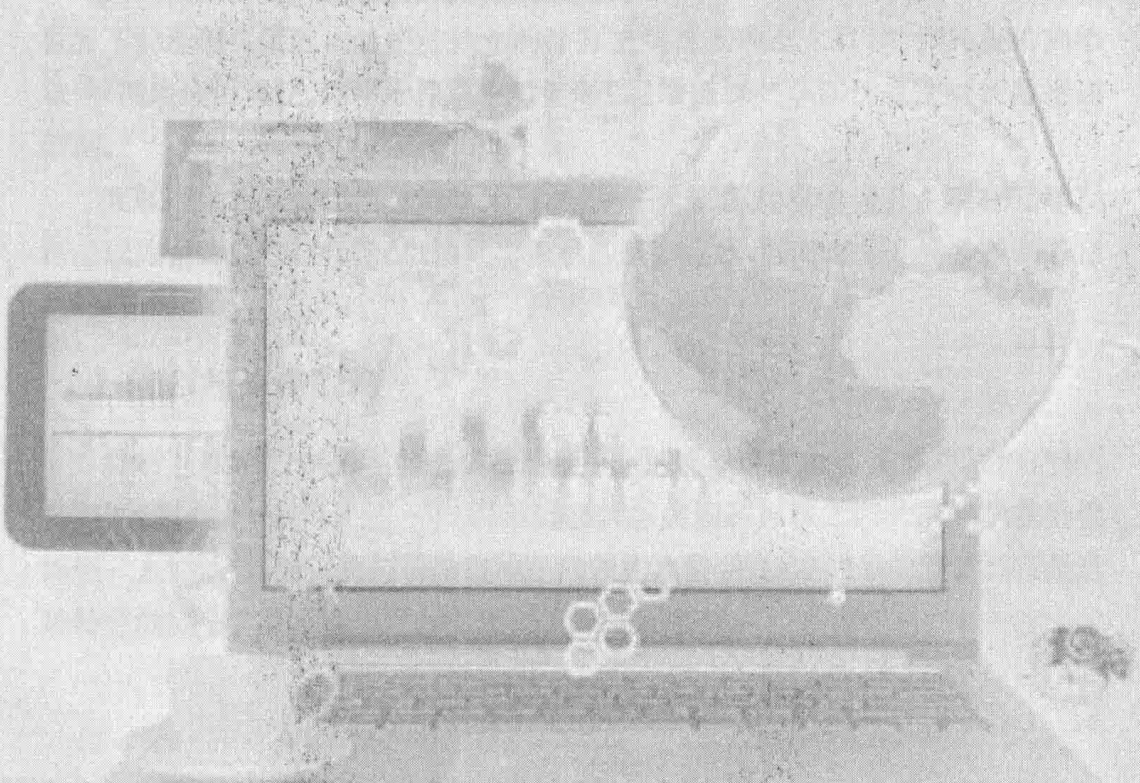
目　录

上篇　互联网与大数据

下篇 社会关系与社会网络

上篇

互联网与大数据



第一章　互联网

纵观互联网的产生与发展历程，一个根本动力是对“互相连接”的需求，这里的互联从最初的计算机与计算机的互联，发展成为全球各地个体与个体之间的互联。实现互联，一方面需要相关技术的支持；各种相关软、硬件技术的发展和成熟，进一步促进了“高效互联”；另一方面，随着互联网技术的成熟，各种新的商业模式不断出现，从而为进入大数据时代奠定了基础。本章回顾了互联网的产生与发展过程，简要介绍了互联网的相关技术，并进一步介绍了互联网发展进入“互联网＋”的特征。

第一节　互联网的产生与发展

互联网（Internet）又称因特网、国际网、网际网，是利用通信设备和线路将全世界不同地理位置的功能相对独立的计算机系统互联起来，以功能完善的网络软件（网络通信协议、网络操作系统等）实现网络资源共享和信息交换的数据通信网。

互联网以相互交流信息资源为目的，是一个资源共享的集合。现在互联网已经成为企业、政府和研究机构共享信息的基础设施，同时也是开展各种商务活动的基础。

一、互相连接的网

分析互联网的产生不能脱离历史背景，它最早可以看作美苏冷战的产物。在美国，20 世纪 60 年代是一个很特殊的时代，美国和苏联之间的冷战状态升温加剧。人们认为，能否保持科学技术上的领先地位，将决定战争的胜负。正是这种特殊的历史背景孕育出了一个不同凡响的新事物。

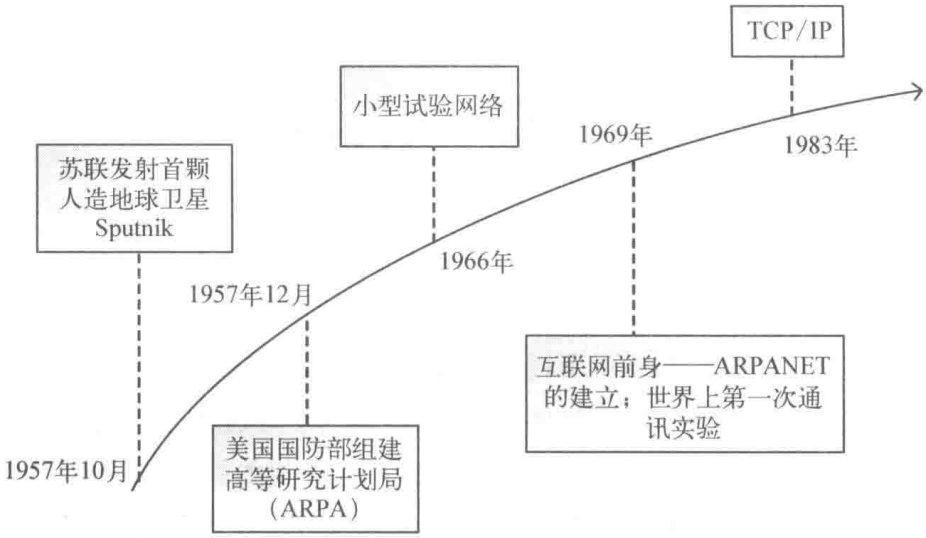


图 1-1　国外互联网发展的时间历程及重要事件

（一）阿帕网（Advanced Research Projects Agency Network，ARPANET）

1957 年 10 月 4 日，在苏联的拜科努尔航天中心，人类第一颗人造地球卫星“史伯尼克”被送入太空。在地球的另一端，“史伯尼克”顷刻间汇成国家安全危机的阴云笼罩了整个美国，因为“史伯尼克”卫星意味着在争霸全球的竞赛中，苏联人终于先行一步。

两个月后，美国总统向国会提出，建立高等研究计划局（Advanced Research Projects Agency，简称 ARPA），办公地点就设在五角大楼内。新生的 ARPA 即刻获得了国会批准的 520 万美元的筹备金及 2 亿美元的项目总预算，是当年中国国家外汇储备的 3 倍。今天网罗了每一个人的互联网，就萌芽在这项拨款中。

ARPA 资助每一个科研项目，为研究者提供功能不同的计算机，它们动辄数十万、上百万美元，这些庞然大物互不兼容，造成经费的极大浪费，并因此产生将计算机连接的念头。这个想法在美国科学界酝酿已久。曾经参与发明第一颗原子弹和第一台电子计算机的科学家万尼瓦尔·布什在 1945 年就提出了记忆延伸的概念，展望了关于信息检索、网络建设的可能前景。ARPA 信息处理技术处处长泰勒的前任利克里德也在 1960 年发表了题为《人机共生》的文章，预言人们通过机器的交流，将变得比人与人、面对面的交流更为有效。

1966 年，泰勒提出由阿帕出面构建一个小型的实验网络，并得到了支持。

于是拉里·罗伯茨、保罗·巴兰(提出“分布式通信系统”理论)、罗伯特·卡恩和温顿·瑟夫(发明“TCP/IP”协议)、伦纳德·克兰罗克(提出“分组交换”理论)等杰出研究人员迅速达成共识,中心是靠不住的,必须构建分布式的网络,因为如果建立一个中心节点把所有机器连起来,那么中心节点会承受巨大的压力,而且如果中心节点出现问题,整个网络都会崩溃。

在1969年10月,ARPA组建的ARPANET第一期工程投入使用,研究员们进行了世界上第一次互联网络的通讯试验,他们从洛杉矶向斯坦福传递一个包含五个字母的单词LOGIN,但由于传输系统突然崩溃,通讯无法继续进行,因此仅仅传送了两个字母“LO”。

ARPANET诞生之时,只连接了5个地点:加州大学伯克利分校、斯坦福大学、加州圣芭芭拉、犹他大学和BBN,1971年,ARPANET技术开始向大学等研究机构广泛普及。到了20世纪70年代,ARPANET已经有了好几十个计算机网络,但是每个网络只能在网络内部的计算机之间互联通信,不同计算机网络之间仍然不能互通。1969年,ARPANET采用的通信协议为网络控制协议(NCP)。[1]

1983年,美国军方正式将其所有的军事基地的各子网都联到了ARPANET上,美国国防部从中分拆出一个子网MILNET(随后更名为国防数据网络,最终成为NIPRNET,即非保密IP路由网络),该子网主要承担非保密的军事通信。同时宣布旧通信协议NCP全部过渡到TCP/IP协议。正是因为通过采用具有扩展性的通信协议TCP/IP,才能够将不同的网络相互连接,标志着互联网的正式诞生。

1986年,美国国家科学基金会(National Science Foundation, United States, NSF)将分布在美国各地的5个为科研教育服务的超级计算机中心互联,并支持地区网络,形成NSFNET。1988年,NSFNET代替ARPANET成为互联网的主干网。1989年,ARPANET解散,互联网从军用转向民用。

(二) 电子邮件

互联网的产生使得人们可以互相连接,电子邮件的出现一方面依赖于互联网,另一方面改变了人们互相交往和连通的方式。电子邮件是在20世纪70年代发明的,在80年代得以兴起。如图1-2所示,虽然ARPANET在1971年已基本建成就绪,但70年代的沉寂主要是由于当时使用ARPANET网络的人太

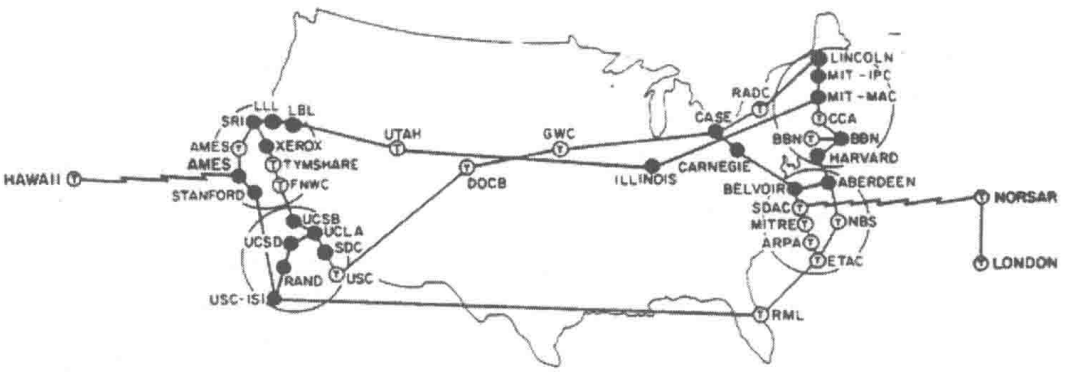


图 1-2　1971 年 9 月横贯美国大陆的 ARPANET 网络建成就绪

少，网络速度也仅为 56 Kbps 标准速度的 1/20。受网络速度的限制，那时的用户只能发送些简短的信息；到 20 世纪 80 年代中期，个人电脑兴起，电子邮件开始在电脑迷以及大学生中广泛传播开来；到 90 年代中期，互联网浏览器诞生，全球网民人数激增，电子邮件被广为使用。

1988 年，美国伊利诺伊大学的学生史蒂夫·多那（Steve Dorner）开始开发电子邮件软件 Eudora，这个软件的出现使得电子邮件成为网络中的主流。由于 Euroda 是第一个有图形界面的电子邮件管理程序，它很快就成为各公司和大学校园内主要使用的电子邮件程序。

随着互联网的兴起，Netscape 和微软相继推出了它们的浏览器和相关程序，微软和它开发的 Outlook 使 Euroda 逐渐走向衰落。

（三）万维网

互联网从军用转到民用后，民众有了使用互联网的权利，但是在相当长的时间里，互联网并不属于普通人，它仍然被限制在专业人士的圈子中，只有专业人士才能通过复杂的代码程序，前往特定的地方，捕捉特定的信息。

1991 年，欧洲粒子物理研究所（CERN）的提姆·博纳斯·李（Tim Berners-Lee）发明了利用超文本格式在互联网上发布信息的 Web 方法——万维网（World Wide Web），互联网开始向社会大众普及，人们正式通过 Web 网站来进行社会互联。1993 年，伊利诺伊大学国家超级计算机应用中心的学生马克·安德里森（Marc Andreessen）等人开发了首个流行的 Web 浏览器——Mosaic 后，互联网开始得以爆炸式普及。自那时以来，Web 网站和 Web 网页的数量便如爆炸般疯长起来。