

联系

BARBARA R. JASNY, LAURA M. ZAHN, ELIOT MARSHALL

翻译：隋巧虹，韩筱璞

中国科学技术大学近代物理系，合肥 230026

我们身处在相互联系、不可避免的网络世界中，无论是那些直接影响我们的，还是间接影响我们的网络体系。

—Martin Luther King Jr.

无论是从混沌状态到复杂系统，还是从人体细胞内分子运动到世界各国的相互联系，我们都是网络的一部分。这种特殊的现象显示出科学家们已经将网络分析推广到各个学科领域。

Barabasi 十年前发表了一篇关于无标度网络的开创性文章，引导我们回顾过去，展望未来(p. 412)。由于认识到误用基本假设所带来的潜在危险，来自不同领域的研究人员纷纷投入到网络分析中去的热潮需要一些缓和(Butts, p. 414)。同时，网络分析也为我们提供了一个平台，帮助我们应对重大社会挑战：**Bascompte**(p. 416)正在研究着气候变化对生态系统影响这一全球性问题，而 **Ostrom** 则在研究能可持续性维持的组织系统。

物理学家们同样也承担起了这一挑战，**Adrian Cho** 在报告中说到(p. 406)，最近物理学家们旨在使用定量的方法来预测索马里发生的种族冲突，观察因特网用户的情绪变化，追踪新的行为范式的出现。此外，**John Bohannon**(p. 409)调查研究了网络分析中一个最具争议性的用途：查明恐怖组织的重要头目并且消灭他们。此外，网络可以帮助我们理解一些潜在的机制，这些机制可影响到我们人类个体和一些社会系统，诸如货币兑换(Schweitzer et al., p. 422)，以及可以导致病毒传播(Vespignani, p. 425)的交通运输系统。

从微观层面上说，分子生物学家正在利用网络系统分析基本的细胞回路，并且利用基本细胞回路(Kim et al., p. 429)来描述如何测量并且模拟出细胞内部结构的相互联系，并基于此原理生成对外界扰动的反应预测。此外，**Science Signaling** 将注意力集中在信号网络的动力学研究上。

在许多问题上，我们需要更多更好的数据。只观察结构斑图是不够的，我们还需要深入研究他们是如何演化和改变的。我们现在所面临的挑战的难度告诉我们需要学习的东西还很多。我们应该如何理解一个复杂系统中的层次关系——从理解 DNA 序列到疾病征兆的转变，或是预测下一次经济衰退的时间？网络分析帮助我们从一个新的高度来理解世界是如何运作的，同时也让我们对未来十年里所要学习研究的内容感动兴奋不已。

无标度网络：过去十年及未来前景

Albert-László Barabási

翻译：隋巧虹，韩筱璞

中国科学技术大学近代物理系，合肥 230026

DOI: 10.1126/science.1173299

【摘要】几十年来，我们一直都默认复杂网络系统中的成分例如细胞、社会，乃至网络都是随机联系在一起的。在过去的十年里，关于复杂网络的雪崩似的研究进展表明，许多真实的网络系统，它们聚合到类似的结构上而不依赖于年龄、功能、规模等特征，而这种普遍性允许来自不同领域的研究者以一种共同的范式研究网络理论。十年前发现的无标度网络是帮助催化网络科学出现的几个重要事件之一，这是一个充满挑战和成就的全新的研究领域。

自然、社会以及许多工程技术领域需要依靠庞大的网络系统来支持，这些网络系统在近几十年来所显示出的重要性及其矛盾性已经证实了该系统所具有的复杂性和带来的理解上的困难性。举一个简单的模型，1959年由数学家 Pál Erdős 和 Alfréd Rényi[1]所引进的模型就引起了人们对于互联网系统的广泛思考。他们认为复杂系统是通过随机联系结合的，这个假设已经被社会学、生物学以及计算机科学所接受。它具有强大的预测能力，能够解释诸如为什么每个人只需要 6 次传递，就可以接触到任何一个人[2-5]，而这个现象最早是在 1929 年观察到的[2]，但是在 Watts 和 Strogatz 将其扩大到社会领域之后才在物理学领域产生共鸣[5]。然而，不可否认，这种随机假设所取得的成功确实也提出了一个问题：真实世界的网络系统是否也是随机联系的？也就是说，诸如细胞或者社会所构成的系统，如果它们的节点、分子或是社会中的人类是随机联系的，那它们之间的相互作用是否也是紧密分布？这个问题同样也激励着我们的工作，十年前无标度网络就应运而生[6,7]。

我们第一条线索来自于十年前在对万维网结构调查得到的情况[8]，暗示真实世界的网络体系明显地显示出非随机分布的特征。我们发现，一个网页有 k 个链接的概率（换句话说，度数为 k ）遵循幂律分布，

$$P(k) = k^{-\gamma} \quad (1)$$

这吃惊地背离了随机网络理论所预言的泊松分布。然而，直到我们认识到公式（1）也能够概括通过电影联系的演员合作网络、通过引用联系的科学文献网络的动力学特征之后[9]，我们才开始怀疑无标度特性[6]并不只限于万维网。1999 年在科学杂志上发表的论文就报道了在具有完全不同自然属性的网络之间所存在的出人意料的相似性，并显示出增长和优选连接这两种机制是其潜在的产生原因（如图（1）所示）。

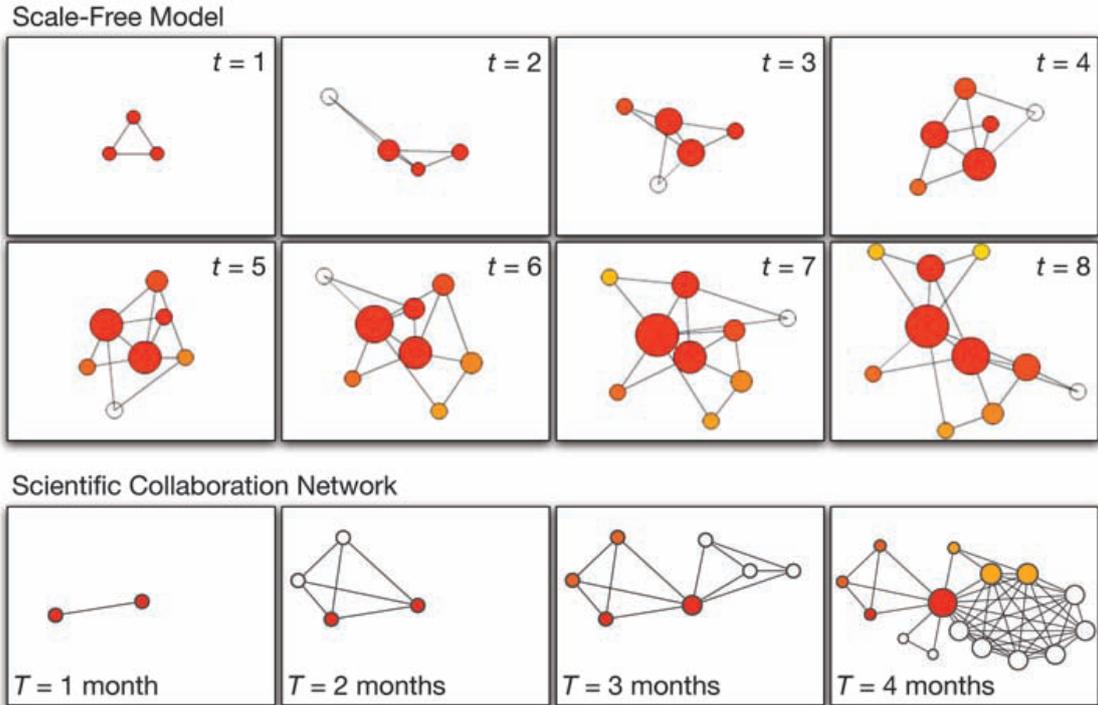


图 1：无标度网络的产生。（上图和中图）十年前，人们引进了能够产生无标度拓扑结构的简单过程，如图中顶部两行所示。从三个相互联系的节点开始，在每个图像中一个新的节点（空圈）被添加进网络。当决定在何处连接时，新的节点会倾向于附加在有着更多连边的节点上，这个过程被看作是优先连接。由于增长和优选机制，我们可以观测到一个“富者更富”的过程，这就意味着较之连接较少的节点，有较多连接的节点就更容易获得更多的连接，导致一些高度连接的中心节点的自然出现。节点大小被选择与节点的度成正比，在演化过程中最终导致最大的中心节点的出现。由此产生的网络度分布遵循幂律分布（方程（1）），指数等于 3（见附件视频 S1 到 S3）。（下图）在物理学家的科学合作网络中关于生长过程的图示。每一个节点对应于一位作者，而如果两名作者共同完成一篇文章，那么就将两个代表这两名作者的节点连接起来。这四幅图显示了在一个月的时间间隔里这种网络的生长过程，标明网络是如何随着时间不断扩展，最终导致核心枢纽的出现。而节点大小仍然同节点的度成正比。[Credit: D. Wang and G. Palla]

1999 年当我们在总结：“我们期待这种标度不变性是许多复杂网络的通用特性。” [7]当然，这更多是预言而非现实，因为大自然本可以选择有多少种网络系统，就有多少种不同的架构。但是，现代网络理论中最令人惊讶的发现可能就是这种网络拓扑的普遍性：许多真实的网络系统，从细胞到因特网，都收敛于类似的结构上，而与它们的年龄、方式和规模等没有关系。正是这种普遍性允许来自不同学科的研究者以一种共同的范式研究网络理论。

今天，对网络系统无标度特性这一关键的科学兴趣，从蛋白质分子相互作用到社会网络，从组成万维网的相互联系的文件链接网络到背后的互联网硬件，都已经毫无疑问地确立了上来。这个根据不仅来自于好的结构图和数据库，而且来自经验数据和能够预测网络结构的模型分析之间的一致性[10,11]。然而，早期发现带来的喜悦也不是没有负面影响，促使一些研究者甚至在缺少好的证据的情况下来标注许多无标度网络。但是，网络研究的结果促使我们更好地去理解构成网络结构的因素。例如，虽然非晶材料中随机绑定的原子形成一个令人着迷的网络，但是现在我们知道它既不能显示出小世界[12]也不能显示出无标度特性，这是由于这些纽带必须遵循一定的化学条件[13]。最后，人们感兴趣的一些网络的拓扑结构，像哺乳动物大脑的神经层次的结构仍然有待进一步的研究，而这意味着在这个领域中我们需要同时进行数据和模型的研究[14]。

无标度特性的意义之一在于认识到网络系统的结构和演化是不可分割的[6]。事实上，传统网络模型旨在以一种巧妙的设置方式连接具有固定数目的节点。这种无标度特性使得我们不得不承认由于节点和连接的出现，网络处在不断地变化之中（如图（1）所示）。换言之，为了解释系统的拓扑结构，我们必须得首先描述拓扑结构是如何产生的。

如果不是发现了一系列的对忽视网络拓扑结构所带来的危险，网络理论的影响很有可能会被限制住。举个例子，Romualdo Pastor-Satorras 和 Alessandro Vespignani发现在无标度网络中，传染病的传播阈值收敛于0[15]。长期以来，人们都认为只有传播速率超过一临界阈值的病毒才可以在人群中生存下来。但是，这种传播速率能够捕捉传输中的动力学行为，这个传播阈值受制于病毒传播的网络拓扑结构。因此，阈值消失意味着在无标度网络中，甚至连弱毒性病毒都可以自由传播，这个发现影响了从艾滋病到计算机病毒传播的所有传播过程。类似地，Shlomo Havlin和他的合作者们[16]发现在无标度网络中，网络整体的连通性并不被随机的节点移除所破坏，解释了真实网络系统在随机节点故障下的鲁棒性[17]。作为新兴理论的证据，这些发现[15,16]被归因于相同的数学特性，即度分布二阶矩的发散性（Eq. 1）——无标度网络的一个新的特征[6]。近来，这些特性引起了人们极大的兴趣，例如，真实网络系统（例如电网、因特网等等）存在着漏洞易遭受攻击，以及中心节点可能会遭到大规模的破坏的事实等等[17,18]。

很明显，无论是在自然界还是技术领域，没有哪一个网络系统是完全随机的，也就是说，超越随机性的机制定性了他们的演化。从度分布[6]到度度相关性[19-21]、构型[22]和社会团体[23-25]，不同拓扑特征的广泛存在性被作为研究不同现象以及做出预测的跳板。同时，网络系统理论从根本上重塑了我们对于复杂性的认识。诚然，虽然我们对于复杂概念的定义仍然缺乏一致的意见，但是网络在这个领域的作用是十分明显的：所有的系统都被认为是复杂的，从细胞到因特网，从社会到经济体系，包含了相当多的通过复杂网络相互影响的成分。可以肯定的是，我们早就注意到了这些网络。但是直到最近，我们才开始通过数据和工具来探索他们的拓扑结构，帮助我们认识到潜在的联系对于系统行为有着重大的影响：除非探讨其网络拓扑，否则没有办法去理解复杂系统。

在许多方面，在未来对于复杂理论的需求是显而易见的：我们需要理解我们所能感知的复杂系统的行为。我们需要能够预测因特网对于攻击以及交通拥堵的反应或者是细胞对于它所在环境变化的反应。为了在这个方向能够取得更大的成就，我们需要攻克下一个前沿问题，就是理解在网络中发生的动力学过程。而问题在于我们已经有了几乎和复杂系统一样多的动力学现象。例如，生物学家研究代谢网络中的反应动力学；计算机科学家观测计算机网络中的信息流；流行病学家、社会学家以及经济学家探索社会网络系统中病毒和思想的传播。尽管是多种多样的，但是否有这样的可能：这些动力学过程有着一些共同的特征？我怀疑这样的共性是存在的；我们只是还没有发现能够解释他们普遍性的框架。如果我们可以做到，同网络拓扑结构的普遍性相比，我们可能很快就可以找到能够构成复杂系统理论基础的这样的一个东西。

我们能否保持这种势头并且在下一个十年或是更长的时间取得这个成果呢？在我看来，这是可能的，主要的瓶颈是数据的驱动。事实上，大型可靠的网络数据图的突然出现在过去的十年里促进了网络理论的发展。如果在近几年能够捕获发生在网络系统中的动力学过程的详细数据，我们的想象力就成为了唯一的限制因素。如果我对接下来的十年做一个大胆的预测，它是这样的：由于我们每天所要使用的电器设备数量的增长，从移动电话到全球定位系统和因特网，它们已经捕获了我们生活中的一切行踪[26,27]，在一个真正量化的方式下，我们首先能够认识的复杂系统不是细胞，也不是因特网，而是我们人类社会。

今天，对网络系统的认识已经成为一系列传统学科的共同目标：细胞生物学家使用网络来研究信号传导和代谢过程，并命名在这一领域的一些应用；计算机科学家正在绘制互联网

和万维网的结构；流行病学家追踪病毒传播的传输网络；脑研究者正在致力于研究连接体——一种神经水平的大脑连接图。虽然复杂系统的许多研究热潮来了又走，但是有一件事日益清楚：对于复杂系统而言，组元之间内在的相互连接是如此的重要——这正是现在我们关注网络的原因。

参考文献

1. P. Erdős, A. Rényi, *Publ. Math. (Debrecen)* 6, 290(1959).
2. F. Karinthy, in *The Structure and Dynamics of Networks*, M. Newman, A.-L. Barabasi, D. Watts, Eds. (Princeton Univ. Press, Princeton, NJ, 2006).
3. S. Milgram, *Psychol. Today* 2, 60 (1967).
4. I. Pool, M. Kochen, *Soc. Networks* 1, 1 (1978).
5. D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998).
6. A.-L. Barabási, R. Albert, *Science* 286, 509 (1999).
7. In a random network, the average node sets the scale of the network, which means that most nodes have about the same number of links as the average node. For networks that follow Eq. 1, for $g < 3$ the second moment of the distribution diverges, which means that the average is not characteristic because the error bars characterizing our uncertainty about its value are infinite. These networks lack a characteristic scale; hence, they are called scale-free. Formally, networks whose degree distribution follows Eq. 1 are called scale-free networks.
8. R. Albert, H. Jeong, A.-L. Barabási, *Nature* 401, 130(1999).
9. S. Redner, *Eur. Phys. J. B* 4, 131 (1998).
10. G. Caldarelli, *Scale-Free Networks* (Oxford Univ. Press, Oxford, 2007).
11. S. N. Dorogovtsev, J. F. F. Mendes, *Evolution of Networks: From Biological Networks to the Internet and WWW*(Oxford Univ. Press, Oxford, 2003).
12. The small-world property refers to the fact that in many networks the average node to node distance is rather small, of the order of $\log N$, where N is the number of nodes in the network.
13. L. A. N. Amaral, A. Scala, M. Barthelemy, H. E. Stanley, *Proc. Natl. Acad. Sci. U.S.A.* 97, 11149 (2000).
14. E. T. Bullmore, O. Sporns, *Nat. Rev. Neurosci.* 10, 186(2009).
15. R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* 86, 3200 (2001).
16. R. Cohen, K. Reez, D. Ben-Avraham, S. Havlin, *Phys. Rev. Lett.* 85, 4626 (2000).
17. R. Albert, H. Jeong, A.-L. Barabási, *Nature* 406, 378(2000).
18. A. E. Motter, *Phys. Rev. Lett.* 93, 098701 (2004).
19. M. E. J. Newman, *Phys. Rev. Lett.* 89, 208701(2002).
20. R. Pastor-Satorras, A. Vázquez, A. Vespignani, *Phys. Rev. Lett.* 87, 258701 (2001).
21. S. Maslov, K. Sneppen, *Science* 296, 910 (2002).
22. R. Milo et al., *Science* 298, 824 (2002).
23. M. E. J. Newman, *Phys. Today* 61, 33 (2008).
24. G. Palla, I. Derényi, I. Farkas, T. Vicsek, *Nature* 435, 814(2005).
25. J. Reichardt, S. Bornholdt, *Phys. Rev. E* 74, 016110(2006).
26. A. Vespignani, *Science* 325, 425 (2009).
27. M. C. González, C. A. Hidalgo, A.-L. Barabási, *Nature* 453, 779 (2008).

Supporting Online Material

www.sciencemag.org/cgi/content/full/325/5939/412/DC1 Movies S1 to S3

我们自己和我们之间的相互作用：一个终极物理问题？

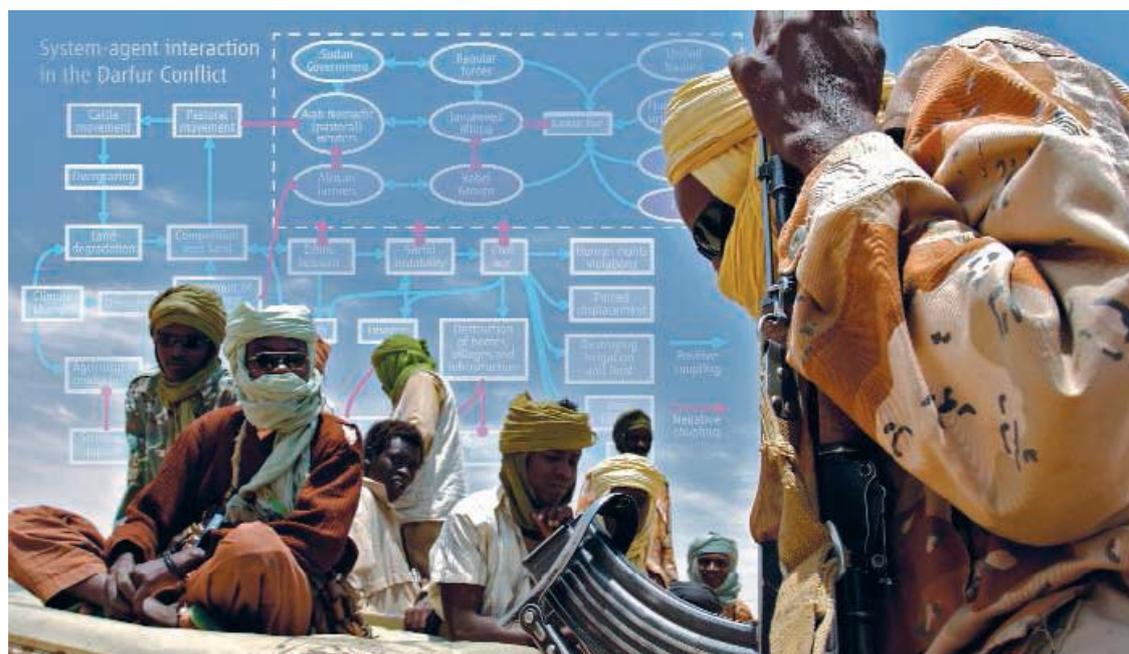
原文出处： Science 24 July 2009:Vol. 325. no. 5939, pp. 406 - 408

原作者：ADRIAN CHO

(译者：林颖婷 中国科学技术大学近代物理系 安徽省合肥市金寨路 96 号)

(姜罗罗校对)

在复杂社会经济学领域，物理学家和其他分析人员将个体看成了可交换的电子。这种方法可以解读这个社会吗？而它的困难在何处呢？



新模式军队. 民兵战士在一个模拟达尔富尔冲突的模型中与其他组织的相互作用

苏黎世，瑞士——Janusz Holyst 的语气听起来有点沮丧。“当我阅读情感心理学的教材时，发现里面既没有一个数字，也没有一个公式。”这位华沙科技大学的理论物理学家慨叹道。不过，Holyst 希望能够完善这类书所欠缺的内容。他打算发展一套能够对情感进行量化分析的方法，并试图用这些方法精确地读取个体的情感。

凭借着欧盟委员会 360 万欧元的资金支持，Holyst 和他的八名合作者们想要编写一个计算机程序来分析网络聊天室里的对话，并显示人们是何时变成兴奋、生气等情绪的。标记关键词是不够的，他说，因为人们使用语言的方式并不相同，比如有些人咒骂的行为和愤怒的情绪无关。最后，Holyst 希望能够解释群体情绪。例如，当球队取得主场胜利时，幸福的愉悦感是如何在全场球迷中蔓延的。“现在有数十种模型来解释舆论的形成，”他说，“却没有一个来诠释情感。”

Holyst 是一批人数不多但队伍在不断扩大的物理学家中的一员，他们的研究对象从原子电

子转向了社会现象，如恐怖行动、城市发展进程、和网络视频普及过程。同社会学家们一起，他们将人群看成是许多个体相互作用的“复杂社会经济系统”，并且借用一些物理学、数学和计算机科学中的概念性工具对此进行分析。上个月，130名不同领域的研究者们聚集在此讨论这项工作。*

“社会物理学”的研究开始于20世纪70年代初。一个著名的例子是，物理学家发现，个体间通过相互作用最终形成公共舆论的过程与相邻原子通过重排各自的磁场形成磁性晶体的行为十分相似；随后，Ising模型被用来分析这个社会现象，而这个模型原本是用来描述上述磁相互作用的。20世纪90年代，许多物理学家转向研究经济问题（见408页边栏），由此形成了经济物理学这一研究分支，尽管它一直处在有争议中。此外，由于复杂系统的研究者们已经在交通、流行病学以及经济领域的研究中做出了实际贡献，因此现在这种趋势有增无减。还有些人正在对付一些更加困难的问题，例如社会标准是如何出现的。

“社会物理学家提出的问题比大多数自然科学家以为的还要复杂的多，但解决起来又不像大多数社会学家想的那样无望。”瑞士联邦理工学院苏黎世分校物理研究出身的社会学家Dirk Helbing说。但是过于简化也是很危险的。因为Ising模型可能并不符合实际情况，“在某些领域，物理学家因滥用Ising模型而恶名远播”，从事流行病研究的弗吉尼亚理工大学物理学家Stephen Eubank指出。

然而，物理学家和社会学家正在共同致力于研究更加符合现实的模型，Helbing说。复杂系统研究方法有助于避免——或者至少能够解释——金融危机，比如现在的全球性经济衰退，他说。“我们花费了数十亿美元试图去理解宇宙的起源，” Helbing说，“而我们却依然不了解维持社会稳定、经济发展或者世界和平所需要的条件。”我们理解人类以及人类社会还需要付出更多的努力。

复杂的复杂系统

科学家无法按照他们定义原子或者基因的方式准确地来定义什么是复杂系统。相反，他们倾向于描述一个复杂系统看起来会是什么样子的，并且产生何种现象。“如果你问一名生物学家，‘什么是生命？’他会给你一张列表，上面写明生命所具有的特征，而不大可能给你一个严格的定义。”物理学家，来自挪威科技大学的Ingve Simonsen说，“复杂系统也是如此。”

首先，复杂系统包含了许多元素，这些元素之间的相互作用是如此的强，以至于它们总

*社会经济系统中的危机处理国际研讨会, 8-12 June 2009.

会用某种方式组成一个集体。而这种“突现行为”使得整体不仅仅是个体之和。汽车也许是复杂的，但是它不是复杂系统，因为它的每一部分与其他部分之间的相互影响是可预知的。但是交通中的汽车形成了一个复杂系统，因为驾驶员对位置的争夺行为，会导致一些意料之外的事情发生，比如会在没有明显原因的情况下，出现“幽灵堵塞”现象。

复杂系统也是十分难以应付的。各部分之间小小的改变就对整个系统产生深远的影响——就像2006年11月4日，德国北部一条电力电缆的断开造成了远至葡萄牙的区域性断电。

这种系统通常非常难以控制，并且不大可能调节到一个最佳工作状态。同时，“复杂系统产生故障的原因实在是太多了，我们根本无法对所有可能都做好防范。”Eubank说。

正如铆工这项职业，这个研究领域也是由所使用的工具来命名的。研究者们从统计物理学家那里借用了相变理论——相变过程中，大量原子的相互作用会导致整个材料产生戏剧性变化，例如水冻结成为冰，高温的铁在降温过程中出现磁力——并用相变理论去探究为何民意会突然改变，去反对一场曾经很得民心的战争。

许多工作也采用这样的网络模型：用点或者“节点”代表个体——如疾病流行过程中的受害者，而个体之间的社会联系则用连线或者“边”表示。研究者们通常采用“基于行为者（agent）的模型”：在计算机模拟中，无数虚拟用户像机器人一样，按照既定的规则进行相互作用。这种模型也许可以帮助揭示一些现象的本质，例如，投资者、对冲基金以及银行三者意见的互动是如何引起市场的震荡的。

在这种模型的背后，有一个假设，至少在某些独特的情况下，人们就像亚原子粒子一样几乎没有差别。“我们总认为我们每个人都是独一无二的，”印第安纳大学布卢明顿分校研究网络模型的物理学家Alessandro Vespignani说，“但是在人际交往中，我们大概有90%的行为未表现得那么与众不同。”这种方法非常符合物理学家们的特点——倾向于推断隐藏在复制现象下的简单机制，Vespignani说：“物理学家们在数学严格性、计算方法以及对问题的直觉三者之间带来了一个特殊的平衡。我们是做近似的艺术家。”

病菌和交通堵塞

在过去十年里，复杂系统研究者们已经取得了一些显著进展。例如，他们已经更深刻地了解了交通流。在物理学家的眼中，交通流就类似于在管道中流动的液体。尽管如此，两者间还是有一些关键的不同。由于液体的原子会互相碰撞，在液体流过一个狭窄通道时，流速通常会增加。而驾驶员则会尽一切可能避免碰撞，所以当道路比较狭窄时，流速必然会降低。

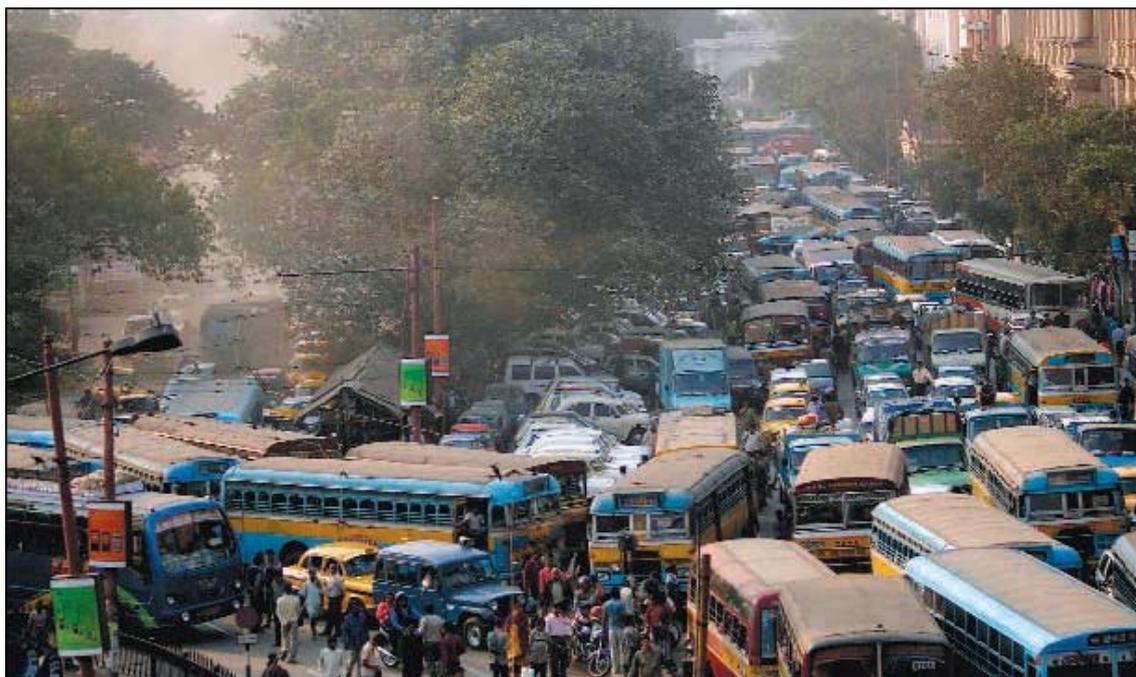
为了抓住交通流的本质，科学家们必须计入车辆之间的“力”——实际上，是驾驶员对

其他车辆的反映。在20世纪90年代，物理以及其他学科科学家们建立了多种考虑了驾驶员反应的数学模型，准确地模拟了公路和城市交通，并且这些模型已被开发为商用软件，供市政规划者等人使用。研究者们还可以解释一些让人非常惊讶的效应，例如行人通过过道时是如何自发形成两条反向通道的，以及在路中央放置障碍物竟然能够加快行人流的速度。

复杂系统专家在流行病学方面也做出了贡献。在2001年，Vespignani和他的同事提出，对一类被称为“无标度网络”的高度连接网络，不论给多少人注射疫苗，都不可能完全阻止疾病的传播。然而，2003年，工作于以色列拉马特甘的巴伊兰大学的物理学家Shlomo Havlin和他的同事发现了简单却比随机注射疫苗更有效的防疫策略。只需要在随机被接种者的朋友当中再次进行随机接种，卫生官员们通常就可以使拥有更多社会联系的人们产生免疫力，从而切断社会关系这个疾病传染的渠道。

这些工作非常及时——甚至是紧迫的。在会议上，Vespignani提出一个猪流感H1N1流行的详细模型，该模型包括了全球航线网络，以及主要大城市详尽的交通地图。这样的投入使得研究者们不仅可以预测疾病的感染率并且还可以预测疾病传播的地理途径，Vespignani说。他的初步结果表明到十月份为止，将有30%到60%的澳大利亚人会感染上H1N1病毒。

这些成果使得他们的研究进入了主流流行病学的行列。医疗官员们已经让建模者们加入到某些问题例如H1N1流行问题的研究中来。和Eubank的建模工作一样，Vespignani的工作也是由美国国立卫生研究院提供基金支持的。“这在一定程度上要归功于那些中层人员，他们不断的努力使高层官员们认识到需要让科学在流行病防治上发挥更大的作用。” Eubank说。



硬社会学科学

交通问题和传染病流行似乎还比较像是物理问题，但是研究者们同样也在探究那些看起来是纯社会学问题的现象。有一些不成文的规矩约束了我们的某些行为，比如说，禁止询问他人收入。为了探究这类社会规则的出现原因，一些人开始对演化博弈理论进行计算机模拟，使极大数量的虚拟参与者进行逻辑竞争。一个经典的设定是，网格上相邻节点上的个体之间面临着“囚徒困境”：若他们选择彼此合作，那么都会得到收益；当他们选择互相欺骗或者说“背叛”对方，那么就都被处罚。但是当成为唯一的背叛者时，就会获得更大的收益；成为唯一的合作者时则会得到更重的惩罚。这样的逻辑迫使了所有人都选择背叛对方。

然而，为了使博弈更加有趣，研究者们使参与者的策略可以进行改变。参与者可以模仿他最成功的邻居，或者迁移到更成功的博弈者的附近。**Helbing**发现，在任一种情况下，背叛者最终都会占据整个体系。但是把模仿和迁移的策略结合，就会导致合作者数量的不断扩大。

更引人注意的是，**Helbing**让两拨参与者分别在不同的规则下进行相同的博弈，结果导致了不同的主导策略。可是，交流会使得来自某个群体的博弈者，学习另外一个群体的博弈者的策略。**Helbing**说，这就像社会规范的产生，交流会使得参与者们改变他们的行为。这个结果似乎和禁止挖鼻孔毫无关系，但是社会规范通常可以是任意内容的，“因此我们认为将其从具体内容抽象出来完全没问题。”**Helbing**说。

另外一些研究者们正在分析历史运动。中世纪的欧洲由多个王国组成，这些王国由不同等级的领主统治，他们从比自己等级低的领主那里收取赋税，再上缴给比自己等级高的领主。这种间接统治体制，使得最高君主能够完全统治他们的国家。这一流程已经被瑞士联邦理工学院苏黎世分校的政治学家**Lars-Erik Cederman**建立模型模拟出来了。

Cederman将每位领主表示为地图上的一个点，以及能够最有效率地从下而上输送资金的树状网络上的一个节点。**Cederman**发现，如果一个领主对自己领地的统治力开始降低，间接统治的方式就根本不会出现。但是随着这个领主控制力降低的影响扩展到遥远的乡村——很有可能是由于科技进步——却会导致领主关系网络逐渐被简化，并且最终消失。“我们断定，这个模型是第一个体现这个历史潮流的模型。”**Cederman**说。

德国汉堡大学政治学家**Jürgen Scheffran**更加雄心勃勃，他想要建立气候变化对地区冲突的影响的模型。他和他的同事们给苏丹达尔福尔地区的种族冲突建立了模型。在这个地区里，不断向南扩展的沙漠，已经迫使阿拉伯牧民迁移到南撒哈拉农民的土地上。“这或许是第一批发现的气候改变导致冲突发生的实例之一。”他说。接下来的五年里，**Scheffran**和他的同事打算建立详细的基于行为者的计算机模型来重现这场冲突。这项艰巨的工作要量化众多参

与者的相互影响，包括农民、牧民、为农民利益而战的叛军、镇压叛军的贾贾威德民兵队、苏丹政府、援助组织以及其他。“若顺利的话，我们会得到减少暴力的更好策略。” Scheffran 说。

复杂系统领域进一步的发展方向还是未知的。由于欧盟委员会近期已经对该领域将来四年的研究工作投入了两百亿欧元，因此复杂系统研究在欧洲发展是最快的。而美国的研究者们则面临着更为困难的基金前景，来自迈阿密大学（佛罗里达）的物理学家Neil Johnson说。他将该校擅长的鱼类数理生态学应用到了恐怖分子关系网的建模中。

不过，有着成功的经验和科学家们应付新问题的创新能力，研究人员相信他们将会得到进一步的重要成果——即使他们无法预测这种特殊努力何时会使他们取得最丰厚回报。

ADRIAN CHO 报道

金融物理：多年来的争议

SCIENCE, 325 卷, 408 页, 2009 年 9 月 24 日

作者: Adrian Cho

(翻译: 丁中俊、中国科学技术大学近代物理系、安徽省合肥市金寨路 96 号)

(姜罗罗 校对)

在 1997 年, 布达佩斯大学的物理学家 Imre Kondor 和布达佩斯科技经济大学的物理学家 János Kertész 组织了一个刚刚崭露头角的领域: 金融物理的会议。这次会议的结果是毁誉参半。金融物理是复杂系统研究领域的最大分支, 并且已经有很多物理学家涌入其中。但是, 许多经济学家认为金融物理是浅薄的。“在这次会议之后, 我去了一家银行工作, 并且再也没有遇到敌意,” Kondor 说, “但是, 学术界的反应与金融业界的反应显著地不同。”

金融物理学家认为传统的经济学理论是有缺陷的。它建立在“典型代理人模型”基础之上, 例如: 一个假设的 Joe 和整个经济力量相互作用。这些模型忽略了相关性。为了严格证明其理论, 经济学家假设市场波动遵循带状的高斯分布, 这低估了大波动的概率。Kertész 强调“传统的经济学主要是创造定义明确, 易处理的, 但是没有任何实际用处的数学模型。”

金融物理学家采取基于大量数据的方法来分析问题。尽管他们尚未取得大的突破, 但他们的贡献正受到越来越多的认可。例如: 金融物理学家引入了能够分析股票之间关联性和高效地优化投资组合的新工具。他们还研究了新类型数据, 例如: 市场中最小行为的高分辨率记录和记录着每一支股票报价行情的限价订单登记簿, 不论这个报价最终是否导致交易。

Kondor 说, 物理学家正在帮助改变研究重点, 例如: 经济学家一直都知道股票利润的涨落并不像他们的模型假设的那样温和。利润的真实分布在极限时存在着“胖尾”, 这意味着现实情况中的巨额收益和亏本比假设的可能性大。但是, 大部分经济学家仍旧认为这件事是物理学家侥幸成功的, 而不是不可避免的。Kondor 进一步指出, 物理学家开始系统地研究胖尾的后果。

金融物理学仍旧没有给经济学家留下好印象。瑞士苏黎世大学的 Ernst Fehr 说: “根据我的观点, 金融物理没有影响力, 并且以后也不会有影响力”。物理学家在用他们的模型解决问题时没有抓住细节, 他说, “我曾经雇佣了一个物理学家, 但确实令我很失望。”

尽管如此, 一位德国 Kiel 大学的经济学家 Thomas Lux 却持相反的看法。他指出, 在欧洲和日本, 越来越多的经济学家正在寻找标准经济学理论的备选方案。Lux 认为“相互作用产生了经济, 但经济学忽略了相互作用, 相反, 被金融物理学家喜欢的多代理人模型在细致研究相互作用时是理想的”。

由于成群的物理学家在金融界, 一些人因为现在的全球金融危机而怪罪他们。金融物理学家则愤愤不平。他们认为 大部分在金融领域工作的物理学家被要求用现成的工具而不是发明更好的工具去筹集和评估证券, 例如臭名昭著的信用违约互换。瑞士联邦科技研究所 Didier Sornette 说: “此外, 物理学家一直对隐含在证券价格, 对冲, 风险分析中的理论很不满意。”



金融物理学家警告：大的市场波动是不可避免的

反恐怖主义的新工具：元网络分析

研究者已经发明了成熟的程序去探索社会交往表层以下的更深层次的东西。

他们在反恐战争上做得如何？

作者：JOHN BOHANNON

翻译人：唐大海

赵明 校对

PALO ALTO, 加利福尼亚——在硅谷的一个舒适的会议室，这个远离伊拉克地狱般恐怖暴力的地方，**Shyam Sankar** 在屏幕上放映了一张巴格达的卫星图片。

“现在让我们来看看严重的恐怖活动的地理分布吧”，这个软件工程师这样说。

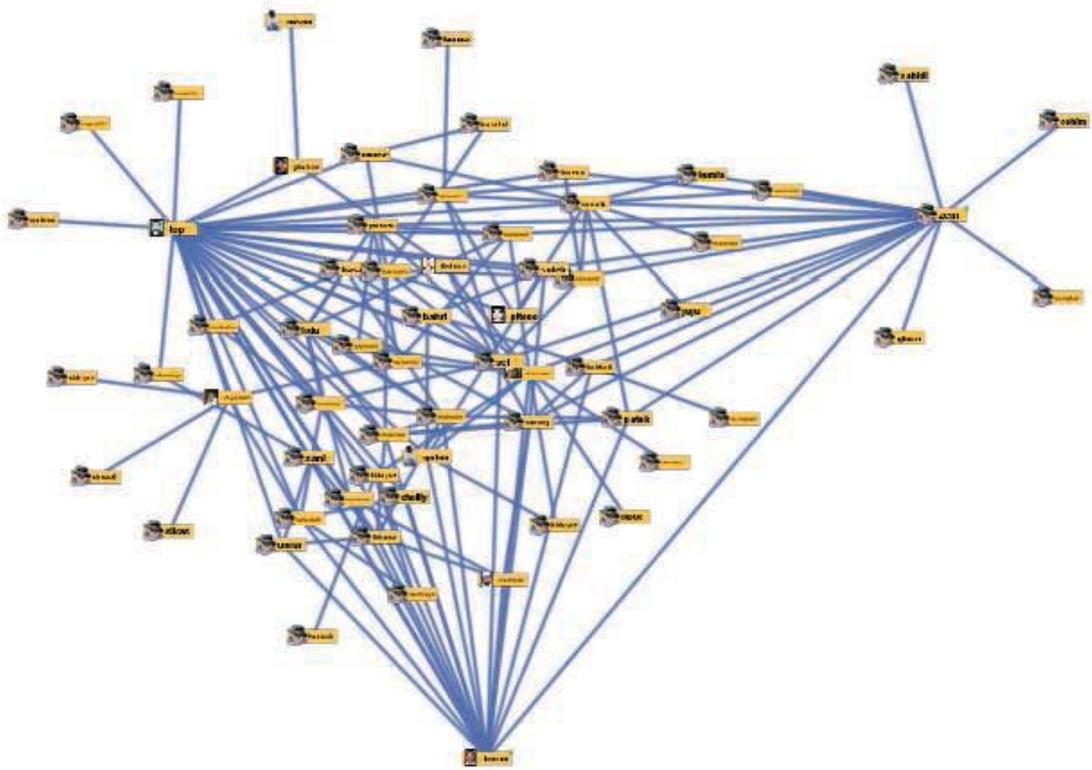
他点了几下鼠标，在地图下面弹出了一张时间表，看起来像一个城市的轮廓在 2008 年的 4、5 月间陆续地冒了出来。但是，那些冒出来的柱形图代表的不是摩天大楼，而是每天屠杀数量的记号：自杀式炸弹袭击、枪击、迫击炮袭击以及简易爆炸装置爆炸袭击。这些数据由普林斯顿大学和纽约州的美国西点陆军军官学校的科学家们的合作所得。他们正在寻找这些已知事件背后的模糊的恐怖分子关系网络。当 **Sankar** 缓慢地把光标扫过时间表时，弹出了一张巴格达的萨德尔城区爆炸的彩图——这是袭击的热成像图。“当建构这一道安全墙时，这些攻击显现出时空关联的性质”，**Asher Sinensky** 一边说一边用鼠标顺着一条横穿这个城区的线划了划。

像这样的洞见对美国和伊拉克武装力量试图预测下一次恐怖活动是关键的。27 岁的 **Sankar** 和 29 岁的 **Sinensky**，都是 **Palantir** 技术公司的“前瞻部署工程师”，该公司是这个数据分析平台的制作人，目前它的生意兴隆。

“这里是与海军研究生院合作做的工作”，**Sinensky** 说，同时，密密麻麻的小方块出现在屏幕上，他们都被线条连起来形成一张网。“这是与诺丁·穆罕默德顶层保持联系的人员的网络，” 这些人是 **Jemaah Islamiyah** 组织成员和东南亚的最受通缉的恐怖分子。这些数据由布鲁瑟尔的国际危机集团提供的非机密数据提取获得，展现了嫌疑人间的已知联系与关系，以及对于已知的恐怖阴谋他们的参与情况。“你拥有这个巨大的球形的东西，这通常没有什么意义，” **Sinensky** 说。他在下拉式菜单中选了一个命令按钮。“所以这些研究生安置了这些插件来使用经典的社会网络分析技术，像介数（**betweenness**）、中心性（**centrality**），以及本征向量中心性（**eigenvector centrality**）。”他执行了一条命令，所有嫌疑人的名字跳了出来，并且通过计算作为这个网络节点的重要性大小，这些名字被排了序。“毫不惊奇，**Noordin** 排在第一位”，**Sinensky** 说，“但是紧接下来的两个排名又是谁呢？也许我应该集中一些资源在这些人上”。

十年前，大部分的社会网络研究都是纯理论和学术性的。但在 9.11 事件以后，在如何应用这些研究到战争问题上“爆发了一阵兴趣”，**Kathleen Carley**——宾夕法尼亚州匹兹堡卡内基梅隆大学的一个计算机科学家——这样说。**Palantir** 公司只是众多争夺军事基金的许多公司中的一个。学术性的网络科学家诸如 **Carley**，也参与进来竞争有丰厚利润的美国军方合同和奖金。

尽管这方面的研究十分火热，但是，在究竟社会网络分析已经对反恐主义产生了多大效用这点上，仍然存在尖锐的分歧。有人担心，在匆忙的捕捉恐怖分子活动中，美国军方过于相信社会网络分析的能效了。一位前美国官员甚至声称在战争区域使用这些方法已经导致了不道德行为。



阴谋者联系图。围绕 Noordin Mohammed 的恐怖分子网络。





致命的攻击的线索。研究者已经努力去模拟伊拉克 IED 袭击背后的反叛分子网络。

纠缠的网络

就在 2001 年美国那次造成 3000 多人丧生的劫机事件发生后的几个星期内，人们就发现那些恐怖攻击并不是某个政府而是一个跨国的恐怖分子团体主导的：19 名劫机者和几十名资金及物流提供者。“情报机关陷于歇斯底里状态，”一位自从 1980 年代以来就替美国政府分析包括这次恐怖活动在内的多个武装分子网络的法医精神病学家——**Marc Sageman** 这样说。美国政府官员“求助于任何人”，只要他们能评估这个新威胁以及防止下一次攻击。

在第一批被挖掘出来的人当中，其中一个就是 **Valdis Krebs**，他是一位研究组织内部社会关系网络的管理层顾问。“一个恐怖组织本质上就是一个项目团队，这跟平常的团队是一样的，”他这样说。媒体提供了“连续不断”的关于 9.11 网络信息——包括他们的会议，住所以及在世界各地的金钱交易——**Krebs**把这些信息映射成一张“快速和肮脏”的社会网络。**Krebs**把他的分析结果在 2001 年 12 月发表 **connections** 杂志上，这是一家国际社会网络分析杂志。其中一个节点

——穆罕默德·阿塔(Mohamed Atta)——在整个网络图中特别突出。“阿塔在所有中心性测度中——度，距离，以及介数得分最高，” Krebs在他的文章中得出结论说。“度揭示了阿塔在网络中的活动强度，距离测度了他接近网络中其他人以及监视事态发生的能力。介数显现了他对网络流动性的控制能力——他实际上扮演了一个中间人的角色。”阿塔实际上是头目，他也是基地组织的一名成员，而基地组织头目本拉登(Osama bin Laden)声明对9.11事件负责。据Krebhs说，此后不久，他和其他的网络研究者“被邀请到华盛顿出席了许许多会议和简报”。

到2003年，美国国防部已经将原来的仅包括9.11事件的恐怖分子网络扩大到包括伊拉克和阿富汗“暴乱分子”的网络。这年的10月，美国军方开启了一个专门致力于对付简易爆炸袭击的研究任务，美国国防部把这描述为“对具有适应性和弹性的恐怖分子和叛乱分子网络的武器”。这个战略的一部分是用网络分析现有数据——从装置的种类到筹资、建造、和部署。

根据Sageman的看法，结果是令人失望的。“网络分析并没有真正起到抓住坏人的作用”，他说。网络分析作用受到限制的部分原因是由于图论作为网络分析的基础数学理论不够灵活。“我们善于建构静态网络，”他说，“但是，像这样的网络总是随时间改变，而我们并没有动态的图理论”。例如，当一个恐怖分子被抓住或被击毙，“他将会被他的堂兄弟取代”，而这个堂兄弟会带来与原来不同的社会联系。“改变一条连接会完全地改变整个图”，他说，但是理论并不能顾及这种情况。理论甚至很难顾及增长。“我们没有一个衰减函数来有效地去除噪音”，Sageman说。结果，网络模型增长的越大，噪声就越多——“能获得的信息就越少”。

根据Krebs的看法，反恐怖主义网络分析的缺陷不只是来源于数学，而是比这更深的层次。“这也牵涉到如何理解社会学”，他如是说。如果研究者不能“把注意力放在正确的事情上面以及收集正确的数据。”，那么一个网络无论多好都不能提供正确的洞见。由于没有考虑交往的内容，社会网络分析进入了“披萨外卖人问题”：把日常交往与有重要影响的联系混淆了。

作为这个问题的介绍，Sageman指出一个关于2005年7月7号伦敦地铁爆炸的报告，该报告由英国情报与安全委员会发表于5月，标题为：是否7月7号可以避免？利用网络分析，研究者跟踪了阴谋者之间的关系，生成了一个混乱的纠缠连接。“他们不能从这张图中获得任何信息”，Sageman说。“这是一团麻！”（英国政府官员拒绝提供给science杂志一个可出版的更丰富内容的版本，理由是出于“安全原因。”）

走到元分析

根据在军事学院主教网络分析的美国军队少校Ian McCulloh，未来的方向是一个被称为“动态元网络分析”的技术。McCulloh从Carley那里学到这个技术，Carley是他的顾问，去年获得了卡内基梅隆大学的博士学位。Carley和McCulloh说他们的模型可以处理随时间变化的恐怖分子网络。经典的社会网络分析的只是那些像“谁”在网络中这样的问题。Carley说，“元网络分析的内容包括人物，时间，性质，地点，原因。”通过掌握这些层面，元网络“开始在文化层面上有所收获，”。“你必须超越通信网络去考虑日常规范、态度和信念等，以及角色性别、年龄和群体的分布，”她说。“这个程序试图去做到超过一般社会网络分析的事情。”

Carley 已开发出自己的计算机程序去做那件事。据她所说，其中之一是组织风险分析程序（ORA），帮助分析师“利用人的信息来连接作为节点的人。”。然后，ORA 检视这个网络，并找出那些点，也就是那些对整个系统造成危险的人。”程序使用网络理论和社会心理计算出人的“认知需求”，“认知需求”被 **Carley** 定义为对一些事物的一种测度，如“他们需要接触多少其他人，他们涉及多少活动，那些活动有多么复杂，以及有多少资源需要处理”。在商业经营活动的例子中，分析师能用 ORA 识别出哪个员工对公司的生存是至关重要的。这个同样适用于反叛分子的网络。

McCulloh 和 **Carley** 用元网络分析技术分析了伊拉克反叛分子制作的 1500 个影视录音。**McCulloh** 说，“反叛分子会把他们的大部分攻击录制成磁带用以宣传”。“截止到 2006 年 3 月，四分之三死亡的美军士兵是由我们杀死的（磁带上说）”。**Carley** 把这些磁带上的数据抽取了下来，他说，“制成了一个大的网络，利用分割算法把这些录像分成不同的小组。当你再回头来看看这些视频时，你会看到类似于法医学迹象一样的证据来确定谁是恐怖组织的成员。”从磁带中如何抽取数据的细节是保密的，“因为我们担心反叛分子会学习我们使用的方法”，**McCulloh** 这样说。他和 **Carley** 同美国军方一起合作把这个技术在伊拉克“运作起来”。那里的美军指挥官接收到太多的信息，但却没有充足的时间来处理这些信息。**McCulloh** 说 **Carley** 的元网络软件帮助他们找到一些暗示——这提高了抓住或者杀死反叛分子的机会。

McCulloh 声称这个技术已经产生了异常的效果。“发生在伊拉克的狙击活动下降了 70%”，他说，他有信心，相信由于 **Carley** 的程序提供了正确的洞见，简易爆炸袭击死亡人数也会下降，尽管他还不能拿出数据来证明。他说，“这是元网络分析的一个简单应用”。

但是 **Sageman** 很怀疑在伊拉克的军事进步归结为网络分析。“我根本不相信（元网络分析）已经提供了有效的帮助”，他这样说。“对狙击事件下降的一个简单解释可能是部落起义”，这抑制了伊拉克的叛乱活动。“我们没有办法知道个究竟，这大体上来说是与这个领域一样大的问题”。**Carley** 则反驳说 **Sageman** “没能很好地理解这个方法”。

即便不是所有的研究者都在反恐中接受了网络分析技术，但美国军方还是接受了。军队已经于 2 年前在马里兰州的 **Aberdeen** 建立了一个网络科学中心。今年，美国军事研究实验室斥资 1.62 亿美元在一个新项目上——网络科学与技术联盟——让学术、产业和军事研究人员在“网络中心战”上工作。

Carley 是申请这个军事基金的一名学者。她说网络分析已经为战争做好了准备。十年前，模型只能一次处理几百个人的简单信息，但现在，她说，“网络分析工具可以处理数百万或千万个节点了”。

调查网络科学：它的黑暗面

作者：JOHN BOHANNON

翻译人：唐大海
赵明校对

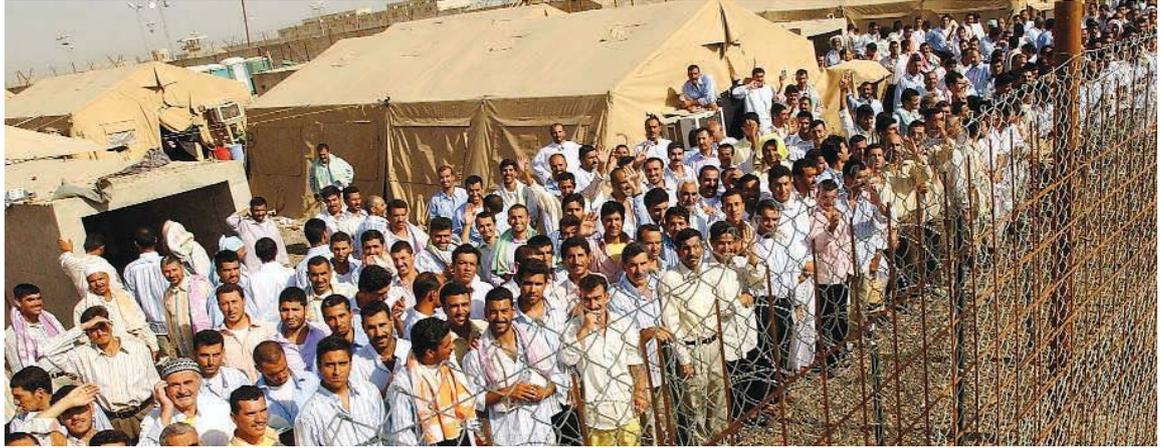
几个月以前，Lawrence Wilkerson，这位前任的美国国务院官员及陆军上校，发表了一个关于美国近年来怎样进行战争的尖刻的评论。他还描绘了关于社会网络科学怎样应用在战争区域的噩梦般的画面。在描述了美国武装力量在2003年进入伊拉克后如何收集情报以试图确定叛乱分子网络后，Wilkerson描绘了这种被他称为“马赛克哲学”的战略。他声称这个战略跟基因组排序一样。但在这里研究者不是收集数以百万计的DNA片段，而是与审问数千计的平民囚犯获得的资料打交道。

Wilkerson于2009年3月17号在华盛顿笔记——一个政治评论网站——上写道，美国军方现在依赖于网络分析“计算机程序”，以便“将节点连接起来从而识破恐怖分子或他们的阴谋”。基于新美国基础这个华盛顿的一个智囊团，Wilkerson评论道：“这种做法并不关心一个被拘留者是否是无辜的”。根据Wilkerson的看法，马赛克方法的的目标是“把任何可能的信息都抽取出来……以期得到尽可能多的关于一个村庄，一个地区或一个人群的信息……那样，尽可能多的人被尽可能久地被拘留起来以便这个方法……运作下去”。

Wilkerson告诉科学杂志说，他的言论是基于他做为美国前国务卿鲍威尔的参谋长“从2000年至2005年间接触的机密文件”获得的。他推测被关押在伊拉克和阿富汗的人数为5万人。他说他并不知道哪些程序或研究者参与进来。美国国防部拒绝对上述言论发表评论。这些言论不能被独立证实。

science杂志接触过的这些网络研究者都很熟悉建立大网络来进行情报收集的基本策略，但是许多人并不相信在阿富汗和伊拉克竟有他说的那么大规模的使用。“如果这是真的那可太可怕了”，MarcSageMan说，他是一位网络研究者，也是美国军方的长期承包商，但是这是“令人难以置信的”。他补充说，即使被试过，这个方法也不可能有效。

那些制作出网络分析程序的研究人员对此也有相似的反应。位于加利福尼亚Palo Alto的Palantir技术公司是业界的领袖，其中的一位软件工程师说他们从未听说过这些滥用的情况发生。并且，Palantir公司CEO Alexander Karp说Wilkerson所提到的“计算程序”也不可能是他们的程序，因为这家公司只是近期才开始争取美军方的合同。“如果我们知道有这样的事情发生，我们会立即退出”，他补充说。



分析素材？一位前美国官员声称在伊拉克和阿富汗无辜的平民被盘问，以便获取建立恐怖分子网络模型所需要的信息。

只有一个同**science**接触过的学者听说过这种马赛克哲学的说法。“那不是用在学术领域的术语，而是用在军事里面”，**Kathleen Carley**这样说，她是宾夕法尼亚匹兹堡卡内基梅隆大学的一位计算机科学家，现在为美国军方做网络分析方面的工作。“我没有参与有关‘马赛克’的工作”。**Carley**并没有排除这种战略是无效的可能性。在不确切知道审问和数据分析是怎样进行的情况下，她说“我认为你无法判断它是否是有用的”。

当被问到是否这种马赛克哲学实际地被美国军方或情报部门运用过时，**Wilkerson**强调说不是的。已经揭示出来的关于美军虐待囚犯的事情已经“使钟摆摆向了另一端”，他说。“我听说唯一应用和可能使用该方法的地方就是国家安全局……在那里，它的用处就是对从**e-mail**和电话通话中获得的海量信息数据库进行挖掘。”

“这是不是网络分析的黑暗面？……我认为可能是，”**Brian Uzzi**说，他是伊利诺斯州埃文斯顿西北大学的一名网络科学家。“所有强有力的方法都伴随一个黑暗面的增长。它们的力量最终会被不负责任地使用。我认为在这里最让我恐惧的事情是，一个以发现新见解著称的方法被错误地使用了。”

重谈复杂网络分析的基础

Carter T. Butts

Department of Sociology and Institute for Mathematical Behavioral Sciences, University of California at Irvine, 3151 Social Science Plaza, Irvine, CA 92697-5100, USA

E-mail: buttsc@uci.edu

翻译: 林颖婷, 韩筱璞

中国科学技术大学近代物理系, 合肥市 230026

摘要: 网络分析方法已经成为研究人际关系、神经元连接、因特网结构等等各种问题的有效方法。然而, 要正确地应用网络分析方法, 依赖于为所研究的问题选择合适的网络表达方式。

在过去的十年里, 研究人员们对网络的研究兴趣得到了迅速的增长, 而且这些研究内容大部分不属于“传统”的数学、计算机科学和社会科学领域[1, 2]。作为一种对关系结构进行描述、测量、建模的形式化方法, 网络分析方法的应用为这些新兴领域(包括物理、生物和医学)的发展铺平了道路。另一方面, 如此快速的发展也产生了曲解或者滥用现有的研究方法的风险, 导致不适当的研究结论或者较差的结果。

标准框架和核心假设

大多数网络研究借用了图论里的一套表述形式。研究者首先确定一些可分辨的实体, 并用顶点集来表示它们。顶点集里面的每一个元素, 通常被称为节点, 代表着可能参与所研究的相互作用中的单个个体。而个体间的相互作用则用连边来表示, 通常被连边连接的节点对可以是无序的(此时连接被称为无向的), 也可以是有序的(此时连接被称为是有向的)。总体说来, 网络就是一个由一个节点集, 以及这些节点之间成对的连接关系的集合构成的图(图 1A)。

这种表示框架非常严格。要将一个系统用网络来描述, 我们首先必须把系统简化为一个由离散组元构成的清晰定义的集合, 而这些组元之间的相互作用本质上应该为严格二元化的。若选择任意两个这样的组元(很可能是有序的), 它们之间的关系只有两种, 要么存在相互作用, 要么不存在。尽管这样的框架可能看上去是如此的严格, 以至于似乎没有什么用处, 但是, 这种框架是为了研究该复杂系统的某个特定性质(例如特定时间标度内某个社区内疾病传播的特点), 而对某个复杂得多的系统采取的近似。此外, 正是简化了的图形拓扑结构, 促进了它在数学 [3] 和相关科学应用 [4, 5] 上的长足发展。

为了描述更加复杂的情况, 人们对基本框架的进行了多种多样的扩展和修改。通过给各边加上不同的权重(就像 *Caenorhabditis elegans* (一种无脊椎动物[译者注]) 的神经元之间连接的不同强度 [Fig. 1B] [6]), 我们就可以避开相互作用二元化的假设。多边关系(例如组员身份)可以用能够连接任意多的节点的“超边”来表示[7]。对各种相互作用的时间特性我们可以这样做, 例如: 对可反复进行截面取样的一些集团结构或者因特网拓扑结构, 可将它们视为时间序列[8]; 对婚姻和雇佣关系的生活史数据, 可以研究其时间间隔[9]; 而e-mail收发或者无线电通讯行为[Fig. 1C], 则可以作为有效的瞬时事件来处理[10]。

许多测量、分析、建模技术是基于标准框架之上的。然而, 对某些感兴趣的系统, 若该框架的基本假设无法作为该系统的合理近似, 我们就有必要改变描述和研究的方式。那么, 在选择网络表述的时候, 有哪些因素必须考虑在内? 一旦我们进行了不恰当的选择,

将会产生什么样的后果呢？

何时节点可称之为“节点”？

假如有一个生物学家，他想要观察动物与寄生植物之间相互作用关系，并采用了网络方法来进行研究。假定有足够的时间、技术和资源，他可能对指定地区进行采样并记录所有相互作用，或许会构建一个该地区每只动物和该动物觅食点之间相互联系的网络。但是要将什么确定为一个觅食点？若将每一株植物看作一个独立觅食点，对于小型植物，这样做还能显得合理；但是，哪怕对于一棵树，这样做也可能会掩盖了这颗树可能会带来的复杂影响。在各解剖学单位类型（例如树皮、树干、树叶）之间或者之中进行进一步划分，可以完善一些细节，但是这里同样需要决定要采用哪种划分方式。生物学家定义潜在觅食点的方式将会非常显著地影响相互作用网络的结构。

一个基本的问题是，如何定义与研究兴趣相关的可区分的实体类。当然，能划分这么一个类的前提是，我们默认这样的类是可以被定义的（并且，这样做是科学有效的）。这里有一些做出了正确假设的例子：研究朋友关系网[11]或者血缘关系网[12]时将个人作为节点，以及研究引用关系网时，将个人的出版物作为节点[13]。另一方面，研究一些集团之间的相互作用时，如团体[14]、家庭 [15]或者组织之间的相互关系，则有可能会出现困难。这是由于相互作用着的单元具有流动性，并且组成大集团的子单元可能会与本集团以及其他集团的子单元产生相互作用。

和生物学的例子类似，如果我们将所有存在潜在的相互作用的元素都归入到一个独立单元内，那么很可能成为对现实的一个不恰当近似。例如，我的研究小组曾经研究过在有组织救灾过程中形成的网络。如果将所有在国家政府的支持下运作的团体都合并成单元，那么类似城市搜救队这样的小单位和大的政府部委部门之间的差异就会被模糊掉，同时还会错误地认为他们之间必然存在资源共享和合作关系。同样，其他的系统也会因为整合而造成信息丢失，包括蛋白质结合点处的分子结构形式[16]以及互联网拓扑结构中的分层结构[17]。

节点集的不同确定方式会从本质上影响到所形成网络的规模和密度，最终影响到随后的分析。例如，基本的网络特性如聚集度（衡量了边连接到少数节点的程度）在定性和定量上都会随着网络规模变化而变化[18]，甚至对于相当简单的网络生成模型也是如此[19, 20]。在存在分层结构的情况下，不同的集团划分方式也会形成具有迥异结构特点的网络（图 1A）。为了避免得到错误的结论，节点集的定义必须包含研究中可能参与到所研究的相互作用中的所有可分辨的个体；而且该定义在跨网络研究中需要保存一致。当无法唯一地确定个体集的情况下（例如，有时候在进行地理分析时要将连续空间划分开来[21]），有穷网络表示法很可能是不恰当的。采用一个替代的框架（例如连续空间表示）会更加有效。在另一些情形下（如多级过程[22]），则可以对同一系统在多种整合层次上同时进行分析。

什么才能成为边？

许多法律关系（如婚姻）是二元化的，只有极少数社会才会允许第三者的同时存在。即使有些关系是可以用所达到的程度来描述，我们依旧可以用“有”或“无”来诠释它们。例如两人之间是否存在朋友关系，给定两个神经元之间是否存在连接。当这种关系反映了相互作用的总趋势或是潜在可能性时，采用二进制表述法会使理论研究和实际测量得到大大简化。举例说来，研究性关系网络就要比列举出性生活情况简单的多，并且实际上一些潜在的相互作用也会对结果产生影响，甚至这些关系其实并未真正发生（就像交易网络中潜在的贸易伙伴[23]或者支配关系中的第三方观察者[24]）。

二元化有时候会将研究引入歧途。许多形式的相互作用本来就是间歇性地产生，并且还是在不同条件下发生的。把这种相互作用的数据二元化，不仅掩盖了这些变化，还要求

研究者确定一个阈值以便二元化，而阈值的选择会在很大程度上改变所形成网络的性质，这种改变有可能是由于选择性断边[26]直接导致的，还有可能是通过改变网络（连接）的密度间接进行的[27]。表现在不同的连接强度上的网络结构是非常大量的[图 1B]。这个问题仅仅通过改进数据收集和统计的技术是无法解决的。恰恰相反，我们必须先确认所研究的相互作用是否足够稳定，保证在感兴趣的时间范围内可以很好地将它近似看成是不变的常数函数；还要确认该函数的值是否可涵盖到每对个体之间的相互作用，确保能够近似采用二元描述。有些情况下，个体之间的关系大不相同（如贸易汇率和迁移率），无法共同使用一个阈值；此时用一个加权图来描述通常会更加合适。评定这种近似方法的效果的进一步工作，特别是那些能提供在某些特定领域上具体的、由实际证明可行的实践方针的研究工作，将是非常有意义的。

时间尺度以及网络演化进程

在考虑如何用合适的节点和边描述系统时，还必须弄明白所感兴趣的演化进程的时间尺度。相对于某些进程，比如仅需要数小时或者数天的信息传播过程，稳定关系如亲属关系或友谊关系（形成需要多年的时间[28]）可以近似认为是基本不变的。在人的一生这个时间尺度上，这种关系网络的结构不可能不发生改变，但是在我们的演化进程的时间尺度内，却可以认为网络结构延续了数十年。同样，快速演化的网络的动力学特性（如紧急情况下的无线电通信（图 1C） [29]），即使对某些快速行为，如信息交流，也具有潜在的重要意义。因此，若对动力学过程认识错误，就会得到非常错误的结论。

什么情况下采用静态网络描述会带来偏差？一个实例就是艾滋病的传播过程。对性行为的研究普遍认为在固定的时间段内，个体拥有的性伴侣数目的分布是不均匀的（平均值远离分布函数的中间值） [30]。早期的一些基于极端的不均匀度分布的网络（例如，幂律分布 [31]）上简单扩散过程研究有力地证明了，艾滋病毒和类似的性传播疾病的传播能力主要受到少数拥有大量性接触的人的行为影响[32, 33]。这个结论曾经具有相当大的现实意义，因为它意味着只有控制中心节点才有希望更有效地降低传播阈值[31, 32]。尽管节点度幂律分布的模型对这些网络的适用性已经被质疑[30, 34]，同样重要的是，对艾滋病传播问题性接触的时间累积网络的假设是有效的。对动力学网络对疾病传染的易感性而言，接触存在的时刻和持续时间[35]是非常关键的，然而这些却在时间累积网络的表述中被掩盖了。这在图 1D里有体现：一个特定网络是否变成全体被感染状态或者全体非感染状态，取决于边存在的持续时间和建立时刻。

对动态网络上传播行为的研究表明，连边发生的同时性也是传播能力的重要预测因素；当各边为长程同期连接时，均匀分布的低平均度网络也有利于疾病传播；而各边为短程且断续连接的高平均度的网络则可能不利于传播[35–37]。以减少同期连边为目的的干预手段并不等同于中心节点控制，因此可以看出，通过分析动态网络得到的公共健康建议，会与通过时间累积得到的静态网络的结论完全不同，尽管后者似乎是建立在合理假设上的。

虽然艾滋病毒的传播是一个具有说服力的例子，但需要注意的是，类似情况也会发生在一些看起来与此完全不同的系统中，例如无线电通信[Fig. 1C]以及点对点通信网络。举例说来，在最近的研究中已经强调了网络成员进入与移出（或“搅动”）对系统演化的影响 [38]；这种情况下，只有考虑了节点集的动力学性质，才能理解连边的动力学（潜在的和实际的数据传输）。

结论

用网络来描述实证现象只是一个理论上的处理方式。它要求人们对存在哪些相互作用的个体、相互作用的形式和发生的时间尺度进行假设。这些假设并不是可以随意进行的，实际上人们很可能会在这里犯错。不管是研究蛋白质相互作用，性关系网络或计算机系

统，要想得到正确的结果，最关键的就是选择好合适的表述方式。

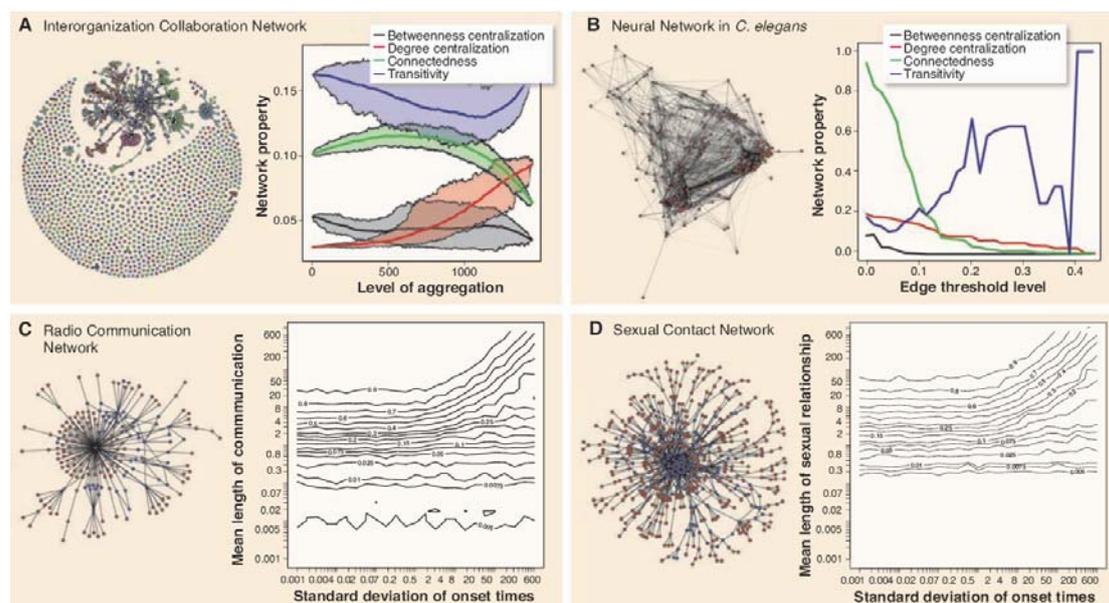


图1: 改变“节点”和“边”定义方式的影响。A、卡特里娜飓风灾害抗灾头13天, 团体合作网络[39], 体现了节点整合的潜在影响; 边代表合作关系。根据整合层次的不同, 每个次级团体都可以看作是一个节点或者是一个上级组织(包含着许多次级节点)。(左)未进行任何整合操作的结构情况, 从属指定团体的所有次级团体用同一种颜色表示。(右)反复整合从原网络中随机选择的团体, 使得它们的上级团体可以形成一系列聚集着的新结构。基于100个整合值系列, 阴影部分显示了其中主要的90%的整合值的范围, 而线代表它们的平均值。整合的程度, 以及整合步骤的顺序(由模拟间隔的宽度体现)都会影响网络基本性质。B、*C. elegans*的神经网络(根据[6]的数据), 左图中连边的颜色深度, 反映了神经元间连接强度的不同。节点代表神经元, 边反映了有向的连接[箭头指示方向]。当不考虑连接强度(阈值=0)时, 所有可能的连接都被显示出来了。选取不同的连接强度阈值(从0到所能观察到的强度最大值的50%)会生成不同结构参数的网络(右)。C和D显示了在动态演化的网络中, 连边时间特性所造成的影响, 参见等高线图(右); 每条线表示的是该时刻扩散影响到的人群比例。(左)所研究的网络的时间累积作用, 包括了观察时段内所有产生的连接。C、911事件中的一个无线电通讯网络, 包括了[29]中所提供的人数最多的人群, 该数据取自在纽新航港局哈德逊河捷运(Port Authority Trans-Hudson)第26频道上进行过的各种电话通讯联系(左)。我们随机选择一个个体, 按照给定的连边参数, 通过指数传播过程建立信息传播网络, 每条线中间的数字(右)是经过250次模拟的平均结果。网络的静态性质没有发生改变, 但是连边的时间特性的变化(通讯时长, 和观察时段内的开始时间)导致扩散能力的变化。D、(右)采用了文献[40]中性接触网络的最大部分进行传播过程的模拟, 得到了与[C]相似的结果, 尽管度分布和集聚特性(多重共同路径连接起来的人的比例)并不相同。每一条边标出了被一个随机个体感染到的比例(250次模拟的平均值), 该个体由是扩散过程的一个参数给出。

参考文献

1. S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, *Science* 323, 892 (2009).
2. M. Newman, A. Barabasi, D. J. Watts, Eds., *The Structure and Dynamics of Networks* (Princeton Univ. Press, Princeton, NJ, 2006).
3. B. Bollobás, *Modern Graph Theory* (Springer, New York, 1998).
4. M. E. J. Newman, *SIAM Rev.* 45, 167 (2003).
5. U. Brandes, T. Erlebach, Eds., *Network Analysis: Methodological Foundations* (Springer-Verlag,

Berlin, 2005).

6. D. J. Watts, S. H. Strogatz, *Nature* 393, 440 (1998).
7. S. Wasserman, K. Faust, *Social Network Analysis: Methods and Applications* (Cambridge Univ. Press, Cambridge, 1994).
8. T. A. B. Snijders, *J. Math. Sociol.* 23, 149 (1996).
9. C. T. Butts, J. E. Pixley, *J. Math. Sociol.* 28, 81 (2004).
10. C. T. Butts, *Sociol. Methodol.* 38, 155 (2008).
11. K. M. Carley, D. Krackhardt, *Soc. Networks* 18, 1 (1996).
12. J. P. Boyd, *J. Math. Psychol.* 6, 139 (1969).
13. N. P. Hummon, P. Doreian, *Soc. Networks* 11, 39 (1989).
14. R. L. Breiger, *Soc. Forces* 53, 181 (1974).
15. M. Murphy, *Eur. J. Popul.* 12, 363 (1996).
16. D. Reichmann, O. Rahat, M. Cohen, H. Neuvirth, G. Schreiber, *Curr. Opin. Sys. Biol.* 17, 67 (2007).
17. L. Subramanian, S. Agarwal, J. Rexford, R. H. Katz, *Proceedings of IEEE INFOCOM* (2002).
18. C. T. Butts, *Soc. Networks* 28, 283 (2006).
19. D. Strauss, *SIAM Rev.* 28, 513 (1986).
20. M. S. Handcock, *Dynamic Social Network Modeling and Analysis*, R. Breiger, K. M. Carley, P. Pattison, Eds. (National Academies, Washington, DC, 2003), pp. 229–240.
21. S. Openshaw, *The Modifiable Areal Unit Problem* (Geo Books, Norwich, 1984).
22. P. R. Monge, N. S. Contractor, *Theories of Communication Networks*. (Oxford Univ. Press, Oxford, 2003).
23. D. Willer, Ed., *Network Exchange Theory* (Praeger, Westport, CN, 1999).
24. I. D. Chase, C. Tovey, D. Spangler, M. Manfredonia, *Proc. Natl. Acad. Sci. U.S.A.* 99, 5744 (2002).
25. H. Whitehead, S. Dufault, *Adv. Stud. Behav.* 28, 33 (1999).
26. J. P. Onnela et al., *Proc. Natl. Acad. Sci. U.S.A.* 104, 7332 (2007).
27. K. Faust, *Sociol. Methodol.* 37, 209 (2007).
28. D. L. Morgan, M. B. Neal, P. Carder, *Soc. Networks* 19, 9 (1997).
29. C. T. Butts, M. Petrescu-Prahova, B. R. Cross, *J. Math. Sociol.* 31, 121 (2007).
30. D. Hamilton, M. S. Handcock, M. Morris, *Sex. Transm. Dis.* 35, 30 (2008).
31. F. Liljeros, C. R. Edling, L. A. N. Amaral, H. E. Stanley, Y. Aberg, *Nature* 411, 907 (2001).
32. Z. Dezso, A. Barabási, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 65, 055103 (2002).
33. R. Pastor-Satorras, A. Vespignani, *Phys. Rev. Lett.* 86, 3200 (2001).
34. J. H. Jones, M. S. Handcock, *Proc. R. Soc. London Ser. B* 270, 1123 (2003).
35. J. Moody, *Soc. Forces* 81, 25 (2002).
36. M. Morris, M. Kretzschmar, *AIDS* 11, 641 (1997).
37. M. Morris, S. Goodreau, J. Moody, *Sexually Transmitted Diseases*, K. K. Holmes, et al., Eds. (McGraw-Hill, New York, ed. 4, 2007), chap. 7.
38. D. Stutzbach, R. Rejaie, *Proceedings of ACM SIGCOMM* (2006).
39. C. T. Butts et al., *31st Annual Hazards Research and Applications Workshop*, Boulder, CO (2006).
40. J. J. Potterat et al., *Sex. Transm. Infect.* 78 (suppl. 1), i152 (2002).
41. 致谢：感谢与K. Faust, M. Morris, J. Moody, C. Marcum, A. Markopoulou和R.

Martin的有益讨论，感谢J. Potterat和S. Muth提供他们的数据。本研究工作由美国国家科学基金会BCS-0827027 以及CMS-0624257，美国海军研究办公室奖N00014-08-1-1015 提供基金支持。

解开生命之网

Jordi Bascompte

Integrative Ecology Group, Estación Biológica de Doñana, Consejo Superior de Investigaciones Científicas, Calle

Américo Vespucio s/n, E-41092 Sevilla, Spain. E-mail: bascompte@ebd.csic.es

(李志华 译 中国科技大学近代物理系 合肥 230026)

(刘建国 校对)

摘要: 对生物多样性的研究通常只集中于物种的丰富程度,但忽略了物种间的相互作用。现有的研究或事先假定物种间的相互作用为均匀分布或一次只考虑两个或几个物种间的相互作用。与此相对照,基于网络的研究方法提供了一个能够描述对于物种间相互作用并且强调它们整体相互依赖性的有效表示方法。为了应对生态系统的整体变化,如何将成对的相互作用的理解扩展到系统整体的关联仍然是一个需要解决的重大挑战。

基于网络的生态学研究强调物种间的相互作用模式,而并非识别构成群落的物种(相互作用网络节点)。对物种间相互作用的复杂网络描述和达尔文对**纠缠的河岸(tangled bank)**的想法一样古老,这说明网络在生态学的研究中是十分重要性的。尽管如此,网络只是最近才被纳入主流的生态学领域。“生命之网”(Web of life)模型描绘了物种的全局相互依赖性,对忽略物种相互作用或假设相互作用均匀分布于物种的生态学理论做了必要的补充。

网络方法得益于从其它领域,如物理和社会学的概念和方法。这种观念的交流使得我们能够把生态网络跟蛋白质相互作用网络或因特网集团结构的连通性相比较。对不同系统来说,建立一个可以相互比较的框架是非常有益的,因为这表明这些网络的形成存在一个共同的机制。并且,这些研究工作有助于对寻找共同的结构特性,以及解决具有干扰的稳定性问题。

生物多样性的结构

食物链网络是生态学的核心,它通过构建一个群落中物种间的捕食作用链接来描述和量化生态系统的复杂性[2-7]。大的网络由小的模体(motif)组合而成,模体是相对于网络随机化节点间重复出现的联结模式[8]e (Fig.2)。例如,现实的食物链网络表现出一个重复出现的3-捕食食物链(其中掠食者吃掉消耗者,后者又依次吃掉食物源; Fig. 2A),而杂食模式(掠食者既吃掉消耗者又吃食物源; Fig. 2B)重复出现在多数食物链网中,但在其它一些网络中并不出现[9]。Robert May 在理论研究的基础上认为,模体如何组成更大规模的网络可能会影响网络全体的稳定性[9]。寻找支持这一理论的现实依据是当前一个十分活跃的研究领域[10]。对食物链网络中的模体分析同样也扩展到包含量化的信息,比如相互作用的强度[11]及掠食者和被食者的体重指数之间的比例,等等。

最近,生态学家研究了捕食-被捕食之外的共生相互作用,比如在植物和它们的授粉动物或种子传播者之间的关系。这种相互作用对促进和保持地球上的生物多样性[13]及把群落集中于一个共生依赖网络起到了重要的作用(Fig. 1A)。这种共生网络有以下特点:(i)是异质的,其中大多数物种与少数几种物种相作用,少数物种具有比单个机会高得多的相互作用数;(ii)是巢状的,其中专一物种与广泛物种所作用的种群的子集相互作用;(iii)是建立在物种间弱的

和不对称的联结 (例如, 在某些情形, 当一种植物强烈依赖于一种动物, 而后者却较少地依赖前者)[14]。因此, 共生网络既非随机组织的, 又非隔离组织的, 而是聚集于一个具有广泛联系的核心物种的周围。

几组物种在时间和空间中共同演化[13], 对这种现象的研究得到了网络方法的促进。如果这几组物种及其相互作用在网络中重复出现, 它们就可以被当作模体。这时, 它们就作为基本的模块被扩展到整个网络。这些模体通常会以一种可以预测的方式在生态系统中变化和发展, 因此, 科学家提出了共同演化的地理镶嵌体的概念 (geographic mosaic) [13]。共同演化的非网络分析中容易产生两个误解: 首先, 共同演化导致高度特定的一对一的相互作用; 其次, 在物种丰富的群落中的共同演化生成弥漫性组合 (diffuse assemblage)。地理变化的网络模体和地理镶嵌体的出现打破了这些假设。

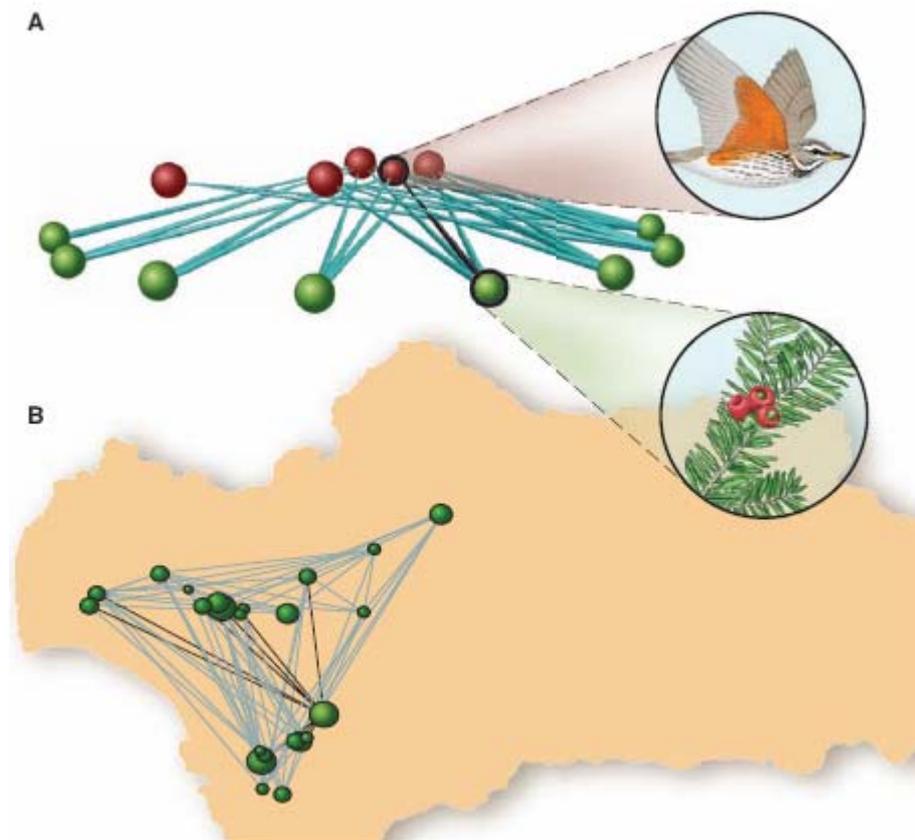


Fig.1.生命之网是一种概括各元素间生态学连接的有力表示。本图展示了几个生态网络如何从成对相互作用扩展到整体的例子: (A) 一个植物动物共生网络, 描绘了植物及其种子传播者之间的互利作用和 (B) 一个表示一种地中海植物空间基因穿越一个复杂地貌中的变异。像 (A) 一样的研究聚焦于群落尺度的共同演化以预测全局的变化如何传播穿越这种网络。像 (B) 一样的研究提供了一个框架来处理所有地段对基因流的同时影响和量化单一地段对整个物种的持续性的重要性。

结构影响鲁棒性

缺少对生态网络结构的理解, 就无法触及网络在受到物种灭绝、栖息地丧失或其它人为影响时的鲁棒性问题。这种网络模型已经能够预期物种的随机灭绝会导致相继的其它物种灭绝, 这是由食物源的缺失造成的。在这些模拟中, 研究人员发现食物链网络对于物种的随机灭绝是鲁棒的, 但这种情况下的鲁棒性依赖于几个少数互相联结的物种的作用——把整个网络粘在一起。如果这些关键物种消失了, 那么可以预见到整个网络会很快崩溃[7,15,16]。这些仿真针对的是物种的数量而非他们的类型。接下来的工作是将植物和动物也置于网络相互作用中。

系统生态的相关性(例如, 具有相同基因的物种)部分地解释了物种间的相互作用模式[17]。因而, 共同演化的雪崩效应趋向于演化到分类学上相关的 (taxonomically related) 物种, 这会造成共同演化树的随机剪断和分类多样性更快的被侵蚀[17]。与此相关, 共同灭绝在食物网中中级联发生趋向于与食性单一的物种有关。即那种在群落中与其他物种较少重叠趋向于被一组捕食者吃掉和/或吃一组食物的物种[18]。因此, 食性的多样性——物种所扮演的食性角色的范围——比基于只关注共同灭绝的物种的数量的预期下降得快得多[18]。将来的研究应该补充级联共同灭绝的仿真。这些仿真可以通过映射一种物种起到的生态系统服务的缺失来实现[19], 比如生物控制中的授粉。

科学家提出了两种方法研究种群动力学, 第一种用假定物种不具有动力学的固定节点, 并利用拓扑的方法仿真物种灭绝的后果。另一种方法赋予每个物种一个种群动力学模型。一般来说, 动力学方法已经运用了最基本的捕食相互作用的描述, 例如一个 3-捕食食物链 (Fig.2A) 这些基本的对食性相互作用的描述(即理论家熟知的捕食模式)在一个简单的成对相互作用和复杂的整体群落之间建立了沟通的的桥梁[20]。事实上, 基于每个物种的充足性以及每个物种是怎样被其他捕食者或被捕食的物种的充足性影响, 捕食模式能够在每个物种上用数学表示出来。但是, 在单个捕食相互作用模型间的稳定性和整个大的食物网络研究中, 这仍旧是一个空白。

网络模体为填补这个空白提供了一个可能的方法, 即按一个食物网络里的相对频率来处理最重要的捕食模式。重复出现的模式 (模体) 逐渐成为理解食物网络的最重要的概念[9-11]。但是, 我们不能假定整个网络的稳定性能够由其分支的稳定性导出, 因为各个部分的和不等于整个网络。因而, 我们需要增加对整个食物网络的稳定性理解的研究。

最近, 生态学家分析了大型食物链网络模型, 其中有两种要素刻画了这些模型。第一种: 研究者在占据多个捕食级别中的许多物种中加入了现实的相互作用网络[6, 7]。第二种: 他们用对物种身体尺寸和新陈代谢率的现实估计[21], 这和随机赋参数值的 Lotka-Volterra 模型形成对照。利用这些模型, 科学家们研究了网络特征的稳定性作用, 比如观测到的捕食者和被捕食者间的身体尺寸比例以及预测了部分物种被去掉的后果[22]。这些预测的准确度随着食物链网络规模的增大而增加, 因而食物链网络越复杂, 对物种灭绝影响的分析就越简单[22]。

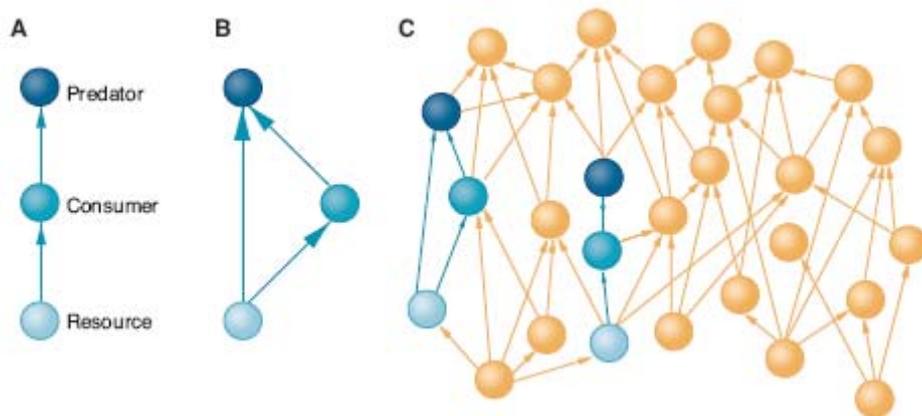


Fig.2.构建生态网络的基本模块。(A and B) 两种捕食模块: 一个 3-捕食食物链 (A) 和一个杂食链 (B), 这两种模块是动力学分析的主题。其中一些模块可以在整个网络中重复出现, 从而可以看作网络模体 (C)。未来的研究有希望能够解释在哪个程度上整个网络的稳定性能够被它的基本单元所决定。

结构影响网络规模

网络结构不仅会影响一个网络的鲁棒性 (在何种程度生物多样性会消失), 而且会影响它原来的规模 (一开始有多少物种可以存在)。测试竞争与共生相互作用在决定可以稳定存在的

最大物种数的扩展理论[23]表明物种丰富度的潜在增长来自于共生的网络结构。对于任何相互作用数，共生网络的巢状结构能最大化共生的物种数[23]。例如，两个植物物种相互竞争吸引共同的授粉者会增强，因为更多的授粉者会被该地区可得的花朵的总数所吸引。在这些情况下，在相反的竞争和促进力下会取得一个平衡，这取决于共生网络的结构[23]。

合理地研究网络结构与功能的关系面临着三个挑战。首先，需要发展新模型，需要提出既包含种群动力学怎样影响网络拓扑又包括拓扑如何影响动力学的模型[6]。其次，对网络的分析需要包含多元相互作用类型，因为稳定性很可能与多元相互作用类型在组合中起到的作用有联系[23]。如果共生和对抗效应是非叠加的，这种研究就尤其是相关的，因为直到目前为止，网络研究主要是独立开展的[24]。第三，物种入侵、气候变化和其他当前对生态和环境的挑战将需要一个网络焦点，因为多元物种可能受到局部和全局两种尺度的正在进行的变化的挑战[25]。例如，共生授粉网络表现出促进入侵植物物种并且网络结构同样似乎能缓冲入侵的影响[26]。

超越物种的相互作用

网络在生态学中的应用并非局限于物种相互作用，而是能够应用于穿越复杂地貌（landscape）的族群迁移。例如，一个节点可作为一块可得的栖息地，而连接两个这样两块的一条边能够表示个体或基因的运动。当栖息地由连续而变为分离的块时（Fig.1B），物种原来在此分块地域栖息的持续性将由地块之间的局域灭绝和迁移间的平衡来决定[27]。的确，网络可能是复杂地貌的一个简化表示，即使参数化的数据无法得到。这些网络的拓扑结构提供了单独地块的对整个地域联通性的相对重要的信息。例如，网络理论已经应用于处于危险的墨西哥斑点猫头鹰研究，通过把原来分离的地块映射成节点并且用最大扩散距离上的信息来处理这两种地块是否会有潜在的联结。这种表示导致的预期是，只要栖息地块的网络子集保持住，即使栖息地发生变迁，此物种仍将能存活下来。

此外，这种方法可用于可视化或分析基因变量在空间中的网络（Fig.1B）。传统的方法应用一个族群对另一个的成对效应的总和，但网络方法使得完全考虑多局域族群在基因变化成型中的影响[29]。最后，当网络理论应用于族群生物学，它能用来处理谁碰见谁的异质性。这种方法可以扩展到社会网络，作为估计在复杂社会中疾病传播[30]和合作演化[31]的一个途径。

结论

网络是一种有用的生态系统描述方法，它能够表示多个单元间相互作用的组成。网络方法在生态系统中的应用提供了一个概念性的框架来了解在群落层级的干扰的影响。这是面临全球环境变化下，生态学朝着更复预见性迈进的第一步。网络同样能够在以前的均匀族群、疾病和社会中引入异质性。最后，网络使得我们找到看似不同的体系间的一般性，即尽管本性不同，他们会有相似的形成过程并且/或者相似的力作用于它们的结构来实现其功能。尽管我们仅仅刚开始理解环境改变怎样影响物种相互作用以及生态系统动力学，通过分析简单的成对的相互作用，网络能够提供一个方法来处理一些问题，类似于过渡捕鱼怎样导致捕食中的级联效应，或者共生的分裂会如何影响族群中族授粉服务[25]。当看似不相关的领域的想法相互之间产生流动（一个复杂系统研究的特征属性），我们可以预见，未来将会出现更加准确地预测对食物网络的干扰反应的模型，这正是当今生态学家面对的主要挑战。

参考文献及注释

- [1]. C. Darwin, *On the Origin of Species by Means of Natural Selection* (John Murray, London, 1859).
- [2]. J. E. Cohen, *Food Webs and Niche Space* (Princeton Univ. Press, Princeton, NJ, 1978).
- [3]. S. L. Pimm, *Food Webs* (Chapman & Hall, London, 1982).
- [4]. G. Sugihara, thesis, Princeton University, Princeton, NJ, 1982.
- [5]. R. M. May, *Nature* 238, 413 (1972).
- [6]. M. Pascual, J. A. Dunne, *Ecological Networks. Linking Structure to Dynamics in Food Webs* (Oxford Univ. Press, Oxford, 2006).
- [7]. J. M. Montoya, S. L. Pimm, R. V. Solé, *Nature* 442, 259 (2006).
- [8]. R. Milo et al., *Science* 298, 824 (2002).
- [9]. D. B. Stouffer, J. Camacho, W. Jiang, L. A. N. Amaral, *Proc. R. Soc. London Ser. B* 274, 1931 (2007).
- [10]. M. Kondoh, *Proc. Natl. Acad. Sci. U.S.A.* 105, 16631 (2008).
- [11]. J. Bascompte, C. J. Melián, E. Sala, *Proc. Natl. Acad. Sci. U.S.A.* 102, 5443 (2005).
- [12]. S. B. Otto, B. C. Rall, U. Brose, *Nature* 450, 1226 (2007).
- [13]. J. N. Thompson, *The Geographic Mosaic of Coevolution* (Univ. of Chicago Press, Chicago, 2005).
- [14]. J. Bascompte, P. Jordano, *Annu. Rev. Ecol. Evol. Syst.* 38, 567 (2007).
- [15]. J. A. Dunne, R. Williams, N. Martinez, *Ecol. Lett.* 5, 558 (2002).
- [16]. J. Memmott, N. M. Waser, M. V. Price, *Proc. R. Soc. London Ser. B* 271, 2605 (2004).
- [17]. E. L. Rezende, J. E. Lavabre, P. R. Guimarães Jr., P. Jordano, J. Bascompte, *Nature* 448, 925 (2007).
- [18]. O. L. Petchey, A. Eklof, C. Borrvall, B. Ebenman, *Am. Nat.* 171, 568 (2008).
- [19]. A. Dobson et al., *Ecology* 87, 1915 (2006).
- [20]. R. D. Holt, in *Multitrophic Interactions in Terrestrial Ecosystems*, A. C. Gange, V. K. Brown, Eds. (Blackwell Science, Oxford, 1997), pp. 333–349.
- [21]. U. Brose, R. J. Williams, N. D. Martinez, *Ecol. Lett.* 9, 1228 (2006).
- [22]. E. L. Berlow et al., *Proc. Natl. Acad. Sci. U.S.A.* 106, 187 (2009).
- [23]. U. Bastolla et al., *Nature* 458, 1018 (2009).
- [24]. C. J. Melián, J. Bascompte, P. Jordano, V. Křivan, *Oikos* 118, 122 (2009).
- [25]. J. M. Tylianakis, R. K. Didham, J. Bascompte, D. A. Wardle, *Ecol. Lett.* 11, 1351 (2008).
- [26]. M. A. Aizen, C. L. Morales, J. M. Morales, *PLoS Biol.* 6, e31 (2008).
- [27]. I. Hanski, O. Ovaskainen, *Nature* 404, 755 (2000).
- [28]. D. Urban, T. H. Keitt, *Ecology* 82, 1205 (2001).
- [29]. R. J. Dyer, J. D. Nason, *Mol. Ecol.* 13, 1713 (2004).
- [30]. J. P. Aparicio, M. Pascual, *Proc. R. Soc. London Ser. B* 274, 505 (2007).
- [31]. H. Ohtsuki, C. Hauert, E. Lieberman, M. A. Nowak, *Nature* 441, 502 (2006).
- [32]. I thank L.-F. Bersier, P. Buston, J. E. Cohen, J. Dunne, M. A. Fortuna, R. D. Holt, P. Jordano, T. Keitt, J. Lavabre, R. M. May, J. Olesen, D. Stouffer, G. Sugihara, J. N. Thompson, J. Tylianakis, and two anonymous reviewers for comments on a previous draft. P. Jordano, A. Aparicio, and M. A. Fortuna provided material for Fig. 1. Funded by the European Heads of Research Councils, the European Science Foundation, and the European Community Sixth Framework Programme through a European Young Investigator Award.

分析社会生态系统持续性的普适性框架

Elinor Ostrom

¹ Indiana University, Bloomington, IN 47408, USA; ² Arizona State University, Tempe, AZ85287, USA.

E-mail: ostrom@indiana.edu

(翻译: 丁中俊、中国科学技术大学近代物理系, 合肥 230026)

(刘建国 校对)

摘要: 在全球范围内存在着一个重大问题: 渔场, 森林和水资源的潜在损失。人们对导致自然资源改善、恶化的原因的理解是有限的, 因为各个学科使用不同的概念和语言来描述和解释复杂的社会生态系统。如果没有一个共同的框架来组织人们的发现, 分散的知识将不会被积累。直到最近, 普遍接受的理论是: 资源的使用者永远不会自组织地维护他们的资源, 政府必须采取相应措施。然而, 很多学科的研究发现一些政府的政策会加速资源的破坏, 反而一些资源的使用者为了达到资源的可持续性会花费时间和精力。我们使用一个普适性的框架来鉴定十个影响自组织可能性的变量。

现在的世界正面临着大量自然资源的破坏和丢失的威胁, 例如渔场、湖泊、森林、生物多样性的减少, 以及巨大的气候变化。所有人类使用的资源都包含在复杂的社会生态系统中。社会生态系统由许多子系统以及这些子系统的内部变量组成, 类似于由器官组成的有机体, 由组织组成的器官, 由细胞组成的组织, 由蛋白质组成的细胞[1]。在一个复杂的社会生态系统中, 一些子系统是相对独立的, 例如资源系统, 资源元素, 使用者以及管理系统, 但是又通过相互作用产生一些结果, 这些结果通过反馈相反地影响这些子系统及其组成部分以及其他社会生态系统。

人类需要科学知识来维护社会生态系统, 但是生态科学和社会科学是独立发展的, 不容易结合起来[2]。此外, 一些学者趋向于发展简单的理论模型来分析资源问题的各个方面, 并且提出普适性的办法。例如, 对由于缺少公认财产系统而提出的自然资源破坏的理论预测曾提出用特殊政策来解决所有的问题, 最后导致失败[3,4]。

当收获资源的人是多种多样的, 互不通讯, 并且没有建立相应规范管理这些资源时, 对资源崩溃的预测在非常大的、高值的、开放系统中是正确的[5]。该预测在下列条件下是无效的: 当收获人和当地领导自组织有效的规范来管理一个资源, 或者当主体在严格的实验环境中采取其它措施避免过度收获[3,6]。

在诊断为什么一些社会生态系统是持久的, 而另一些会崩溃时存在着一个核心的挑战: 确认和分析这些复杂系统在不同的时空尺度上各个层次之间的关系[7-9]。完全理解一个复杂系统需要了解具体的变量以及它们的组成部分是如何联系的[10]。因此, 我们必须学习如何研究和利用复杂性, 而不是把它从这些系统中消除[11]。因为不同的学科运用不同的框架、理论和模型来分析这个复杂的多层次整体的各自部分, 因此, 这个过程是非常复杂的, 需要一个共同的分类框架使各个学科朝着更好地理解社会生态系统方向努力。

我将展现一个最新的多层次嵌套的框架来分析在社会生态系统中取得的成果[12]。图 1 给出了这个框架的概况, 并且显示了四个一级核心子系统之间的关系。这些子系统除了之间会互相影响, 还会和与其相连的社会、经济、政治环境以及相关的生态系统相互影响。这些子系统是 (i) 资源系统 (例如, 一个包含确定领土并标明受保护的公园, 在这块领土里有森林、野生动物和水系统); (ii) 资源元素 (例如, 树林、灌木、公园里其他植物、各种野生动物、水量和水势); (iii) 管理系统 (例如, 管理公园的政府和其他组织、使用公园的相

关规定、和如何制定这些规定)；(iv) 用户（例如，为了食物、娱乐和商业等目的以各种方式使用公园的个体）。每个核心子系统是由多个二级变量组成的（例如，资源的大小、资源元素的移动性、管理级别、用户的资源知识）（表 1），这些二级变量进一步由更高级别的变量组成。

这个框架有助于识别单个社会生态系统的相关变量，例如在缅因州海岸上的捕虾渔场和依靠捕虾生活的渔民[13]。它还提供了一套研究其他具有相似性质的社会生态系统的变量，例如威斯康辛州北部的湖泊（例如，为什么一些湖泊的污染程度比其他的严重？）[14]，世界各地的森林（例如，为什么当地管理的森林比政府保护的森林旺盛？）[15]，用水制度（什么因素影响农民有效地管理灌溉系统？）[16]。如果没有一个框架组织在理论和实证研究中识别的相关变量，那么由生物物理学家和社会科学家从不同的国家的不同的资源系统中获得的离散的知识就不可能被积累。

提供一套相关变量及其子系统的研究框架很有意义的。这些变量及其子系统可以用来设计数据收集器，进行实地作业，以及分析关于社会生态系统持续性的研究。它有助于确定一些影响某些特殊政策的因素。这些特殊政策只会增强某种类型和大小的资源系统而不会增强其他系统的持续性。表 1 列举了许多实证研究中已经确认的会影响相互作用和结果的二级变量。选择相关的二级或更高级别的变量来做分析取决于正在研究的特殊问题、社会生态系统的类型、以及时空尺度。

为了说明社会生态系统框架的用处，我将关注这样一个问题：什么时候一个资源的使用者会花费时间和精力去防止重蹈覆辙？Garrett Hardin [17]早些时候认为用户会陷入到过度使用的泥潭中，绝不会花费时间和精力使自己脱身。如果这个回答已被研究证明了，将不需要用社会生态系统框架来分析上面的问题。各个学科的广泛实证研究发现许多资源的用户为了增强资源系统的持续性曾设计和实行了昂贵的管理系统[3,6,7,18]。

这个问题的一个理论解释是：当管理一个资源的期望收益高于用于投资更好的规范的花费时，用户自组织的可能性更高。尽管能够创造联合收益，但自组织地维持一个资源将会花费时间和精力，导致短期经济收入的下降。这些代价，某些用户不按照规则获取收益的风险会导致用户避免进行代价较大的变化，依旧按原来的方式获取收益[6]。由于获得预期收益和成本的正确可靠的测量值代价很高、很困难，使得根据用户的预期净收益来验证理论的正确性变得很困难。

被实地研究人员观察和测量的许多变量被设定为影响用户从群体行动到自组织的可能性。通常表一中带星号的十个二级变量（）用来确定用户自组织地管理资源可能性的正面或负面影响[3,6,19,20]。为了说明这些变量对理解持续性和解决什么时候会发生自组织行为问题很重要，我先简要地讨论它们如何影响预期收入和成本。

资源系统的大小。对于与土地相关的资源系统(例如森林)，由于确定边界、检测使用方式和获得生态知识需要巨大的成本，使得非常大的领土不可能达到自组织状态。非常小的领土不可能生产出大量的有价值的产品流。因此，适中的领土最有可能达到自组织[15]。同样，始终从中等大小的沿海地区、湖泊和河流捕鱼的渔民[13]比在大海里寻找高价值鱼的渔民更有可能达到自组织[5]。

系统的生产能力。一个资源系统的当前生产能力通过所有部分的自组织产生影响曲线。如果一个水资源或渔场已经耗尽或者非常丰富，用户不会认为为了未来而需要管理它。只有当用户看到缺乏时，才会自组织地进行投资[19]。

系统动力学的可预测性。系统的动力学需要具有足够的可预测性，使得当用户打算制定特殊的收获规则或划定禁止进入的领域时能够估计将会发生什么。森林系统比水系统容易预测。一些渔场系统接近于数学混沌，对用户和政府官员来说特别具有挑战性[21]。小尺度系统的不可预测性导致田园系统的用户在更大的尺度上组织起来来增加整体可预测性[22,23]。

资源元素的移动性。由于观察和管理一个系统需要成本，移动的资源元素像野生动物和流淌着的水具有自组织的可能性比静止的元素例如树木和湖泊里的水要小[24]。

用户的数量。由于让大量用户聚集到一起并且赞同改变需要花费更高的代价，因此组织规模的大小对自组织执行成本的影响是负面的[19,20]。如果管理资源的成本很高，例如管理印度的大量公共森林，大规模组织在调动必需的人手和其他资源方面更能胜任[25]。因此，组织的大小具有一定的相关性，但是它对自组织的影响取决于其它的社会生态系统变量和管理任务的类型。

领导人。当资源系统的某些用户具有企业家的素质，并且由于之前组织过其它事情被尊为当地领导时，更有可能存在自组织[19,20]。例如，从印度的卡纳塔克邦和拉贾斯坦邦的48个灌溉系统样本的调查显示，大学毕业生和有影响力的长辈的出现对建立灌溉组织有很强的正面影响[16]。

规范/社会资本。如果资源系统的所有用户遵守关于如何在一个小组里行为的道德伦理标准并且足够相信其他用户能够保持一致，则将降低达到统一意见和监管的成本[20,26,27]。

社会生态系统的知识。当用户共享相关社会生态系统的属性时，他们的会相互影响。在其他社会生态系统中使用的规则等知识时，他们将会降低自组织成本[7]。如果一个资源系统的再生速度比较慢，人口反而增长的很快，例如在东部岛屿，用户不会理解应该考虑资源的容量，导致自组织的失败和资源的破坏[28]。

资源对使用者的重要性。在成功的自组织案例中，用户主要依靠资源进行生活或对资源的持续性有很高的依赖性。否则，组织和维持一个自管理的系统的代价将不值得用户去努力[3,7,15]。

集体选择规则。当用户(例如墨西哥的塞里渔民[29]和尼泊尔的森林用户小组[30])在集体选择水平上有自主权去制作和加强自己的规则时，他们降低了执行和防止资源被侵犯的代价[5]。

获得这个十个变量的测量值是分析用户是否自组织的第一步。由于任何一个变量的作用取决于其它的社会生态系统变量的值，因此，对这些关系的数据分析是很有挑战性的。在大部分复杂系统中，这些变量以非线性形式相互作用[8-10]。更进一步，尽管社会生态系统的长期持续性最初依赖于用户或政府制定的规则，但这些规则在长期运行中是不够的[7,18]。

如果最初由政府或用户制定的一套规则和当地的条件不一致，将很难达到长期的可持续性[8,9,18]。对灌溉系统[16,26]、森林[25,31]和海岸渔场[13]的研究表明长期的持续性依赖于与资源系统属性、资源元素属性、用户属性相匹配的规则。因为虾的肚子里的卵容易被看见，禁止捕捞怀孕的雌鱼的规定容易在虾身上实施，并且对维持虾场非常重要[13]。然而，这个规定在怀孕的鱼身上很难实施，因为对于鱼是否存在鱼卵是很难评估的。

传统科学家已经研究并记录了长期存活的资源系统中使用的规则，并比较、记录了在世界各个地区使用的各种规则[21]。简单的楷模政策是没有用的。例如，加拿大政府为西海岸建立的总捕鱼配额政策导致出现渔民普遍倾倒不需要的鱼，对捕鱼政策的误解导致在1995年关闭了基础渔场[32]。为了弥补最初的错误，政府再次开放了渔场，但是将海岸分为50多个部分，分配了可转让的配额，并且要求所有的渔船有中立的监察人员在船上记录所有的捕鱼过程[32]。

此外，规则的长期持续性依赖于监管、执行以及没有被更大的政府的政策破坏。对多个国家森林的最新研究显示，规则的长期效用依赖于用户互相监管收获行为的积极性[15,31,33,34]。更高级别的管理系统既可能促进也可能毁坏处于社会生态系统级别的管理系统。例如，在非洲、亚洲、和拉丁美洲的殖民力量不承认已经发展好几个世纪的当地习俗，反而强加上自己的规则，这种做法的结果不是造成资源的破坏就是过度使用[3,7,23]。

人们正努力修改和发展在这里呈现的社会生态系统的框架，目的是为了建立可比较的数

据库来发现、收集相关的流程，这些流程会影响森林、牧场、海岸、水系统的持续性。跨学科的研究将加快积累的速度，并且增加关于如何增强社会生态系统持续性方面所需的知识。我们需要关于社会生态系统的核心变量的定量和定性数据使得学者们建立和测试关于政府、社区、个人之间成本和利益的理论模型，以及提出更好的政策。

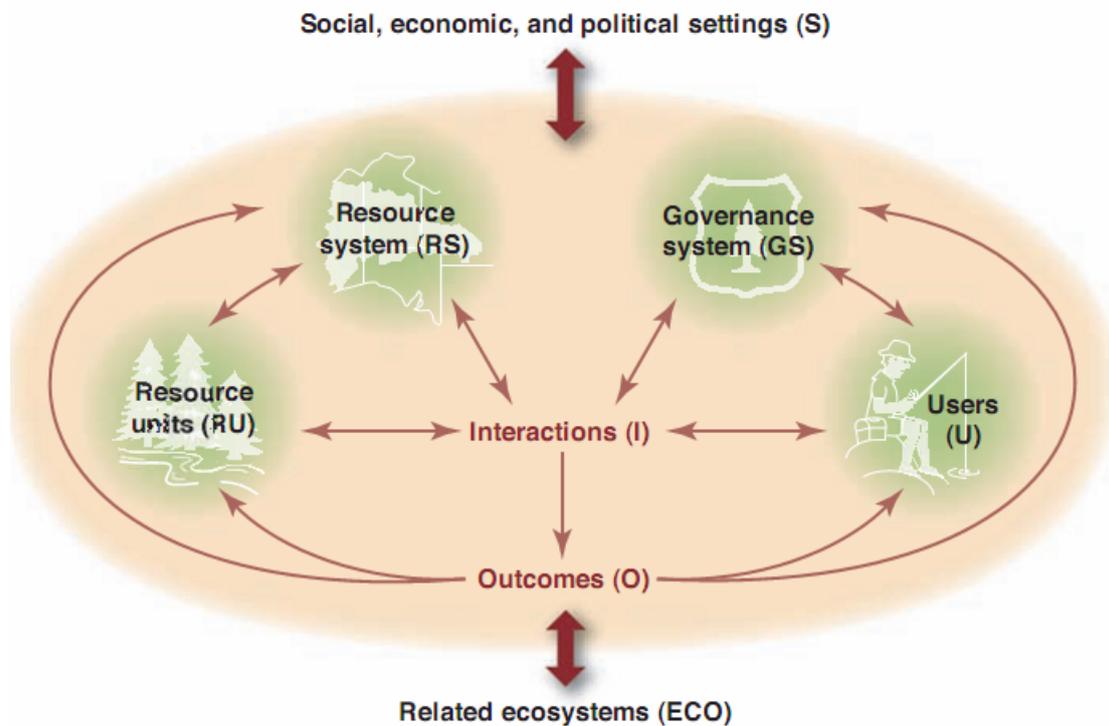


图 1. 分析社会生态系统的框架的核心子系统

表 1. 分析社会生态系统的框架的一级核心子系统下的二级变量的例子 (S, RS, GS, RU, U, I, O 和 ECO)。这个框架没有按照重要性对变量进行排序，因为它们的重要性随着不同的研究而变化。

社会、经济、和政治环境 (S)

- S1 经济发展。 S2 民主趋势。 S3 政治稳定性。
S4 管理资源的政策。 S5 市场刺激。 S6 媒体。

资源系统 (RS)	管理系统 (GS)
RS1 组成部分 (水、森林、牧场、鱼)	GS1 政府组织
RS2 系统边界的确定性	GS2 非政府组织
RS3 资源系统的大小*	GS3 网络结构
RS4 人造设施	GS4 财产权利系统
RS5 系统的生产能力*	GS5 操作规则
RS6 平衡性质	GS6 群体选择规则*
RS7 系统动力的可预测性*	GS7 宪法规则
RS8 储藏特征	GS8 监管和批准过程
RS9 位置	

资源元素 (RU)

- RU1 资源元素的移动性*

用户 (U)

- U1 用户数量*

RU2 增长和替代的速率	U2 用户的社会经济学属性
RU3 资源元素之间的相互作用	U3 用户的历史
RU4 经济价值	U4 位置
RU5 资源元素的数量	U5 领导/企业家*
RU6 与众不同的标志	U6 规范/社会资本*
RU7 时空分布	U7 社会生态系统知识/智力模式*
	U8 资源的重要性*
	U9 使用的技术

相互作用 (I)	→	结果 (O)
I1 不同用户的收获级别		O1 社会性能的测量 (例如, 效率、公平、义务、可持续性)
I2 用户共享的信息		O2 经济性能的测量 (例如, 过度收获、弹性、生物多样性、持续性)
I3 商讨过程		O3 作为其他社会生态系统的外部因素
I4 用户之间的冲突		
I5 投资活动		
I6 劝说活动		
I7 自组织活动		
I8 网络活动		

相关的社会生态系统 (ECO)

ECO1 气候模式。ECO2 污染模式。ECO3 进入和离开社会生态系统的流。

参考文献:

- [1]. E. Pennisi, *Science* 302, 1646 (2003).
- [2]. R. B. Norgaard, *Conserv. Biol.* 22, 862 (2008).
- [3]. National Research Council, *The Drama of the Commons*(National Academies Press, Washington, DC, 2002).
- [4]. L. Pritchett, M. Woolcock, *World Dev.* 32, 191 (2004).
- [5]. F. Berkes et al., *Science* 311, 1557 (2006).
- [6]. E. Ostrom, R. Gardner, J. Walker, *Rules, Games, and Common-Pool Resources* (Univ. of Michigan Press, Ann Arbor, MI, 1994).
- [7]. F. Berkes, C. Folke, Eds., *Linking Social and Ecological Systems* (Cambridge Univ. Press, Cambridge, 1998).
- [8]. M. A. Janssen, *Complexity and Ecosystem Management* (Edward Elgar, Cheltenham, UK, 2002).
- [9]. J. Norberg, G. Cumming, Eds., *Complexity Theory for a Sustainable Future* (Columbia Univ. Press, New York, 2008).
- [10]. S. A. Levin, *Ecology* 73, 1943 (1992).
- [11]. R. Axelrod, M. D. Cohen, *Harnessing Complexity* (Free Press, New York, 2001).
- [12]. E. Ostrom, *Proc. Natl. Acad. Sci. U.S.A.* 104, 15181 (2007).
- [13]. J. Wilson, L. Yan, C. Wilson, *Proc. Natl. Acad. Sci. U.S.A.* 104, 15212 (2007).
- [14]. W. A. Brock, S. R. Carpenter, *Proc. Natl. Acad. Sci. U.S.A.* 104, 15206 (2007).
- [15]. A. Chhatre, A. Agrawal, *Proc. Natl. Acad. Sci. U.S.A.* 105, 13286 (2008).
- [16]. R. Meinzen-Dick, *Proc. Natl. Acad. Sci. U.S.A.* 104, 15200 (2007).

- [17]. G. Hardin, *Science* 162, 1243 (1968).
- [18]. T. Dietz, E. Ostrom, P. Stern, *Science* 302, 1907 (2003).
- [19]. R. Wade, *Village Republics: Economic Conditions for Collective Action in South India* (ICS, San Francisco, CA, 1994).
- [20]. J.-M. Baland, J.-P. Platteau, *Halting Degradation of Natural Resources* (Oxford Univ. Press, New York, 2000).
- [21]. J.M.Acheson, J.A. Wilson, R.S. Steneck, in *Linking Social and Ecological Systems*, F. Berkes, C. Folke, Eds. (Cambridge Univ. Press, Cambridge, 1998), pp. 390–413.
- [22]. P. N. Wilson, G. D. Thompson, *Econ. Dev. Cult. Change* 41, 299 (1993).
- [23]. E. Mwangi, *Socioeconomic Change and Land Use in Africa* (Palgrave MacMillan, New York, 2007).
- [24]. E. Schlager, W. Blomquist, S. Y. Tang, *Land Econ.* 70, 294 (1994).
- [25]. A. Agrawal, in *People and Forests: Communities, Institutions, and Governance*, C. C. Gibson, M. A. McKean, E. Ostrom, Eds. (MIT Press, Cambridge, MA, 2000), pp. 57–86.
- [26]. P. B. Trawick, *Hum. Ecol.* 29, 1 (2001).
- [27]. E. Ostrom, *Understanding Institutional Diversity* (Princeton Univ. Press, Princeton, NJ, 2005).
- [28]. J. A. Brander, M. S. Taylor, *Am. Econ. Rev.* 88, 119 (1998).
- [29]. X. Basurto, *J. Soc. Nat. Resour.* 18, 643 (2005).
- [30]. H. Nagendra, *Proc. Natl. Acad. Sci. U.S.A.* 104, 15218 (2007).
- [31]. E. Ostrom, H. Nagendra, *Proc. Natl. Acad. Sci. U.S.A.* 103, 19224 (2006).
- [32]. C. W. Clark, *The Worldwide Crisis in Fisheries: Economic Models and Human Behavior* (Cambridge Univ. Press, Cambridge, 2006).
- [33]. G. C. Gibson, J. T. Williams, E. Ostrom, *World Dev.* 33, 273 (2005).
- [34]. E. Coleman, B. Steed, *Ecol. Econ.* 68, 2106 (2009).
- [35]. Supported in part by NSF grants BCS-0624178 and BCS-0601320. I thank T. K. Ahn, R. Axtell, X. Basurto, J. Broderick, E. Coleman, C. Eavey, B. Fischer, C. A. González, E. Jameson, B. de Leon, D. Porter, M. Schlueter, D. Sprinz, and J. Walker for comments and suggestions.

经济网络：新挑战

Frank Schweitzer^{1*}, Giorgio Fagiolo², Didier Sornette^{1,3}, Fernando Vega-Redondo^{4,5},

Alessandro Vespignani^{6,7}, Douglas R. White⁸

¹,ETH Zurich, D-MTEC, Kreuzplatz 5, 8032 Zurich, Switzerland.

²,Laboratorio di Economia e Management (LEM), Scuola Superiore Sant'Anna, Piazza Martiri della Liberta 33, 56127 Pisa, Italy.

³,SwissFinance Institute, c/o University of Geneva, 40 Boulevard Du Pontd'Arve, 1211 Geneva

⁴, Switzerland. ⁵Economics Department, European University Institute, Via della Piazzuola 43, 50133 Firenze, Italy.

⁵Instituto Valenciano de Investigaciones Economicas, Calle Guardia Civil, 22 esc. 2 no 1, 46020 Valencia, Spain.

⁶,School of Informatics and Pervasive Technology Institute, Indiana University, 919 East 10th Street, Bloomington, IN 47408, USA.

⁷,Institute for Scientific Interchange, 10133 Torino, Italy.

⁸,Institute of Mathematical Behavioral Sciences, University of California, 3151 Social Science Plaza, Irvine, CA 92697, USA.

(陈博奎 译 中国科学技术大学近代物理系理论物理所 安徽省合肥市金寨路 96 号)

(刘建国 校对)

摘要：当前的经济危机说明对经济网络的结构和动力学进行全新的、基础性的解释是非常必要的。现有的经济体系越来越多的在相互依赖的基础上，通过跨国借贷、投资网络、贸易关系或供应链进行实施，而这些方式很难进行预测和管理。因此，我们需要建立经济理论的新范式，这种范式应强调经济网络系统的复杂性，并且能够对其进行修改和扩展。通过进一步健全经济网络，会为科学家或政策制定者制定政策带来方便。比如制定政策减少个人兴趣和全球效率之间的冲突，以及如何制定政策减少全球性失败风险，等等。

经济系统和其它复杂系统一样，反映了大量不同的个体之间的动态相互作用，而不是只反映几个关键的个体。总的来看，经济系统最终的行为通常是不能预测的，正如目前的经济危机一样，不能用几个主要个体的失败来进行解释。因此我们需要更基本的洞察系统的动力学，以及他们究竟是怎样追述到基本相互作用网络（interaction network）的结构特征的。

经济网络的研究工作需要从两个方面入手，一个是从经济和社会关系角度入手，另一个是对于对物理学和计算机科学中复杂系统的研究。在上述两个观点中，节点都表示不同的可分辨的个体，它可以用来表示公司、银行或者国家。而不同节点之间的连线表示它们之间的相互合作，比如贸易、所有权、研发结盟、或者是债务信用关系。在相同条件下不同的个体也许有不同的行为，并且有策略上的

相互作用[1]。这些相互作用的演化可以用网络动力学来表示，它受时间和空间的约束，并能够随着环境改变，也能够同个体共同演化[2]。网络的形成和演化是通过增加或是删除个体和个体间连接来实现的。

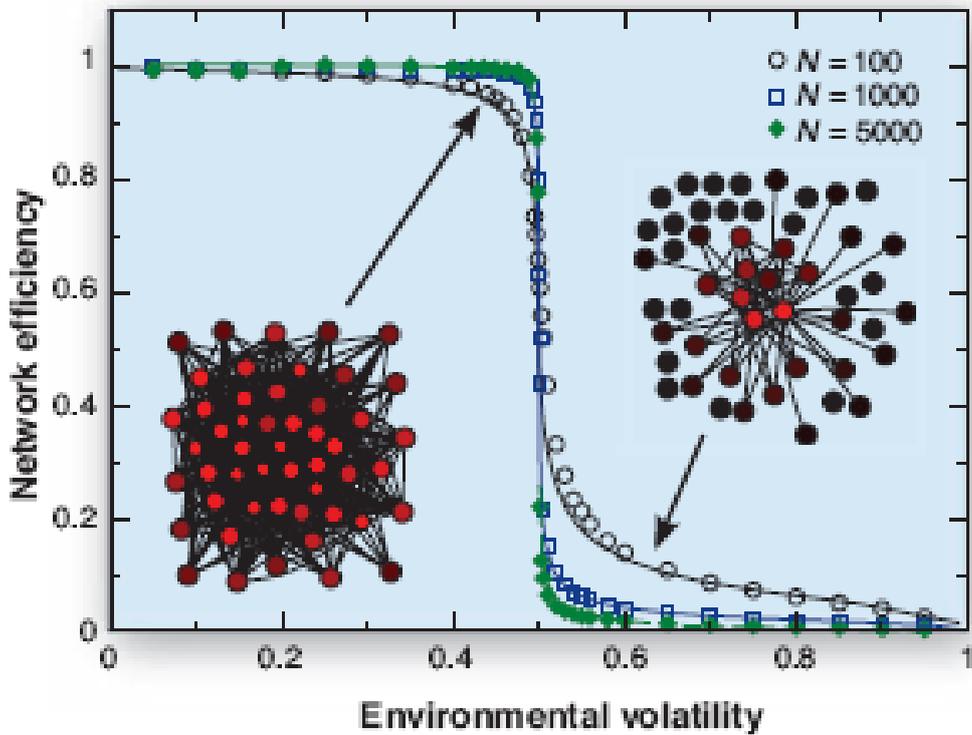
社会经济学着重研究相对简单的网络结构（或者是互惠类型（reciprocally shapes））是怎样影响作用个体的决策行为。一个典型的例子就是星形网络。在这样非常集中的组织中，最主要的“hub”中心与所有的个体间都存在一条通路。从微观的角度出发，我们应该将系统分解，细致地研究网络内部的联系。相比之下，对于大规模的系统而言，则采用一种宏观的观点，这种观点致力于研究将网络作为一个整体的统计规律。这两种方法都有其自身的优势和劣势。微观观点的研究归根于对网络结构和个体行为的过于简单假设[3]。例如，微观方法也许会强调在公司内部非正式联系的发展中个体动机的影响，也许不能成功预测实际的动态结果。而宏观方法对于大尺度系统的性质有着更好的解释，但是不能把这些性质与可分辨个体的经济动因联系起来[4]。

在最新的微观方法中，人们通常把经济网络看作是个体间相互合作、竞争的网络博弈，其中，个体可以是在联合研发项目上有着合作关系的公司[5]或者共享工作机会的工人[6]。作为他们要试图获得个体最大收益的必然结果，他们的联系或是增加或是减少。在这种情况下，个体必须有能力预测其它个体可能做什么（一般而言是在不完全且不对称（imperfect and asymmetric）的方式下进行的）；利用他们所处环境的信息（这种信息可能是有限的），在必要的时间边界范围内构造问题，并且从过去吸取经验。如果在将来遇到类似的情形时，这会形成一种带有偏见的经验。

这些考虑必将导致有大量的选择，个体必须在有限信息的基础上做出选择。对经济网络的微观分析有赖于博弈规则理论，这些理论以等同于纳什均衡为目标（即，在没有个体有偏离动机的情况下，这种均衡具有战略性稳定的意义）。经济网络的微观分析也可以依赖于对运筹学方面的研究，在这个领域里运算法则的寻找和优化被进一步发展。然而，由于节点的数量和可能的连接在成倍增长，一些问题变得非常难以解决，经典分析法不再适用。

以前博弈论方面的文献特别强调内在动机和社会经济网络诱导行为的关键作用（3，7，8）。然而，这种微观的方法还没有特意与在工作中能够识别复杂系统的效力的宏观方法结合。如果没有这些信息，我们不能充分理解一些重要的问题，比如个人喜好和总体福利间的矛盾，或者他们对网络性能整体效益的影响。此外，如果潜在的环境遭受持续的波动，则问题会变得更加复杂，比如由于改革内在的无常性[9]，又例如，如果个体失去平衡也会导致问题更加复杂。如果这是原因之一，那么我们有理由推断：个体会服从一种简单有限理性的规则。这种规则可以根据他们的经验来修改。然而，在这样的条件下，个体不能达到高效率

的配置，尽管他们不断的努力适应每一个变化。而且，波动环境中的每一个小变化都会在系统整体的构造中产生严重的后果。（如图1）



（图 1：环境波动：均衡网络的结构和有效性在很大程度上取决于能使那些目光短浅的个体取得连接的条件。我们从计算机模拟可以看出，个体喜欢与那些具有较高中心度邻居的邻居相连。网络效率是通过个体们的总集中性（aggregate centrality）来衡量的。环境的波动性测量风险，这种风险是指如果有单一的个体遭受外来的冲击，它被迫删除一个连接。如果失去的连接使网络的效率下降并且环境的波动性打破了以前一些关键的平衡，则坚固的同等级网结构会被分解成稀疏的、具有层次结构的、类似于一个核心外围的结构，并且会导致网络效率的下降。）

经济网络上观察到的统计规律无法复现以往研究方法的结果，而这一点恰恰证明了我们提出的这种复杂系统方法可能会对大尺度网络提供一些预测的观点。这些预测由对随机规则的实验产生，这些规则考虑到个体除了某些随机性外的典型特征，并且影响连边的形成，例如用节点的连通度（连边的数量）或他们的中心度量节点的重要性，这个结点反过来又会被它对其它节点的连边影响。

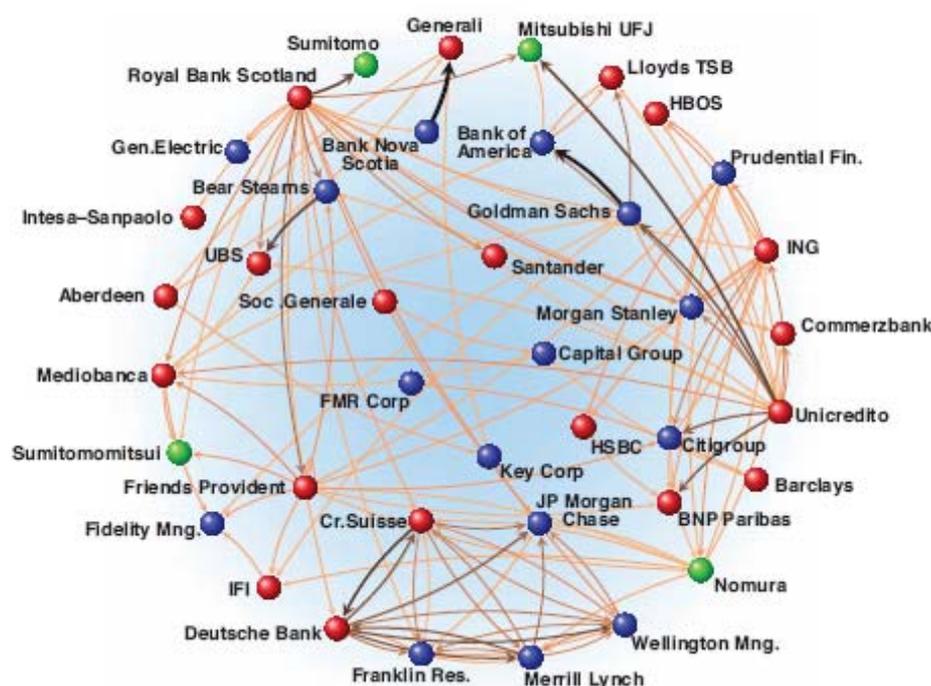
然而，复杂系统方法假定规则是源于外部的，并且不能明确地预测出这些规则是如何有可能以个体经济动因为基础的。因此，不同于理解单个个体的内在行为，复杂系统方法着重理解网络形成规则是如何系统地影响其连接结构[4]。

网络产生于不同的随机性运算法则，比如随机的，无标度或小世界网络，都可以用来与现实的复杂网络相比较。这些现实的复杂网络包括在生物领域，即新陈代谢和遗传网络；在基础设施方面，即公路网和行政区域；在通信方面，即互

联网和移动电话；和在社会联系方面，即敌友关系[1, 2, 10]。通过比较贯穿不同学科的网络结构，我们会发现经济网络也反映一个相似的普遍性[11]。事实上，在国际银行网络中银行之间的联系显示出大量的胖尾和无尺度系统的特征[12, 13]，这说明只有少数几家银行与其它众多银行有联系。在这个例子中，类似的投资行为将集中在银行网络上。类似的规律也可以在许多其它例子中发现，比如国际贸易网络（ITN）[14, 15]，或者区域投资或所有权网络[16]。

在复杂网络中，连接不是二元的（现有的或者非现有的），而是根据当下的经济交易进行加权。[例如，在世界主要非银行金融机构的网络中，如图 2]。此外，连接也表示贸易量，投资资本量等等，并且他们的权重是随着时间的推移而变化的。我们考虑不同层次上的网络。例如直接与非直接的连边，加权或非加权的边，这有助于阐明其拓扑性质的演化。

Pushing Networks to the Limit



（图 2：国际金融网络的一个样本，其中的节点代表主要的金融机构。单向加权的连接代表它们之间存在的最强关联。节点的颜色代表不同的地理区域：欧洲成员国用红色表示；北美用蓝色表示；其它国家用绿色表示。即使在图中表示出减少的连接数，与真实的世界经济相比较，网络仍显示出金融机构的高度相连性，这些金融机构都相互持有股份并且在包含少数几个点中形成闭环。这暗示金融部门是高度的相互依赖，它会影响市场的竞争、系统的危险并且会使网络变得不稳定。）

当外国直接投资（FID）欧洲的公司时，欧洲公司显示出有向网络的特性，

并且具有无标度特性。这种标度依赖于投资和被投资公司的员工数目,和两家公司即将获得和将要投出的资金数目[16]。该网络允许基于公司间的联系和活动,对于那些将要收到或将要被投资地区的时间点进行预测。

类似的结构转变在国际贸易网络中也能发现。全球每增加一个该国家的金融贸易伙伴,就对该国进行加权,而建立这种贸易链接的概率正比于该国家的权重。这样,不同的国家,其中心权重随时间的演化方式就不同了,这恰恰印证了全球经济的区域一体化的分歧,并且比传统的国际贸易和宏观经济统计要做的好[17]。这也是后者没有说明整体网络的拓扑性质而仅仅考虑双边直接贸易连接的原因。例如,在1980年和2005年之间东亚国家凭借他们丰富的经验,他们的中心地位得到巨大提升,但许多拉丁美洲国家的中心地位却有所下降。然而,在这些地区的贸易统计却显示出相似的范式。换句话说,这些令人惊讶的不同的发展记录不能仅仅通过国际贸易和宏观经济统计进行追踪。因此,网络型方法也许能提供提供一个强有力的途径去管理、监督和控制复杂经济系统。

然而只关注于网络的中心性或者其它性质只能提供一个一阶分类,如果这些个体指代的是国家或者公司,它强调波动性和随机性的作用而不能预测个体的潜在动力学。我们可以预见,下一代研究工作将能够测量任何背离普遍性,并且能使我们认同与可分辨个体动力学和他们做决策过程有关的习性。这些新的研究方向应该从合并对可分辨个体策略的描述与它们协同进化的相互作用网络开始。我们应该预测并提倡支持网络构造的经济政策,这种网络构造对网络冲击有很强的稳定性并且它能促进一体化或贸易。下面我们简短描述一下解决这些尝试都需要些什么:

大量的数据分析。我们能够获得越多越好的数据,这将有利于从定性研究到定量研究的转变,从而形成实证科学。随着计算能力的不断增强,大尺度网络的数据能从经济的不同水平中获得(比如公司、工业界和国家),并且模型可以通过大量的同代人、合成物和数据集来检测。新的处理方法应该打开一个范围广泛的业务数据,并且能开发互联网通讯。随着时间的推移,它有可能在具体的合作中收集个性化数据,比如雇工流量、研发合作,还有在商业或者公司银信贷市场合作中。这些大量的数据需要更有力的工具来处理,而操作这种大规模的可用信息恰恰反映了个体之间的相互作用和网络属性。包含这些信息的资料库的出现有可能因此使理论上的经济网络实验[18, 19]和以实验为依据的经济网络研究相互补充[20, 21],并提供实时的大尺度的实时观测[22]。

时间和空间。经济网络属性允许依赖时间,我们将能够摆脱单次快拍数据逼近。这将允许研究人员通过合作补充信息的来源来鉴定网络的进化途径。这种假定的一个很好的实例就是人类应用生物学领域里的研发网络[20],该网络遵从从一个可预测的生命周期,该周期与知识的交换和合作的时机有关。

结构鉴定。在特定的总体经济数据里，从报道的数据中提取网络的拓扑结构是非常困难的。尤其在银行部门更是如此，尽管对总数据的理论分解已经有所研究，但是他们债务信用关系的明细账目是不公开的[13]。因此，分析工作就类似于预测未来，我们能够显示的只是以前的知识或者可预测的信息。利用大规模的数据集合可以识别经济网络中的统计规律，但在评估各种测量的相关性时仍然存在困难。在一个进化的经济网络中，我们需要关于个体角色的作用和影响力的信息[23]。我们需要新的方法来识别模式，引入新的概念来量化直接和间接的影响（比如通过所有权）。在国际贸易网络（ITN）中识别这些在网络中具有相似地位的个体的研究工作表明：我们已经实施了有效的步骤来识别由相互影响的个体所发挥的功能角色。这些个体联系着他们多关系的互动网络中纽带结构里的具体模式[24]，利用统计出的对独特角色的最优配置策略，再像在世界体系模型中那样横切它们每一个链接的核心、半边缘、边缘（但是要更精确些），利用小网络描绘大网络。这种绘制为新世界—非洲—东亚对抗北中欧亚联盟而成立的国际贸易联盟实例中提供了一个数据统计对应。

超越简单化。所有的经济网络关于它们的个体与合作力度都是混杂的，并且能随时间强烈的变化[25]。以前对效率和平衡（或者是战略上的稳定）网络的研究都假定是同性质的。然而，正如在国际贸易网络（ITN）中，加权与非加权网络性质所显示的差异一样（14），任何相变的预测在这种简单的假设下都有可能失效。事实上，个体的异质性都能防止相变的发生，即变为稳定的来源。

系统的反馈。尽管个体都怀着最好的意图，然而，简单的放大机制（就像放牧）往往就能主导网络的动力学。如果一个节点失效，经济网络会服从扩充，这将导致负荷量的再分配（比如电网的电力或者银行网络的信贷债务）。如果仅有一个节点失效，它也会加快其它节点的失效，最终会导致一系列的失效并且会使系统崩溃，这些都会被记为系统性风险。这一点尤其适用于金融网络，其中两个有联系的非银行金融机构之间的连接，代表他们长期的债务和债权。然而，它仍然不能很好的被解释金融网络的结构是怎样影响系统失效的可能性的。尽管这是一个热门的学科，大多数理论和以实验为依据的方法都不适合预测级联的网络效应。主流观点假定一个密集网络允许个人失败风险的多元化[26]。然而，研究结果显示出系统风险的增加有赖于节点之间的耦合强度[27]。此外，最稳定的动态会聚网络模型只能解释在每一时步个体加入或是撤离该网络的行为。个体的整个群体在网络中的增加或减少（比如，作为系统失效的一部分）也许会导致一个更大的、更难预测的、更不稳定的系统。

总之，我们期望在经济网络中会有一个富有挑战性的研究议程，这种议程是建立在能够尽力捕捉到个体行为与他们内部动力学作用所形成的丰富流程的基础上。然而，为了使其有效，我们需要以分析大量数据为实证来洞察经济网络；

需要形成一种理论能够涵盖经济个体与其相互作用的合理描述; 需要从一个系统性的角度去理解这个全球性的效应是来自于经济网络相互作用的改变。我们预测这样的研究将创造更加统一的经济网络, 它将提高我们的认识, 并引导我们对它的进一步研究。尽管我们还远远不能就此方法给出一个令人满意的解释, 也无法将许多因素统一起来, 但目前的研究发展却展示了一个良好的开端。

参考文献

- [1]. F. Vega-Redondo, *Complex Social Networks* (Econometric Society Monographs, Cambridge Univ. Press, Cambridge, 2007).
- [2]. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, Cambridge, 2008).
- [3]. M. O. Jackson, A. Wolinsky, *J. Econ. Theory* 71, 44 (1996).
- [4]. R. Albert, A.-L. Barabasi, *Rev. Mod. Phys.* 74, 47 (2002).
- [5]. J. Hagedoorn, *Res. Policy* 31, 477 (2002).
- [6]. M. Granovetter, *Getting a Job: A Study of Contacts and Careers* (Univ. of Chicago Press, Chicago, 1995)
- [7]. V. Bala, S. Goyal, *Econometrica* 68, 1181 (2000).
- [8]. M. D. König, S. Battiston, M. Napoletano, F. Schweitzer, *Netw. Heterog. Media* 3, 201 (2008).
- [9]. M. Marsili, F. Vega-Redondo, F. Slanina, *Proc. Natl. Acad. Sci. U.S.A.* 101, 1439 (2004).
- [10]. S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, *Science* 323, 892 (2009).
- [11]. R. M. May, S. A. Levin, G. Sugihara, *Nature* 451, 893 (2008).
- [12]. G. Iori, G. De Masi, O. Precup, G. Gabbi, G. Caldarelli, *J. Econ. Dyn. Control* 32, 259 (2008).
- [13]. M. Boss, H. Elsinger, M. Summer, S. Thurner, *Quant. Finance* 4, 677 (2004).
- [14]. G. Fagiolo, S. Schiavo, J. Reyes, *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 79, 036115 (2009).
- [15]. D. Garlaschelli, M. I. Loffredo, *Phys. Rev. Lett.* 93, 188701 (2004).
- [16]. S. Battiston, J. F. Rodrigues, H. Zeytinoglu, *Adv. Complex Syst.* 10, 29 (2007).
- [17]. J. Reyes, S. Schiavo, G. Fagiolo, *Adv. Complex Syst.* 11, 685 (2008).
- [18]. M. Kosfeld, *Rev. Netw. Econ.* 3, 20 (2004).
- [19]. S. Callander, C. Plott, *J. Public Econ.* 89, 1469 (2005).
- [20]. W. Powell, D. White, K. Koput, J. Owen-Smith, *Am. J. Sociol.* 110, 1132 (2005).
- [21]. B. Kogut, G. Walker, *Am. Sociol. Rev.* 66, 317 (2001).
- [22]. D. Sornette, F. Deschatres, T. Gilbert, Y. Ageon, *Phys. Rev. Lett.* 93, 228701 (2004).
- [23]. T. A. Snijders, G. G. van de Bunt, C. E. Steglich, *Soc. Networks*, in press; published online

26 March 2009(10.1016/j.socnet.2009.02.004).

[24]. J. Reichardt, D. White, *Eur. Phys. J. B* 60, 217 (2007).

[25]. A. Kirman, *J. Evol. Econ.* 7, 339 (1997).

[26]. F. Allen, D. Gale, *J. Polit. Econ.* 108, 1 (2000).

[27]. S. Battiston, D. Delli Gatti, M. Gallegati, B. Greenwald, J. Stiglitz, *J. Econ. Dyn. Control* 31, 2061 (2007).

技术社会系统行为的预测

原文出处: Science 24 July 2009: Vol. 325. no. 5939, pp. 425 - 428

原作者: Alessandro Vespignani

原作者单位: Center for Complex Networks and Systems Research, School of Informatics and Computing, and Pervasive Technology Institute, Indiana University, Bloomington, IN 47408, USA; and Institute for Scientific Interchange, Turin, Italy.

原作者: E-mail: alexv@indiana.edu

译者: 彭丹 中国科学技术大学近代物理系 安徽省合肥市金寨路96号 (230026)

(周涛 校对)

我们生活在一个联系日益紧密的技术社会系统的世界,其中,由不同的技术层面组成的基础设施与各种社会成份相互作用,并推动自身的使用和发展。例子有因特网,万维网,WiFi 通讯技术,运输和移动基础设施。多标度特性和复杂性是理解和处理这些网络至关重要的特征。新数据的获得、理论的发展和复杂网络建模正在提出一个完整构架,使得我们更接近获得对技术社会系统行为的真实预测能力。

现代技术社会系统由大尺度物质基础设施(例如运输系统和电网)组成,这些设施嵌入在由通信和计算基础设施组成的密集网络中,而这些网络的动力学与演化则由人类行为限定和驱动。为了预测这些系统的行为,我们应该对真实世界数据中发现的模式进行数字化描述开始。这些描述构成模型的基础,以用于趋势预期、风险评估甚至最终对未来事件的操控。如果结合正确的数据反馈,在非常复杂的环境中,计算建模方法可提供达到要求水平的可预测性。最成功的例子是天气预报,先进的超级计算机设备被用于将当前数据和海量的历史天气模式资料库结合起来,进行大规模计算机模拟。虽然我们经常抱怨每日天气预报的准确性,但是我们必须记得数值天气模型和预报,使得我们可以预测飓风的路径和强度、风暴和其他剧烈气象事件,而且很多情况下,对这些事件的预计及准备挽救了数千生命。

在天气预报上已经取得了数十年的成功,那么,对接下来的流行病时空图以及从中国到印度今后数十亿的新增上网人口对互联网生长和稳定性的影响的定量预测中,为什么我们没有取得同样的成功呢?根本的不同在于,在技术社会系统中预测现象,基于我们有限的社会和人类行为的知识——我们对自身规律的认识远远少于支配流体和气体的物理规律。换句话说,尽管制出大气层紊流的卫星图像已经成为可能,但我们还是没能获得有关人类移动、人群总体风险感知的以及人们采取特定社会行为的趋势的世界范围的大规模定量知识。但是近年来,在数据收集、新的信息学工具的发展以及计算能力的提高等方面已经取得了巨大进步。一个结合了建立在基础设施上的社会人口统计学和行为学方面的巨大的、量化的数据流渐为可用(1-6)。类比物理学中的情形,最终我们能做到从对“社会原子”或“社会分子”(也就是,小的社会团体)的分析走向对社会集合态的定量分析,正如上世纪初社会学家展望的那样(7)。这里,可以在空间(地理的和社会的)和时间中刻画的数百万个体组成的大规模社会系统,我称之为“社会集合态”。从对少量元素的研究转到对大尺度集合体行为的研究,如同从原子分子物理转到物态物理。聚集在一起的大量相同元素,如何在受到各种外力以及其他元素作用下,引起不同的宏观的动力学行为——对该问题的理解指引了通往定量计算方法和预测能力的道路。然而同时,在对社会聚集态的研究中我们再次遭遇了物态物理研究中遇到过的所有挑战,包括从湍流到多标度行为。

事实挖掘和代言网络¹

¹ 这里的代言网络(proxy networks)和下文中的代言数据(proxy data)是指那些本身并非我们直接的兴趣所

技术社会系统信息流的水平不仅仅归功于现代计算机处理器的数字运算能力的提升。对人和技术之间关联的特性以及计算机网络虚拟世界和我们真实世界之间边界模糊性的洞察正在改变我们对数据的获取能力，导引我们走向“事实挖掘”。“事实挖掘”定义为通过机器能够感知到的与人类社会行为相关的环境数据的集合 (2)。一个最好的人/技术互连的例子可以在对人类迁移的分析中找到。过去，获知人类相互作用和迁移主要依赖于人口普查和调查数据，这些通常不够完整或限于特定背景。尽管关于人类交通的研究有了进步(8,9)，数据的缺乏阻碍了个体层次上基于动力学法则的关于人类活动的一般性框架的构建，这可以从小团体到大市区和国家，以自下而上的观点将空间尺度架接起来。尽管如此，在一个开创性的工作中，Brockmann等人展示了通过货币追踪的热门网站（例如<http://en.eurobilltracker.com>和www.wheresgeorge.com）收集的可以代言人类迁移活动的货币流通记录的大量数据。这个工作开创了一个一般性的方法去挖掘人类相互作用和活动的代言数据(10)。类似的，现代移动电话和个人掌上电脑，结合了诸如蓝牙、全球定位系统和WiFi等尖端科技，常常提供我们日常活动的痕迹(2,11)。例如，最近的研究中，Gonzalez等人(6)使用移动电话数据追踪100,000个人的活动，时间跨度超过6个月。此外，使用传感器和标识去获得微观尺度一对一相互作用数据现在已经成为可能(1,2)。

虽然我们面临着一系列伦理的私人的问题，这几类数据，以及在技术社会系统中产生、获取和交换信息成本的降低正在改变我们对很多现象的理解(12-17)。人类流行病的空间动力学，决定于那些携带疾病到本未感染人群的个体的迁移。类似地，人类迁徙和活动给很多生态系统带来了原本不在其中的新的生物。人口混合和群体的合并与隔离同样促使语言和方言的演化。最后，因特网空间的日常活动详细给出了我们对信息的开发和攫取。

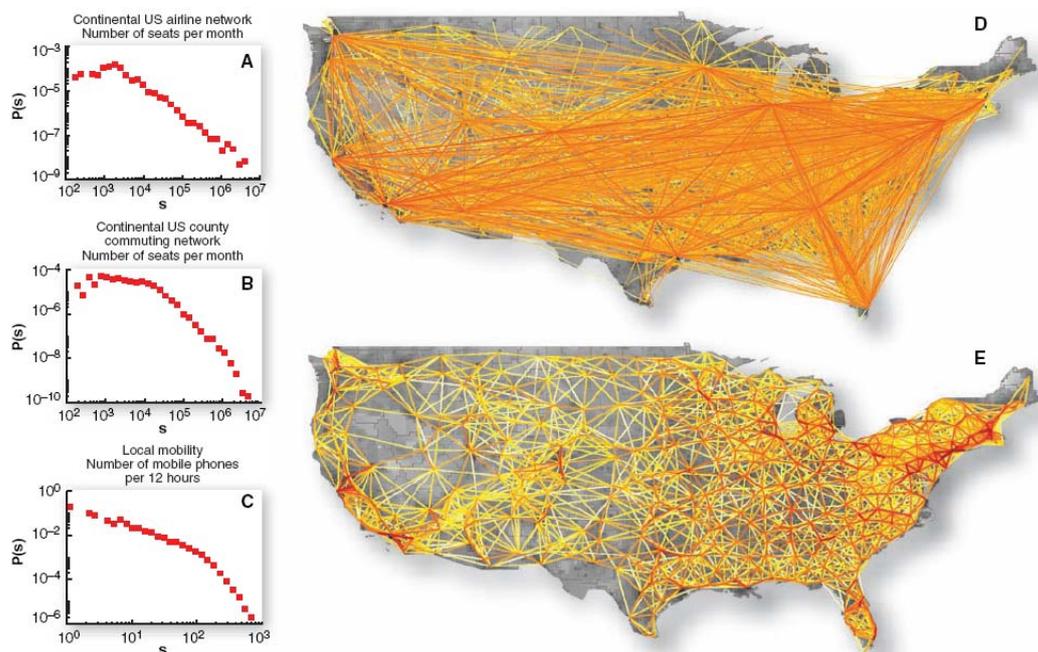


图1. 迁移网络的多标度特性。在左边，我们报道三个不同网络的交通概率分布 $P(s)$ ，以任意给定连边上移动个体的数目来计量：(A)美国大陆航空网络，(B)美国大陆县郡交通网络以及(C)大城市地区在电话基站之间的迁移。所有情形下，分布都是高度倾斜的，跨越了3到7的数量级。在右边，我们给出美国航空网络的图示(D)和主要人口普查区之间的交通网络(E)。从黄到深红的颜色表征按对数标度的交通流量。与交通网络

在，但可以反映和表现研究目标性质的网络和数据。例如我们可以通过货币流通数据和手机空间位置的移动了解人类个体可能的运动轨迹，而货币和手机本身，并不是我们的兴趣所在。——校者注

的格子状结构相比较，航空网络主要由长程边构成。交通网络的日平均流量比航空网络的大一个数量级。

网络思维

因特网和虚拟世界是我们每天都浏览和探索的网络(17-19)。人类相互作用模型基于社会网络，其中，节点描述相互作用的个体，连边表示可能的相互作用(20)。迁移、社会生态学和流行病学模型依赖于因人群间可交换性而互联在一起的全部人口组成的种群网络(21)。大量工作显示大部分真实世界网络展现了动力学的自组织（就是说，无需外力干涉，随着时间演化它们变得更复杂）以及统计上的明显的异质性——这些正是复杂系统的典型特点(22-24)。刻画这些网络的各种统计分布（包括节点连接概率和连边强度）通常是胖尾及倾斜的，而且在好几个数量级内变化。复杂网络提出的最初挑战在于它们的内部互联程度（网络的网络）和多标度特性。图1描绘了作为不同尺度上人类迁移的例子三个网络，范围从跨洲航空旅行到移动电话基站间的市内迁移。理想地，对由人类迁移驱动的过程做出预测，我们需要将这些数据和它们的不同粒度（从几百米几小时到数千公里数天）结合到一个巨大的多标度网络中。

因此，技术社会系统的复杂性要求“网络”思维。一个简单的例子是流行病传播的大尺度描述。蔓延于14世纪的瘟疫（黑死病）的传播主要是空间扩散现象。有关历史的研究确定疾病的传播遵循一个简单的模式，这个模式可以完全数学化描述于各项代表扩散的连续微分方程的框架内。正如在1933年预期的那样(27)，当今世界，传染病（例如SARS或最近的猪流感）在人群中大范围和地理的影响主要由于商业的空中旅行。一个始于东南亚的流行病将很快抵达北美和欧洲（图2）。因而，这张图不能简单地按照扩散现象来描述；而是应当和现代交通网络的空间结构结合起来。例如，航空交通网络的胖尾特性解释了为什么仅仅交通管制对全球性的流行病无效，除非全球的迁移率减小一个数量级(29-31)。

现代网络思维另一个至关重要的方面是动力学的自组织，这导致不依赖于人类规划和系统工程的大规模基础设施中特定模式的涌现。动力学自组织系统最好的例子也许是因特网，不过大多数通讯基础设施、道路和运输系统、供给网络和电网也是动态增长网络。例如，道路修建显然是有计划的，成本、效率和效用最优化的考虑对规划工作有影响，这就毫不奇怪。因而，可以一般性地预期道路网络展现高度的规则性。然而日常经验告诉我们这不是实情，尤其是在经过了长时期发展的城镇。为此，研究者构造了简单的道路形成模型(32)，试图抓住激励计划者的最优化观念和影响设计者的有限时间空间范围之间的关系。



图2. 从对起源于越南河内的流行病模拟得到的流行病入侵树。结点表征散布全世界的3200个人群，而有向边显示流行病从一个人群移动到另一个的路径。从暗红到深蓝的颜色图对应流行病入侵的时间次序。模拟

结果来自针对世界范围的流行病和迁移模型(38)。

尽管如此，提出一个多标度网络的整体描述，最大挑战在于必须同时处理多个时间和长度尺度。任一尺度上系统最终的动力学行为是发生在所有尺度上的事件的产物。传播疾病的单个个体和因特网的单个节点显然不被网络的多标度特性影响，正如单个分子不在意湍流的多标度特性。尽管如此，集体的动力学的行为和我们和技术社会系统进行数学计算分析的能力被系统的多标度特性约束。在网络和技术社会系统的情况中，多标度难题因为大规模数据集的可获取性正变得越来越明显。因此，我们不得不发展合适的形式和技巧，正如过去研究多标度物态系统（流体，固体，宇宙中物质扩散，等等）的研究者所做的那样(33)。为了获得对技术社会系统解析的理解和计算的方法，我们必须找到不同的策略去处理那些在非常不同的特征尺度上起作用但还彼此影响的动力学行为或方程。这样的方法最终将允许分层式计算方法的定义，在其上，不同的模型假设和粒度在一个一般性的多标度框架下可以一致。

多标度网络的应用

有关网络特性的知识开创了发现和理解支配与社会系统耦合在一起的大基础设施系统的新的统计规律和动力学规律的道路。此外，空间分布式人群的大量相互连接和复杂性以及多标度网络的强异质性，是构建一个从头开始的计算模型的关键；和传统的平均场或自上而下的策略正相反，在这个模型中系统行为能以自下而上的角度来理解。这出现在很多情形下，范围从城市规划(34)到流行病模型(35-38)。著名的例子是TRANSIM和EPISIMS工程(35)，包含数百万个个体的模型，分别被用来模拟整个城市的动力学和交通以及生物个体的扩散。

某些情况下，对复杂网络的理解提出了违反直觉且令人惊讶的方法去设计和管理复杂技术社会系统。例如，在电力网和其它载流网络上，单个节点或连边的失效能引发多米诺骨牌效应（“级联故障”），这时由于流量再分配导致的超载可能引发网络的整体失效。利用多标度网络连边上载流的异质性，A.E.Motter(39)提出一个基于移除一定量节点以引发特定失效的适应性防御机制。虽然这个机制看起来似乎违反直觉，但是经过恰当选择的节点引起的特定失效并没有放大级联过程，正相反，可能减轻最终的损害。换句话说，现在能提出一个基本原理去理解那些构成技术社会系统行为大多数有趣特征的临界点和非线性特性。

最棘手的难题

尽管很多观念上的基本问题仍然没有解决，界定技术社会系统内在的预测极限的最大障碍是它们的敏感性和对社会适应性行为的依赖。在一个没有外力的系统中，稳定态可以达到。这时，社会行为和物质基础设施之间的反馈决定网络行为和有趣的动力学过程的细节。我们可以设想利用稳态数据去预测这样“非正常”情形下的系统行为。尽管如此，灾难事件的情形下（例如，在大规模流行病或大型自然灾害这样的突发事件期间社会秩序的瓦解），技术社会网络的行为远离平衡进入未知的领地。

一个有关预测和控制技术社会网络中灾难性事件演变的有趣的并在伦理上引起挑战的方面是，系统会对公共可用的预测作出相应的改变。社会行为也会对预测本身作出反应和适应。与物理系统中的情形正相反，预测本身成为了系统动力学的一部分。另外，预测可能意味着不道德的控制和偏好社会特定群体的预期策略。最后，错误预测的风险可能导致代价惨重的或不道德的社会控制机制而没有任何实际的利益。尽管如此，通过改进模型的精确性和可靠性，上面的一些问题可能得到部分解决，很明显，社会对预测的适应性改变给我们提出了新的方法上和伦理上的难题。

处理这些问题要解决三个主要的科学难题。首先是收集有关信息传播的大规模数据和发生在危机期间的社会反应。凭借在特殊灾难或危机事件时运行的大规模移动通信数据库（例

如移动电话, Twitter博客和社会Web工具), 这目前还做不到。第二个难题是形式模型的公式化表述, 这使得量化个体的风险感知和认知现象对技术社会网络结构和动力学的影响成为可能。第三个难题涉及实时控制中影响计算模型的监测设备能力的分配。复杂系统和网络理论、数学生物学、统计学、非平衡态统计物理和计算机科学在对付这些难题中都起了关键的作用。尽管这样一套完整的方法可能还处在发展初期, 现在似乎可以设想创建计算的预测基础设施, 帮助我们设计更好的能量分配系统、规划交通畅通城市、预知英特网连通性的需求或管理突发卫生事件期间资源的部署。

参考文献和注释:

1. D. Lazer et al., *Science* 323, 721 (2009).
2. A. Pentland, in *The Global Information Technology Report 2008-2009* (World Economic Forum, Geneva, 2009), p. 75.
3. J. P. Onnela et al., *Proc. Natl. Acad. Sci. U.S.A.* 104, 7332 (2007).
4. D. Brockmann, L. Hufnagel, T. Geisel, *Nature* 439, 462 (2006).
5. D. J. Watts, *Nature* 445, 489 (2007).
6. M. C. Gonzalez, C. A. Hidalgo, A.-L. Barabasi, *Nature* 453, 779 (2008).
7. G. A. Lundberg, M. Lawsing, *Am. Sociol. Rev.* 2, 318 (1937).
8. S. Erlander, N. F. Stewart, *The Gravity Model in Transportation Analysis* (VSP, Utrecht, Netherlands, 1990).
9. J. de Dios Ortuzar, L. G. Willumsen, *Modelling Transport* (Wiley, New York, 2001).
10. D. Brockmann, F. Theis, *Pervasive Comput.* 7, 28 (2008).
11. F. Giannotti, D. Pedretti, *Mobility Data, Mining and Privacy* (Springer, Berlin, 2008).
12. S. Riley, *Science* 316, 1298 (2007).
13. C. Viboud et al., *Science* 312, 447 (2006); published online 29 March 2006 (10.1126/science.1125237).
14. L. Hufnagel, D. Brockmann, T. Geisel, *Proc. Natl. Acad. Sci. U.S.A.* 101, 15124 (2004).
15. V. Colizza et al., *Proc. Natl. Acad. Sci. U.S.A.* 103, 2015 (2006).
16. G. M. Ruiz et al., *Nature* 408, 49 (2000).
17. B. A. Huberman, L. Adamic, *Lecture Notes in Physics* (Springer, Heidelberg, Germany, 2003).
18. R. Pastor-Satorras, A. Vespignani, *Evolution and Structure of the Internet* (Cambridge Univ. Press, Cambridge, 2004).
19. M. E. Crovella, B. Krishnamurthy, *Internet Measurements: Infrastructure, Traffic and Applications* (Wiley, Chichester, UK, 2006).
20. S. Wasserman, K. Faust, *Social Network Analysis* (Cambridge Univ. Press, Cambridge, 1994).
21. I. A. Hanski, O. E. Gaggiotti, *Ecology, Genetics and Evolution of Metapopulations* (Academic Press, San Diego, CA, 2004).
22. A.-L. Barabasi, R. Albert, *Science* 286, 509 (1999).
23. M. E. J. Newman, *SIAM Rev.* 45, 167 (2003).
24. A. Barrat, M. Barthelemy, A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge Univ. Press, Cambridge, 2008).
25. A. Barrat et al., *Proc. Natl. Acad. Sci. U.S.A.* 101, 3747 (2004).
26. J. D. Murray, *Mathematical Biology* (Springer, New York, 1993).
27. A. Massey, *Epidemiology in Relation to Air Travel* (H. K. Lewis, London, 1933).
28. J. S. M. Peiris, K. Y. Yuen, K. Stohr, *N. Engl. J. Med.* 349, 2431 (2003).

29. T. D. Hollingsworth, N. M. Ferguson, R. M. Anderson, *Nat. Med.* 12, 497 (2006).
30. J. M. Epstein et al., *PLoS One* 2, e401 (2007).
31. V. Colizza, A. Vespignani, *J. Theor. Biol.* 251, 450 (2008).
32. M. Barthelemy, A. Flammini, *Phys. Rev. Lett.* 100, 138702 (2008).
33. G. I. Pavliotis, A. M. Stuart, *Multiscale Methods: Averaging and Homogenization* (Springer, Heidelberg, Germany, 2008).
34. M. Batty, *Science* 319, 769 (2008).
35. S. Eubank et al., *Nature* 429, 180 (2004).
36. N. M. Ferguson et al., *Nature* 442, 448 (2006).
37. T. C. Germann et al., *Proc. Natl. Acad. Sci. U.S.A.* 103, 5935 (2006).
38. V. Colizza et al., *PLoS Med.* 4, e95 (2007).
39. A. E. Motter, *Phys. Rev. Lett.* 93, 098701 (2004).
40. I thank V. Colizza, D. Balcan, B. Goncalves, M. Gonzalez, and H. Hu for help with the figures and M. Gonzalez for the data used in Fig. 1C. I am partially supported by NIH, NSF, the Defense Threat Reduction Agency, the Lilly Endowment Foundation, and the Future Emerging Technologies projects Epiwork and Dynanets.

转录调控回路：从字母表中预测数字**

Harold D. Kim,^{1*} Tal Shay,^{2*} Erin K. O'Shea,¹ Aviv Regev^{2†}

¹Howard Hughes Medical Institute, Harvard University Faculty of Arts and Sciences Center for Systems Biology, Departments of Molecular and Cellular Biology and Chemistry and Chemical Biology, Cambridge, MA 02138, USA.

²Department of Biology, Massachusetts Institute of Technology, and Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA.

**These authors contributed equally to this work.*

†To whom correspondence should be addressed. E-mail: aregev@broad.mit.edu

(译者：谢悦 中国科学技术大学近代物理系，合肥市 230026，中国)

(张勇 刘建国 校对)

摘要：转录调控回路决定了顺式和反式因子如何将信号转换成信使 RNA(mRNA)的表达水平。随着在不同条件定量探测基因表达状态的和高通量技术的发展，用于构建和测试转录调控的模型的数据正以飞快的速度产生。这里，我们回顾了用于获取详细的定量小尺度回路模型和大尺度上粗略的全基因组模型的实验和计算方法。为了理解转录调控回路的工作方式和连接，我们讨论了组合小尺度和大尺度方法的可能。

基因组学的下一个前沿是如何系统地将基因组和细胞的功能部分集成成回路，从而把信号转换成细胞反应。这些回路包括信号转导，代谢和转录的通路。该工作中，一个重要问题就是在给定细胞种类和一定条件下，如何预测每个基因的表达水平。本文重点讨论了调控回路，这里我们把反式调控因子和顺式序列决定因素作为调控回路的输入信息，把基因表达水平作为输出信息(如图 1A 所示)。

科学家已经从小尺度方法(small-scale approach)发展出了很多细致的定量模型来控制一个或多个基因的调控回路。与之相对应的大尺度方法(large-scale approach)则利用计算方法重构基因组尺度的回路。利用这些方法可以得到综合的，略显粗糙的模型。这两种方法自身都在相互影响[1,2]，而且目前具有合并的趋势。合并的模型可以提供回路连接的全基因组模型，该模型具有实际的生物化学意义[3]。

小尺度方法：原核生物的成功案例

通常，小尺度方法使用热力学或动力学模型解释实验数据，它关注单一线路(single circuit)及其顺式调控(cis regulation)问题。目前，原核生物体内的数个调控回路的输入和输出已经成功地在小尺度上进行了度量[4-7]，其中“输入”包括了反式调控因子活动的诱导物[6,7]，或者反式因子本身[4,8,9]。这些研究依赖于结合在具有良好特征的启动子（该启动子的活性具可以用可控方式连续调整）的转录因子(TF)。一般用蛋白质产生率、活性或定态蛋白质水平表示输出，这些输出用某种标示物进行度量，比如，一个荧光蛋白基因或者编码某种酶的基因（酶的活动作为显示物），并且该基因处于所研究的启动子的控制之下。这种输入和输出关系被称为基因调控作用，通常只度量一群细胞中的输入输出关系，有时候该方法也被用于探测单个细胞。

根据分子系统的先验知识，通常用直接探测方法推测一个定量的基因调控模型。基因调控作用常常用 Hill 函数来近似表示（如图 1B 所示）。Hill 函数包括临界值、敏感度和最大表达水平等参数。为了理解表观参数和实际基因调控的生化参量之间的关系，需要知道以下几

个方面问题: 转录因子是否为低聚物; 转录因子与顺式元件在启动子相互作用的强度有多大; 当与促进物结合时, 它们是否与自身或其他因子协同作用; 还有它们如何与 RNA 聚合酶相互作用。大量对原核生物系统 (比如, lac 和 λ 启动子) 的研究已经对关于单个输入基因调控作用的 Hill 参数的生化基础提供了深刻的见解。

科学家已经构建了适用于更复杂回路的定量模型, 例如基于 lac 启动子的回路能进行“与”和“或”的简单逻辑运算模型[6,10]。回路模型能以由底至上的方法预测更复杂的调控网络的行为, 分析大肠杆菌中的人工基因网络时使用的就是这种方法[3]。总之, 原核生物的基因表达数据和相应模型的吻合得益于模型的可操作性非常强, 另外还得益于与其相关的生物化学和生物物理的知识积累。

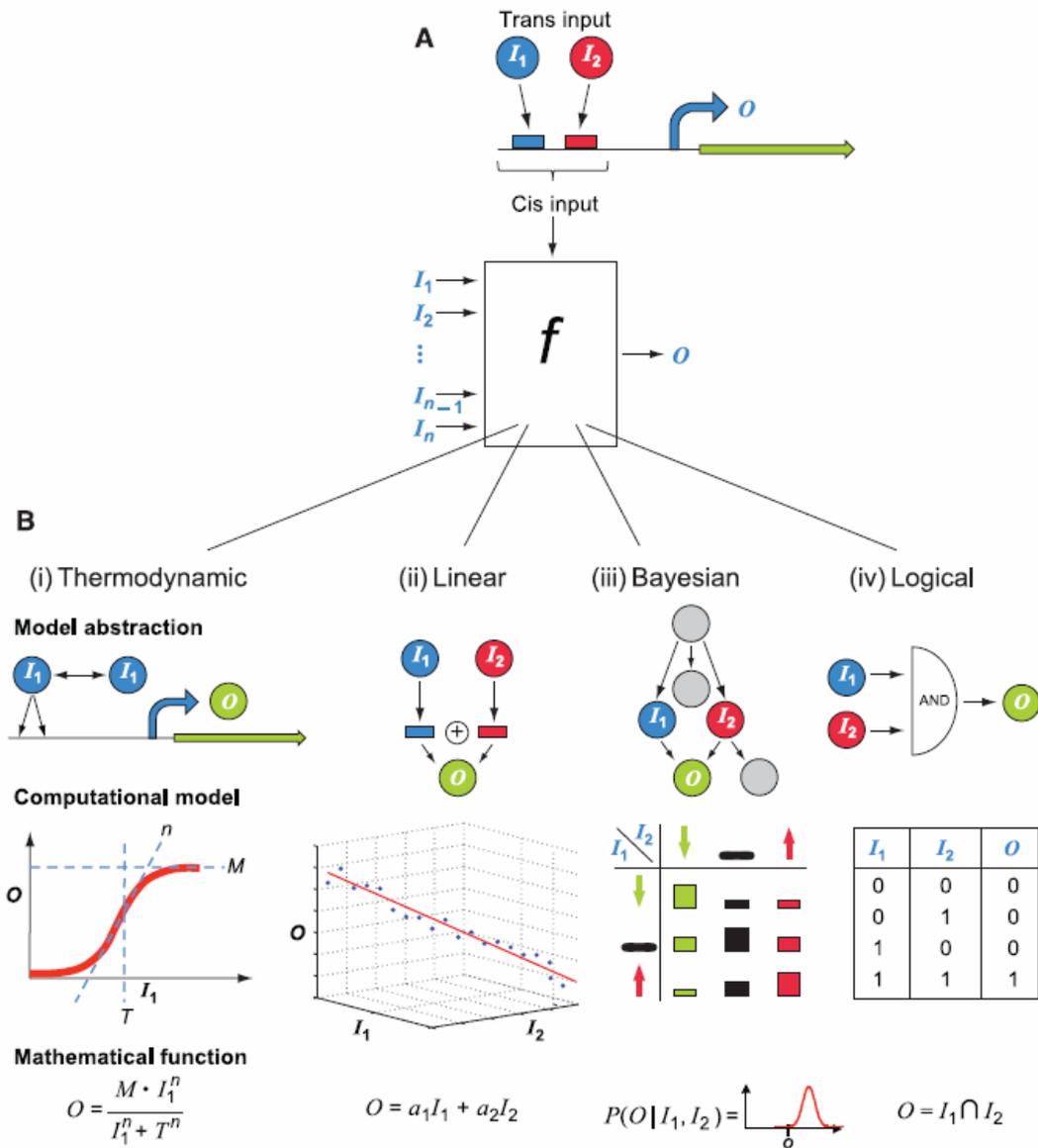


图 1. 基因调控功能。(A) 基因调控作用描述反式因子输出 (比如转录因子 (I_1, I_2, \dots, I_n)), 和正调控输入 (比如调控成分) 如何被转换到基因的 mRNA 层次 (O)。 (B) 不同的函数描述应用不同数学方法的方案。大多数的小尺度方法依靠热力学激发模型 (i), 它解释了对单个可调反式因子输入的依赖性, 并适合一个 Hill 函数。被抽象出来的 Hill 参量 (包括最大表达水平 (M)、临界值 (T) 和灵敏度 (n)) 成为存在的正反馈调控输入变量的函数。大尺度方法采用了许多方式。(ii) 线性模型假设输出是各个正输入或反输入的线性组合。(iii) 贝叶斯网络给出在给定输入条件下, 输出的概率分布。(iv) 逻辑 (布尔) 回路把输出当作在输入进行逻辑操作的结果。(顶部) 各个模拟方法的假设; (中间) 对应模拟方法的基因调控作用; (底部) 等价的数学形式。所有的模拟方

法都可用于小尺度和大尺度的研究中。

我们可以采用很多方法证明和完善这些模型。例如，可以比较从该模型提取的生物化学参数（转录因子与 DNA 之间的平衡解离常数）和在体外测量的相应参数。另一个方法通过改变正调控元素的 DNA 序列和观察数据是否适于该模型扰动平衡下的转录因子。这些测量也对调控回路的可塑性和演化提供了深刻的见解[10]。这个模型可以通过测量附加的变量（如基因表达中的噪音）来作进一步限制[13]。

当一个模型未能被数据证实时，我们可以预先找到那些可能与调控回路有联系的、未知的事件或组成部分，这样可以正确识别回路连接。例如，一个小于期望的最大表达水平表明存在另一个路径或因子[7]（它可限制转录因子与启动子的结合或调控结合的活动）。高于期望值的灵敏度可能指向隐藏的协同机制（比如转录因子间的物理相互作用），或指向由其他因子调控的间接相互作用[7,14]。依次连接的回路也可导致更高的灵敏度[15]。另外，隐藏的反馈圈能明显的影响基因调控作用[16]。

对真核生物的基因调控作用的小尺度方法

真核系统有比原核生物更为复杂的调控回路。例如，因为核小体通常会妨碍转录因子的结合，真核生物的正调控元素无法被转录因子等同获取。此外，核小体不是转录因子的被动妨碍者，而是被依赖于 ATP 的染色质重塑因子从 DNA 中移除的。另外，核小体与 DNA 有不同的结合能力，这依赖于它们的化学修饰和所结合的 DNA 序列[17]。

真核生物的正调控编码更加多样，并且有许多复合转录因子的结合点[18]。另外，反调控方面的复杂性作为定位功能和蛋白质翻译后修饰功能影响着转录因子的活性，在度量和建模两方面都提出了挑战。此外，识别特异性序列的转录因子和各种染色质重塑因子、组蛋白修饰复合体之间也相互作用。但是，我们对这些因子及它们在基因调控上的定量效应的特性却认识有限。因此，即使对研究比较成熟的模型真核启动子，我们也不能完全理解在回路中组成成分，连接和生化相互作用等方面的相互作用。

根据酵母[8,19-21]、线虫[22]和海胆[23]中的启动子构造的研究成果，科学家已经对基因表达层的相互作用进行了建模，度量了合成的和基因组的启动子酵母的基因表达水平。这些启动子在转录因子结合点（处于不同的环境）的成分和组织是不同的。在转录因子与 DNA 结合和转录因子相互结合的平衡态上，酵母基因表达水平能被成功地模拟[19,20]。文献[24]强调了决定基因表达水平方面核小体的重要性。文献[8]通过使用不同的正调控元素组合，对测试酵母中磷酸盐调控启动子的基因调控作用进行了研究，该研究识别了核小体在分离基因调控函数的临界值和最大表达水平中的作用。

大尺度：从单基因到全基因模型

小尺度方法注重在组成部分和连接已知的情况下研究回路中的调控作用。相对而言，大尺度方法没那么精确，但是可被用于推断回路的成分和连接。这些研究成果在研究高级生物体（其大部分回路是非典型的）的基因调控时有特别重要的意义。

预测全基因输出的大尺度方法的出现与分析基因组尺度的回路输入、输出和基因规模的连接具有紧密的联系。例如，mRNA 水平的输出能够用基因芯片或测序进行度量[25]。此外，举一个正调控输入的例子：由于计算技术的发展和整个基因组序列的可获性，我们可以更为容易地探测启动子的序列[26]。反调控输入可以当作 DNA 与蛋白质直接相互作用来测量，测量内容包括转录因子[18,27]、组蛋白修饰状态[28]和体内核小体的位置[29]。转录因子启动子的相互作用也可从体外分析来进行推断（如蛋白质结合芯片[30]）。

正调控作用的全基因重构

在全基因测序和描绘的研究工作已经大大促进了重构正调控作用方法的发展，这里正调控作用包含了启动子中正调控元素的种类、数目和组织以解释被观察的表达水平。目前，科学家已经用线性的、贝叶斯的和热力学的方法研究全基因重构，每一种方法都反应了在转录

调控的生化基础上的不同假设(图 1B)。线性模型认为表达输出信号是正调控输入的线性组合[32-34]。基于概率的贝叶斯方法能发现具有干扰数据的大尺度数据中的规律,并且通过模拟基序(motifs)、相关距离和方向的组合[35],可以获取启动子的组合逻辑和组织。一些研究结合这两种方法的优点,在概率背景下,构造一个线性模型[31]。最近,科学家提出了一些更符合实际的热力学模型。例如,通过计算正调控序列上反调控因子的所有可能分布概率,并将它们对表达的贡献求和,预测了果蝇细胞分裂中的表达模式[36]。核小体和反调控因子的序列偏好性已经被用来预测酵母的表达[37]。然而,热力学模型还不能应用于基因组尺度上的度量,这是因为正调控作用的全基因推断严重依赖于 DNA 序列中正调控元素的精确、全面的测量——这是一个非常困难的问题,在高级组织中更是如此。

评估模型在预测基因表达时成功与否的是一项具非常有挑战性的任务。理想情况下,大尺度模型不仅可以描述某一数据集,并且能够被检验推广到其它未知的数据。然而,大部分数据集是规模有限的,并且也缺乏进一步的系统实验。尽管已经有很多工作尝试使用各种测量评估模型预测质量,然而,这些研究并没有报导这些模型预测的成功比率。例如,预测各个基因模块配置的模型,在酵母研究的一般报导中,同时参与调控的基因或拥有同一功能的基因进行成功配置的成功率在 30%[38]到 73%[35]范围内变化。其它的工作报导了检验模型质量的通用数据存在的可能性,但是这样的测度很难在不同复杂度模型之间去比较。迄今为止,已报道的最好的这种数据是用来预测每个基因的实际表达模型;他们一般报导了由该模型解释的基因表达偏差的百分比。利用正调控序列预测基因表达的贝叶斯模型和线性模型都报导了酵母数据具有很高的成功率(例如,文献[35]中的 51%,文献[33]中的 52%到 72%)。然而,类似的方法在哺乳类细胞中的成功率却很非常低(例如,文献[31]中的 6%,文献[32]中的 11%到 24%)。表达的总额也是十分重要的,我们期望用一个特殊模型来解释其总额。最近,对酵母中人工启动子库的研究估计:正调控最多能解释表达偏差的 65%,热力学小尺度模型能解释 44%到 59%[19]。在尾索动物(Urochordate)海鞘(Ciona)中的类似工作解释了在正调控层次上偏差的 30%到 89%。下一个重要目标就是创建标准的研究方法和数据集,这样我们可以对不同模型的表现进行公平比较。DREAM 工程的目的就是通过给界内提出挑战以实现这一目标[42]。

发现方向: 推断在基因组尺度上的反调控

单独的正调控输入并不能预测基因表达,因为动态的表达模式随着环境调控、细胞种类和细胞周期阶段而改变。这三种基因组尺度策略上的变化都伴随着反调控因子的相应变化。

物理相互作用的基因组尺度测量允许直接将反调控因子合并到回路模型。例如,在体内[18,41]或体外[43]探测到的转录因子与 DNA 的相互作用被用于获得酵母中转录因子结合的详细图谱。这些图谱可以与表达输出的测量相结合构建一个调控回路模型[44]。然而,由于受到测量条件不一致等因素影响限制了这方面的研究。反之,刺激前后转录因子结合的测量能帮助区别直接和间接的目标[1,18,45]。我们注意到,反调控输入量的规模和品质依然被所需的试剂、数据和成本限制(例如,对所有转录因子构建带免疫标签的版本,并在许多条件下的染色质免疫沉淀反应中使用它)。

在另一种方法中,反调控因子的水平、活力和连接可以从 mRNA 水平中被推断,这是因为许多反调控因子是被嵌入在转录反馈调控圈中[46,47],因此反调控因子在 mRNA 水平的变化是可测的。我们做出这样的推断是因为目前还不能平行地分析大量调控因子的活力。翻译及翻译后调控能在反调控因子的 mRNA 水平和活动蛋白质水平之间产生非常大的偏差,尽管如此,这些方法在大肠杆菌、酵母、老鼠和人类数据分析方面还是成功的[40,48-50]。在其中一种方法中通过收集瞬时的表达数据,描述细胞分裂与哺乳类系统[46,51]对环境刺激反应的在内过程,这显示了由不同转录因子控制的转录序列波传播转录程序。然而,从输出水平对调控活力所进行的推断限制了这些模型区别因果关系与相互关系的能力。

调控器的扰动可以被用于推断和证实反调控功能。例如，263 个转录因子在酵母中被系统地敲除，且每个敲除都被用来与全基因表达的原始类型进行比较[52]。RNA 干扰技术能瞄准具体的基因和 RNA 的表达，该技术使得在更小的规模下用类似方法研究哺乳类细胞成为可能[31]。上位性分析可以被用来推断多输入回路的详情情况，如下所示：通过比较一个、两个或三个反调控因子被删除的各品系，可以研究酵母渗透压反应的全基因效应。

基因组尺度上的合并正反调控的模型

下一个明显的研究方向是在不同条件下系统地同时度量和模拟反调控因子与正调控成分之间的关系[51]。然而，在精细干扰和测度对回路的正反调控输入研究方面，我们的能力是有限的，原因有三：成本高、高级组织中基因工具的相对缺乏和能产生启动子序列变种的人工方法的缺乏[19]。

一个设计精细干扰的可选择策略就是更多地关注自然的基因多态性现象(它在基因表达中有潜在的变化)。基于正反调控因子的基因变化，我们现在可以解释人类自然基因变种表达差异，也能够对如何表达定量特征的调控回路进行重构[53,54]。这个方法之所以十分具有吸引力是因为它包含了对多回路组成部分的微扰所产生的共同效应，并且使用基因联接和联合的能力去决定因果关系。结合联接分析，概率性的和线性的模型已经将基因型作为预测器包含进来，以推断酵母、老鼠和人类中的反调控因子和调控网络[55-57]。

类似地，研究杂交生物体的表达能够发现两个相关物种的基因差异，这样可以去评估相关的正反式因子对调控的贡献(如酵母和果蝇[54,58]中所观察的一样)。尽管不同环境条件下表达调控的变化伴随着反式因子的变化，但在酵母中物种间基本的表达差异可以用正调控变化来解释[58]。在老鼠细胞(将人类染色质注入其中)的相关研究也已经研究表明正调控效应要强于反调控效应[59]。

验证工作给所有的大尺度研究提出了特别的挑战。具体的预测可以用小尺度方法来测试，比如个体反式因子或正元素的灵活运用，接着是全基因描绘(例如，文献[31,47])。然而，这些策略通常只是被用于验证小部分的模型(“最佳选择”)，并集中在系统的联接上而不是定量方面。在少数例子中，大尺度验证是通过比较模型的预测来进行的(例如，从正调控数据推断调控)，模型的预测有独立的全基因数据(例如，对于所有的转录因子，转录因子与 DNA 的结合[18])。然而，大部分情况下的模型验证工作都是不具有说服力的，因为模型的组成部分很少被测试。因此，两个关键挑战是：如何在基因组尺度上对预测进行实验测试，如何将收集的实验结果合并起来。一个立即可行的研究方向是根据测试数据检验模型如何推广，另一个更雄心勃勃的目标是利用数据对模型进行反复提炼。

小尺度方法研究中的回路方法能否应用于研究高级组织中的详细基因和分子操控技术是一个巨大挑战。例如，哺乳动物中反调控因子的基因敲除，在代价和耗时方面大约是酵母和细菌中的 100 倍，还有启动子操控也是一个类似的挑战。由于许多反调控因子是多效的，他们的删除是致命的，所以要求有更为尖端的方法。小干扰 RNA 的技术已经使基因沉默(压制)成为可能，但它倾向于非特异性的偏离目标起作用(这还未被完全理解)。未来对这些非特异效应的深入研究将极大地促进哺乳类转录回路的研究。哺乳类细胞中联合压制干扰多蛋白的能力是十分有限的。在斑马鱼、海胆和非洲爪蟾方面吗啉代寡核苷酸微注射的引入已经很成功了，但是在哺乳类系统中这个方法还没被建立。

迄今为止，已经在海胆中重建了动物中最综合和成功的模型，该模型包括基因调控网络(它具体说明了小分裂球的联接)的完整模型[61]。在这个模式生物中，系统的操控和反调控因子、启动子中的正调控元素与 mRNA 输出的测量被整合以设计一个详细且被证实的基因调控模型。使用基于吗啉的压制的能力，启动子设计方法，和原位测量 mRNA，这三个因素对该模型的有着巨大贡献。

小尺度和大尺度方法的联合

尽管基因调控的小尺度和大尺度方法都取得了进展,大部分研究还不能弥补这些方法之间的裂痕。正如先前注意的,小尺度方法能产生精细且现实的细节和广泛的证据,但是被限定在一些基因(常常是一个基因)上。而大尺度方法能考察很多基因,但常常依赖于生物学上不可行调控作用(例如,布尔逻辑或线性),还缺乏证实。

最近研究强调联合两种方法的希望。例如,有代表性的遗传学方法是将小尺度研究与线性模拟及 mRNA 描绘相结合,这个方法允许酵母中控制对高渗透性反应的调控回路的详细定量重构[1]。另外,大尺度研究的发现可以被小尺度方法详细地跟进。例如,在 mRNA 图谱中观察到的酵母调控子的行为,可以用在控制表象目标基因的单输入回路方面的小尺度研究来解释。贝叶斯网络方法起初是为研究 mRNA 剖面图[48]而发展起来的,该方法已经被成功地应用到信号路径的单个细胞的测量上去了[62]。

随着最近技术的发展(小尺度方法的实验工具被用到基因组层次),小尺度和大尺度方法之间的差异变得更为模糊。这些差异包括,在多细胞生物发展[63]、大尺度序列操控(允许启动子和转录因子的设计和干扰)、大尺度可调反式因子干扰期间(这在高级生物[31]中也包括),单个细胞中描绘多个基因的能力。因为细胞成分的大尺度干扰变得越来越可行,所以需要基因描绘的样本数量也变得异常的大。这提高了基因组尺度剖面分析方法的需求。相对于如同微阵列的方法,这个方法更便宜和快捷,需要的细胞数量和工作量大体上更少。总之,这样的改良方法将能够构造基因调控作用(该作用更接近小尺度系统中使用的现实的模型)的大尺度模型。

另一个主要进展的方向是将这些方法和时间序列实验(它进行动态基因调控作用的研究)相结合。目前的许多研究依赖于定态的度量,但是在细胞分化和生病时,对应环境的变化,转录响应也随时间变化。此外,瞬时的关系能帮助从因果关系中将关联区别开来。大部分计算方法没有运用这个能力,但是最近的研究证明它的潜能[31,51]。一个具有里程碑意义的工作就是构造定义完善的一个子网络(“简单模型”),研究其上的调控和动力学[20,66],并作为评价其它模型的基准。

在所有的研究方法中,最根本的问题或许是能否构造可以推广到新情况的模型。几乎所有主流的模型都是在一定条件下进行预测表达。由于可能状态的空间非常大,预测的成功程度严重依赖于对基因调控作用中非线性和组合效应潜在特征的理解。

在这篇综述中,我们仅仅考虑了调控转录层次上的回路,这里的输入是直接影响转录的正调控序列和反调控因子(例如,转录因子、核小体或染色质修饰),输出是转录水平。然而,这些回路是由许多其他相互作用成分构成的更大且更复杂的细胞网络的一部分。例如,仅仅是 RNA 本身就接受编辑、化学修饰、RNA 结合蛋白和其他非编码 RNA 的影响,这影响其降解的速率、是否能被翻译及蛋白质翻译的速率。更一般地,转录回路被紧紧耦合到信号、新陈代谢和定位系统中(它们是复杂的三维细胞和机体的组织的一部分)。复杂系统以这种方式处理信息和完成功能,并最终决定表达类型。

致谢

本文得到了 Amit 的帮助。H.D.K.得到了 Burroughs Wellcome Fund Career Award at the Scientific Interface 基金的资助。E.K.O.得到了 HHMI 和 NIH GM51377 的支持。A.R.得到了 Burroughs Wellcome Fund Career Award at the Scientific Interface 基金和 NIH DP1-OD003958-01 的支持。

参考文献和后记:

- [1] A. P. Capaldi et al., Nat. Genet. 40, 1300 (2008).
- [2] L. Cai, C. K. Dalal, M. B. Elowitz, Nature 455, 485 (2008).
- [3] N. J. Guido et al., Nature 439, 856 (2006).
- [4] N. Rosenfeld, J. W. Young, U. Alon, P. S. Swain, M. B. Elowitz, Science 307, 1962 (2005).
- [5] L. Bintu et al., Curr. Opin. Genet. Dev. 15, 125 (2005).

- [6] Y. Setty, A. E. Mayo, M. G. Surette, U. Alon, *Proc. Natl. Acad. Sci. U.S.A.* 100, 7702 (2003).
- [7] T. Kuhlman, Z. G. Zhang, M. H. Saier, T. Hwa, *Proc. Natl. Acad. Sci. U.S.A.* 104, 6043 (2007).
- [8] H. D. Kim, E. K. O'Shea, *Nat. Struct. Mol. Biol.* 15, 1192 (2008).
- [9] A. Becskei, B. B. Kaufmann, A. van Oudenaarden, *Nat. Genet.* 37, 937 (2005).
- [10] A. E. Mayo, Y. Setty, S. Shavit, A. Zaslaver, U. Alon, *PLoS Biol.* 4, 555 (2006).
- [11] M. E. Wall, W. S. Hlavacek, M. A. Savageau, *Nat. Rev. Genet.* 5, 34 (2004).
- [12] A. D. Johnson et al., *Nature* 294, 217 (1981).
- [13] M. Thattai, A. van Oudenaarden, *Proc. Natl. Acad. Sci. U.S.A.* 98, 8614 (2001).
- [14] J. T. Wade, T. A. Belyaeva, E. I. Hyde, S. J. W. Busby, *EMBO J.* 20, 7160 (2001).
- [15] S. Hooshangi, S. Thiberge, R. Weiss, *Proc. Natl. Acad. Sci. U.S.A.* 102, 3581 (2005).
- [16] U. Alon, *Nat. Rev. Genet.* 8, 450 (2007).
- [17] S. D. Taverna, H. Li, A. J. Ruthenburg, C. D. Allis, D. J. Patel, *Nat. Struct. Mol. Biol.* 14, 1025 (2007).
- [18] C. T. Harbison et al., *Nature* 431, 99 (2004).
- [19] J. Gertz, E. D. Siggia, B. A. Cohen, *Nature* 457, 215 (2009).
- [20] T. Ellis, X. Wang, J. J. Collins, *Nat. Biotechnol.* 27, 465 (2009).
- [21] J. Gertz, B. A. Cohen, *Mol. Syst. Biol.* 5, 244 (2009).
- [22] C. D. Brown, D. S. Johnson, A. Sidow, *Science* 317, 1557 (2007).
- [23] C. H. Yuh, H. Bolouri, E. H. Davidson, *Science* 279, 1896 (1998).
- [24] F. H. Lam, D. J. Steger, E. K. O'Shea, *Nature* 453, 246 (2008).
- [25] A. Mortazavi, B. A. Williams, K. McCue, L. Schaeffer, B. Wold, *Nat. Methods* 5, 621 (2008).
- [26] P. D'haeseleer, *Nat. Biotechnol.* 24, 959 (2006).
- [27] D. S. Johnson, A. Mortazavi, R. M. Myers, B. Wold, *Science* 316, 1497 (2007).
- [28] T. S. Mikkelsen et al., *Nature* 448, 553 (2007).
- [29] N. Kaplan et al., *Nature* 458, 362 (2009).
- [30] M. F. Berger, M. L. Bulyk, *Methods Mol. Biol.* 338, 245 (2006).
- [31] The FANTOM Consortium and Riken Omics Science Center, *Nat. Genet.* 41, 553 (2009).
- [32] D. Das, Z. Nahlé, M. Q. Zhang, *Mol. Syst. Biol.* 2, 2006.0029 (2006).
- [33] D. H. Nguyen, P. D'haeseleer, *Mol. Syst. Biol.* 2, 2006.0012 (2006).
- [34] T. S. Gardner, D. di Bernardo, D. Lorenz, J. J. Collins, *Science* 301, 102 (2003).
- [35] M. A. Beer, S. Tavazoie, *Cell* 117, 185 (2004).
- [36] E. Segal, T. Raveh-Sadka, M. Schroeder, U. Unnerstall, U. Gaul, *Nature* 451, 535 (2008).
- [37] T. Raveh-Sadka, M. Levo, E. Segal, *Genome Res.* 19, 1480 (2009).
- [38] E. Segal, R. Yelensky, D. Koller, *Bioinformatics* 19, i273 (2003).
- [39] E. Segal, Y. Barash, I. Simon, N. Friedman, D. Koller, *RECOMB '02: Proceedings of the Sixth Annual International Conference on Computational Biology, Washington, DC, 18 to 21 April 2002 (Association for Computing Machinery, New York, 2002)*, pp. 263–272.
- [40] D. Pe'er, A. Tanay, A. Regev, *J. Mach. Learn. Res.* 7, 167 (2006).
- [41] F. Gao, B. Foat, H. Bussemaker, *BMC Bioinformatics* 5, 31 (2004).
- [42] G. Stolovitzky, D. Monroe, A. Califano, *Ann. N. Y. Acad. Sci.* 1115, 1 (2007).
- [43] C. Zhu et al., *Genome Res.* 19, 556 (2009).
- [44] C. Ye, S. J. Galbraith, J. C. Liao, E. Eskin, *PLOS Comput. Biol.* 5, e1000311 (2009).
- [45] C. T. Workman et al., *Science* 312, 1054 (2006).
- [46] I. Amit et al., *Nat. Genet.* 39, 503 (2007).
- [47] E. Segal et al., *Nat. Genet.* 34, 166 (2003).
- [48] D. Pe'er, A. Regev, G. Elidan, N. Friedman, *Bioinformatics* 17, S215 (2001).
- [49] J. J. Faith et al., *PLoS Biol.* 5, e8 (2007).

- [50] K. Basso et al., *Nat. Genet.* 37, 382 (2005).
- [51] S. A. Ramsey et al., *PLOS Comput. Biol.* 4, e1000021 (2008).
- [52] Z. Hu, P. J. Killion, V. R. Iyer, *Nat. Genet.* 39, 683 (2007).
- [53] R. B. Brem, G. Yvert, R. Clinton, L. Kruglyak, *Science* 296, 752 (2002).
- [54] P. J. Wittkopp, B. K. Haerum, A. G. Clark, *Nat. Genet.* 40, 346 (2008).
- [55] S. I. Lee et al., *PLoS Genet.* 5, e1000358 (2009).
- [56] V. Emilsson et al., *Nature* 452, 423 (2008).
- [57] Y. Chen et al., *Nature* 452, 429 (2008).
- [58] I. Tirosh, S. Reikhav, A. A. Levy, N. Barkai, *Science* 324, 659 (2009).
- [59] M. D. Wilson et al., *Science* 322, 434 (2008).
- [60] D. R. Corey, J. M. Abrams, *Genome Biol.* 2, reviews1015.1 (2001).
- [61] P. Oliveri, Q. Tu, E. H. Davidson, *Proc. Natl. Acad. Sci. U.S.A.* 105, 5955 (2008).
- [62] K. Sachs, O. Perez, D. Pe'er, D. A. Lauffenburger, G. P. Nolan, *Science* 308, 523 (2005).
- [63] D. Dupuy et al., *Nat. Biotechnol.* 25, 663 (2007).
- [64] J. Du et al., *Nat. Biotechnol.* 27, 77 (2009).
- [65] G. K. Geiss et al., *Nat. Biotechnol.* 26, 317 (2008).
- [66] I. Cantone et al., *Cell* 137, 172 (2009).