

译者序：社会网络分析：探索关系背后的科学与技术

- 在一个社交网络中，谁是明星人物？谁是边缘人物？谁是八卦传播者？谁是沟通两个不同网络的搭桥者？谁是真正的实权人物，在社交网络中沉默却拥有极大权利？
- 为什么基地恐怖分子发动攻击时，每组人数不能超过六人？为什么网上的病毒传播的关键临界节点是7%，当一条微博的渗透率突破它，才从线性增长变为病毒传播式增长？
- 为什么脸书一步一步起家时，都先要在一个小的范围内，吸引社群关注达到6%的饱和度时，才考虑从一个小的社群——如哈佛，跃迁到一个更大的社群——如常青藤高校中去呢？

本书就是帮助理解我们这些现象背后的科学与技术。对各类关系谜题的研究，构成了人类学术史波澜壮阔的一面。始于上个世纪三十年代，由社会心理学家莫雷诺创立，社会学、经济学研究者发扬光大的学派——社会网络分析（Social Network Analysis, SNA），开启了一条将关系图论化的新道路。

社会网络分析将社会关系看作是节点（nodes）和连带（ties）组成的种种网络（networks）。节点表示网络内的独立行动者（Actor），网络则用来表示行动者（Actor）与行动者（Actor）之间的关系。行动者主体可以从个体、组织到国家；同样，关系也可以是人际关系、组织关系与国家关系。如果说传统回归分析等统计技术关心的是个体层面各类属性的统计技术，社会网络分析则致力于分析关系数据。

推特脸书等社交媒体，已成为人们生活的一部分。每时每刻，天南海北，都在产生大量实时关系数据。第一次在历史上提供了对人类关系图谱，在时间与空间上大尺度的取样。社会网络研究者可以轻易地下载到分布在全世界的数千万甚至上亿红男绿女长达数年的喜怒哀乐。其中蕴涵的价值，令每一位社会网络爱好者激动不已。

然而，社会网络分析庞大的学术积累已制造了一个较高的入门门槛。社会网络分析新增的三块砖头：小世界网络、无标度网络与随机网络，更使得它渐渐与物理科学的复杂性研究、计算机科学的网络科学领域交错在一起，走向“可计算的社会科学”（Computational Social Science）。

初学者如何更好地迈过这个门槛？《社会网络分析入门》是一本帮助你快速掌握社会网络分析技术要点、核心概念与典型算法示例的好书。本书与市面上既有的图书不同，它具备一个鲜明特点：基于开源软件NetworkX讲授。

截至今日，还有不少社会网络分析研究者使用UCINET等基于Windows窗口的软件。我们可以将UCINET理解为社会网络分析中的类Excel软件。海量增长的数据、日新月异的计算模型，使得这类商业软件弊端越来越明显。使用开源软件与商业软件更是在社区支持、知识再生产上差异极大。基于R或Python的开源社会网络分析包，如R语言的SNA或IGRAPH、Python的NetworkX，在设计新模型、开发新算法、分析大数据量与进行可重复的分析等方面目前已经完胜。

本书以基于Python的网络分析包NetworkX作为社会网络分析工具,但并不是一本NetworkX使用手册。作者将重点放在如何从庞大的社会网络分析学术积累中,挑选最精要的与实用的知识点,以帮助读者形成关于社会网络分析的知识谱系图。全书可以分为四部分。第一、二章是基础知识,主要介绍社会网络分析的背景信息与图论的基础知识。第三、四、五章是第二部分,主要介绍如何分析社会网络,分别从个体与群体两个层面,介绍了社会网络的主要测量指标与分析方法。其中第三章重点介绍社会网络节点层面的四个核心指标:

- 1) 程度中心性: 哪些是明星人物? 哪些是边缘者? 程度中心性回答类似问题。这是最为人理解的社会网络测量指标。以微博为例, 程度中心性就是粉丝的数量, 那些程度中心性高的人就是微博中的明星。
- 2) 亲近中心性: 亲近中心性通过点与其他点的距离来测量。那些在社交网络中经常与人互动、人际关系颇好的人, 比如公司中的八卦传播者, 往往亲近中心性得分较高。
- 3) 居间中心性: 节点的居间程度, 表示一个网络中经过该点的最短路径的数量。在网络中, 节点的居间程度越大, 那么它在节点相互之间的信息传播起到的作用也就越大。在两个社会网络之间的人, 比如跨界者往往拥有较高的居间中心性。
- 4) 特征向量中心性: 那些在社交网络中沉默却拥有极大权利的人物, 如《教父》中的主人翁柯里昂。社会网络研究者将他们称之为“灰衣主教”。特征向量中心性就是找出他们的办法。基本原理是一个有着高特征向量中心性的行动者, 与他建立连接的很多行动者往往也被其他很多行动者所连接。在社交网络中, 有着非常多明星做朋友的人, 即使他沉默不语, 也可能是一位重要人物。

社会网络分析不仅仅在节点层面测量。第四章、第五章介绍了如何分析群体。其中, 第四章主要介绍社群划分的一些基础知识: 如何将庞大的社会网络划分为小的组块; 如何利用社会网络中的结构洞牟利; 如何进行三元组普查与分析? 例如, 如何通过岛屿方法逐步找出推特上埃及革命成千上万条转发中的核心人物? 又如, 如何评估埃及革命中一个人的信息传播能力? 显然, 如果你的朋友们相互信任, 将比那种一个明星发言, 粉丝们单纯收听的星形网络传播能力更强。第五章主要介绍二模网络与多模网络的基础知识。关系还会存在于不同类型的主体之间, 比如公司雇佣员工、投资者购买公司股票, 人们占有信息与资源等等。这些关系被称之为二模关系。现实生活中的关系往往是二模或多模的。比如我们在微博上, 通过你的兴趣、地域、使用的标签来给你推荐新的朋友, 或者基于你对一些公共事件的看法, 将你划分到特定政治群体中, 往往是基于二模或多模网络的分析得出。

第六章是全书最精彩的部分, 关注信息如何传播, 初步展示了分析动态社会网络发展的建模技巧。一条微博如何从一两个人关注, 突然成为流行用语? 作者在实验中发现当网络密度接近7%的时候, 将从线性增长(每次增加一条链接)转化为病毒式扩散——也就是说, 如果转发一条微博、加入一个网络社群等的人的比例达到7%, 其他人将会在关键阶段马上跟进。这是一个推动脸书走出哈佛的神奇数字。脸书一步一步地跃迁, 总是遵循一个规则——在一个社群里到达饱和点之后才移入一个更大的社群。作者通过手写的Python算法, 为读者打开了动态社会网络与网络仿真的大门——我们如何用算法来模拟人类社会各类关系的变迁? 有了自己手写算法的经验, 未来读者们使用netlogo等网络仿真软件, 将更加得心应手。

对于初学者来说, 四、五、六这三章有一定难度, 需要同时理解社会科学与编程技巧两方面的知识。第七章则轻松多了, 主要介绍获取网络数据的入门知识。如果希望深入了解, 可以阅读

作者推荐的相关资源。附录A介绍了收集社会网络分析所需数据的传统方法、伦理准则与相关API。附录B介绍了如何安装本书相关软件，如[NetworkX](#)、[matplotlib](#)等。

总而言之，作为一本非常新的入门读物，《社会网络分析入门》通俗易懂，基于Python进行分析使得其灵活性变得更高。可以说，本书令学习者从一开始就具有上手实践的能力，除介绍网络数据获取技巧、网络抽样方法、网络在个体与群体两个层面的基本属性之外，还涉及到目前日益热门的网络模拟方法，融合基础理论与算法于一身。简约却不简单，上升空间非常大！无论你是对社会网络感兴趣的大众读者，还是社会网络的专业研究者、开发者，相信本书都会在社会网络的理论与实践两方面给予你启发！

有趣的是，全书通过基于[github](#)网站协作，使用[markdown](#)语法完成。全书第一、二章由中科院心理所硕士、北京语言大学心理学教师王颖翻译；第三章由中科院心理所硕士、安人公司同事王薇翻译；第四、五、六章由香港城市大学媒体与传播系博士候选人、互联网挖掘实验室研究员王成军翻译；第七章由香港城市大学媒体与传播系博士候选人刘璟翻译。附录一由北京大学教育学硕士、安人公司同事曲元周翻译；附录二由吉林大学心理系学士、安人公司同事牟百会翻译。全书由阳志平审校成书、统一定稿。

感谢各位参与翻译本书的友人，更感谢协助校对的豆瓣网友孔明、BOOMER与哈哈。最后特别感谢在组织翻译的过程中，机械工业出版社华章公司编辑吴怡老师与秦健老师。在去年杂事甚多的情况之下，他们一再包容与理解我的拖稿。他们的细致工作，更使得本书增色不少。希望这本小书的面市，能够帮助更多人迈入一个广阔迷人的社会网络分析世界。

始生之物，其形必丑。水平有限，不当之处，敬请读者指出。全书如有任何翻译疑问或者关于社会网络分析软件的使用问题，欢迎与我联系。未来有关勘误，也敬请留意华章公司相关网站。我的联系方式是：

- Email: im@yangzhiping.com
- website: yangzhiping.com

如果读者希望深度研读社会网络分析，可以阅读我们整理的图书资源：

- 阳志平：[社会网络分析入门书目导读](#)
- 王成军整理的豆列：[社会网络和复杂网络研究](#)
- 阳志平整理的豆列：[社会网络分析技术：探索复杂性与自组织性](#)

阳志平 2013年3月6日于北京