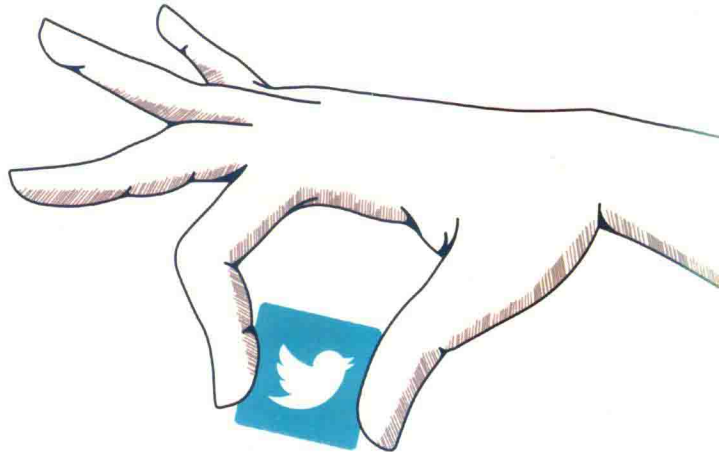
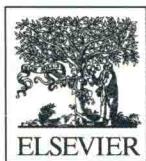
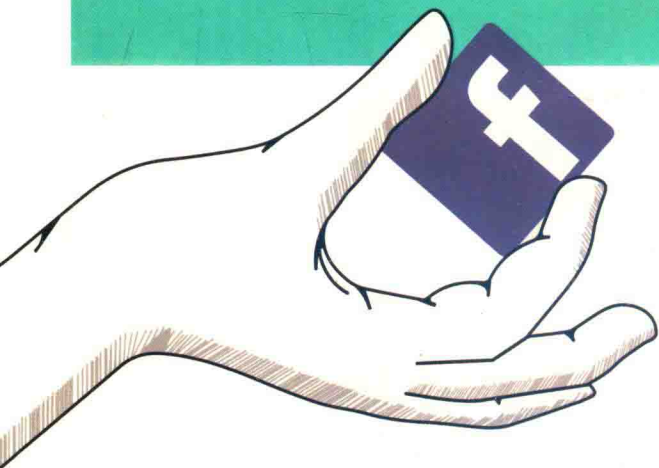




在线社交网络：

在 **Facebook** 和 **Twitter** 个体关系网中
发现的人类认知约束

Online Social Networks:
Human Cognitive Constraints
in Facebook and Twitter Personal Graphs



Valerio Arnaboldi
Andrea Passarella
Marco Conti
Robin I.M. Dunbar

著

凌 非 程 之 译



中国工信出版集团



电子工业出版社
PUBLISHING HOUSE OF ELECTRONICS INDUSTRY
<http://www.phei.com.cn>

在线社交网络： 在 Facebook 和 Twitter 个体 关系网中发现的人类认知约束

Valerio Arnaboldi, Andrea Passarella,
Marco Conti, Robin I. M. Dunbar 著
凌非 程之 译

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

Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs

Valerio Arnaboldi, Andrea Passarella, Marco Conti, Robin I.M. Dunbar

ISBN: 978-0-12-803023-3

Copyright ©2015 by Elsevier. All rights reserved

Authorized Simplified Chinese translation edition published by the Proprietor.

Copyright ©2017 by Elsevier (Singapore) Pte Ltd. All rights reserved.

Published in China by Publishing House of Electronics Industry under special arrangement with Elsevier (Singapore) Pte Ltd. This edition is authorized for sale in China Mainland. Unauthorized export of this edition is a violation of Copyright Act. Violation of this Law is subject to Civil and Criminal Penalties.

本书简体中文版由 Elsevier (Singapore) Pte Ltd. 授予电子工业出版社在中国大陆发行与销售。未经许可之出口, 视为违反著作权法, 将受法律之制裁。

本书封底贴有 Elsevier 公司防伪标签, 无标签者不得销售。

版权贸易合同登记号 图字: 01-2016-4609

图书在版编目 (CIP) 数据

在线社交网络: 在 Facebook 和 Twitter 个体关系网中发现的人类认知约束 / (意) 瓦莱里奥·阿尔纳博尔迪 (Valerio Arnaboldi) 等著; 凌非, 程之译. —北京: 电子工业出版社, 2017.3

书名原文: Online Social Networks: Human Cognitive Constraints in Facebook and Twitter Personal Graphs

ISBN 978-7-121-30869-7

I. ①在… II. ①瓦… ②凌… ③程… 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 019845 号

策划编辑: 张 迪 (zhangdi@phei.com.cn)

责任编辑: 张 迪

印 刷: 北京盛通印刷股份有限公司

装 订: 北京盛通印刷股份有限公司

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: 8.25 字数: 92 千字

版 次: 2017 年 3 月第 1 版

印 次: 2017 年 3 月第 1 次印刷

印 数: 4 000 册 定价: 49.00 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 zltz@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: (010) 88254469, zhangdi@phei.com.cn。

译者序

20 世纪末，大量研究人员发现，真实世界并不像看上去那么随机，而且可以将其对应到图论中，利用图论的知识来分析和解释现实生活的现象。因此，复杂网络的研究应运而生。

在复杂网络中，节点可以代表复杂系统中的独立元素，而节点间的连接关系则可以用连边来映射。例如，在社交网络中，如果把网络中的用户看成社交网络中的节点，那么这些用户则是通过相互之间的关注或其他交互行为构成了整个社交网络。网络自身带有的拓扑结构和网络特性等特征可以作为研究人员了解和揭示网络本质的途径。

现如今，人们的生活已经离不开互联网了，而互联网上的在线社交网络尤其与人们的生活密切相关。其中，Facebook 和 Twitter 作为世界上最大的两个在线社交网络，为研究人员了解和揭示网络中的有趣现象提供了很好的研究对象，并且还可以通过研究结果不断地改进和优化现有的应用。

本书的作者 Valerio Arnaboldi 等，从复杂网络中用到的图论知识和分析方法入手，讨论了 Facebook 和 Twitter 中自我网络的结构属性，并与其他社交网络进行了比较；另外，从动力学的角度分析了 Twitter 中的自我网络，并为在线社交网络的改进和优化提出了一些建议。最后，还讨论了如何发现在线社交网络中的结构属性以拓

展对社交网络的分析并建立在线服务。

在本书的翻译过程中，我的师弟程之投入了巨大的精力，在此表示感谢。最后，感谢电子工业出版社给了我们这次机会，感谢

感谢作者 Valerio Arnaboldi 和他的团队，为我们带来如此美妙的研究方向和分析方法；感谢我的师弟程之为翻译付出的宝贵时间和精力；感谢 SwiftGG 的 riven 和其他小伙伴们的帮助；感谢电子工业出版社的张迪编辑以及其他幕后工作者的辛勤劳动；最后还要感谢家人，感谢你们的理解和支持。

在本书的翻译和校对过程中，译者虽已尽力将作者的意思以通俗易懂的方式表述出来，但毕竟能力有限，问题和疏漏在所难免，恳请各位读者批评指正。

凌非
2017 年 1 月

前 言

毫无疑问，在线社交网络（OSNs），如 Facebook 和 Twitter，正在改变我们的沟通方式，同时它们也管理着我们的社交生活。智能手机的普及使网络与现实世界（Cyber-Physical World）的社交联系变得难以区分，且互相依赖。

在这种场景下，在线社交网络的分析成为了有趣且重要的话题，原因有二。（1）透过在线社交网络中用户的行为可以深刻理解人类的社交行为。众所周知，人们的社交容量是受限于他们的认知和时间资源的。同时，在线社交网络在这方面的影响仍然是未知的。

（2）在线社交网络是用户与信息在网络与现实交织的世界中首要的沟通方式。发现在线社交网络中人类关系的关键特性可能对设计以用户为中心的服务大有帮助。

本书中，我们研究并分析了个人社交网络（自我网络，Ego Networks）的结构属性。本书采用多学科的方法来研究社交网络，并讨论了该领域内最新的研究成果。结果显示，Facebook 和 Twitter 中的自我网络与先前离线环境下发现的结构属性相同。这就意味着，尽管在线社交网络使人们在生活中有巨大的改变，但也不能提升人们的社交容量，因为这显然受限于人的大脑容量。此外，由于 Facebook 和 Twitter 的大量数据，因此与离线社交网络相比，在线社交网络可能会发现自我网络中的新属性。这意味着，我们可以采用

大规模在线沟通数据集的研究方法来深化人类社交行为的知识。实际上，在线数据呈现了一种研究人类行为的微观视角。

最后，本书也讨论了如何发现在线社交网络中的结构属性以拓展对社交网络的分析和建立未来的在线服务。我们也讨论了几个分析信息扩散的例子，并在本书中提出了基于全新沟通平台的研究成果，即特定的在线社交网络的结构属性是如何影响该类服务的关键特征的。

致 谢

Valerio Arnaboldi 感谢家人在写书过程中对其的支持。

Macro Conti 感谢他的妻子 Laura 对其的支持和启发。

Andrea Passarella 感谢他的妻子 Erica 长久以来的理解、鼓励和陪伴。

Valerio Arnaboldi、Macro Conti 和 Andrea Passarella 在本书中的相关工作也在大数据分析和社交挖掘的欧洲实验室（European Laboratory on Big Data Analytics and Social Mining, SoBigData, <http://www.sobigdata.eu>）中实施，该实验室包含了 IIT-CNR 和社交挖掘领域内的一系列机构。SoBigData 在 H2020 内也是处于领先地位的，而且 SoBigData 研究机构也是唯一一个欧盟资助的关于大数据和社交数据挖掘的研究机构。

Robin I.M. Dunbar 的研究是由欧洲研究委员会（European Research Council）授权支持的。

目 录

第 1 章 引言	1
1.1 离线社交网络和在线社交网络	1
1.2 网络与现实交织下的在线社交网络	3
1.3 自我网络分析和社交大脑假说	6
1.4 本书的目的	7
1.5 本书结构	8
第 2 章 人类社交网络	9
2.1 引言	9
2.2 无权社交网络的宏观属性	10
2.2.1 复杂网络中的指标	10
2.2.2 社交网络分析中得出的关键结论	14
2.2.3 生成网络图的网络模型	23
2.3 从社交网络图到互动网络图	26
2.4 社会网络的微观属性	29
2.4.1 自我网络中的层级结构	31
2.4.2 扩展自我网络与结构洞	33
2.5 桥接社交网络中的微观和宏观属性	35
2.6 本章小结	38

第 3 章 Facebook 中的自我网络结构及连边强度	40
3.1 引言	40
3.2 在 Facebook 中对连边强度进行建模	41
3.2.1 基础工作	41
3.2.2 Facebook 数据及分析的方法论	43
3.2.3 Facebook 中连边强度的组合	45
3.2.4 预测连边强度的模型	48
3.3 Facebook 中自我网络的结构	53
3.3.1 用于在线自我网络分析的大规模数据集	54
3.3.2 Facebook 中自我网络的层级结构	59
3.3.3 验证	63
3.3.4 讨论	65
3.4 本章总结和讨论	66
第 4 章 Twitter 中自我网络的结构	68
4.1 介绍	68
4.2 对于 Twitter 自我网络分析的数据集	68
4.2.1 Twitter 用户的活跃时长	69
4.2.2 Twitter 的社交相关用户	71
4.3 Twitter 中自我网络的结构属性	75
4.3.1 Twitter 自我网络的尺寸对认知限制的影响	75
4.3.2 Twitter 中自我网络层级	77
4.3.3 Facebook 和 Twitter 之间的比较	80
4.4 总结和讨论	81

第 5 章	Twitter 自我网络的进化动力学	83
5.1	个人资料	83
5.2	社交网络图的演化属性	83
5.3	Twitter 自我网络的动态分析方法	87
5.4	Twitter 自我网络的动态特性	89
5.4.1	个人社交关系的演化	89
5.4.2	用户离开 Twitter	91
5.4.3	无向交流	92
5.4.4	社团活跃度	94
5.4.5	社交容量和自我网络层级的动态性	95
5.4.6	自我网络更换频率	98
5.5	总结和讨论	102
第 6 章	结论	105
6.1	介绍	105
6.2	自我网络结构和信息扩散	105
6.3	研究方向	109
6.4	本书里程碑	112
参考文献		114

第1章 引言

1.1 离线社交网络和在线社交网络

从传统定义来说,“社交网络”表现为:在特定社会环境(如工作场所、国家、科学界)中,一系列的人及这些人之间存在的双向社会关系^[1]。社交网络分析旨在通过观察网络结构的属性来解释被观察的环境中产生的社会现象(如工作场所中新观点的产生,疾病的大规模传播或科学家之间合作的产生)。

最近,社交媒体(如 Facebook 和 Twitter)的出现为分析社交网络提供了一个崭新的机会。现如今,社交媒体的广泛使用,以至于它们都可以代表大部分人的整个社交世界了。因此,针对这些社交媒体的分析可以为我们人类的社交行为提供新的视角。不同于许多传统的交流方法(如面对面交谈、电话交谈),社交媒体正在逐步建立一种新型的“在线”社交环境,在这个环境中的社会关系虽然可以不是映射自现实中存在的关系,但也可以是只在虚拟世界中产生并维持的新关系。为了突出不同社交环境之间的区别,我们把“在线”社交网络(Online Social Networks)定义为:特定社交媒体中的用户,以及存在于这些用户之间的联系而组成的社交网络;而“离线”社交网络指的是其他所有不使用社交媒体而产生的社交网络(如通过面对面或电话交流形成的网络)。我们对在线社交网络的定义侧重于社交媒体把人映射到在线交流的虚拟世界的能力,这是一种其

他交流方式都无法实现的能力。本书会大量分析和讨论不同社交环境的区别，特别是“在线”和“离线”社交网络的差异。

毫无疑问，Facebook 和 Twitter 是当前世界上最大且最重要的在线社交网络之二，同时，本书也将这两者作为主要的讨论对象。对于不太熟悉这两个社交网络的读者，我们对 Facebook 和 Twitter 的主要特点做了如下简单的描述，同时也介绍了一些在书中可能用到的专业术语。

Facebook 是世界上使用人数最多的在线社交网络服务，在 2015 年的第一季度就拥有了超过 13 亿的月活跃人数^[2]，它建立于 2004 年，到目前为止已经投入使用了 13 个年头。Facebook 为方便社交提供了许多特性。每个用户都有一个可以自定义的个人资料（profile）。通过个人资料可以查看该用户一个叫 wall 的消息栏，该消息栏记录了所有用户自己创建的（状态更新）或从其他用户收到的消息（posts）。消息（posts）可以包含图片、URL 和视频等多媒体信息。用户可以在对消息进行评论的同时，与其他用户进行交流或对消息添加其他信息。为了能与其他任一用户 B 交流（如在用户 B 的消息栏上发布消息，评论用户 B 的消息或照片），用户 A 必须和用户 B 是朋友关系。朋友关系指的是用户双方互相接受对方成为自己的朋友。用户可以在一个叫 news feed 的页面看到所有朋友发布的动态。这个页面实时显示朋友们发布的消息、新增的评论和图片等。Facebook 的用户可以与在消息栏上写消息的其他用户直接交流。消息也可以包含对多个用户的引用（译者注：类似于新浪微博的@功能）。用户可以通过消息管理器（messenger）来与其他用户进行私聊。Facebook 同时也提供其他在线交流的方式，如音频、视频

电话等。在 Facebook 中最广泛使用的就是对喜爱的内容（如消息，图片）进行点赞的特性了。

Twitter 也是在线社交网络且提供了微博服务，它于 2006 年建立，并在 2014 年的第二季度拥有了大约 3 亿的月活跃用户^[3]。在 Twitter 上，用户可以发布短消息（tweets，最多 140 个字符）。用户 A 只要关注（follow）了用户 B，那么用户 A 都会对用户 B 每次发布新的短消息时自动接收到提示。关注用户 A 的用户 B 称为用户 A 的被关注者（followers，译者注：类似于微博中的粉丝），同时，关注用户 B 的人为用户 B 的朋友（friends）。

多媒体（如 URL、视频和图片）和一些特殊的标记丰富了 Twitter 短消息的内容。具体来说，短消息可以用特殊的标记来引用其他用户（译者注：类似于新浪微博的@功能）。即使被引用到的用户并不是短消息作者的被关注者，也会自动接收到通知。当然，用户也可以回复短消息。在这种情况下，短消息的作者就会收到一个已回复的通知。

在 Twitter 中，用户可以转推（retweet）短消息，换句话说，就是把短消息推送给他的被关注者。每一条短消息都可以通过一个特殊字符“#”来指定它的主题。Twitter 使用“#”来划分所有的短消息并得到热门话题（trending topics），这些热门话题可以通过一个特定的页面来展示和搜索。当一个单词、一个词组或一个主题被非常频繁地提到时，它就成为了热门话题。

1.2 网络与现实交织下的在线社交网络

毫无疑问，像 Facebook 和 Twitter 这样的在线社交网络已经深深

影响了人们互相交流的方式，不管是年轻人还是老人。更令人惊奇的是，它们所带来的文化上的改变可能已经远远超过了原有表达自己及与他人交流的方式。只要提几个简单的例子，我们就能看出现在每个社交上的动作都可以在在线社交网络上完成了，如找工作、做广告及组织活动等。另外，我们也总是能随时利用我们口袋里的手机来访问在线社交网络。

手机等设备的使用影响了我们人类的发展，我们只须要通过手机就可以将网络与现实世界连接起来。现实世界中发生的事情只要通过手机记录下来，就可以发送到网络世界中；反过来，虚拟应用和服务也能影响现实世界。这就逐渐产生了一个网络与现实交织的世界（cyber-physical world, CPW）^[4]。这种交织的情况为创造革命性的应用打下了坚实的基础，应用也可以通过挖掘现实中用户的社交模式来提升网络世界的服务质量。

在这样一个网络与现实交织的世界中，现实中的事件会影响网络世界中用户个人的和社会领域的信息处理方式。人之所以处在这个过程的核心，主要是因为人可以通过智能设备将现实中发生的事件以图片、视频、文字等内容发送到网络世界中。社交媒体提供了一种很好的方式来支持以用户为中心的交流模式，而人们也凭此为创造和扩散信息做出了积极贡献，这些信息都受现存的社会结构影响。这就将在线社交网络放在了 CPW 情景的核心了。分析在线社交网络的重要意义主要有两个原因：一方面，它有利于我们理解人在崭新的虚拟世界中的社交行为，以及新环境下产生的社会现象；另一方面，利用从研究在线社交网络 获得的知识，它有助于建立以人为本的服务和应用。

举一个例子来说明，通过研究在线社交网络的结构是如何帮助我们理解在线社交中出现的现象的，我们也可以考虑在线社交网络对信息扩散的影响。迄今为止，针对在线社交网络全局结构的研究表明其具有标准的“小世界”特性，表现为用户间的较短平均距离和较高的聚类系数。此外，在线社交网络中每个用户的社交人数呈现长尾分布（即大多数人的社交人数很少，而少数人的社交人数很多）。另外，几乎所有用户都能被网络中的其他用户联系到，因此这让网络形成了一个“巨片”（giant component）。这一“巨片”为信息传播提供了良好的基础条件，也让在线社交网络成为首选的广告渠道，并迅速取代了电视、收音机等传统方法。尽管针对网络结构属性的研究还处在初级阶段，但利用这种方法肯定可以预测出更多未知的领域。

另外，从分析在线社交网络的角度来看，已经发掘出了许多在线社交网络的全局属性（本书其余部分会详细阐述）。然而，从用户个人的角度来看，我们仍然没有对在线社交网络产生的影响有一个清晰的认识，特别是在个人的社交网络结构和我们处理社会关系的能力上。毫无疑问，在线社交网络在与人交流联系方面是一个强有力的工具。例如，可以和老同学、海外的朋友等那些用传统方法成本较高才能联系到的人进行联系。我们通过在线社交网络增加了所能维持的社交关系的总数，但这仍难以评价在线社交网络是否提高了我们的社交能力。可能在线社交网络只是另一个维持我们社交关系的工具，它虽然非常有用，但也许不能从根本上改变我们由于认知行为等约束早已形成的社会系统结构。不管怎么样，针对在线社交网络用户所在的个人社交网络的分析已经自然而然地成为了研究

的热点，这个个人社交网络又名自我中心社会关系网络或自我网络（Ego Networks）。

1.3 自我网络分析和社交大脑假说

自我网络掌控着用户（自我）与其社交伙伴（alters）之间的关系。因此，自我网络作为人类社交网络中的一个基本构建块确定着人们的社会行为。在离线社交网络中，人们已经发现自我网络的结构属性是受到限制的。具体地说，我们的社交能力受限于人类大脑的大小及有限的时间，而这两者可以用来管理我们的社会关系。以上这些发现为社交大脑假说（Social Brain Hypothesis, SBH）的构成提供了基础，它认为大脑进化的原因主要是为了应对日益增长的社交系统中的“计算”需求；举个例子，事实上人类不得不把建立越来越大的社交网络作为升职加薪道路上的关键策略，而这需要更多的“计算资源”，因此更大的大脑诞生了^[5]。这一假说与过去几个世纪里的传统智慧完全不同，过去认为大脑进化只是（或主要是）为了应对生态问题，如如何制造工具。而与其他假说截然不同的是，社交大脑假说可以解释为什么人类仍然要留着这个消耗占比每天摄入量 20% 的大脑。虽然真正驱动大脑变大的原因似乎是亲密的社会关系，但动物通过表达复杂的社交行为，如战术欺骗和联盟的形成等，而产生了更大的大脑^{[6], [7]}。这尤其适用于与推理和意识相关的那部分大脑——大脑皮层。研究结果为 SBH 提供了正确的证据，而在灵长类动物中，呈现出了一种大脑皮层大小和社会群体大小之间的关系，其中社会群体的大小可以体现出社会系统的复杂性和社会