

是可以预测的 揭开人类行为背后隐藏的模式

Hidden Pattern behind Everything We Do



爆发

BURSTS

大数据时代预见未来的新思维

全球复杂网络研究权威

Albert-László Barabási

[美] 艾伯特-拉斯洛·巴拉巴西◎著 马慧◎译

中国人民大学出版社
China Renmin University Press



爆发

大数据时代预见未来的新思维

BURSTS

The Hidden Pattern behind
Everything We Do

[美] 艾伯特·拉斯洛·巴拉巴西 (Albert-László Barabási) 著 马慧◎译



中国人民大学出版社
·北京·

图书在版编目(CIP)数据

爆发:大数据时代预见未来的新思维/(美)巴拉巴西著;马慧译. —北京:中国人民大学出版社, 2012

ISBN 978-7-300-15474-9

I. ①爆… II. ①巴… ②马… III. ①网络经济—经济发展趋势—研究 IV. ①F062.5

中国版本图书馆CIP数据核字(2012)第063488号

版权所有, 侵权必究

本书法律顾问 北京诚英律师事务所 吴京菁律师
北京市证信律师事务所 李云翔律师

爆发:大数据时代预见未来的新思维

[美] 艾伯特·拉斯洛·巴拉巴西 著

马慧 译

Baofa: Dashuju Shidai Yujian Weilai de Xinsiwei

出版发行 中国人民大学出版社

社 址 北京中关村大街31号 邮政编码 100080

电 话 010-62511242 (总编室) 010-62511398 (质管部)

010-82501766 (邮购部) 010-62514148 (门市部)

010-62515195 (发行公司) 010-62515275 (盗版举报)

网 址 <http://www.crup.com.cn>
<http://www.ttrnet.com> (人大教研网)

经 销 新华书店

印 刷 北京中印联印务有限公司

规 格 170 mm × 230 mm 16开本 版 次 2012年8月第1版

印 张 21.25 插页2 印 次 2012年8月第1次印刷

字 数 270 000 定 价 59.90元

版权所有

侵权必究

印装差错

负责调换

BURSTS

THE HIDDEN PATTERN BEHIND
EVERYTHING WE DO

推荐序一

生活抵触随机运动吗

胡泳

北京大学新闻与传播学院副教授，洞察中国社会数字化进程第一人

人类的日常行为模式不是随机的，而是具有“爆发性”的。

预测人类行为是一个经久不衰的梦想。科学家乃至伪科学家们为了解开人类行为之谜，已经努力了数百上千年。美国东北大学教授艾伯特－拉斯洛·巴拉巴西，作为全球复杂网络研究权威，“无尺度网络”概念的提出者，畅销书《链接：网络新科学》(Linked)的作者，似乎有足够的资格，也来尝试实践一下这个梦想。

原因很简单，巴拉巴西拥有前边的追寻者所不具备的利器，那就是：当今世界的数字化，已然通过互联网、社会化媒体、电子邮件和移动电话等，将我们的社会变成了一个巨大的实验室。人类在这个实验中留下的电子踪迹，比

如打上时间印迹的文本、声音和图像，互联网搜索，社交网络中的种种关系等加在一起，合成了史无前例的海量数据集，记录了我们的活动、我们的决定以及我们的生活本身。

这使得下述想法听上去就激动人心：对这些电子踪迹的分析，会不会对人类行为的秘密提供深刻的洞见？巴拉巴西穷根溯源，宣布自己找到了被长期认为是完全偶然的人类行为之下的有序模式：他将这一模式命名为“爆发”，就是说，我们的工作和娱乐及其他种种活动都具有间歇性，会在短期内突然爆发，然后又几乎陷入沉寂。用巴拉巴西的比喻来说：“长时间休息之后就会出现短时间的密集活动，就像贝多芬音乐中悦耳的小提琴声被雷鸣般的鼓声打断一样。”

巴拉巴西在结语中论断道：

当我们将生活数字化、公式化以及模型化的时候，我们会发现其实大家都非常相似。我们都具有爆发模式，而且非常规律。看上去很随意、很偶然，但却极其容易被预测。

这个论点与前两年的一本热门书《黑天鹅》恰成鲜明的对比。《黑天鹅》的作者塔勒布认为人类行为是随机的，都是小概率事件，是不可以预测的。也因此，塔勒布相信，没有什么比一种随机的智慧对我们的生存更加重要。塔勒布其实反对长期流行的一种见解：前人的经验会给予后人教益。而该见解符合中国的传统史观：以史为鉴，可以知兴替。

如今，恰如时尚的流行风随水转，巴拉巴西提出，人类行为 93% 是可以预测的。你的生活只是看上去随机而偶然；但实际上，无论你访问网页还是访问女友，都是以爆发的方式完成的，因而也是可预测的。这其中的关键在于，无论在自然界还是人造世界，许多事情遵循幂律分布：一旦幂律出现，爆发点

就会出现。

幂律在巴拉巴西的上一本畅销书《链接：网络新科学》中已被谈论得很多，现在大家都熟知巴拉巴西在研究网络时的一项重要发现：**互联网是由少数高链接性的节点串联起来的。极少数节点拥有海量点击，而绝大多数网站只有寥寥可数的人造访。幂律决定了网络的结构和网络的走向。**

现在巴拉巴西要证明，**幂律也主宰着我们的真实活动的节奏。**为什么会存在爆发模式？因为我们工作任务太多而时间却太少。当我们遇到这种情况时，我们的应对之道是确定优先次序。我们会先干最紧要的事情，忘掉其他次序靠后的事情。一旦某件事情被忘掉，那它被忘掉可能不是一时半会儿，而是经年累月。幂律就在这种优先次序的排定中产生。

巴拉巴西说：“时间是我们最宝贵的不可再生资源，如果我们尊重它，就必须设定优先级。一旦优先级设定了，幂律规律和爆发的出现就不可避免。”巴拉巴西把爆发看成某种生命的推动力：

生命远不是流畅或随机的，而是在所有时间尺度内都是爆发的——从几毫秒到几小时的细胞活动；从几分钟到几周的人类活动；从几周到几年的疾病来袭；还有从几千年到几百万年的进化过程。爆发是生命奇迹的必要因素，表明生物为了适应和存活会进行不懈的斗争。

这样看来，偶然性中还是存在某种神奇的规律。“其兴也勃焉，其亡也忽焉”，中国人的历史智慧，其实说的不就是历史的爆发性吗？所以，巴拉巴西的看法与塔勒布相反，而与中国史家相近：如果了解人的过去，那么其未来就不会存在多少令人惊讶之处。

作为一个科学家，巴拉巴西的颠覆是大胆的。他批评科学家们仍默然接受人类行为科学的基本范式：我们的行为实际上是随意的、不可预测的、偶然的、无法确定的、不可预知的，以及无规无序的。但这一假定的唯一问题在于，它完全错了。生活如此抵触随机运动，渴望朝更安全、更规则的方向发展。

巴拉巴西在本书中试图论证的是，对于幂律的认知最终会导致对人类行为的精准预测，但他似乎并不能完全驳倒塔勒布式的世界观。

- 一方面，我们当然知道，人类是习惯的产物，所以，人类所作所为有很多是可以预测的。
- 另一方面，人类的生活中又充满波动性和分叉点，在这个意义上，个体的生活和群体的行动又是不可预测的。

当把随机性等同于不完全的信息时，塔勒布实际上提出了人类知识的脆弱性问题，这和前启蒙时代的思想家是一脉相承的，他们相信人类的理解具有不可靠性。相比之下，巴拉巴西更像一个启蒙后的科学家，高估自己的知识，低估不确定性（也就是低估未知事物的范围）。

最终的问题还在于，使人类行为完全可预测是不是一件可欲之事。试想，如果人类世界也像自然现象一样，可以被理解、量化、预测和控制，那将是一件多么可怕的事情！我们已经用认知自大毁坏了自然，如果按照巴拉巴西理论的潜台词，人类能够从我们的经验中学习的话，那么，面对人类行为的问题，我们更需要认知谦卑。当然，在塔勒布看来，人们是不会具有这种认知谦卑的，所以，“黑天鹅”总会跳出来毁掉许多长久的努力。

BURSTS

THE HIDDEN PATTERN BEHIND
EVERYTHING WE DO

推荐序二

在科学和历史之间

姜奇平

中国社科院信息化研究中心秘书长，《互联网周刊》主编

美国物理学会院士巴拉巴西的《爆发》，在最新的时间、最新的领域，讨论了一个最古老的问题：到底应该用决定论的观点，还是用非决定论的观点，看待人类行为？作为复杂网络研究的权威，巴拉巴西在大数据的新背景下，认为数据、科学以及技术的合力，会使人类变得比预期中容易预测得多。

这本书写得非常富有趣味和悬念，我是像看金庸的武侠小说那样，一口气读下来的。这本书由科学和历史两部分交织组成。其中历史的部分差不多就是一部小说，它是如此吸引人，以至于看到一半，我开始扔掉每章的前半部分，专看“小说”部分，一直看到主人公塞克勒被处死的大结局。这种感觉就像看电视连续剧，被“欲知后事如何”吸引着，结果顾不得按部就班一集一集地

看，直接到百度去搜分集剧情简介了。

故事讲的是匈牙利十字军的一段历史，主人公塞克勒受主教之命，征召农民组成十字军。在出征半路上，与贵族军发生“误会”，大水冲了龙王庙，内部打了起来。这一事变，将塞克勒激反，走上起义之路。泰勒格迪曾准确预言事情的走向。最终，塞克勒兵败被擒，坐上“燃烧的御座”，被带上烧红的铁制王冠。

这本书虽然可以当历史小说来看，但它的重心还是在每章的前半部分，即讨论科学的部分。历史故事只是为了证明作者的观点：人类行为 93% 是可以预测的，就像泰勒格迪做到的那样。在日趋精密的数字技术条件下，有了从四处搜集来的信息，我们不会再把人类的行为视为互不相关、随意偶然的独立事件。相反，它们应该是相互依存的奇妙大网的一部分，是相互串联的故事集中的一个片段。它们会在不经意时显示次序，在意想不到之处偶然出现。人类行为遵循着一套简单并可重复的模型，而这些模型受制于更加广泛的规律。

爆发，被作者视为宇宙运行的科学。作者认为，当我们将生活数字化、公式化以及模型化的时候，我们会发现其实大家都非常相似。我们都具有爆发模式，而且非常规律。看上去很随意、很偶然，但却极其容易被预测。

巴拉巴西的观点虽然独特，但并没有从哲学的谱系中逃逸。作者采用的“科学－历史”叙事框架，构成了本书所涉及问题的元问题本身。当科学和历史被“主义”化为科学主义和历史主义时，二者的对立构成的是一个老问题：

自然是决定论的，历史是非决定论的。

- 如果把社会发展当做一个自然过程，作为科学的对象，它是有规律、可预测的。
- 如果把社会发展当做一个历史过程，作为历史学的对象，它是无规律、不可预测的。
- 介于二者之间的观点认为，社会发展是一个自然历史过程。

按照巴拉巴西的观点，自然与历史的比重，大约应是 93% 与 7% 的关系。这就不难理解，作者对波普尔的观点，即否定历史决定论的观点，基本是否定的。由此可以看出巴拉巴西的观点在哲学上的大致定位。

作为《链接》一书的作者，巴拉巴西还是一位知名的网络问题专家。《爆发》一书的新意，并不在于提出了新的历史观，而在于结合科学技术发展新的事实，对特定历史观进行了重新论证。作为网络研究者，我对此也很感兴趣，尤其其中提到的“大数据”，正是下一步互联网发展的重要方向，而 facebook 推出的时间轴（Timeline）与本书反复提到的 Lifeline 都是当前网络前沿的时尚。因此我想就此谈谈我个人的判断。

书中说到的技术事实的部分是可信的，例如谈及手机运营商掌握着我们的实时通信信息和行踪等。目前，这种可供分析人类行为的参考信息主要有几大类：

- 个人在地球上的全部运动轨迹（通过 LBS 采集）；
- 个人的全部支付记录（通过在线支付采集）；
- 个人的全部交往记录（通过 SNS 采集）；
- 个人的全部言行记录（通过邮件、文档、Timeline、视频监控等采集）。

商家通过这些数据，确实可以预测客户的行为，从而提供有别于他人的个性化服务。从这个意义上来说，巴拉巴西的预言是有道理的。

但是，这些事实是否足以改变“决定论－非决定论”这个水平上的哲学结论，每个人可能都会有自己的判断。人具有自由意志，这是不同于分子随机运动之处。我们有可能成功预测到一个沉默的人在某个时刻突然爆发，但要猜透这个人的斯芬克斯之谜，光有科学和技术可能还是不够的。

BURSTS

THE HIDDEN PATTERN BEHIND
EVERYTHING WE DO

推荐序三

不在爆发中爆发，就在沉默中沉默

周涛

电子科技大学教授，互联网科学中心主任

诸位捧起这本书的时候，千万不要以为这只是一本妙趣横生又充满噱头的科普小册子，事实上，这是一本野心极大的著作——它只是在面厚墙上凿开了一个小洞，但如果你贴上去，使劲儿张望，就可以隐约看见隐藏其后的宏大世界。巴拉巴西，复杂性科学最富盛名的国际领军人物，藉由此书向大家传递的，不是一个现象、一种观点或若干模型，而是一整套理解人类行为时空模式的观念和理论。这让我想起了十多年前帕·巴克的名著《大自然如何工作》：一样才华横溢的作者，一样独领风骚的理论，一样路人皆知的野心！不同的是，巴克对外探求自然界运作规律的蛛丝马迹，巴拉巴西却试图解剖人类自由意志也不能甩脱的可预测的模式——两者尽皆美妙！

霍金说：“21 世纪将是复杂性科学的世纪！”大量典型的复杂系统，都直接或间接和人发生关系。

- 从人的社会属性出发，我们研究社会经济系统中通过人的相互作用和策略博弈涌现出来的复杂性。
- 从人的生物属性出发，我们研究人体内的神经系统、代谢系统、基因调控系统等如何协同工作。

然而，前者把人看得太小，后者又把人看得太大，对于与人相关的复杂系统的整体认识，尚缺失一个环节，就是人类自身行为在时间和空间上表现出来的复杂性。《爆发》一书迈出了弥补这一缺失环节的坚实的一步：它向我们展示了人类行为自身的复杂与多样，又从这复杂与多样中总结出了若干简单的规律，最后再告诉我们可能导致这些规律的背后机制。

爆发，英文写做“Burst”，科研文献中往往翻译为一个更低调的词：阵发。单从时间这个维度上看，“爆发”可以理解为某类事件在一段较短的时间内密集发生，之后是很长的一段沉默期，然后同类事件再次以很高的频率在短时间内多次发生。

譬如某人在一段时间内疯狂打游戏，然后会有一段时间不打游戏，下一次开始又会很高频率地打游戏。如果计算相继两次登陆游戏的间隔时间，其分布会近似符合幂律。

爆发不仅仅在人类行为中普遍存在，在自然界中也是常见的。譬如一次大地震前前后后有很多较小的地震，这些地震都在短期内密集发生，而两次大地震之间往往会有很长的一段沉默期。即便我们不考虑余震，只观察大地震发生的时间序列，我们依然会发现同一地区相继两次大地震发生的间隔时间分布很不均匀——总是较长的沉默期隔开了事件密集出现的集簇。

爆发仅仅反映了人类行为时间上的异质性，实际上，人类行为在空间上也具有异质性：我们会频繁地访问一些地点，其他地方只是偶尔到达；我们在有些地方会流连忘返，其他地方只是走马观花；大多数时候我们只是在近处逛逛，但偶尔也会安排长途旅游。人自身的特征也具有极大的异质性：有的人腰缠万贯，有的人一贫如洗；有的人朋友遍天下，有的人却形单影只。这些表面上看绝无关联的事物，却都被幂律所统治：

- 打游戏的间隔时间是幂律的；
- 地震的间隔时间是幂律的；
- 地方访问频数分布是幂律的；
- 地点停留时间分布是幂律的；
- 旅行距离分布是幂律的；
- 个人财富分布是幂律的；
- 社交网络上朋友的数目是幂律的；

.....

这本书从时间的异质性出发，想表述的却不仅仅是时间规律，还包括一切异质性中的普适特征；这本书从人类行为出发，想刻画的却不仅仅是人类行为，还包括宇宙运行的万千规律。

这本书向前迈出了一步，而更远的征程上的美景或许来自于各位读者的贡献。不管是叫爆发、阵发还是异质性，我们都会无数次遇到以幂函数为代表的一类非常广阔的分布函数，这些函数中的一大部分都具有发散的二阶矩。对这类分布函数进行抽样分析、参数估计、检验以及置信区间的确定等，都需要更加完善的数理统计理论基础。既然事件到达的时间间隔是异质的，无法用均匀过程或者泊松过程来刻画，那么以前大量从泊松过程得来的排队论的结论都需要重新审视。同样，间隔时间分布中发散的高阶矩再次成为完美理论的拦路虎。事实上，面对这样异质的序列，计算事件发生的记忆性这样简单的任务都变得不同寻常。我们有理由相信，这些普适而美妙的统计规律的发现，会给概率论、数理统计以及随机过程这些传统的学科提出新的问题，而刻画爆发特性的分布函数有一天将成为概率统计与随机过程教材中的常客。

本书的很多结论在学术界仍然存在很大的争论，本书也提出了若干亟待解决的问题。这些争论和问题包括：

- 人类行为的时间规律真的可以分为两个具有不同幂指数的普适类吗？
- 任务优先级的模型能够在多大程度上刻画人类行为，其他因素如人类活动的周期与节律，人类行为的兴趣因素等，又分

别起到多大作用？

- 社会相互作用会对人类行为爆发模式起到重要影响吗？
- 人类个体的出行距离分布真的符合幂律吗？

在科学出版社 2012 年出版的著作《社会动力学》中，我和我的同事韩筱璞、闫小勇、杨紫陌、赵志丹，对这些现象、模型和结论进行了仔细的分析，并介绍了目前存在的几乎所有的分歧和挑战。这些内容可以作为本书的有益补充。

我记得《爆发》这本书英文版问世的时候，荣智海和王煜全分别向我推荐过，后来辗转得到一本，那时候就觉得应该有一个中文的译本以飨读者。非常感谢湛庐文化以及译者的努力，使得这个愿望成为现实。该译本不仅仅是一个简单的翻译，还对原书的结构进行了优化重组。高手下棋，不仅敏于子力和实地，还讲究行棋的节奏，所以看国手下象棋、围棋，有欣赏音乐之感。本书结构精巧，读起来亦有品乐之感，既有节奏舒缓的慢板，又有力量集中的华彩，实为译著中难得的佳品！

很多具有爆发性质的事件也具有很强的记忆性：爆发后易再爆发，沉默后易再沉默。希望本书能够像一个引爆物，激发中国学术界一连串的爆发。

BURSTS

THE HIDDEN PATTERN BEHIND
EVERYTHING WE DO

目录

推荐序一

生活抵触随机运动吗 I

胡泳

北京大学新闻与传播学院副教授

洞察中国社会数字化进程第一人

推荐序二

在科学和历史之间 V

姜奇平

中国社科院信息化研究中心秘书长、《互联网周刊》主编

推荐序三

不在爆发中爆发，就在沉默中沉默 IX

周涛

电子科技大学教授，互联网科学中心主任

引言

历史不会重演，却自有其韵律 001

虽然万事皆显出自发偶然之态，但实际上它远比你想象中容易预测。马克·吐温曾说过：历史不会重演，却自有其韵律。忘掉那些将生命看做掷骰子或是巧克力盒的比喻，把自己想象成处于自动驾驶状态的做着美梦的机器人，你就会更加接近真相。

【选举新教皇】 017