

文章编号: 1007-6735(2011)05-0409-09

复杂系统结构及动力学现象的统计物理研究

汤雷翰^{1,2}, 赵 鸿³, 刘宗华⁴, 周 涛⁵, 汪秉宏^{6,7}

(1. 北京计算科学研究中心, 北京 100864; 2. 香港浸会大学 物理系, 九龙; 3. 厦门大学 物理系, 厦门 361005;

4. 华东师范大学 物理系, 上海 200062; 5. 电子科技大学 互联网科学中心, 成都 610054;

6. 中国科学技术大学 近代物理系, 合肥 230026; 7. 上海理工大学 复杂系统科学中心, 上海 200093)

摘要: 近年来生物学、信息学和社会科学海量数据的积累, 推动了各类定量分析手段的发展. 统计力学作为联系多体系统微观与宏观描述的数学理论, 在复杂性科学的发展历程中正发挥着作用. 本文结合复杂网络和非平衡态系统的若干研究成果, 对统计物理方法在复杂现象的描述、主要变量及其相互依赖关系的确定、模型的提出与分析、及普适性规律的发现等方面的进展作一简单介绍, 并探讨统计物理学与相应学科深层次交叉所面临的机遇和挑战.

关键词: 复杂系统; 统计物理; 多层次结构与动力学; 海量数据分析与建模

中图分类号: N 94 **文献标志码:** A

Statistical physics approach to the structure and dynamics of complex systems

TANG Lei-han^{1,2}, ZHAO Hong³, LIU Zong-hua⁴, ZHOU Tao⁵, WANG Bing-hong^{6,7}

(1. Beijing Computational Science Research Center, Beijing 100864, China;

2. Department of Physics, Hong Kong Baptist University, Kowloon, China; 3. Department of Physics, Xiamen University, Xiamen 361005, China; 4. Department of Physics, Eastern China Normal University, Shanghai 200062, China; 5. Web Sciences Center, University of Electro Science and Technology of China, Chengdu 610054, China; 6. Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China; 7. Complex System Science Center, University of Shanghai for Science and Technology, Shanghai 200093, China)

Abstract: The rapid accumulation of massive amounts of data in biology information technology and various branches of social science in recent years has prompted the development of novel quantitative methods. Statistical mechanics, as a mathematical theory that connects microscopic and macroscopic descriptions of systems with many degrees of freedom, can play an important role in this process. A short review of recent progress in the application of statistical physics methods to a broad range of complex phenomena is presented here. From the examples discussed one may obtain a glimpse of the power of the statistical mechanical approach in capturing the system dynamics at different levels of representation and on many different time scales. With the help of simplified models and approximate treatments, the key variables and their interdependencies can be elucidated. Further theoretical considerations and comparative studies may lead to the discovery of unifying principles. The opportunities and challenges in a closer integration of statistical physics and complex systems science are briefly discussed.

收稿日期: 2011-10-24

作者简介: 汤雷翰(1961-), 男, 教授. 研究方向: 凝聚态与统计物理、计算物理、非线性动力学及系统生物学.

E-mail: lhtang@csrc.ac.cn

汪秉宏(联系人), 男, 教授. 研究方向: 统计物理、非线性科学、复杂系统理论. Email: bhwang@ustc.edu.cn

Key words: *complex system; statistical physics; multi-scale structure and dynamics; large-scale data analysis and modeling*

现代人生活的自然和社会环境在任何意义下都会被划分为一个或者多个相互缠绕的复杂系统. 永不停息的技术进步带来了庞大并不断细化的社会分工, 信息革命又将个体的活动通过多条通讯渠道关联起来, 资本市场和政府的宏观调控影响着大小公司的运作和行业的兴衰, 而某个农场的偶然一次病毒基因突变却可以通过使用现代交通工具的人体传遍地球的每个角落, 威胁整个人类的生存. 如此庞大复杂的系统, 不断地翻腾变化, 吐故纳新, 何是因? 何是果? 可否驾驭? 规则如何设定? 成效如何评估? 危机能否避免? 这些问题的研究, 已不仅仅属于兴趣和好奇心驱动的学术与认知范畴, 而可影响人类社会的未来发展.

伴随着以上提到的技术进步与社会结构的不断复杂化, 人类采集数据、观察现象的能力飞速提升. 以互联网为标志的信息革命的到来, 为大规模数据的记录、储存和分享提供了前所未有的便利. 大量高精度数据的涌现, 为我们定量分析系统的内在关联、寻找各类变量间的相互依赖关系、确定系统行为如何受外界因素影响等性质提供了可能. 自然科学与人文科学的交融, 正将此可能逐步转变为现实, 为复杂性科学这一被誉为 21 世纪的科学的兴起带来了无限生机.

复杂性科学从根本上来说还是一个非常年轻的学科. 复杂系统研究的第一个全球范围的浪潮可追溯到上世纪七八十年代, 那时候一批有代表性的理论体系得以建立, 其中有些思想从上世纪 60 年代就开始萌芽, 有些成果的影响力一直持续到现在. 这些有代表性的理论体系包括: 突变论、微循环论、耗散结构理论、协同学、涌现学说等. 那个时候的一个主要趋势是寻找构建适合于一切复杂系统的普适理论——往往不是从现象, 而是从类似第一性原理出发, 构建理论体系. 很多重要的理论成果, 特别是耗散结构理论和涌现学说, 现在依然具有相当强的生命力和广泛的影响力. 与此同时, 我们也注意到, 这些学说希冀成为复杂系统普适理论的目标, 可以说没有实现. 总结起来, 由于当时的研究对象主要是均匀驱动的物理化学系统, 动力学相对简单, 呈现的时空结构有其特殊性, 因此所构建的理论体系的适用范围有它的局限性, 与其它系统和理论体系的互补性和一致性也没有仔细和严谨的评估, 在一些关键

问题的把握上便不可避免地出现偏差.

21 世纪初的复杂系统研究则呈现了完全不同于上一个世纪的态势. 很多年轻学者不再投入建立具有普适价值的理论体系, 而是从海量数据出发, 首先积累现象和统计规律, 然后建立模型去解释这些现象和规律. 相当一部分研究只停留在现象或者粗糙的规律层面, 另外一部分研究上升到具体理论模型的层面, 很少有研究关注完整理论体系的建立. 当然, 有些学者已经认识到了建立基础理论的重要性, 譬如 Ginestra Bianconi 一直致力于网络系统的讨论, 在更大的参数空间上分析网络结构和功能的关系, 以期找到一般性的规律.

对于一个年轻又充满活力的学科来说, 螺旋式的发展模式有它的必然性. 回顾物理学的发展, 基本概念和方法论的建立往往需要经过一个漫长的迭代过程. 对研究的问题做出清晰界定需要实验观察和数据积累到一定水平, 相应的理论框架的建立也需要经过多方面的尝试、比较和优化. 简化模型的提出和分析, 有助于核心问题的提炼和更有针对性的实验的设计和开展, 从而加快研究进程. 在包含大量变量和复杂动力学的系统的研究中, 统计物理学可以为模型的分析提供严格和系统的方法, 建立系统微观和宏观性质之间的联系. 通过对模型的定量分析, 研究者可以深入了解并重新审视研究对象, 推测系统随着参数的变化可能出现的改变并加以实验验证, 同时也可以拓展的参数空间里分析比较系统的性质和行为. 模型研究的另一个功能是通过模型的数学处理和变换, 将在不同背景下提出的问题联系起来, 起到相互启发和借鉴的作用. 这是新观点和新方法乃至普适理论提出的一个重要源泉.

近十年来, 统计物理学的基本思想、理论、方法和概念在复杂现象的研究中已经有了广泛的渗透. 举例来说, 人类行为和经济波动中的异常检测, 以及信息系统中关联程度的刻画, 都可根植于相应的统计系统; 社会系统中恐慌情绪的传播、信息系统中某一谣言的大爆发, 以及突发性的金融震荡, 都可以用相变与临界现象理论加以近似刻画; 生物调控信号的传输, 信息和社会系统中意见和信息的传输都可以在某种程度上抽象为具有相互作用的多体系统的传播动力学. 可以说, 在近年来的生物信息学、人类动力学、复杂网络等新兴交叉学科方向的飞速发展

过程中,统计物理学在新概念、新方法的提出等方面都发挥了积极作用.

统计物理学与复杂性科学的深度交叉,不仅为相应学科的发展提供了理论基础和工具,同时也对本学科的发展开辟了众多全新的课题,将过去对复杂性的抽象讨论引向深入.与经典统计物理学所关注的均匀体系不同,现实世界中,如生物、工程和社会中的复杂系统表现出很多特殊的属性,例如系统开放、受能量驱动处于非平衡态且不断演化、结构化但不完全有序、具有记忆、适应甚至主动改变外部环境等功能.这类系统的复杂性突出表现在系统的多层次结构、宽广和鲁棒的响应特性、复杂的内在逻辑和复杂的动力学,在宏观层次上具有多种功能,以适应外界和自身对其“处理问题能力”的要求.因此,功能的有效实现的相关机制在这类系统的研究中占有主导地位.

本文结合统计物理学在复杂系统研究中近年取得的进展,对相关的工作做一个简单的回顾.讨论按四个方面展开:a.低维和有限体系的非平衡态统计;b.复杂网络的多层次动力学;c.信息挖掘的统计力学研究;d.生物系统海量数据的整理和建模.由于涉及的问题和研究内容非常广泛,文献量也十分巨大,本文的目的并不是对已有的工作进行系统的总结,而是希望通过代表性问题的陈述及相应研究方法和成果的比较,对领域目前的研究现状加深认识,为制订未来的研究目标和研究路线图提供依据.

1 低维和有限体系的非平衡态统计

将低维和有限体系划分为复杂系统,主要是由于其特有的热涨落性质、对杂质的敏感性、及其呈现的各类非线性动力学特性.相对于其它复杂系统,它们相对简单但具有较明确的物理背景.作为实现各类器件和生物功能的载体,此类系统在纳米材料研究、细胞内分子过程的生物物理研究中占主导地位,实验科学领域有大量的成果和知识积累,同时也为理论研究提出了实际而且迫切的科学问题.从上世纪80年代开始,制备和观察手段的发展开辟了低维小尺度系统研究的新时代,除了微纳颗粒、纳米线、纳米管、石墨烯等低维材料以外,自然界中的生物大分子以及血液循环和污水处理等领域涉及到的微纳尺度流体等都属于这一范围.低维材料在光、电、热、磁、声及力学性能等方面往往表现出不同于块体材料的性质,其动力学行为也由于小尺度而偏离平衡态系统特别是热力学极限下已知的规律.在实际应用中,系统往往处于外界驱动的非平衡稳态下,其统计性质的描述也需要

在非平衡统计物理的框架下开展^[1-5].

平衡态系统由于满足玻尔兹曼分布,其物理性质可以通过统计平均直接获得.当系统处于非平衡态时,以上性质便不成立,目前因没有一般的推导此类系统所处状态的统计权重的方法,计算系统的各类特性便需从解动力学方程出发,这样不仅计算量大大增加,统计物理学众多强有力的计算统计平均的严格或近似的方法也都没有了用武之地.因此,如何有效地计算非平衡态的统计分布函数便是非平衡统计研究最核心的问题.虽然这个问题远远没有解决,但近十几年来,以涨落定理(fluctuation theorems)和大偏差理论(large deviations theory)为代表的理论工作却在回答这个问题的进程中开启了可喜的大门.在与实验研究的结合方面,理论的进展为低维材料的热传导,热扩散,能量传输的不对称现象,低维系统中微粒物质的热迁移,微纳米尺度流体,分子马达等课题的实验和数值模拟研究提供了有力的支撑.

以 Jarzynski 等式为例,自 1997 年此等式被提出以来,其研究进展非常迅速^[6-11].我们知道如果使某个统计物理系统的某个参数在两个给定值间转变,通常将经历一个非平衡态过程. Jarzynski 等式表明,在此过程中对系统所做的功,取决于当系统处在两个不同系统参数下的平衡态(温度相同)的自由能之差.这是关于非平衡态过程少有的具有一般意义的严格结果之一,也在 RNA 的单分子折叠拉伸等实验中得以应用.后续理论工作包括 Crooks 的涨落定理,量子系统中的 Jarzynski 等式及其推广等. Jarzynski 等式的讨论也为发展全新的平衡态及平衡态之间过渡的计算方法提供了启示.

已低维系统热传导的研究,早在上世纪六七十年代就开始了,最近 20 年来对各类模型中出现的非线性激发的细致分析,使得人们对傅立叶热传导定律在低维系统中的适用性有了深层次的了解,认识到了动量守恒或平移不变性在决定低维系统热传导规律方面的重要角色,并且提出了热传导规律对空间维数依赖关系的猜测^[12-26].低维热传导研究目前存在几个焦点.

a. 反常热传导的微观机制是什么? 1998 年胡斑比、李保文、赵鸿提出了动量守恒导致低维系统反常热传导的观点.目前已经清楚,尽管这一原则具有一定的普遍性,但是也存在个别反例,因此反常热传导的存在需要进一步探寻更基本、更深层次的微观机理.

b. 低维系统热传导规律对于系统空间维度的依赖关系如何? 一些理论猜测认为具有平移不变性的

系统热传导系数随着系统尺寸幂指数发散(一维),对数发散(二维),保持不变(三维).这一结果意味着虽然低维系统具有反常热传导,三维材料则仍然遵从傅立叶热传导定律.围绕这一猜测展开了一系列的数值检验乃至实验验证,目前基本认可了这一结论,同时对于具体的依赖方式仍然存在争论.

c. 反常热传导和反常扩散之间的关联.正如昂萨格倒易关系建立了不同非平衡态物理过程系数之间的联系,一些研究者试图建立反常热传导发散系数和反常能量扩散系数之间的普适性联系并且提出了两个不同的公式.但是,这两个公式的正确性仍然处在质疑之中.

d. 反常热传导和反常能量扩散的微观机制的探讨.这方面的争论更加激烈,例如赵鸿等在研究了非线性一维晶格中孤波-孤波、孤波-声子、声子-声子散射的定性定量规律的基础上,通过对有限温度晶格中能量涨落时空关联行为的研究,提出一维晶格中的反常热传导和反常能量扩散是由于孤波的特殊动力学属性造成的^[16-17],而李保文等则以有效声子方法(Bogoliubov 变换)为出发点讨论这一问题,认为有限温度下只存在有效声子,并且认为有效声子能给出某些宏观量对温度依赖规律的更准确预测^[18].

e. 低维材料热传导性质的实验测量.由于实验技术的发展,纳米线、纳米管的热传导,包括与尺寸的依赖关系,已经有了一些测量工作^[19],二维材料石墨烯的热传导实验也已经初步开始^[20].

在低维和有限系统的讨论中,反常统计特性,肥尾幂指数分布,自相似分形结构,奇异高次统计矩等,都是常见的特征.这些特异现象的研究,不仅为深入发展非平衡统计物理和非广延统计物理提供了依据,而且可以为统计物理复杂系统的其他课题提供理论参考、研究方法和工具.

2 复杂网络的多层次动力学

时空系统中的自组织现象及其产生的各种规则和不规则的结构长久以来一直是非平衡统计物理学所关注的问题之一.以往的研究集中于以格点模型为代表的空间均匀系统,而经复杂网络定义的单元间相互作用的系统近十几年来成为研究热点,大量的研究成果极大地丰富了动力学系统现象学的知识,为统计物理学向其他学科的渗透提供了大量有价值的参照模型.以下就复杂网络上的动力学研究近年来在描述人类行为和各类社会经济系统的统计特性和演化规律、大脑神经元系统同步现象的定量刻画等方面所取

得的进展做一简述.

人类行为具有高度的复杂性.针对人类行为统计特性的研究起始于 2005 年 Barabasi 在《Nature》杂志上发表的针对人类行为时间特性的开创性的工作^[27]和 2006 年 Brockmann 等发表在《Nature》杂志上的针对空间特性的工作^[28].从 2005 年算起,人类动力学的研究,仅在《Nature》,《Science》,《PNAS》,《Nature Physics》,《Physical Review Letters》,《PLoS ONE》等有广泛影响力的期刊上,就发表了超过 50 篇研究论文,涉及人类时空行为实证分析、理论模型和具体应用等方面,已经成为复杂性科学领域的一个重要新兴领域.研究的内容涉及人类行为对社会系统形成、社会的结构和发展的效应等,在传染病传播、城市交通、社会安全、国民经济发展等重大民生课题的科学探讨中也取得了若干初步成果,深化了人们对问题背后各种因数间相互影响和博弈的认识.

流行病的传播及爆发几率与其载体——人群的流动和接触模式有密切关联.传统上,流行病的研究主要是假定感染个体与健康个体是均匀混合的,流行病的传播是通过平均场来进行的.这种做法与真实发生的过程显然有相当大的距离,比如人类活动是通过交通工具进行的,而且具有其自身的独特特征如社区性、旅行目的性及对场所的时间段占据特性等.考察真实的流行病传播过程如萨斯的传播,人们发现其传播轨迹并不是从一个区域逐渐向外传播的,而是具有网络特征,因此流行病传播的研究必须在复杂网络上来进行.这方面的一个重要成果是,由于社会网络是无标度网,具有大量联接度的中心节点的存在将导致流行病传播的阈值几近于零^[29],也就是说,一个非常小的传染率就能使流行病在人类社会传播开来.理论的一个重要的应用是在流行病的预防与控制中.计算发现,从无标度网络上随机选择少量节点进行免疫能极大地减少流行病的传播.网络上流行病传播的最新进展是将网络上的静态个体拓广到移动的情形,从而将静态模型拓广到个体随机行走的反应-扩散模型^[30]及非随机行走(即目的性旅行)的加速传播模型等^[31].然而流行病的研究是一项长期而艰巨的任务,还有许多重要问题需要进一步解决,如不同种群间的流行病的传播、流行病与时变网络之间的相互作用、多层网间的流行病传播等.最新研究表明人类流动模式在长时间尺度下具有一定的可预测性,研究这种可预测的流动模式如何影响流行病的传播,或者更广泛的舆论、谣言、恐慌等在社会网络上的传播,将有助于预防各种突发事件和维持社会的稳定.

在社会经济领域,金融市场也是高度复杂的动力

学系统.2008年金融危机对世界经济造成的巨大冲击,使得人们在重新审视主流金融经济理论的同时,寻找非传统的方法和理论,探讨可能认识和解决金融经济问题的新思路和新途径.2008年以来,Bouchaud^[32],Lux等^[33],Schweitzer等^[34],Farmer等^[35]纷纷在《Nature》和《Science》撰文指出,人们应该可以从实际金融经济数据出发,探索市场的运动规律.而新的金融经济理论,需要考虑异质主体的相互作用,以及金融经济网络的结构、功能和演化规律.这样构建的金融经济理论,是以实际市场数据为基础的实证唯象学说,有别于以逻辑为基础的公理体系.近年来,物理学家应用统计物理学中研究多体系统的概念和方法,从微观层面研究金融市场的集体行为,从而建立相应的多体微观模型,寻找普适和非普适的运动规律,是人们对金融市场从另一角度和另一层次的认识和探索,具有重要的科学意义和潜在的应用价值.同时,物理学家也从金融问题的研究中得到启发,推动复杂系统统计物理本身的发展.

复杂网络动力学的另一个例子是大脑神经网络,它是由几百亿个神经元通过极其复杂的、多层次连接而形成的大脑皮层神经系统,是自然界中所知的最为复杂的动力学网络体系.它的结构与动力学直接关系到大脑的各种功能及相应的精神疾病及认知障碍.大脑从下到上可以分为7个层次:分子、神经元、神经元群、神经网络、大脑皮层、功能分区和神经中枢,其中神经元、神经元群、神经网络和大脑皮层4个层次上都发现了同步现象.然而,神经系统中的不同时空尺度的复杂同步行为是如何实现的,它与大脑神经元之间发放的输运及信息处理之间的关系如何,这些问题可以通过研究振子网络中复杂的同步行为而获得重要理解.最近十多年来,由于脑造影技术的进步,人们已在系统层次对大脑的连接及活动积累了非常多的有益数据.如何分析理解这些数据从而获得对大脑大规模的复杂结构、动态活动及认知功能之间的关系,必将是未来研究大脑会蓬勃发展的新方向,呼唤着新的统计物理方法来揭示出大脑大尺度的结构和功能网络与认知过程的关联.

3 信息挖掘的统计力学研究

随着信息存储与处理技术以及互联网与万维网的发展,我们可以获取的信息总量呈现爆炸性增长的趋势.与此同时,我们处理信息的能力却没有相应的增长.目前信息科学面临的最大挑战是如何解决信息过载问题,也就是如何帮助普通用户获取其所感兴趣

的信息.信息挖掘是极有希望推动并最终解决信息过载问题的关键方法.从广义上说,信息挖掘包括两部分:一是信息检索,二是信息发现.前者着力于将已有信息以一种更好的方式进行组织,并以更易于理解和获取的方式展现给用户——典型的例子是搜索引擎.后者则是通过分析,自动地推断一些可能的关联,这种关联本身可以看作新的信息——典型的例子是推荐引擎.搜索引擎是以被动的方式服务(需要用户输入搜索词),是对已知信息的再组织;推荐引擎是以主动的方式服务,是对未知信息的发现.

信息的组织和发现是重要的科学问题,事实上,信息挖掘的研究荟萃了计算机科学、物理学和数学的很多思想、方法和技术.与此同时,我们也注意到,尽管信息挖掘一直是一个非常活跃的研究领域,这方面的研究一直停留在具体方法的层面,也就是针对一个具体的问题或一类问题,提出具体的算法并比较算法在某些数据集上的表现.事实上,国际学术界对于信息挖掘方法和技术的研究甚嚣尘上,却一直缺乏对基础理论的探索.统计力学的系综理论有望成为信息挖掘的理论基础,而平均场分析方法有望成为分析估计算法表现的有力武器,这方面的研究进展,不仅可为信息科学带来重大突破,也能拓广统计物理对其他学科的渗透,对于建设具有广阔视野的统计物理学科平台有重大助益.

传统的研究主要集中于信息挖掘技术的研究,大多停留在算法设计与高效实现的层面.典型的信息挖掘问题包括链路预测和个性化推荐,典型的信息挖掘技术包括:基于相似性的挖掘方法、基于群集智能的协同挖掘方法、基于内容的挖掘方法、基于潜在语义空间的挖掘方法、基于概率模型的挖掘方法、以及混合算法和集成学习方法等.此外,多种数据分析技术,如数据聚类、Bayesian网络、关联规则、数据分类、K-means方法、最大熵方法、云模型、多示例学习、神经网络、线性回归等均被用于信息挖掘.这方面的研究进展可以参考文献[36-39].

最近,张翼成及其研究小组将统计物理的方法引入到信息推荐领域中来,利用经典的物理学方法(物质扩散和热传导等)来设计个性化推荐算法.他们的工作主要集中在两方面:基于网络结构的推荐和基于协同过滤的推荐.在网络结构方面,张翼成等首次提出将热传导理论引入推荐系统^[40].周涛等对基于网络的信息推荐算法进行了比较系统的研究,提出个性化推荐在一定程度上等价于二部分图向一部分图的含权投影问题,并依此提出了基于复杂网络资源分配的推荐算法^[41].周涛等提出了一系列刻画推荐结果

多样性和新颖性的指标,建立了比较完整的包含多样性、新颖性和精确性的评价体系^[42-43],并通过结合物质扩散和热传导的混合算法,解决了困扰学术界的多样性-精确性兼顾两难问题^[44].更复杂的包括社会标签、噪音、负面评分等因素的讨论,最近也被纳入到了物理方法的框架中,并得到了明显优于以往方法的结果^[45].

总的来说,信息挖掘是具有重大理论和应用价值的多学科交叉的活跃研究领域.由于目前各类方法缺少坚实和统一的理论基础,因此无法很好地回答一些基本问题.利用统计力学的思想、方法和概念,建立信息挖掘的理论基础,将有可能为该领域的研究带来新一轮在方法论和算法设计层面的创新成果,并有可能取得重大突破.以二部分图上电子商务商品推荐^[46]为例,目前的信息挖掘手段面临3个重要问题,分别是冷启动问题——新用户新商品无法进行精确推荐;稀疏性问题——用户和商品的直接关系很少,信息不充分,推荐效果差;多样性-精确性难以兼顾——精确的推荐往往都倾向于推荐热销流行的产品,新颖性、多样性和个性化程度都低,用户体验差.这3个问题影响了信息挖掘的应用效果.推动解决或部分解决这些问题,可望引导产生创新高技术产品,带来可观的社会经济价值.

4 生物系统海量数据的整理和建模

生命现象作为一类特殊的复杂系统,以它高度优化的遗传和繁衍模式,经过数亿年的进化和分化,在地球生物圈这一特定的物理生态环境下,将结构与功能的互动发挥到了极致.生存竞争,资源猎取,优胜劣败,其中有数不尽的玄机 and 叹息.这类现象的科学研究,近年来由于知识的大量积累、高通量数据采集等生化分析手段的发展、以荧光显示为代表的活体观察显微技术的出现和大规模推广、以及与大规模 DNA 测序同步发展的生物信息学的诞生,人类对生命的认识正经历着一个全方位的飞速提升.

生命现象的多层次和千变万化也为科学地研究和提取其中的规律提出了极大的挑战.相对说来,细胞内的基本生命运动模式在生物进化的早期就相对地固定了下来,因此更有利于系统优化过程的定量研究.细胞作为一个生命单元,有着非常丰富的功能和相应的调控系统,其中细胞代谢的转录调控系统是一个具有代表性的功能化复杂系统.代谢是生命现象中的一个基本过程,细胞生长所需的生物质量(包括氨基酸、核酸、磷脂分子等多种化合物)和化学能是由酶

催化的代谢反应所生成的,代谢反应与代谢物形成一个庞大的二分网络,支撑代谢物的相互转化.经过大半个世纪的积累,人们对代谢网络的组成和架构已有较详尽的了解(见 KEGG 数据库 <http://www.genome.jp/kegg/>).Palsson 组运用流平衡和线性优化方法,建立了包括大肠杆菌和酵母菌在内的多种细胞的代谢模型,用以计算在特定生长环境下的细胞生长速率,结果部分得到了实验的证实^[47-48].给定生长条件下代谢流的分布近似于树状,兼具有明显的区域性,单个反应的速率依赖于相应酶的数目和活性.而转录调控作为细胞内部的一个包含多个输入节点和多重反馈机制的复杂网络,可以综合外界养分及内部需求的信息,控制代谢通路的开关及流量,优化物质和能量的利用率.随着近来基因测序和基因芯片等高通量技术的发展,有关的实验数据正在快速积累.虽然不同物种的代谢网络大致相同,酶蛋白表达水平的调控系统却千变万化,具有复杂的拓扑结构和多样的调控逻辑^[49-51].粗略来说,这可以归结于细胞组分、生长环境、及细胞周期的差异.另一方面,不同物种在进化过程中也可能是找到了同一个问题的不同解,或者当某个参数发生变化的时候,解的结构也发生了变化.不同物种转录调控网络所呈现的丰富的行为,为结构与功能关系的研究提供了大量的实例,但此类海量数据的整理尚缺乏系统性和合适的理论基础.

建立和认识功能化复杂系统的结构与功能关系,需要引入一整套的分析方法,这与建立系统微观相互作用与宏观性质关系的研究有可类比之处,但由于功能这一目标函数的引入,结构的讨论为统计物理学提出了新的问题,包括微观到宏观定量关系的建立(关键变量的引入和准确定义,系统层次的建立)、数据挖掘、方法的评估和系统宏观特性(功能)的确立与划分等关键科学问题.整体说来,以上关键问题的解决依赖于对有关正问题(forward problem,即给定动力学规则来预测系统行为)^[52-53]和反问题(inverse problem,即从现象和数据出发确定背后的动力学规则)^[54-55]认识的不断深入.以细胞代谢的转录调控系统为例,目前有多种实验手段通过给予细胞刺激或控制个别基因表达来观测其形貌和内部状态演变,获取大量数据.例如基因组表达谱,蛋白质数量和代谢中间物浓度随时间的变化等.对数据进行适当的预处理后,运用建立在信息熵等概念上的各类算法来反推调控因子与受控基因间的作用关系,这是研究系统动力学机理的第一步.而正问题则是研究系统结构如何产生和影响功能,将分子网络动力学模块与其实现的生物功能直接联系起来,通过系统的非线性动力学、噪声传

播及稳定性分析,了解系统的运作机理.在分析大规模生物网络动力学的研究中,引进多层次结构,合理地将网络分解成不同层次的功能单元是降低系统复杂性的一个基本和重要的手段^[56-57].

分子网络结构和功能相互依赖的演化过程也是一个解码生命现象的基本问题,甚至可以说是最基本的问题,因此受到广泛的关注^[58-59].网络结构变化包括个体数目的增减、节点间相互作用的强度及方式的变化等;而功能的变化则可体现为执行特定任务时“效率”的提高、稳定性的增强以及自适应能力的增加等.对于互演化问题的研究不仅将进一步加强我们对复杂系统中结构和功能之间联系的认识,同时也将直接提高我们优化和设计分子线路的能力.从网络系综的角度出发,统计物理的方法可以应用到网络演化过程中过渡态及停留时间等问题的定量讨论中去^[60-61].

5 结 论

近10年的大量研究成果表明,统计物理已经成为复杂系统研究的重要理论依据和方法来源,展示了它的普适性和实用性.在海量数据的挖掘和整理方面,简化统计物理模型的提出和运用在数据的深度分析中发挥了指导性的作用,为有效地寻找和准确地表述数据背后的规律提供了有力的支撑.量化描述和关系的引入,一方面大大丰富了人文科学的研究内容,由定性的甚至是哲理层次的讨论扩展为定量的分析,另一方面也迫使学者们重新审视现有的认知体系,建立与信息爆炸相适应的理论.实际上,一百多年来,统计物理学积累了大量处理多自由度系统的模型和方法,这一知识宝库可成为复杂系统未来发展的取之不尽的资源.

与此同时,交叉学科研究中涌现的问题常常在时间和空间上呈现多尺度和多层次,结构和动力学具有非线性、有限尺度、复杂关联、小概率事件等特征.与经典统计物理学对均匀系统物质结构和宏观物理态的所取得的成就相比,如何有效地运用统计物理学的基本思想和基本方法来处理复杂系统中的各类现象尚有很多问题需要解决.例如,在生物界或现实生活中人们所关心的功能化复杂系统往往属于一般复杂系统中的特例,从非功能化系统研究中总结出的统计规律通常不适用于此类系统.蛋白质折叠的统计物理研究历程就提供了一个很好的例子.多肽链的折叠需在特定的生理环境下完成,且折叠时间也必须控制在秒的数量级,因此相应的氨基酸

序列必须具有某类特性以满足功能上的要求.以此类推,在更大的尺度上如细胞内部各类生化通道和分子过程的研究中,物理理论的建立需要解决以下几个问题:a.定义目标函数;b.建立物理化学约束及有关动力学;c.优化资源管理以提高效率和竞争力;d.在基本动力学基础上,嵌入与环境变化相关的调节机制.显然,这些问题的解决不仅需要有关学科知识的积累,也需要多层次上相应的统计物理学工具包括理论分析与计算方法的发展.在目前的认知水平下,一条可行的研究策略是从简单到复杂,从具体系统出发逐步寻找功能化系统的一般规律.

综上所述,复杂系统的研究,亟待以统计物理为基础的一套坚实的理论支撑,而这又有待于解决统计物理自身面临的若干问题,包括非平衡态统计问题、有限系统的分析方法、反问题的分析方法等.我国已经形成了有相当数量和质量的学者,在相关方向上做出了有一定国际影响力的工作.研究的进一步深入,需要更多的具有良好统计物理学知识的年轻学者的加入,同时紧密地结合相应学科的实验研究,用最新的数据启发和推动理论工作的开展.在统计物理理论的发展上,我们建议关注以下三方面的工作:a.发展统计物理基础理论以处理非均匀系统,特别是处于非平衡态的受驱系统中的涨落现象;b.深入研究某些具体但具有代表性的复杂系统,逐步建立基于统计物理思想和概念的理论和方法,譬如信息挖掘和其他反问题的统计系综理论;c.发现和积累多尺度、多类别的复杂系统的新现象和新规律,加强对功能化模块的分析和讨论.实际复杂系统多层次时空结构的特性,要求研究者不仅对某一层次现象的分析和讨论做得深入,对不同层次的衔接也有充分的认识,这样才可以较准确地把握研究方向,最终解决人们所关心的问题.通过领域同行的共同努力和国家层面上全方位的学科发展和人才培养计划的实施,我们希望看到在今后的5~10年内,统计物理与复杂系统的这一多学科交叉的基础领域在我国能够不断发扬光大,研究队伍在数量和质量上都有较大的提高,领域的学者们在具有重大理论和应用价值的问题的研究上活跃于国际学术舞台并取得瞩目的成绩,同时为解决与现代社会人类生活与社会发展息息相关的若干重大问题做出应有的贡献.

参考文献:

- [1] 郝柏林,于渌.统计物理学进展[M].北京:科学出版社,1981:411-418.

- [2] 霍裕平,郑久仁. 非平衡态统计理论[M]. 北京: 科学出版社, 1987.
- [3] 陈式刚. 非平衡统计力学[M]. 北京: 科学出版社, 2010.
- [4] ABE S, RAJAGOPAL A K. Revisiting disorder and tsallis statistics[J]. *Science*, 2003, 300(5617): 249 – 251.
- [5] MARCONI U, PUGLISI A, RONDONI L, et al. Fluctuation-dissipation: Response theory in statistical physics[J]. *Physical Report*, 2008, 461(4 – 6): 111 – 195.
- [6] JARZYNSKI C. Equalities and inequalities; irreversibility and the second law of thermodynamics at the nanoscale[J]. *Annual Review of Condensed Matter Physics*, 2011, 2: 329 – 351.
- [7] BUSTAMANTE C, LIPHARDT J, RITORT F. The non-equilibrium thermodynamics of small systems[J]. *Physics Today*, 2005, 58(7): 43 – 48.
- [8] LIPHARDT J, DUMONT S, SMITH S B, et al. Equilibrium information from nonequilibrium measurements in an experimental test of jarzynski's equality[J]. *Science*, 2002, 296(5574): 1832 – 1835.
- [9] COLLIN D, RITORT F, JARZYNSKI C, et al. Verification of the crooks fluctuation theorem and recovery of RNA folding free energies[J]. *Nature*, 2005, 437(7056): 231 – 234.
- [10] LIU F, OU – YANG Z C. Generalized integral fluctuation theorem for diffusion processes[J]. *Physical Review E*, 2009, 79(6): 060107.
- [11] QUAN H T, YANG S, SUN C P. Microscopic work distribution of small systems in quantum isothermal processes and the minimal work principle[J]. *Physical Review E*, 2008, 78(2): 021116.
- [12] LEPRI S, LIVI R, POLITI A. On the universality of anomalous one-dimensional heat conductivity[J]. *Physical Review E*, 2003, 68(6): 067102.
- [13] DHAR A. Heat transport in low-dimensional systems[J]. *Advances in Physics*, 2008, 57(5): 457.
- [14] MAI T, DHAR A, NARAYAN O. Equilibration and universal heat conduction in Fermi-Pasta-Ulam chains[J]. *Physical Review Letters*, 2007, 98(18): 184301.
- [15] DELFINI L, LEPRI S, LIVI R, et al. Nonequilibrium invariant measure under heat flow[J]. *Physical Review Letters*, 2008, 101(12): 120604.
- [16] ZHAO H, WEN Z, ZHANG Y, et al. Dynamics of solitary wave scattering in the Fermi-Pasta-Ulam model[J]. *Physical Review Letters*, 2005, 94(2): 025507.
- [17] ZHAO H. Identifying diffusion processes in one-dimensional lattices at thermal equilibrium[J]. *Physical Review Letters*, 2006, 96(14): 140602.
- [18] LI N, LI B, FLACH S. Energy carriers in the fermi-pasta-ulam β lattice: solitons or phonons? [J]. *Physical Review Letters*, 2010, 105(5): 054102.
- [19] CHANG C W, THOMPSON C V, GAN C L, et al. Effects of microvoids on the linewidth dependence of electromigration failure of dual-damascene copper interconnects[J]. *Applied Physics Letters*, 2007, 90(19): 193505.
- [20] CAI W, MOORE A L, ZHU Y, et al. Thermal transport in suspended and supported monolayer graphene grown by chemical vapor deposition[J]. *Nano Letters*, 2010, 10(5): 1645 – 1651.
- [21] WANG L, HE D H, HU B. Heat conduction in a three-dimensional momentum-conserving anharmonic lattice[J]. *Physical Review Letters*, 2010, 105(16): 160601.
- [22] CHANG C W, OKAWA D, MA J, et al. Solid-state thermal rectifier[J]. *Science*, 2006, 314(5802): 1121 – 1124.
- [23] LI B, WANG L, CASATI G. Thermal diode: Rectification of heat flux[J]. *Physical Review Letters*, 2004, 93(18): 184301.
- [24] DESHPANDE V V, CHIU H Y, POSTMA H W C, et al. Carbon Nanotube Linear Bearing Nanoswitches[J]. *Nano Letters*, 2006, 6(6): 1092 – 1095.
- [25] BARREIRO A, RURALI R, HERNÁNDEZ E R, et al. Subnanometer motion of cargoes driven by thermal gradients along carbon nanotubes[J]. *Science*, 2008, 320(5877): 775 – 778.
- [26] SQUIRES T M, QUAKE S R. Microfluidics: fluid physics at the nanoliter scale[J]. *Reviews of Modern Physics*, 2005, 77(3): 977 – 1026.
- [27] BARABÁSI A-L. The origin of bursts and heavy tails in human dynamics[J]. *Nature*, 2005, 435(7039): 207 – 211.
- [28] BROCKMANN D, HUFNAGEL L, GEISEL T. The scaling laws of human travel[J]. *Nature*, 2006, 439(7075): 462 – 465.
- [29] PASTOR-SATORRAS R, VESPIGNANI A. Epidemic Spreading in scale-free networks[J]. *Physical Review Letters*, 2001, 86(14): 3200 – 3203.
- [30] COLIZZA V, PASTOR-SATORRAS R, AND VESPIGNANI A. Reaction-diffusion processes and meta-population models in heterogeneous networks[J]. *Nature Physics*, 2007, 3(4): 276 – 282.
- [31] TANG M, LIU Z, LI B. Epidemic spreading by objective traveling[J]. *Europhysical Letters*, 2009, 87(1): 18005.
- [32] BOUCHAUD J P. Economics needs a scientific revolution[J]. *Nature*, 2008, 455(7217): 1181.
- [33] LUX T, WESTERHOFF F. Economics crisis[J]. *Nature Physics*, 2009, 5(1): 2 – 3.
- [34] SCHWEITZER F, FAGIOLO G, SORNETTE D, et al.

- Economic networks: the new challenges[J]. *Science*, 2009, 325(5939): 422 – 425.
- [35] FARMER J D, FOLEY D. The economy needs agent-based modeling [J]. *Nature*, 2009, 460 (7255): 685 – 686.
- [36] 刘建国, 周涛, 汪秉宏. 个性化推荐系统的研究进展[J]. *自然科学进展*, 2009, 19(1): 1 – 15.
- [37] 许海玲, 吴潇, 李晓东, 等. 互联网推荐系统比较研究[J]. *软件学报*, 2009, 20(2): 350 – 362.
- [38] ADOMAVICIUS G, TUZHILIN A. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions [J]. *IEEE Transactions on Knowledge and Data Engineering*, 2005, 17(6): 734 – 749.
- [39] LÜL, ZHOU T. Link prediction in complex networks: a survey[J]. *Physica A*, 2010, 390(6): 1150 – 1170.
- [40] ZHANG Y C, BLATTNER M, YU Y K. Heat conduction process on community networks as a recommendation model [J]. *Physical Review Letters*, 2007, 99 (15): 154301.
- [41] ZHOU T, REN J, MEDO M. Bipartite network projection and personal recommendation[J]. *Physical Review E*, 2007, 76(2): 046115.
- [42] ZHOU T, JIANG L L, SU R Q, et al. Effect of initial configuration on network-based recommendation[J]. *Europhysics Letters*, 2008, 81(5): 58004.
- [43] ZHOU T, SU R Q, JIANG L L, et al. Accurate and diverse recommendations via eliminating redundant correlations [J]. *New Journal of Physics*. 2009, 11 (12): 123008.
- [44] ZHOU T, KUSCSIK Z, LIU J G, et al. Solving the apparent diversity-accuracy dilemma of recommender systems[J]. *Proceedings of the National Academy Science of USA*, 2010, 107(10): 4511 – 4515.
- [45] ZHANG Z K, ZHOU T, ZHANG Y C. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs [J]. *Physica A*, 2010, 389 (1): 179 – 186.
- [46] SHANG M S, LÜ L, ZHANG Y C, ZHOU T. Empirical analysis of web-based user-object bipartite networks. *Europhysics Letters*, 2010, 90(4): 48006.
- [47] REED J L, VO T D, SCHILLING C H, et al. An expanded genome-scale model of *Escherichia coli* K – 12 [J]. *Genome Biology*, 2003, 4(9): R54.
- [48] FEIST A M, HERRGARD M J, THIELE I, et al. Reconstruction of biochemical networks in microorganisms [J]. *Nature Reviews Microbiology*, 2009, 7 (2): 129 – 143.
- [49] KELLIS M, PATTERSON N, ENDRIZZI M, et al. Sequencing and comparison of yeast species to identify genes and regulatory elements [J]. *Nature*, 2003, 423 (6937): 241 – 254.
- [50] GIANCHANDANI E P, JOYCE A R, PALSSON B O, et al. Functional states of the genome-scale *Escherichia coli* transcriptional regulatory system [J]. *PLoS Computational Biology*, 2009, 5(6): e1000403.
- [51] LEWIS N E, HIXSON K K, CONRAD T M, et al. Omic data from evolved *E. coli* are consistent with computed optimal growth from genome-scale models [J]. *Molecular Systems Biology*, 2010, 6: 390.
- [52] CHIN C S, CHUBUKOV V, JOLLY E R, et al. Dynamics and design principles of a basic regulatory architecture controlling metabolic pathways [J]. *PLoS Biology*, 2008, 6(6): e146.
- [53] LI F, LONG T, LU Y, et al. The yeast cell-cycle network is robustly designed [J]. *Proceedings of the National Academy Science of USA*, 2004, 101 (14): 4781 – 4786.
- [54] COHEN A A, GEVA-ZATORSKY N, EDEN E, et al. Dynamic proteomics of individual cancer cells in response to a drug [J]. *Science*, 2008, 322(5907): 1511 – 1516.
- [55] ERNST J, VAINAS O, HARBISON C T, et al. Reconstructing dynamic regulatory maps [J]. *Molecular Systems Biology*, 2007, 3(1): 74.
- [56] ALON U. Network motifs: theory and experimental approaches [J]. *Nature Reviews Genetics*, 2007, 8 (6): 450 – 460.
- [57] TYSON J J, NOVAK B. Functional motifs in biochemical reaction networks [J]. *Annual Review of Physical Chemistry*, 2010, 61: 219 – 240.
- [58] LI H, JOHNSON A D. Evolution of transcription networks-lessons from yeasts [J]. *Current Biology*, 2010, 20(17): R746 – R753.
- [59] MA W Z, TRUSINA A, EL-SAMAD H, et al. Defining network topologies that can achieve biochemical adaptation [J]. *Cell*, 2009, 138(4): 760 – 773.
- [60] DESAI M M, FISHER D S. Beneficial mutation-selection balance and the effect of linkage on positive selection [J]. *Genetics*, 2007, 176(3): 1759 – 1798.
- [61] VISSER J A G M, PARK S C, KRUG J. Exploring the effect of sex on empirical fitness landscapes [J]. *American Naturalist*, 2009, 174(1): S15 – S30.