

复杂网络视角研究反应系统

韩定定¹ 朱亮²

(1 上海市多维度信息处理重点实验室 200241; 2 中国科学院上海应用物理研究所 201800)

DOI:10.13405/j.cnki.xdwz.2015.03.005

网络科学作为统计物理与复杂性科学的一个重要分支,在近十几年来取得了长足的发展。巴拉巴希(A-L. Barabási)等人提出的“无标度”模型和瓦茨(D. Watts)等人提出的“小世界”模型用简洁优雅的规则开创性地描述了现实世界的复杂系统。网络的思想源于图论这一研究顶点和边所组成图形的数学分支,它将系统中的个体看作节点,个体间的关系看作边,将节点连边的数量称为度值。“无标度”模型认为,新加入的节点在连边时服从“偏好依附”规律,即更加倾向于连接度值大的节点,最终形成的网络所有节点的度值呈 $P(k) = k^{-a}$ 形式的幂律分布(k 为度值, a 为幂律分布指数,常见网络的 a 值在2到3之间),这种分布下 k 的取值可以相差悬殊,无法用一个值标度 k 的范围,故模型被称为“无标度”。“小世界”模型解释了著名的“六度分离”现象(图1),发现现实世界中许多系统介于规则网络与随机网络之间,具有较短的平均最短路径(两个节点之间的边数)和较高的聚类系数(节点的两个邻居相互也是邻居的概率)。一个远方的朋友,或是一个出国的亲戚就能大大扩展社交范围;人与人之间往往只需要通过几个人就能联系上,相对于几十亿的人口,我们所处的世界真不大。这些研究工作将图论与统计物理、计算机、生物乃至社会学等诸多学科进行了交叉,将科学的疆域拓展到了新的广度与深度。

在各种复杂系统中,有一类个体关系比较常见但又相对特殊,它就是反应关系,如化合物反应、蛋白质代谢。反应关系具有物质转变这一独特的属性,有别于其他研究对象,如社会中的人际关系、计算机网络中的设备互联、交通系统中的运输集散,这些系统的个体相对独立存在,个体本身没有消耗、生成等具有动力学特征的关系。复杂网络关于反应关系的研究开始于生物系统,并逐步扩展到各个领域。

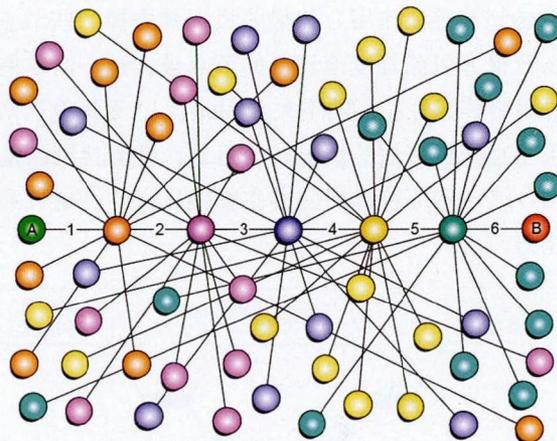


图1 “六度分离”理论由哈佛大学心理学教授米尔格拉姆(S. Milgram)于1967年提出,他通过信件传递实验发现,人类社会是一个小世界,人与人之间的通信平均只需要6个中间人的距离(图片来源: http://en.wikipedia.org/wiki/File:Six_degrees_of_separation.svg)

生物代谢网络

2000年,郑(Jeong)等人首先从复杂网络的视角研究了大规模代谢网络,该工作发表在《自然》杂志上,揭示了生物体网络的一种优化的组织结构。他们分析了涉及生命三大领域的43种有机体,如古细菌、大肠杆菌(细菌)、秀丽隐杆线虫(真核细胞),将代谢涉及的物质作为节点,如核糖、ATP(三磷酸腺苷)、酶,将这些物质的反应关系作为边,连接成43张网络。以大肠杆菌为例绘制的代谢网络如图2所示,包括9种反应物、产物和3种酶,这12个节点只是大肠杆菌网络的一小部分,每种有机体的完整网络都有成百上千个节点。尽管组成要素、代谢通路不同,这些代谢网络都具有相同的“无标度”属性,与一些非生物系统的组织结构有着惊人的相似性。虽然43个网络的节点总数各不相同,有100多的,有800多的,但是它们的网络直径(所有最短路径中的最大值,是衡量网络传导性能的一个指标)却近乎恒定,约为3.3,直径的值与网络节点数呈对数关系,这说明生物体在

增加复杂度的同时会增加内部连接，以维持较短（高效）的代谢通路，也就是说，100多个节点和800多个节点的网络，他们的内部传导能力都很强。另外，网络的“无标度”性意味着较高的鲁棒性，也称可靠性、容错性、抗毁性。网络的结构为少数大度值节点连接着大多数度值较低的节点，大度值节点对网络传导性能起着重要作用。如果随机选择去掉一些节点，那么有很大可能是度值较低的节点被去掉，而这些节点对网络传导性能影响不大。因此，“无标度”网络对于节点随机失效的情况较为可靠，能维持网络基本结构。这一网络特性对应着生物体良好的稳定性和适应能力。

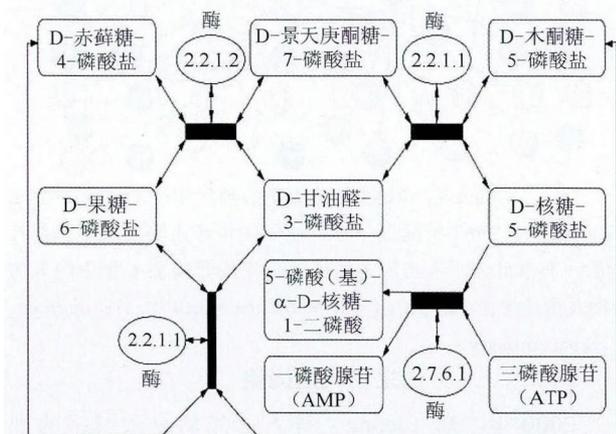


图2 大肠杆菌的部分代谢网络，每种反应物、产物以及酶都对应于网络的一个节点，代谢途径对应于网络的边，根据反应关系相互连接。例如：D-赤藓糖-4-磷酸盐和D-果糖-6-磷酸盐在酶2.2.1.2的催化下反应生成D-景天庚酮糖-7-磷酸盐和D-甘油醛-3-磷酸盐（Jeong H, Tombor B, Albert R, et al. *The large-scale organization of metabolic networks*[J]. *Nature*, 2000, 407(6804): 651 ~ 654.）

郑等人使用统计物理的方法研究生命科学中的系统还得到了很多其他有价值的结果，这和传统的以还原论为主导的生物学研究大相径庭——细胞或微生物中的物质、能量如何产生，信息如何传递，人们通过探索细胞级别的行为来理解这些过程。然而，随着研究工作的深入，越来越多的结果表明很多生物功能无法由单一细胞完成，而需要多个细胞相互协作才能实现，这就要求研究者从整体论的角度着手了，如基因调控网络、蛋白质作用网络、信号传递网络。复杂网络理论的诞生与快速发展正是契合了这一科学研究趋势。

天体物理、化学网络

在上述研究工作的启发下，格莱斯（Gleiss）等人于2001年首先将反应关系网络的研究对象从生物拓展到化学领域。行星大气中的化学反应不同于有酶参与的代谢反应，反应物、产物和反应条件都大相径庭，然而诸如地球、木星、土星等行星的大气化学反应网络其节点度值也都具有 $P(k)=k^{-a}$ 形式的“无标度”幂律分布，幂指数 a 约为1.9。2004年，索罗（Sole）等人进一步推进了这个工作，根据UMIST动力学数据库中更完整的化学反应数据，构建了行星大气和星际介质中的化学物质反应关系网络。他们发现虽然很多网络都具有“小世界”特性，但是地球网络的失配系数显著高于其他行星——失配现象是指节点与度值不同的其他节点相连，在复杂网络中较为常见。地球大气的高失配说明网络中的层级结构特别明显，存在较多小度值节点连接到大度值节点的情况，类似于计算机网络中的集线器（hub）。一个具体的例子可以参看图3（b）和（c），它们分别为火星和地球的大气化学物质反应关系网络。地球的网络度分布是幂指数约为1.16的“无标度”分布，而其他行星的度值更准确地服从于一些接近幂律的其他分布。地球网络能够自行调节气候，使大气在宏观上保持平衡，这样的机制和代谢网络类似，得益于行星早期水分的留存，拥有了有机物参与的大气。这一发表在《欧洲物理快报》上的研究有助于理解行星大气和星际介质在行星演化中的作用。

2010年，乔利（Jolley）等人在星际介质的研究上更进一步，他们根据反应物的消耗和产物的生成确定了连边的方向，结合模拟计算将反应流强（反应率和反应物丰度的乘积）看作网络连边的权重，得到一个有向加权的反应关系网络，并对其做了多种统计分析。对于所得到的结果，他们认为是对传统模拟计算的一个很好的补充，能够由此建立起定量的基准，方便地找到属性不寻常的、对全局系统有重要影响的节点。

构建网络的方法

索罗等人的研究中还提到了反应关系网络的构造方式，在把系统映射为网络的过程中，这是一个值得

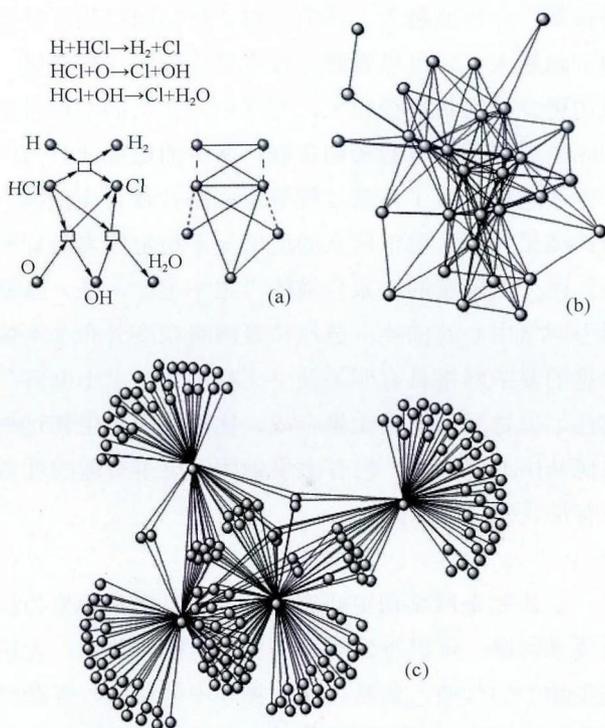


图3 (a) 将3个反应映射为网络的例子：有向二分网络（左）和无向网络（右）；(b) 火星大气化学物质反应关系网络；(c) 地球大气的化学物质反应关系网络 (Sole R V, Munteanu A. *The large-scale organization of chemical reaction networks in astrophysics*[J]. *EPL (Europhysics Letters)*, 2004, 68(2): 170.)

探讨的问题，关系到系统中信息的保留程度。如图3(a)所示，将3个反应映射为网络时，将网络表示为有向二分图可以最大限度地保留信息——反应物、产物作为一类节点（用圆圈表示），反应方程作为另一类节点（用方块表示），连边的箭头对应于反应物的消耗或产物的生成。虽然通过这种方法得到的网络可以把所有反应都完整地表示出来，但是对于二分图，现有的分析工具还略为匮乏。索罗等人采用了一种大为简化的方法，他们只保留了反应物、产物节点，将同一个反应中的反应物与产物相互连接而不区分方向，构成一个反应物、产物的无向网络。即便如此，他们依然得到了许多有价值的结果。那么，将系统抽象成网络的过程中，保留多少信息是足够的呢？

对此，霍姆 (Holme) 等人总结了各种构造网络的方法进行了比较和评价。如图4所示，(a)是一组反应方程，包含 p、q 两个反应，我们至少可以用 (b) ~ (e) 这4种构网方式对反应系统进行抽象。

在复杂网络中，若干节点形成团簇是较为常见的现象，人们使用“模块度”这一指标来衡量连边包含于一个团簇而不是处于两个团簇之间的倾向。他们的评价方法是比较经过构网抽象后的网络模块度和系统真实模块度的差别。通过计算发现，能最好地保留网络模块度信息的构网方式是 (b) 和 (d) ——网络模块度值最接近系统真实值。不难看出，(b) 是反应物和产物相互连边，(c) 是反应物、产物各自内部相互连边，(d) 是 (b) 和 (c) 的叠加，(b) 和 (d) 是近乎等价的方法；(c) 的方法去掉了反应物与产物间的重要连边，很难保留模块度信息，而 (e) 是以反应方程为节点的网络，保留了更少的信息——构网抽象后丢失了原系统中较为重要的信息，这种构网方式就显得不太理想了。当然，构网方式不局限于上述4种，除了二分图，近年来多层网络的研究受到了越来越多的关注。所谓多层网络，即同一组节点，具有多种连边。当我们对单一连边构成的网络进行考察，和对整体网络进行分析时，很可能会得到显著不同的结果，其中的差异值得科研工作者们深入探讨。

核素网络

结合反应关系网络的研究发展，我们把涉及 8000

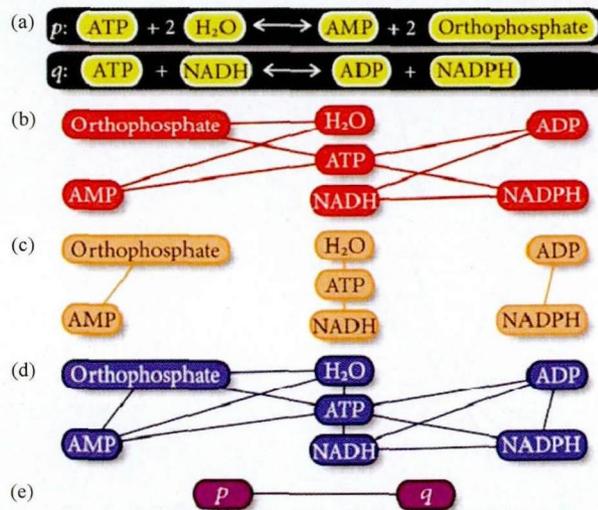


图4 (a) 一组反应方程，包含 p、q 两个反应；(b) 对于每一个反应，相应的反应物和产物相互连边；(c) 反应物之间相互连边，产物之间相互连边；(d) 一个反应中的所有反应物和产物均相互连边；(e) 两个反应因为有共同的物质而相互连边 (Holme P., *Model validation of simple-graph representations of metabolism*[J]. *Journal of The Royal Society Interface*, 2009, 6(40): 1027 ~ 1034.)

多种核素的 8 万多个核反应转化为一张有向加权的
多层网络。构造网络时，参与反应的核素连接至产物
核素，根据反应所依赖的不同粒子，连边分为 4 种类
别：中子、质子、 α 粒子和其他。连边权重对应于反
应率，该值依赖于环境温度。如图 5 所示，横轴为
中子数，纵轴为质子数，即每行都是同一种元素的
同位素，核素根据自身中子、质子数排列于格点，
节点大小对应于其介数值（网络中所有经过该节点
的最短路

径数量，介数值越大，该节点对于提高网络传导性
能的贡献越大）。可以看到，核素网络相对比较规
则，长边稀少，网络直径较大，网络中规律性的结
构可能和元素的周期性有着密切关联；网络的边
缘部分与中子俘获过程、质子俘获过程等反应路
径重合度很高。统计结果显示，稳定核素的度值
与不稳定核素差异明显；由介数确定的核素合成
的“主干道”，是一条蜿蜒于网络中心的曲线。虽
然核素网络在度分布上不像常见的复杂网络具有
明显的“无标度”、“小世界”属性，但是复杂
的重元素合成、化学反应、生物过程如何从中涌
现出来，仍有大量谜团。更多有趣的性质还有待
我们进一步探索。

总结与展望

运用复杂网络理论研究反应系统，将其抽象为
反应关系网络，可以得到对系统更为全面的认识。
无论是生物中的代谢，还是化学、物理中的反应，
复杂网络这一统计物理与复杂性科学研究的新工
具都能胜任分析工作。从无向无权网络到有向加
权网络，从单层网络到多层网络，更精准的构网
方式能够在简洁的系统描述和完整的系统信息保
留之间取得平衡。反应系统的结构特征、鲁棒性、
主要通路都得以展现在人们面前。

可以预见，更好的构网方式，在各种网络上使
用有效的统计工具进行分析，评价不同初始值或
环境变量对系统层面的影响，都是本研究领域特
别关注的方向。在充分的横向研究对比的基础上，
一个通用的反应系统模型，或是若干反应网络度
量指标也许就会出现在不远的未来，为更深刻理
解发生在世界上的各种复杂反应提供帮助。

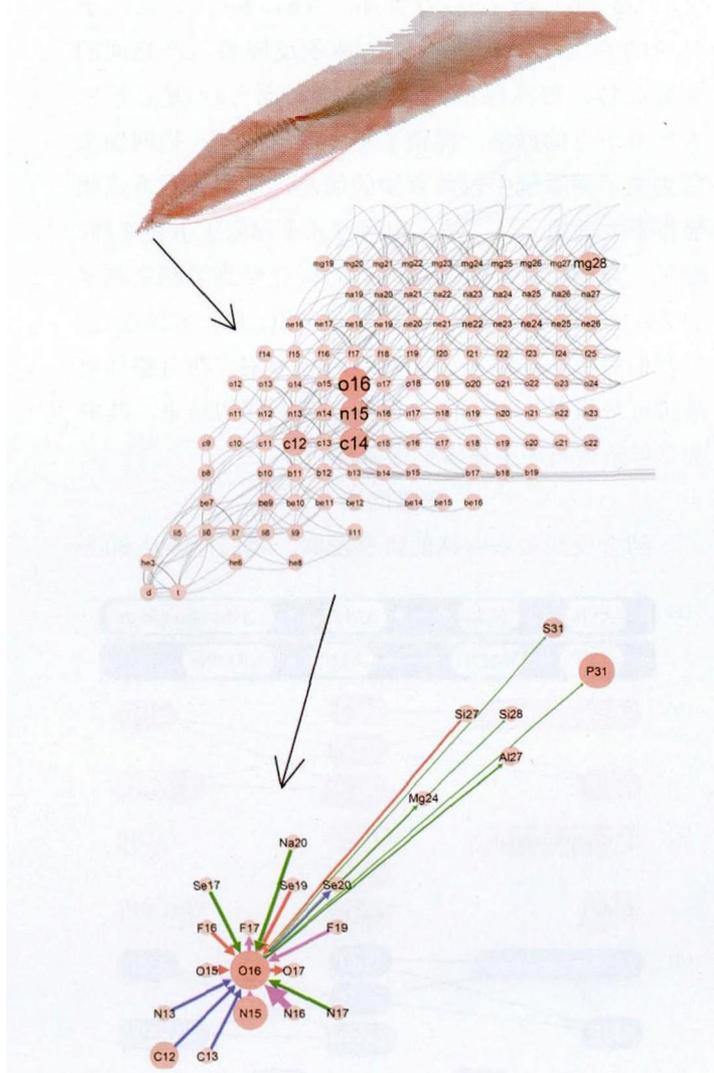


图 5 核素网络的可视化：8000 多种核素作为节点，节点大小对应于其介数值；反应核素连接至产物核素，即连边的方向，连边权重对应于反应率；根据反应所依赖的不同粒子，连边分为 4 种类别用不同颜色表示，中子：红色，质子：紫色， α 粒子：蓝色，其他：绿色，以氧 16 为例， $^{12}\text{C} + \alpha \rightarrow ^{16}\text{O}$ 这个反应对应于碳 12 到氧 16 的蓝色边，不同类型的连边在网络中扮演着不同的角色，看似错综复杂的宏观系统需要多层网络这一工具进行层层剖析

