

文章编号:1007-6735(2012)02-0103-15

人类行为、复杂网络及信息挖掘的统计物理研究

汪秉宏^{1,2}, 周涛³, 周昌松^{4,5}

(1. 中国科学技术大学 近代物理系, 合肥 230026; 2. 上海理工大学 复杂系统科学研究中心, 上海 200093;
3. 电子科技大学 互联网科学中心, 成都 610054; 4. 香港浸会大学 物理系, 香港; 5. 香港浸会大学 非线性研究中心, 香港)

摘要: 通过实证统计与理论模型分析相结合对复杂系统进行研究是一种全新的认识和探索. 建议从人类行为的统计特性、复杂网络同步与复杂神经网络、信息挖掘与复杂网络链路预测 3 个方面, 基于大量的实证统计和分析, 结合有效的动力学模型, 针对人类自身行为的规律特性、社会个体之间的相互作用、神经系统的动力学演化、信息的有效推荐和网络演化的有效预测等重要问题, 运用统计物理理论进行全方位的探索, 深入挖掘各种决定复杂系统演化过程的基本机制与规律.

关键词: 人类行为动力学; 复杂网络; 信息物理

中图分类号: N 94 **文献标志码:** A

Statistical Physics Research for Human Behaviors, Complex Networks, and Information Mining

WANG Bing-hong^{1,2}, ZHOU Tao³, ZHOU Chang-song^{4,5}

(1. Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China;
2. The Research Center for Complex Systems Science, University of Shanghai for Science and Technology, Shanghai 200093, China;
3. Web Sciences Center, University of Electro Science and Technology of China, Chengdu 610054, China;
4. Department of Physics, Hong Kong Baptist University, Hongkong, China;
5. Center for Nonlinear Study, Hong Kong Baptist University, Hongkong, China)

Abstract: Integrating empirical statistics and theoretical models is a novel and promising way to study complex systems. This survey summarized recent progress on human dynamics, complex networks and information filtering, suggesting that to apply the perspectives and methods from statistical physics based on extensive empirical data and build effective dynamical models may solve some long-standing challenges, such as uncovering the hidden regularities of human behavior, revealing the rules governing the interactions between social individuals, characterizing the dynamical evolution of neural systems, digging out personalized tastes, predicting missing information, and so on.

Key words: human behavior dynamics; complex network; information physics

收稿日期: 2012-01-18

基金项目: 国家自然科学基金资助项目(10975126, 91024026); 安徽省高校自然科学基金研究重大项目基金资助项目(KJ2011ZD07); 高校博士点基金资助项目(20093402110032)

作者简介: 汪秉宏(1944-), 男, 教授. 研究方向: 统计物理、非线性科学、复杂系统理论. E-mail: bhwang@ustc.edu.cn

在自然界、工程、经济和社会的各个领域都有大量复杂体系. 十二五提出的可持续和谐发展的长远规划对复杂体系, 尤其是对与人相关的复杂系统的研究提出了前所未有的要求. 特别是近十几年来, 科学研究的一个巨大变化, 就是越来越多地对复杂体系进行系统层次的综合统计物理学分析. 这一趋势的形成由科学研究在几个方面长期积累的合力引起, 包括: a. 复杂性研究的进展. 几十年以来, 非线性和复杂性统计物理研究通过模型的理论研究发展了一套概念和方法来分析耦合体系中复杂的涌现行为, 比如人和动物的集群行为. 这些行为只出现在系统层次, 为分离的系统单元所不具有. 这些理论进展导致了科学概念的革命性变化, 出现了新兴的学科方向, 如系统生物学等. 特别是过去 10 多年来统计物理从传统研究相对独立的多粒子体中标度与普适现象转移到研究大规模复杂体系中各单元或子系统之间相互作用的复杂结构关系, 形成了一个跨学科的、影响至深的复杂网络领域. b. 海量数据采集和处理技术的发展. 计算机技术的进步使得大规模的数据采集取得突飞猛进的变化, 提供了大量研究复杂体系结构与演化的实证数据, 如计算机网络、万维网、交通运输网等. 大量有关人类个体和集体行为的数据也为定量研究复杂社会体系提供了前所未有的契机.

目前正处于一个统计物理大革新的时代, 许多史无前例的、事关社会经济发展的复杂体系正迫切等待发展新的统计物理分析手段和理论模型来对它们进行全面深入的认识. 复杂体系的系统层次统计物理研究虽然已成潮流, 然而它却面临巨大的挑战. 最大的问题是理论模型及计算研究与大规模真实数据之间还存在一个巨大的鸿沟. 复杂性统计理论模型往往没有考虑到许多真实体系, 特别是与人个体或群体有关的生命和社会体系是变化过程中经过选择优化的结果, 对它们的多重约束在这样一个优化问题中没有得到足够的重视和认识. 其次, 作为开放系统, 它与环境交换能量和物质的边界条件对优化结果的选择起到至关重要的作用. 这些是研究真实复杂体系的真正挑战所在. 当模型的构建和分析没有足够考虑这些重要方面时, 理论结果往往与真实系统很难相符, 甚至于不相关. 对这些内在约束和外边界条件的了解有赖于对具体的复杂体系进行细致深入的实验实证研究.

虽然技术的进步为在系统层次观测和收集数据提供了前所未有的机会, 大规模的数据积累也非常

之快. 然而, 有组织、有针对性地从事实验实证数据探索系统的内外约束条件还处于初步阶段. 因而, 如何分析应用这些数据来全面理解真实系统的行为却进展甚微. 关键的问题是缺乏有效的理论框架, 所获得的数据在数量上可能很大, 却不一定有很强的针对性, 信息比较散杂. 数据中大量貌似杂乱可往往包含有用的系统组织信息, 由于缺乏有效的理论指导, 在数据挖掘中很可能当作噪声被丢弃了. 所得的系统信息不易串联在一起形成有效的关于体系在受到多重约束和边界条件下的组织图像. 如在认知神经科学领域, 大脑自发的复杂活动通常作为噪声处理, 大大限制了对这样一个动态复杂体系结构和功能的认识. 经济、社会领域的数据积累也往往缺乏系统的理论指导.

由于对这些根本性的挑战还缺乏比较全面深入的认识, 复杂体系的统计物理研究还处于初步阶段. 我国有不少学者在复杂性一般模型和理论方面开展了比较活跃的研究, 但还缺乏有意识、有组织和有规模的在理论和数据有机结合上的系统化发展. 因此, 很有必要基于当前国内研究团队的分布, 组织国内的科研合作力量, 把理论研究和实证分析紧密结合起来, 缩小理论及计算研究与大模型实验实证数据之间的鸿沟, 从而在根本上推动我国在复杂系统统计物理这一新兴领域的进步, 在国际学术竞争中占一席之地, 甚至处于领军位置.

下面将对人类行为、复杂网络、信息挖掘等 3 种复杂系统的统计物理研究最新进展和主要问题分别进行详细阐述和分析.

1 人类行为动力学

人类行为具有高度的复杂性. 研究人类行为中的规律, 对于经济学、社会学、管理学的研究和应用有着极为重要的价值. 长期以来, 对人类社会行为的研究主要为心理学所关注, 通过心理学实验的方法, 研究人类在各种环境下的心理反应是其主要的研究手段.

最近数十年来, 人们在复杂系统研究领域内取得了令人瞩目的成就. 复杂系统的研究具有着天然的综合性和交叉性, 它所涉及的问题几乎遍及人们所研究的绝大多数领域, 不但包括物理学、生物学等自然科学学科, 也包括了经济学、社会学等社会科学学科. 随着对复杂系统研究的不断深入, 该领域的研究对各种社会学科的渗透越来越强烈, 其理论影响和实际应用也越来越广泛. 近年来, 对人类行为的统

计研究已经成为复杂系统中的一个重要议题。

不同于传统心理学实验方法,复杂系统研究者对于人类行为的研究主要通过统计物理方法.通过对大量人类行为事件进行定量统计,研究其中所隐藏的统计性规律,并根据所研究的问题,提出基本假设,建立理论模型,来探索这些规律的产生机制和可能的动力学影响.基于这样的研究方法,近年来人们发现了人类行为中所存在的大量特殊现象和规律,这些发现引发了人们更进一步地探索人类社会行为的热潮.特别是从2005年以来,仅在《Nature》、《Science》、《PNAS》、《PRL》等强影响因子期刊上就已经发表了30余篇文章.

1.1 人类行为时间统计特性的实证研究

以往一些对社会、经济系统的研究中,常常把单个人的行为简化为可以使用泊松过程描述的稳态随机过程.这种假设必然导致的推论是人的行为的时间统计特征应该是较为均匀的,两个相继行为之间存在极大时间间隔的概率很小.但是,自2005年以来,通过对电子邮件发送与回复、邮件通信等人类行为的时间间隔的实际统计,人们发现这些行为存在与上述假设极为不同的特性^[1-2]:长时间静默与短期内高频率的爆发同时呈现在这些人类行为中,其时间间隔分布存在满足反比幂函数的胖尾,也就是说,这些行为的发生过程是不能用泊松过程描述的.这一出人意料的研究结论提示人们,人类的个体行为可能存在复杂的动力学机制,而随之而来的一个重要问题是这种非泊松特性在人类行为中是不是普遍存在,人们对这一问题进行了极为广泛的研究.

通过各种不同的数据收集方法,人们的研究涉及市场交易^[3-6]、网站浏览^[7-8]、电影点播^[9]、欣赏网络音乐^[10]、手机通讯^[11]、在游戏及虚拟社区中的行为^[12-13]、计算机指令的使用行为^[14]等,包含了商业行为、娱乐行为、日常使用习惯等众多的人类行为.在这些行为中,普遍发现有类似的偏离泊松过程的特性.这些现象显示出,除了受到生理周期强烈影响的部分行为外,时间间隔统计所显示的非泊松特性可能是在人类行为中普遍存在的.

除了时间间隔分布,部分人类行为事件前后时间间隔的相关性也得到了研究者的注意.研究发现,这些人类行为相邻时间间隔的相关性并不明显,而其它同样存在爆发性和长期静默性的自然现象(如地震等)却常常存在正的相关性^[15].这一项研究初步把人类行为和其它复杂系统中的行为特性进行了比较,暗示可能存在统一的深层机制.

1.2 人类行为动力学模型研究

上述统计特性说明人类的众多行为不能使用泊松过程来描述,那么一个重要的问题是这种胖尾分布行为特征的来源是什么.目前一种重要的解释是基于任务队列理论的^[1,16-18],它把人的各种日常行为视作处理一系列的任务,并根据日常生活经验假设对这些待处理任务进行优先级划分.首先处理高优先级者,指出这种具有优先权的行为模式是造成胖尾分布的重要原因.这种基于任务队列的理论模型可以合理地解释很多人类行为中的非泊松特性,例如电子邮件和水陆邮件的发送等,而且可以相当容易地推广到存在多个个体之间交互的情况^[19],在解释人类行为时间统计胖尾分布特征方面取得了很大的成功.

此外,由于人类行为的复杂性,影响人类行为的因素是多种多样的,所以,有部分研究从不同于任务队列的方面出发,提出了多种非排队论模型.例如,有的工作考虑了人类行为中的记忆效应^[20],有的研究了行为的周期性和季节性对非泊松机制的影响^[21],近期的一种理论从多重泊松分布的角度解释了人类的行为特性^[22].

最后,国际上有少数工作研究了人类行为的非泊松特性对网络传播、通讯等动力学过程的影响.例如,发现相比于一般的泊松特性,这种非泊松特性可以给系统带来一些特殊性质,如更快的传播速度等^[23].由于这一领域发展时间很短,在这一问题上还有海量的工作等待开展.

1.3 人类行为空间统计特性的研究

除了发现人类行为的时间间隔分布中广泛存在有非泊松特性,最近也发现在人类行为的空间分布中也存在有非泊松特性等复杂现象.2006年,通过统计帐单传递^[24],人们间接地发现了人类的旅行行程分布存在接近于幂律的胖尾.2008年,Gonzalez等^[25]通过统计移动电话用户在不同基站区域的漫游过程,更进一步地研究了人的旅行行程分布,同样发现该分布具有无标度特性,与早期的结果基本一致.更为直接的基于GPS数据的统计结论^[26]也支持人类行程分布中存在无标度特性.此外,在生物学观测方面也发现大量的动物物种的运动具有类似幂律形式的行程分布^[27-29].由于这种幂律形式的行程分布存在较高频率的远程运动,它无法通过经典的随机行走进行描述.这种行程分布的广泛性,使得人们需要去思考它背后的动力学机制是什么.虽然对于动物行为中的幂律行程分布已经提出了觅食效率优

化^[30-31]、嗅觉梯度机制^[32]、确定性行走^[33]等,目前对人类这种行程分布模式产生机制的解释方面的研究仍然是空白.另一方面,由于这类非泊松特性常常会使得系统出现若干特殊性质,那么这种人类行为的分布上的非泊松特性同样可能影响到城市交通、人流控制、紧急避险等系统的运作,可能会使其带有若干特殊性质,这些问题目前尚未被研究,也值得研究者的注意.

1.4 人类行为动力学对传播的影响

人类行为的特性,不仅帮助人们更好地了解自己的行为特征,进一步挖掘这些统计数据背后所隐藏的人类特性,还关系到了人们对于多个方向的模拟和理解,其中讨论最多的,应属人类动力学对于疾病在人群中传播的影响.在这里主要从时间和空间两个方向来介绍人类行为对于传播速度、波及范围、预防策略等的影响.

经典疾病传播模型都基于一些有悖于人类真实行为的假设:a.人活动的时间间隔相同,即所有人都均匀地在每个时间步活动一次,这里的“活动”,是指传播或康复的行为;b.所有人活动的频率在群体中是无差别的,即每个个体的活动密度一样.然而,参考前文人类时间间隔实证统计发现,这些假设与人类行为具有的阵发性、记忆性和活跃性有着很大的出入.

为关注人类活动的阵发性对于病毒传播的影响,Vázquez等用人们查收E-mail的两组数据,分别为3188个用户之间发送的129135封邮件,1729165个用户之间互发的39046030封邮件,在实证数据的网络上,根据真实时间间隔,模拟了网络上计算机病毒的传播情况^[23].以天和小时为单位的统计结果,都表明人类行为的时间间隔满足幂律分布,会极大程度上减慢病毒的传播.人类活动的阵发性对于病毒传播具有明显的延迟作用.

文献^[34]也运用SIR模型通过模拟,讨论了时间分布的阵发性对于传播过程的影响,设置了存在等待时间的模型,并得到结论:时间的异质性越强,病毒存活概率越小.

在人类活动时间间隔满足阵发性的基础上,为探讨时间序列的记忆性对传播过程的影响,Karsai等基于电话网络的实证数据,时间跨度为9个月,在规模为 $N = 4.6 \times 10^6$ 的网络上模拟传播^[35].通过对比网络是否含权重、网络是否有向、通话时间分布是否有记忆性,模拟结果表明,人类行为时间和空间上的异质性会一定程度地减慢传播过程.

时间的异质性又体现在个体与群体两个层面

上,也有工作进一步比较了两者对于疾病传播速度的影响^[36].所谓群体层面的时间异质性,是指每个人活动的时间序列间隔平均,而人与人之间的频率有较大差别,满足幂律分布.但是个体层面的时间异质性,表现为人与人之间活动的频率相同,而单人活动的时间间隔满足幂律分布.群体层面上时间的异质性对传播速度的影响非常大,相比之下,个体层面上的对传播影响很小.

为了揭示人类的行程分布对于传播过程的影响,Ni等利用连接概率与度成正比,与欧氏距离成反比的机制,构造网络,并在网络上模拟传播过程^[37].研究发现,行程分布的几何特征越鲜明,即人们更倾向于去到距离自己欧氏距离比较近的地方,病毒传播所波及的范围就越小,持续时间也就越短.

相比于个体形成分布的活动特征,更多人关注宏观意义上人类的长程旅行对于病毒在城市间扩散的影响.早在2004年,Hufnagel等就在美国航空网络上,假设人口密度随机分布,建立了SARS病毒传播的模型^[38].该模型考虑了局部病毒传播和城市间由于飞行造成病毒传播两种因素,模拟出的结果与实证很好地吻合,并提出了有效预防及抑制疾病传播的策略.针对以上模拟及解析结果,Hufnagel等还提出了相应的预防策略.通过模拟,比较了减少局部地区个体接触和减少长程出行两种预防策略.得出孤立城市,即减少城市间的长程旅行可以更有效地防治疾病扩散,并给出了集中早期进行疫苗注射的显著效果^[38].此后,2006年Colizza等专门就人类长程出行的拓扑结构对于病毒扩散造成的影响进行了研究^[39].为了进一步探究网络的拓扑结构对于疾病人数分布的影响,还提出了病毒分布熵的概念,来刻画病毒流行的地区异质性.通过与实际网络病毒传播情况的对比,发现以前很多刻画人类行程拓扑结构的模型在细节上还需改进.

1.5 国内相关研究进展

该领域兴起也引起了国内研究者的注意.目前,中国科学技术大学复杂系统课题组、上海理工大学管理学院、上海交通大学自动化系等都已经有了相关的研究论文在国内外学术期刊上发表.这些工作可以简述如下.

在实证方面,中国科学技术大学复杂系统课题组的周涛等与韩国成均馆大学及瑞典皇家学院合作研究了电影点播中的人类行为模式以及与个体活动性之间的关系^[9,40];洪伟等研究了人类短消息通讯中的时间间隔分布^[41],发现了多种无标度特性;上

海理工大学课题组的张宁、李楠楠和周涛合作分析了鲁迅、钱学森等名人的邮件通讯数据^[42-43];上海交通大学的胡海波等人研究了网络在线音乐的收听行为^[10].在理论模型方面,中国科学技术大学的韩筱璞等提出可自适应调节的兴趣机制来解释人类行为的非泊松特性^[44-45].此外,上海理工大学方面还发表了针对人类动力学的中文综述^[46];上海理工大学的郭进利等和中国科学技术大学的周涛等合作编写出版了专著《人类行为动力学模型》^[47];中国科学技术大学的周涛、韩筱璞、汪秉宏也在世界科学出版社出版的专著《Science Matters: Humanities as Complex Systems》中撰写了关于人类动力学研究的一个专门章节^[48].

1.6 面临的主要问题

由于该领域的发展时间短暂,目前仍存在大量问题有待于深入研究.

a. 已有的实证统计主要针对个体行为,但仍然存在大量的个体行为特性未被研究,已有的研究结果难以根据统计特性区分个体行为的主要类别,而针对团体行为的实证研究更几乎是空白.事实上,人类的行为常常受到社会关系的影响,这方面定量的实证研究仍然非常欠缺.另外,一些最近发展的理论,例如人类动力学的普适类假说,受到了新的实证数据的挑战,更清晰和令人信服的图景需要更多和更深入的实证分析.

b. 除了人类的个体行为,目前所做的一些最新统计也发现,一些社会团体的宏观行为也具有类似的非泊松特性,如国家之间战争的时间间隔分布等.由于目前的实证统计有限,对于社会团体而言,这些特性在多大范围内存在,是否与人类个体行为具有相似的生成机制,都仍然是未知问题,需要进行深入的研究.

c. 在研究人类行为的空间分布方面,目前的实证数据都是根据帐单、手机漫游等数据间接获得,缺少对人类行为空间分布的直接观察,而其产生机制和动力学效应方面的研究目前几乎没有.

d. 目前的理论模型研究,虽然已经提出了多种唯相机制来解释人类行为中的非泊松特性,但是这些机制难以覆盖全部人类行为中的非泊松特性现象,需要提出新的更具有普适性的模型.

e. 人类行为特性对各种社会系统动力学效应的影响研究,尽管已经出现了少数成果,但因涉及问题众多,导致许多研究空白,大量工作需要深入进行.例如人类行为的空间分布特性是如何影响城市交通等.

2 复杂网络动力学——同步与神经动力学

2.1 复杂网络:科学与技术的新前沿

过去的10多年,人们见证了由一个影响深远的交叉学科“复杂网络”的出现带来的复杂系统研究上的重大革新(见文献^[49-50]).在很多真实世界中的复杂系统中,系统基本元素之间相互作用构成了复杂的拓扑连接^[51-53].来自不同学科领域的这些复杂网络既不是规则连接也不是随机连接,而是具备两个共同的特性.这两个共同的特性分别是由较短路径长度所描述的小世界特性^[54]和具有很大连接(度)的中心节点的无标度特点^[55].

复杂网络领域的主要方向是图理论分析工具的发展以及将这些工具应用于刻画不同领域里的复杂系统^[49-50].在复杂网络领域的研究方法中,复杂系统的基本元素是由节点来表示,元素之间的相互作用是由边来表示,而元素和相互作用的一些细节特点通常被忽略.

在对复杂系统的研究中,上述过度简化的研究方式限制了方法的效能.一个很大的挑战是,对整个网络的全局统计量(如平均路径长度、簇系数等)进行测量时,网络的结构并不能够真正决定系统的行为.例如,由脑皮层区域之间长程连接所构成的大脑网络的小世界和无标度特点^[56-57]确实能对大脑的功能表现提供一些视角,但是要想对大脑功能作出更重要的深入理解需要基于对神经元和复杂网络结构的重要特性综合之后对大脑的动力学进行深入细致的分析.

将网络的结构和复杂系统行为联系起来的重要一步是对在网络上发生的动力学过程进行研究^[50,58].在很多复杂系统的动力学研究中,特别是对于神经系统^[59-60],复杂网络振子同步作为共同行为自组织现象的重要机制已经成为深入探索的一个重要课题(见文献^[61]).

2.2 复杂网络同步研究的成就及重大挑战

在振子同步性问题上,研究的焦点集中考虑网络拓扑结构方面,特别是小世界和无标度特性的影响上^[61-71].在整个网络的全局同步方面,很多工作要么考虑全同振子实现的完全同步或者是非同振子由于锁相实现的协作振动^[61].之前已经有很多研究主要使用主稳定性方程来探索完全同步状态的稳定性,而在这个方法中,根本不考虑振子的特殊属性,而是将网络同步性与网络的特征谱挂钩^[65-70].

文献主体部分是通过调整权重和耦合强度的拓扑结构进行调整来加强振子的全局同步性^[61,67-71]。

虽然完全同步非常方便进行稳定性分析,但是它不是真实的复杂系统中最自然的状态。相反,大尺度的强同步对应的是系统病态情况,如社会灾难、癫痫发作,真实系统是不希望出现这些状态的。真实系统特别是神经系统,为了实现正常功能需要不同层次的同步,使得系统通过分割成不同动力学模块在各自的子系统里实现特殊的功能。同时,这些专门的模块间相互作用有效地实现了分割处理的信息整合。

最近,网络研究中将网络分成不同拓扑模块的研究方法已经引起了广泛关注^[72-75]。除了小世界和无标度特性,模块化是真实网络另一个普适特征。例如在神经系统中,数十亿神经元被耦合形成不同层次的(从单神经元连接组成的神经柱,由不同神经柱连接再形成的功能区域)模块网络。通过模块之间不同层次上的合作同步,模块结构为系统实现功能的分类和整合提供了一个天然基础。虽然关于模块研究的最近一些工作仍然考虑完全同步态的稳定性问题^[76],但是主流趋势已经转移来研究系统在并未完全同步时动力学模块的形成,这个方法同时用于探测网络结构模块^[77-81]。

值得注意的是,很多之前著名的研究仍然将网络随时间变化的动力学(如网络演化^[49])和动力学行为在网络上的表现(如同步,信息传播^[50,58])分开考虑。然而在大量真实的复杂网络系统中,结构与动力学的相互影响是至关重要的。不只是结构决定了动力学斑图,反过来动力学也使得结构进行了与之相适应的调整,后者在众所周知的神经系统的学习机制中表现得尤为明显。很多真实系统自组织了网络结构,并且在两者共同演化中实现了对功能的优化。有些工作开始关注在适应性网络上对结构和功能的相互影响进行分析^[82-84]。

综上所述,在振子网络的理论分析和模型研究中的主要挑战一方面刻画复杂同步斑图的复杂性程度与网络不同层次结构之间的关联,另一方面是探索在各个层次里通过结构模块和动力学模块的相互影响实现两者的自组织。这个方向的探索刚刚起步,将长期影响对可以用振子及其同步来根本表现其功能的真实复杂系统,特别是神经系统的理解^[59-60]。

2.3 神经网络与神经动力学

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网络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和认知障碍。最近10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了解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蓬勃发展的新方向。

在过去几十年内,非线性与复杂性物理科学的各种理论和方法得到长足的发展,并将其应用于神经科学领域。特别是复杂网络方法的应用已经勾勒出一个关于大脑的新图景,使得可以从解剖学上的神经连接层面以及动力学,也就是大脑功能上的区域相互关联层面上来研究这个复杂的神经网络结构^[58,85-86]。在大脑的系统层面上,根据之前的研究^[26,87],由哺乳动物脑区间的长程连接所构成的脑皮层网络已经展示出小世界和无标度特性。这个发现的意义是巨大的,它暗示着每个脑区的活动都可能同时被其它脑区的活动影响。由测量到的不同脑区活动相关性所得到的大脑功能区之间相互作用的功能网络也展示了这个复杂的大脑斑图,这一点甚至当大脑处于静息态,也就是说大脑在没有外感觉输入,只有完全自发的自组织活动的状态下也同样存在^[88-90]。传统的认知神经科学在过去的10多年也开始转而研究大脑自发活动在功能上的影响^[91]。然而,认知科学方面的分析还是主要局限在考虑由少量脑区所组成的自下而上(前馈)或者自上而下的(反馈)机制^[60]。与只研究少量脑区相比,从复杂网络角度同时研究神经系统中各部分的相互作用将会带来关于动力学和功能的相互关系方面更多的信息,但是该如何解读这些信息又是需要考虑的问题。

大脑的复杂网络假设呼唤着新的方法来揭示出大脑大尺度的功能网络和认知过程的关联。如今基于网络的分析和测量工具不足以应付研究复杂的大尺度动力学相互作用斑图,并将它们与只在50~100 ms的短窗口里发生的各种认知过程相联系。现在这些复杂的大脑网络特征分析只是局限于讨论网络的全局统计量,如簇系数、平均路径长度等。而对于功能网络的分析也主要是考虑在长时间尺度下统计意义上的脑区活动相互作用斑图。普适网络理论对于结构和动力学关系的探索只是局限在一些理想的情况下,如复杂网络中振子同步问题已经被大量研究(见文献^[77-81])。但是,这些研究主要是集中在全局同步上^[61-64,92-94],也就是整个网络的完全同步。通过理论的稳定性分析可知,大尺度完

全同步并不允许信息的分割处理,这一点与神经系统中需要信息的分隔处理与整合的恰当平衡相抵触,因而与神经系统的疾病状态相对应,如癫痫发作^[92-94].

因此,对大脑复杂的连接结构、活动的统计分析和模拟研究的挑战是如何把复杂网络和复杂动力学系统的一般理论同真实神经系统的特异性有机结合,发展新的理论方法并把它应用到真实的数据分析和模型的建立上.这需要从事复杂性统计物理研究的学者与神经科学家之间的紧密合作.

3 信息挖掘的统计物理分析

随着因特网的迅猛发展,接入因特网的服务器数量^[95]和 World Wide Web 上网页^[96]的数目都呈现出快速增长的态势.用户可得信息量的激增使得人们的生活变得多元化,但与此同时,也带来了信息过载的问题.例如,Netflix 上有数万部电影,Amazon 上有数百万本书,Del.icio.us 上面有超过 10 亿的网页收藏.如此多的信息,别说找到自己感兴趣的部分,即使是浏览一遍标题也是不可能的.信息过载,简单来说,就是信息量的激增使得信息的利用率反而降低.高效准确的信息推荐技术,特别是针对不同用户不同喜好的个性化推荐技术,是解决信息过载问题最有前途的方案^[97].另外需要注意的是链路预测技术,这是一种不完整信息重构的有效手段.下面,将从信息推荐理论和算法,以及链路预测理论和算法两方面进行叙述.

3.1 关于信息推荐

个性化推荐系统,本质上讲是根据用户对相关产品的历史评价,代替用户评估他尚未接触的产品的一种工具^[98].这些产品包括书、电影、CD、网页,甚至可以是饭店、音乐、绘画等.个性化推荐系统作为一个独立的概念,在 20 世纪 90 年代已经被提出.由于 Web 2.0 技术的发展和成熟,用户可以方便地针对网上服务提供反馈信息,这些反馈信息可以反映用户的喜好,从而被网络服务提供商用来进行推荐.因此,最近几年,个性化推荐系统的研究得以迅猛发展.

虽然早在 1992 年开发的 Tapestry 系统就已经是真正意义上的信息推荐系统了,但推荐系统的概念直到 1997 年才由 Resnick 和 Varian 正式定义^[98],推荐技术也才逐步发展为一个独立的研究领域.《Communications of ACM》分别于 1992 年和 1997 年出版了两期推荐系统的专刊,国际著名期刊

《Journal of Information Technology and Tourism》、《ACM Transactions on Information System》、《ACM Transactions on Computer-Human Interaction》、《International Journal of Electronic Commerce》、《IEEE Intelligent Systems》、《AI Communications》等也分别于 2003 年、2004 年、2005 年、2007 年出版了推荐系统的专刊.除这些专刊外,多个顶级的国际会议,如 CHI、ACM SIGIR、ECAI、AAAI、ReColl 等都设立专门的推荐系统的工作组.由于推荐系统的重要地位,ACM SIGIR 设立了专门的推荐系统会议,其第一届会议于 2007 年在美国明尼苏达大学召开,第二届会议于 2008 年在瑞士洛桑理工大学召开.

一般而言,信息推荐系统通常包括 3 个组成要素^[97]:推荐候选对象、用户和推荐方法.其中,推荐方法是整个推荐系统中最核心、最关键的部分,在很大程度上决定了推荐系统的性能.目前,针对推荐方法的分类也有好几种,其中多数研究者将推荐技术分为 3 类:基于协同的推荐(协同过滤)、基于内容的推荐和混合推荐技术.

a. 协同过滤系统.协同过滤系统是最早被提出并得到广泛应用的推荐系统,其核心思想可以分为两部分^[99].首先,利用用户的历史信息计算用户之间的相似性;然后,利用与目标用户相似性较高的邻居对其它产品的评价来预测目标用户对特定产品的喜好程度.系统根据这一喜好程度来对目标用户进行推荐.协同过滤推荐系统最大的优点是对推荐对象没有特殊的要求,能处理音乐、电影等难以进行文本结构化表示的对象.Grundy 被认为是第一个投入应用的协同过滤系统^[100],该系统通过建立用户兴趣模型给用户推荐相关的书籍;Tapestry 邮件处理系统人工确定用户之间的相似性^[101];GroupLens 建立用户信息群,群内的用户可以发布自己的信息,依据社会信息过滤系统计算用户之间的相似性,进而向群内的其他用户进行协同推荐^[102];Ringo 利用社会信息过滤方法向用户推荐音乐^[103].其它利用协同过滤方法进行推荐的系统还有 Amazon 的书籍推荐系统^[104]等.虽然协同过滤推荐系统得到了广泛应用,但是也面临很多问题,如新用户或新产品推荐问题(冷启动问题)、打分稀疏性问题、算法可扩展性问题等.

b. 基于内容的推荐.基于内容的推荐不是依据用户对项目的评价意见,而是依据用户已经选择的产品内容信息来计算用户和产品之间的匹配度,进而进行相应的推荐.基于内容的推荐算法的根本之

处在于信息获取和信息过滤^[105],关键在于内容信息的获取和匹配.因为文本信息获取与过滤方面的研究较为成熟,多数基于内容的推荐系统都建立在对产品的文本分析上.在大多数的基于内容的推荐系统中,产品内容常常被描述成关键词.Fab系统^[106]就是一个典型的例子,它用一个网页中最重要的100个关键词来表征这个网页;而 Syskill & Webert 系统^[107]则用128个信息量最多的词表示一个文件,系统根据文本相似性推荐与用户过去喜欢的产品最为相似的产品^[106-107].基于内容的推荐系统中,用户的配置文件构建与更新是其中最为核心的部分之一,也是目前研究人员关注的焦点.例如 Somlo 和 Howe^[108]以及 Zhang 等^[109]提出了利用自适应过滤技术更新用户配置文件;Chang 等^[110]通过区分长期感兴趣与短期感兴趣的关键词,赋予短期感兴趣的关键词更高的权重,在此基础上建立新的关键词更新树,从而大大减少了更新配置文件的代价. Degemmis 等^[111]利用 WordNet 构建基于语义学的用户配置文件,配置文件通过机器学习和文本分类算法得到,里面包含了用户喜好的语义信息,而不仅仅是关键词集.基于内容的推荐系统不可避免地受到信息获取技术的约束,例如自动提取多媒体数据(图形、视频流、声音流等)的内容特征具有技术上的困难,使得这方面的相关应用受到了很大限制.

c. 混合推荐技术.如前所述,协同过滤和基于内容的推荐算法在投入实际运营的时候都有各自的缺陷.因此,实际运营的推荐系统常将两种甚至多种推荐算法进行结合,即采用混合推荐算法.针对实际数据的研究显示这些混合推荐系统普遍具有比上述独立的推荐系统更好的准确率^[112-114].建立混合推荐系统的方法之一是独立地实现协同过滤和运用基于内容的推荐算法,然后将两种推荐结果结合起来,利用预测打分的线性组合进行推荐^[115];又或者,只推荐某一时刻在某一个评价指标下表现更好的算法的结果.例如, Daily Learner 系统^[116]就选择在某一时刻更可信的结果进行推荐,而文献^[117]选择一个与用户过去的打分相一致的结果进行推荐.另外常见的方法是在一种推荐算法的框架中嵌入另外的算法作为某一个辅助部分.例如, Melville 等^[118]利用基于文本分析的方法在协同过滤系统中用户的打分向量上增加一个附加打分,附加分高的用户的信息优先推荐给其他用户; Aciar 等^[119]利用文本挖掘技术分析用户对产品的评论信息,提出基于知识和协同过滤的混合推荐系统.

除了上面3类方法以外,多种数据分析技术,如数据分类、数据聚类、Bayesian 网络、关联规则、K-means 方法、神经网络、线性回归、最大熵方法、云模型、多示例学习等均被用于推荐系统,此处不再赘述.此外,用户的行为特征、个性化的领域知识等也被用于个性化推荐系统.

3.2 关于链路预测

链路预测(link prediction)问题是指通过对已知网络结构的分析,包括一些可能的节点的其他信息,来评估尚不相连的两个点之间产生链接的可能性,进而实现预测^[120].该问题具有重要的应用价值,并且可以对网络科学的理论研究,特别是网络演化规则和节点相似性指标的评判问题起到重要的贡献.下面从实际应用和理论意义两个方面叙述.

很多生物网络,例如蛋白质相互作用网络和新陈代谢网络,节点之间是否存在链路,或者说是否存在相互作用,需要通过大量实验结果进行推断.仅以蛋白质相互作用网络为例,酵母菌蛋白质之间80%的相互作用不为人们所知^[121],而对于人类自身,知道的仅有可怜的0.3%^[122-123].由于揭示这类网络中隐而未现的链接需要耗费高额的实验成本,如果能够在已知结构的基础上设计出足够精确的链路预测算法,再利用预测的结果指导试验,就有可能非常明显地降低试验成本并加快揭开这类网络真实面目的步伐.实际上,社会网络分析中也会遇到数据不全的问题,这时候链路预测同样可以作为准确分析社会网络结构的有力辅助工具^[124-125].除了帮助分析数据缺失的网络,链路预测算法还可以用于分析演化网络.举例来说,近几年在线社会网络发展非常迅速^[126],链路预测可以基于当前的网络结构去预测哪些现在尚未结交的用户“应该是朋友”,并将此结果作为“朋友推荐”发送给用户.如果预测足够准确,显然有助于提高相关网站在用户心目中的地位.另外,链路预测的思想和方法,还可以用于在已知部分节点类型的网络(partially labeled networks)中预测未标签节点的类型——这可以用于判断一篇学术论文的类型^[127]或者判断一个手机用户是否产生了切换运营商(如从移动到联通)的念头^[128],以及用于纠正观察到的网络结构中可能存在的错误^[129],因为很多构建生物网络的实验中存在暧昧不清甚至自相矛盾的数据^[130].

链路预测的研究可以从理论上帮助认识复杂网络演化的机制.针对同一个或者同一类网络,很多模型都提供了可能的网络演化机制^[50,131].由于刻画网

络结构特征的统计量非常多,很难比较不同的机制孰优孰劣,链路预测机制有望为演化网络提供一个简单统一且较为公平的比较平台,从而大大推动复杂网络演化模型的理论研究.另外,如何刻画网络中节点的相似性也是一个重大的理论问题^[132],这个问题和网络聚类等应用息息相关^[133].类似地,相似性的度量指标数不胜数,只有能够快速准确地评估某种相似性定义是否能够很好刻画一个给定网络节点间的关系,才能进一步研究网络特征对相似性指标选择的影响.在这个方面,链路预测可以起到核心技术的作用.链路预测问题本身也带来了有趣且有重要价值的理论问题,也就是通过构造网络系综并藉此利用最大似然估计的方法进行链路预测的可能性和可行性研究.这方面的研究对于链路预测本身以及复杂网络研究理论基础的建立和完善,可以起到推动和借鉴的作用.

近几年,基于节点相似性的链路预测框架受到了广泛的关注,在该框架中,两个节点之间相似性(或者相近性)越大,就认为它们之间存在链接的可能性越大.尽管这个框架非常简单,但是相似性定义本身内涵丰富,它既可以是简单的共同邻居的个数,也可以是包含了复杂数学物理内容的诸如随机游走的平均通讯时间^[134]或者矩阵森林数目^[135].因此,这个简单的框架事实上提供了无穷无尽的可能性.Liben-Nowell和Kleinberg^[136]将相似性指标分为基于节点和基于路径两类,并分析了若干指标对社会合作网络中链路预测的效果.他们发现,在仅考虑节点邻居信息的若干指标中,Adamic-Adar参数^[137]表现最好.周涛、吕琳媛和张翼成^[138]在6种不同网络中比较了9种局部相似性指标在链路预测中的效果,并提出了两种新指标:资源分配指标(resource allocation index)和局部路径指标(local path index).研究发现,新提出来的这两种指标具有明显好于包括Adamic-Adar参数在内的9种已知指标的预测能力.最近其它小组的研究结果显示,新提出来的相似性指标在进行群落划分和含权网络权重设置^[139]的时候也比原有指标好.吕琳媛、金慈航和周涛^[140]进一步在噪音强度可控的网络模型与真实网络中细致分析了局部路径指标的性能,发现这个指标具有与依赖于网络全局结构信息的指标(如Katz参数^[141])可匹敌的预测能力,甚至在噪声较大的情况下可以比Katz参数预测得更加准确.局部路径指标是一个计算量非常小的局部参数,其应用前景可观.刘伟平和吕琳媛^[142]比较研究了一些基于随机

游走的相似性指标,并提出了两种局部随机游走指标,他们发现有限步的随机游走反而可以给出超过全局收敛后的预测精度,而最优的游走步数受到网络平均距离的强烈影响.另外,Huang等的实验结果暗示^[143],在得到节点间的直接相似性后,利用协同过滤技术对相似性指标进行一轮加权处理,一般而言可以得到更好的结果.

最近,最大似然估计方法被尝试应用于链路预测中.Clauset,Moore和Newman^[144]认为很多网络的连接可以看作某种内在层次结构的反映.基于此,他们提出了一种最大似然估计的算法进行链路预测,这种方法在处理具有明显层次组织的网络,如恐怖袭击网络和草原食物链,具有较好的精确度.Guimera和Sales-Pardo^[129]假设观察到的网络是一个随机分块模型(stochastic block model^[145])的一次实现,在该模型中节点被分作若干的集合,两个节点间连接的概率只和相应的集合有关.Guimera和Sales-Pardo^[129]提出了基于随机分块模型的最大似然估计方法,将其用于链路预测,可以得到比Clauset,Moore和Newman更好的结果.

另外一个需要特别注意的趋势,是随着一些原来从事复杂网络研究的学者对链路预测问题的关注,很多复杂网络,特别是社会网络分析中遇到的理论与方法被应用到链路预测中.例如吕琳媛和周涛^[146]发现在针对某些含权网络进行链路预测的时候,权重很小的边反而起到了比高权重边更大的作用,这与社会网络研究中广为人知的“弱连接理论”^[147]有深刻的关联.Leskovec,Huttenlocher和Kleinberg^[148]则注意到了近期“社会平衡理论”的量化研究成果^[149-150],并在此启发下设计了可以预测网络中的正负(友敌)链接的算法.

由于一方面受阻于网络节点外在属性在获取上的难度,另一方面受益于复杂网络研究的快速发展,链路预测问题的主要研究热点逐渐从依赖于节点属性的方法转移到只利用网络结构信息的方法上.显然,后者在理论上也更优美简洁.不过,这个方面的研究主要集中在社会网络上,尚欠对于大量算法在各种不同网络中预测能力的系统分析和总结.另外,目前还没有算法性能和网络结构特征之间关系的较深入的研究.对于比较复杂的网络,例如含权网络、有向网络和多部分网络的讨论虽然有,但非常少,也不系统,相关的研究应该是近几年该方向的主流.

网络系综理论和与之关联的网络熵的概念以及最大似然估计方法有望推动形成复杂网络的统计力

学理论基础. 这方面研究存在一个问题是熵的精确计算复杂性非常大, 对于大规模网络而言往往不能实现. 最近的一些链路预测算法已经应用了网络系综和最大似然的概念, 但是这些算法计算复杂性很大, 精确性也不是很高^[129], 例如文献[144]的方法目前只能处理数千节点的网络, 且其预测效果对于不具有明确层次结构的网络并不好. 作者认为以下两个问题应该是目前国际上相关研究小组比较关注的: 一是如何以网络系综理论为基础, 建立网络链路预测的理论框架, 并产生对实际预测有指导作用的理论结论, 例如通过对网络结构的统计分析估算可预测的极限, 指导选择不同的预测方法等; 二是如何设计高效的算法来处理大规模网络的链路预测问题. 网络系综理论和链路预测的深度结合很可能成为西北大学研究组最近关注的焦点.

最近10年, 复杂网络研究在很多科学分支, 包括物理、生物、计算机等领域掀起高潮^[151], 其中相当一部分研究立足于揭示网络演化的内在驱动因素. 仅以无标度网络(scale-free networks)为例^[152], 已经报道的可以产生幂律度分布的机制就包括了富者愈富(rich-get-richer)机制^[56]、好者变富(good-get-richer)机制^[153]、优化设计(optimal design)驱动^[154]、哈密顿动力学(Hamiltonian dynamics)驱动^[155]、聚生(merging and regeneration)机制^[156]、稳定性限制(stability constraints)驱动^[157]等. 可是, 由于刻画网络结构特征的统计指标非常多, 很难比较和判定什么样的机制能够更好再现真实网络的生长特性. 利用链路预测有望建立简单的比较平台, 能够在知道目标网络演化情况的基础上量化比较各种不同机制对于真实生长行为的预测能力, 从而可以大大推动复杂网络演化机制的相关研究. Guimera 和 Sales-Pardo 在提到网络重建(network reconstruction)的时候已经表达了相近的思想, 但是这方面的研究尚未见报道. 尽管有论文讨论了如何将链路预测的方法和思想与一些应用问题, 例如部分标号网络的节点类型预测^[158]与信息推荐问题^[159]、相联系的可能性与方法问题. 但是, 目前尚缺乏对于大规模真实数据在应用层面的深入分析和研究. 这方面的研究不仅仅具有实用价值, 而且有助于揭示链路预测这个问题本身存在的优势与局限性.

参考文献:

- [1] Barabási A L. The origin of bursts and heavy tails in human dynamics[J]. *Nature*, 2005, 435(7039): 207

- 211.

- [2] Oliveira J G, Barabási A L. Human dynamics: Darwin and Einstein correspondence patterns[J]. *Nature*, 2005, 437(7063): 1251.
- [3] Plerou V, Gopikrishnan P, Amaral L A N, et al. Economic fluctuations and anomalous diffusion[J]. *Phys Rev E*, 2000, 62(3): 3023 - 3026.
- [4] Masoliver J, Montero M, Weiss G H. Continuous-time random-walk model for financial distributions[J]. *Physical Review E*, 2003, 67(2): 021112.
- [5] Politi M, Scalas E. Fitting the empirical distribution of intertrade durations[J]. *Physica A*, 2008, 387(8/9): 2025 - 2034.
- [6] Jiang Z Q, Chen W, Zhou W X. Scaling in the distribution of intertrade durations of Chinese stocks[J]. *Physica A*, 2009, 387(23): 5818 - 5825.
- [7] Dezső Z, Almaas E, Lukács A, et al. Dynamics of information access on the web[J]. *Physical Review E*, 2006, 73(6): 066132.
- [8] Goncalves B, Ramasco J J. Human dynamics revealed through Web analytics[J]. *Physical Review E*, 2008, 78(2): 026123.
- [9] Zhou T, Kiet H A T, Kim B J, et al. Role of activity in human dynamics[J]. *Europhysics Letters*, 2008, 82(2): 28002.
- [10] Hu H B, Han D Y. Empirical analysis of individual popularity and activity on an online music service system[J]. *Physica A*, 2008, 387(23): 5916 - 5921.
- [11] Candia J, González M C, Wang P, et al. Uncovering individual and collective human dynamics from mobile phone records[J]. *Journal of Physics A: Mathematical and Theoretical*, 2008, 41(22): 224015.
- [12] Henderson T, Nhatti S. Modelling user behavior in networked games[C]//Proc 9th ACM Int Conf on Multimidia. New York: ACM Press, 2001: 212.
- [13] Grabowski A, Kruszewska N, Kosiński R A. Dynamic phenomena and human activity in an artificial society[J]. *Physical Review E*, 2008, 78(6): 066110.
- [14] Baek S K, Kim T Y, Kim B J. Testing a priority-based queue model with Linux command histories[J]. *Physica A*, 2008, 387(14): 3660 - 3668.
- [15] Goh K I, Barabási A L. Burstiness and memory in complex systems[J]. *Europhysics Letters*. 2008, 81(4): 48002.
- [16] Vázquez A. Exact results for the Barabási model of human dynamics[J]. *Physical Review Letters*, 2005, 95(24): 248710.
- [17] Vázquez A, Oliveira J G, Dezső Z, et al. Modeling

- bursts and heavy tails in human dynamics[J]. *Physical Review E*, 2006, 73(3): 036127.
- [18] Gabrielli A, Caldarelli G. Invasion percolation and critical transient in the Barabási model of human dynamics [J]. *Physical Review Letters*, 2007, 98(20): 208701.
- [19] Oliveira J G, Vázquez A. Impact of interactions on human dynamics[J]. *Physica A*, 2009, 388(2/3): 187 - 192.
- [20] Vázquez A. Impact of memory on human dynamics[J]. *Physica A*, 2007, 373: 747 - 752.
- [21] Cesar A, Hidalgo R. Conditions for the emergence of scaling in the inter-event time of uncorrelated and seasonal systems [J]. *Physica A*, 2006, 369(2): 877 - 883.
- [22] Malmgren R D, Stouffer D B, Motter A E, et al. A Poissonian explanation for heavy tails in E-mail communication [J]. *PNAS*, 2008, 105(47): 18153 - 18158.
- [23] Vázquez A, Rácz B, Lukács A. Impact of non-Poissonian activity patterns on spreading processes [J]. *Physical Review Letters*, 2007, 98(15): 158702.
- [24] Brockmann D, Hufnagel L, Geisel T. The scaling laws of human travel [J]. *Nature*, 2006, 439(7075): 462 - 465.
- [25] Gonzalez M C, Hidalgo C A, Barabási A L. Understanding individual human mobility patterns [J]. *Nature*, 2008, 453(7196): 779 - 782.
- [26] Rhee I, Shin M, Hong S, et al. On the levy-walk nature of human mobility [C]// *Proceedings of INFOCOM 2008*. Phoenix: IEEE Press, 2008: 924 - 932.
- [27] Bartumeus F, Peters F, Pueyo S, et al. Helical Lévy walks: adjusting searching statistics to resource availability in microzooplankton [J]. *PNAS*, 2003, 100(22): 12771 - 12775.
- [28] Ramos-Fernandez G, Mateos J L, Miramontes O, et al. Lévy walk patterns in the foraging movements of spider monkeys (*Ateles geoffroyi*) [J]. *Behavioral Ecology and Sociobiology*, 2004, 55(3): 223 - 230.
- [29] Sims D W, Southall E J, Humphries N E, et al. Scaling laws of marine predator search behaviour [J]. *Nature*, 2008, 451(7182): 1098 - 1102.
- [30] Viswanathan G M, Buldyrev S V, Havlin S, et al. Optimizing the success of random searches [J]. *Nature*, 1999, 401(6756): 911 - 914.
- [31] Bartumeus F, Catalan J, Fulco U L, et al. Optimizing the encounter rate in biological interactions: Lévy versus Brownian strategies [J]. *Physical Review Letters*, 2002, 88(9): 097901.
- [32] Reynolds A M. Scale-free movement patterns arising from olfactory-driven foraging [J]. *Physical Review E*, 2005, 72(4): 041928.
- [33] Santos M C, Boyer D, Miramontes O, et al. Origin of power-law distributions in deterministic walks: the influence of landscape geometry [J]. *Physical Review E*, 2007, 75(6): 061114.
- [34] Zanette D H. Dynamics of rumor propagation on small-world networks [J]. *Phys Rev E*, 2002, 65(4): 041908.
- [35] Karsai M, Kivela M, Pan R K, et al. Small but slow world: how network topology and burstiness slow down spreading [J]. *Phys Rev E*, 2011, 83(2): 025102.
- [36] Yang Z, Cui A X, Zhou T. Impact of heterogeneous human activities on epidemic spreading [J]. *Physica A*, 2010, 390(23/24): 4543 - 4548.
- [37] Ni S, Weng W. Impact of travel patterns on epidemic dynamics in heterogeneous spatial metapopulation networks [J]. *Phys Rev E*, 2009, 79(1): 016111.
- [38] Hufnagel L, Brockmann D, Geisel T. Forecast and control of epidemics in a globalized world [J]. *PNAS*, 2004, 101(42): 15124 - 15219.
- [39] Colizza V, Barrat A, Barthelemy M, et al. The role of the airline transportation network in the prediction and predictability of global epidemics [J]. *PNAS*, 2006, 103(7): 2015 - 2020.
- [40] 周涛. 在线电影点播中的人类动力学模式 [J]. *复杂系统与复杂性科学*, 2008, 5(1): 1 - 5.
- [41] Hong W, Han X P, Zhou T, et al. Heavy-tailed statistics in short-message communication [J]. *Chin Phys Lett*, 2009, 26(2): 28902.
- [42] 李楠楠, 张宁, 周涛. 人类通信模式中基于时间统计的实证研究 [J]. *复杂系统与复杂性科学*, 2008, 5(3): 43 - 47.
- [43] Li N N, Zhang N, Zhou T. Empirical analysis on temporal statistics of human correspondence patterns. [J]. *Physica A*, 2008, 387(25): 6391 - 6394.
- [44] 韩筱璞, 周涛, 汪秉宏. 基于自适应调节的人类动力学模型 [J]. *复杂系统与复杂性科学*, 2007, 4(4): 1 - 5.
- [45] Han X P, Zhou T, Wang B H. Modeling human dynamics with adaptive interest [J]. *New Journal of Physics*, 2008, 10(7): 073010.
- [46] 李楠楠, 周涛, 张宁. 人类动力学基本概念与实证分析 [J]. *复杂系统与复杂性科学*, 2008, 5(2): 15 - 24.
- [47] 郭进利, 周涛, 李季明, 等. 人类动力学模型 [M]. 香港: 上海系统科学出版社, 2008.
- [48] Zhou T, Han X P, Wang B H. Towards the understanding of human dynamics [M]// Burguete M,

- Lam eds L. *Science Matters—Humanities as Complex Systems*. Singapore: World Scientific Publishing, 2008: 207 – 233.
- [49] Albert R, Barabási A L. Statistical mechanics of complex networks[J]. *Rev Mod Phys*, 2002, 74(1): 47 – 97.
- [50] Boccaletti S, Latora V, Moreno Y, et al. Complex networks: structure and dynamics [J]. *Physics Reports*, 2006, 424(4/5): 175 – 308.
- [51] Jeong H, Mason S P, Barabási A L, et al. Lethality and centrality in protein networks[J]. *Nature*, 2001, 411(6833): 41 – 42.
- [52] Jeong H, Tombor B, Albert R, et al. The large-scale organization of metabolic networks[J]. *Nature*, 2000, 407(6804): 651 – 654.
- [53] Adamic L A, Huberman B A. Power-law distribution of the world wide web [J]. *Science*, 2000, 287(5461): 2115.
- [54] Watts D J, Strogatz S H. Collective dynamics of small-world networks [J]. *Nature*, 1998, 393(6684): 440 – 442.
- [55] Barabási A L, Albert R. Emergence of scaling in random networks[J]. *Science*, 1999, 286(5439): 509 – 512.
- [56] Scannell J W, Burns G A P C, Hilgetag C C, et al. The connective organization of the cortico-thalamic system of the cat[J]. *Cerebral Cortex*, 1999, 9(3): 277 – 299.
- [57] Sporns O, Chialvo D R, Kaiser M, et al. Organization, development and function of complex brain networks [J]. *Trends in Cognitive Sciences*, 2004, 8(9): 418 – 425.
- [58] Motter A E, Matias M A, Kurths J, et al. Dynamics on complex networks and applications [J]. *Physica D*, 2006, 224(1/2): 0612068.
- [59] Engel A K, Fries P, Singer W. Dynamic predictions: oscillations and synchrony in top-down processing[J]. *Nature Reviews Neuroscience*, 2001, 2(10): 704 – 716.
- [60] Fries P. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence [J]. *Trends in Cognitive Sciences*, 2005, 9(10): 474 – 480.
- [61] Arenas A, Díaz-Guilera A, Kurths J, et al. Synchronization in complex networks [J]. *Physics Reports*, 2008, 469(3): 93 – 153.
- [62] Kori H, Mikhailov A S. Entrainment of randomly coupled oscillator networks by a pacemaker[J]. *Phys Rev Lett*, 2004, 93(25): 254101.
- [63] Kori H, Mikhailov A S. Strong effects of network architecture in the entrainment of coupled oscillator systems[J]. *Phys Rev E*, 2006, 74(6): 066115.
- [64] Kiss I Z, Rusin C G, Kori H, et al. Engineering complex dynamical structures: sequential patterns and desynchronization[J]. *Science*, 2007, 316(5833): 1886 – 1889.
- [65] Barahona M, Pecora L M. Synchronization in small-world systems [J]. *Phys Rev Lett*, 2002, 89(5): 054101.
- [66] Nishikawa T, Motter A E, Lai Y C, et al. Heterogeneity in oscillator networks: are smaller world easier to synchronize[J]. *Phys Rev Lett*, 2003, 91(1): 014101.
- [67] Hwang D U, Chavez M, Amann A, et al. Synchronization in complex networks with age ordering[J]. *Phys Rev Lett*, 2005, 94(13): 138701.
- [68] Chavez M, Hwang D U, Amann A, et al. Synchronization is enhanced in weighted complex networks [J]. *Phys Rev Lett*, 2005, 94(21): 218701.
- [69] Zhao M, Zhou T, Wang B H, et al. Enhanced synchronizability by structural perturbations[J]. *Phys Rev E*, 2005, 72(5): 057102.
- [70] Lu Y F, Zhao M, Zhou T, et al. Enhance synchronizability via age-based coupling [J]. *Phys Rev E*, 2007, 76(5): 057103.
- [71] Donetti L, Hurtado P I, Muñoz M A. Entangled networks, synchronization, and optimal network topology[J]. *Phys Rev Lett*, 2005, 95(18): 188701.
- [72] Newman M E J, Girvan M. Finding and evaluating community structure in networks [J]. *Phys Rev E*, 2004, 69(2): 026113.
- [73] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society [J]. *Nature*, 2005, 435(7043): 814 – 818.
- [74] Danon L, Díaz-Aguilera A, Duch J, et al. Comparing community structure identification [J]. *Journal of Statistical Mechanics: Theory and Experiment*, 2005, 2005(9): P09008.
- [75] Farkas I, Ábel D, Palla G, et al. Weighted network modules [J]. *New Journal of Physics*, 2007, 9(6): 180.
- [76] Huang L, Park K, Lai Y C, et al. Abnormal synchronization in complex clustered networks [J]. *Phys Rev Lett*, 2006, 97(16): 164101.
- [77] Gómez-Gardeñes J, Moreno Y, Arenas A. Paths to synchronization on complex networks [J]. *Phys Rev Lett*, 2007, 98(3): 034101.
- [78] Gómez-Gardeñes J, Moreno Y, Arenas A. Synchronizability

- determined by coupling strengths and topology on complex networks[J]. *Phys Rev E*, 2007, 75(6):066106.
- [79] Boccaletti S, Ivanchenko M, Latora V, et al. Detecting complex network modularity by dynamical clustering [J]. *Phys Rev E*, 2007, 75(4):045102.
- [80] Oh E, Rho K, Hong H, et al. Modular synchronization in complex networks [J]. *Phys Rev E*, 2005, 72(4):047101.
- [81] Oh E, Choi C, Kahng B, et al. Modular synchronization in complex networks with a gauge Kuramoto model [J]. *Europhys Lett*, 2008, 83(6):68003.
- [82] Ito J, Kaneko K. Spontaneous structure formation in a network of chaotic units with variable connection strengths[J]. *Phys Rev Lett*, 2001, 88(2):028701.
- [83] Gross T, Dommar D L C, Blasius B. Epidemic dynamics on an adaptive network[J]. *Phys Rev Lett*, 2006, 96(20):208701.
- [84] Zhou C S, Kurths J. Dynamical weights and enhanced synchronization in adaptive complex networks [J]. *Phys Rev Lett*, 2006, 96(16):164102.
- [85] Bassett D S, Bullmore E. Small-world brain networks [J]. *The Neuroscientist*, 2006, 12(6):521 - 523.
- [86] Bullmore E, Sporns O. Complex brain networks: graph theoretical analysis of the structural and functional systems[J]. *Nature Reviews Neuroscience*, 2009, 10(3):186 - 198.
- [87] Kaiser M, Hilgetag C C. Edge vulnerability in neural and metabolic networks [J]. *Biological Cybernetics*, 2004, 90(5):311 - 317.
- [88] Stam C J. Functional connectivity patterns of human magnetoencephalographic recordings: a small-world network[J]. *Neurosci Lett*, 2004, 355(1/2):25 - 28.
- [89] Stam C J, de Bruin E A. Scale-free dynamics of global functional connectivity in the human brain[J]. *Human Brain Mapping*, 2004, 22(2):97 - 109.
- [90] Salvador R, Suckling J, Coleman M R, et al. Neurophysiological architecture of functional magnetic resonance images of human brain[J]. *Cereb Cortex*, 2005, 15(9):1332 - 1342.
- [91] Fox M D, Snyder A Z, Vincent J L, et al. The human brain is intrinsically organized into dynamic, anticorrelated functional networks [J]. *PNAS*, 2005, 102(27):9673 - 9678.
- [92] Womelsdorf T, Schoffelen J M, Oostenveld R, et al. Modulation of neuronal interactions through neuronal synchronization[J]. *Science*, 2007, 316(5831):1609 - 1612.
- [93] Saalman Y B, Pigarev I N, Vidyasagar T R. Neural mechanisms of visual attention: how top-down feedback highlights relevant locations [J]. *Science*, 2007, 316(5831):1612 - 1615.
- [94] Buschman T J, Miller E K. Top-down versus bottom-up control of attention in the prefrontal and posterior parietal cortices[J]. *Science*, 2007, 315(5820):1860 - 1862.
- [95] Zhang G Q, Zhang G Q, Yang Q F, et al. Evolution of the Internet and its cores[J]. *New Journal of Physics*, 2008, 10(12):123027
- [96] Broder A, Kumar R, Moghoul F, et al. Graph structure in the web[J]. *Computer Networks*, 2000, 33(1):309 - 320
- [97] 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.
- [98] Resnick P, Varian H R. Recommender systems [J]. *Communications of the ACM*, 1997, 40(3):56 - 58.
- [99] Herlocker J L, Konstan J A, Terveen L G, et al. Evaluating collaborative filtering recommender systems [J]. *ACM Transactions on Information Systems*, 2004, 22(1):5 - 53.
- [100] Rich E. User modeling via stereotypes[J]. *Cognitive Science*, 1979, 3(4):329 - 354.
- [101] Goldberg D, Nichols D, Oki B M, et al. Using collaborative filtering to weave an information tapestry[J]. *Communications of the ACM*, 1992, 35(12):61 - 70.
- [102] Konstan J A, Miller B N, Maltz D, et al. GroupLens: applying collaborative filtering to usenet news [J]. *Communications of the ACM*, 1997, 40(3):77 - 87.
- [103] Shardanand U, Maes P. Social information filtering: algorithms for automating "word of mouth"[C]//CHI'95 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. New York: ACM Press, 1995:210 - 217.
- [104] Linden G, Smith B, York J. Amazon.com recommendations: item-to-item collaborative filtering [J]. *IEEE Internet Computing*, 2003, 7(1):76 - 80.
- [105] Belkin N, Croft B. Information filtering and information retrieval [J]. *Communications of the ACM*, 1992, 35(12):29 - 37.
- [106] Balabanovic M, Shoham Y. Fab: content-based, collaborative recommendation[J]. *Communications of the ACM*, 1997, 40(3):66 - 72
- [107] Pazzani M, Billsus D. Learning and revising user

- profiles; the identification of interesting web sites[J]. *Machine Learning*, 1997, 27(3): 313 - 331
- [108] Somlo G L, Howe A E. Adaptive lightweight text filtering [J]. *Lecture Notes in Computer Science*, 2001, 2189: 319 - 329
- [109] Zhang Y, Callan J, Minka T. Novelty and redundancy detection in adaptive filtering[C]//*Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM Press, 2002: 81 - 88.
- [110] Chang Y I, Shen J H, Chen T I. A data mining-based method for the incremental update of supporting personalized information filtering [J]. *Journal of Information Science and Engineering*, 2008, 24(1): 129 - 142
- [111] Degenm M, Lops P, Semeraro G. A content-collaborative recommender that exploits WordNet-based user profiles for neighborhood formation [J]. *User Modeling and User-Adapted Interaction*, 2007, 17(3): 217 - 255
- [112] 高滢, 齐红, 刘杰, 等. 结合似然关系模型和用户等级的协同过滤推荐算法[J]. *计算机研究与发展*, 2007, 44(6): 1044 - 1052.
- [113] Yoshii K, Goto M, Komatani K, et al. An efficient hybrid music recommender system using an incrementally trainable probabilistic generative model [J]. *IEEE Transactions on Audio Speech and Language Processing*, 2008, 16(2): 435 - 447.
- [114] Girardi R, Marinho L B. A domain model of Web recommender systems based on usage mining and collaborative filtering [J]. *Requirements Engineering*, 2007, 12(1): 23 - 40.
- [115] Pazzani M J. A framework for collaborative, content-based, and demographic filtering [J]. *Artificial Intelligence Review*, 1999, 13(5/6): 393 - 408.
- [116] Billsus D, Pazzani M J. User modeling for adaptive news access [J]. *User Modeling and User-Adapted Interaction*, 2000, 10(2/3): 147 - 180.
- [117] Tran T, Cohen R. Hybrid recommender systems for electronic commerce [C]//*Proc Knowledge-Based Electronic Markets*. Palo Alto: AAAI Press, 2000: 78 - 83.
- [118] Melville P, Mooney R J, Nagarajan R. Content-boosted collaborative filtering for improved recommendations[C]//*Proceedings of the 2001 SIGIR Workshop on Recommender Systems*. Palo Alto: AAAI Press, 2002: 187 - 192.
- [119] Aciar S, Zhang D, Simoff S, et al. Informed recommender; basing recommendations on consumer product reviews[J]. *IEEE Intelligent Systems*, 2007, 22(3): 39 - 47.
- [120] Getoor L, Diehl C P. Link mining; a survey [J]. *ACM SIGKDD Explorations Newsletter*, 2005, 7(2): 3 - 12.
- [121] Yu H, Braun P, Yildirim M A, et al. High-quality binary protein interaction map of the yeast interactome network [J]. *Science*, 2008, 322(5898): 104 - 110.
- [122] Stumpf M P H, Thorne T, de Silva E, et al. Estimating the size of the human interactome [J]. *PNAS*, 2008, 105(19): 6959 - 6964.
- [123] Amaral L A N. A truer measure of our ignorance [J]. *PNAS*, 2008, 105(19): 6795 - 6796.
- [124] Schafer J L, Graham J W. Missing data; our view of the state of the art [J]. *Psychological Methods*, 2002, 7(2): 147 - 177.
- [125] Kossinets G. Effects of missing data in social networks [J]. *Social Networks*, 2006, 28(3): 247 - 268.
- [126] Kumar R, Novak J, Tomkins A. Structure and evolution of online social networks [C]//*Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2006: 611 - 617.
- [127] Gallagher B, Tong H, Eliassi-Rad T, et al. Using ghost edges for classification in sparsely labeled networks [C]//*Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York: ACM Press, 2008: 256 - 264.
- [128] Dasgupta K, Singh R, Viswanathan B, et al. Social ties and their relevance to churn in mobile telecom networks [C]//*Proceedings of the 11th International Conference on Extending Database Technology: Advances in Database Technology*. New York: ACM Press, 2008: 668 - 677.
- [129] Guimera R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks [J]. *PNAS*, 2009, 106(52): 22073 - 22078.
- [130] von Mering C, Krause R, Snel B, et al. Comparative assessment of large-scale data sets of protein-protein interactions [J]. *Nature*, 2002, 417(6887): 399 - 403.
- [131] Dorogovtsev S N, Mendes J F F. Evolution of networks [J]. *Advances in Physics*, 2002, 51(4): 1079 - 1187.
- [132] Leicht E A, Holme P, Newman M E J. Vertex similarity in networks [J]. *Phys Rev E*, 2006, 73(2): 026120.
- [133] Pan Y, Li D H, Liu J G, et al. Detecting community structure in complex networks via node similarity [J]. *Physica A*, 2010, 389(14): 2849 - 2857.
- [134] Fouss F, Pirotte A, Renders J M, et al. Random-walk

- computation of similarities between nodes of a graph with application to collaborative recommendation[J]. IEEE Transactions on Knowledge and Data Engineering, 2007, 19(3): 355 - 369.
- [135] Chebotarev P, Shamis E. The matrix-forest theorem and measuring relations in small social groups[J]. Automation and Remote Control, 1997, 58 (2): 1505 - 1514.
- [136] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks [J]. Journal of the American Society for Information Science and Technology, 2007, 58 (7): 1019 - 1031.
- [137] Adamic L A, Adar E. Friends and neighbors on the web [J]. Social Networks, 2003 25(3): 211 - 230.
- [138] Zhou T, Lü L Y, Zhang Y C. Predicting missing links via local information [J]. The European Physical Journal B, 2009, 71(4): 623 - 630.
- [139] Wang Y L, Zhou T, Shi J J, et al. Empirical analysis of dependence between stations in Chinese railway network[J]. Physica A, 2009, 388 (14): 2949 - 2955.
- [140] Lü L Y, Jin C H, Zhou T. Similarity index based on local paths for link prediction of complex networks [J]. Phys Rev E, 2009, 80(4): 046122.
- [141] Katz L. A new status index derived from sociometric analysis[J]. Psychometrika, 1953, 18(1): 39 - 43.
- [142] Liu W P, Lü L Y. Link prediction based on local random walk[J]. Europhys Lett, 2010, 89(5): 58007.
- [143] Huang Z, Li X, Chen H C. Link prediction approach to collaborative filtering [C]// Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries. New York: ACM Press, 2005, 141 - 142.
- [144] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks[J]. Nature, 2008, 453(7191): 98 - 101.
- [145] Holland P W, Laskey K B, Leinhardt S. Stochastic blockmodels; first steps[J]. Social Networks, 1983, 5 (2): 109 - 137.
- [146] Lü L Y, Zhou T. Link prediction in weighted networks; the role of weak ties[J]. Europhys Lett, 2010, 89(1): 18001.
- [147] Granovetter M S. The strength of weak ties [J]. American Journal of Sociology, 1973, 78 (6): 1360 - 1380.
- [148] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks [C]// Proceedings of the 19th International Conference on World Wide Web. New York: ACM Press, 2010: 641 - 650.
- [149] Antal T, Krapivsky P, Redner S. Dynamics of social balance on networks [J]. Phys Rev E, 2005, 72 (3): 036121.
- [150] Marvel S, Strogatz S, Kleinberg J. Energy landscape of social balance [J]. Phys Rev Lett, 2009, 103 (19): 198701.
- [151] Barabási A L. Scale-free networks; a decade and beyond[J]. Science, 2009, 325 (5939): 412 - 413.
- [152] Caldarelli G. Scale-free networks; complex webs in nature and technology [M]. New York: Oxford Press, 2007.
- [153] Garlaschelli D, Capocci A, Caldarelli G. Self-organized network evolution coupled to extremal dynamics[J]. Nature Physics, 2007, 3(11): 813 - 817.
- [154] Valverde S, Cancho R F, Sole R V. Scale-free networks from optimal design [J]. Europhys Lett, 2002, 60 (4): 512.
- [155] Baiesi M, Manna S S. Scale-free networks from a hamiltonian dynamics [J] Phys Rev E, 2003, 68 (4): 047103.
- [156] Kim B J, Trusina A, Minnhagen P, et al. Self organized scale-free networks from merging and regeneration [J]. The European Physical Journal B, 2005, 43 (3): 369 - 372.
- [157] Perotti J I, Billoni O V, Tamarit F A, et al. Emergent self-organized complex network topology out of stability constraints [J]. Phys Rev Lett, 2009, 103 (10): 108701.
- [158] Sen P, Namata G, Bilgic M, et al. Collective classification in network data[J]. AI Magazine, 2008, 29(3): 93 - 106.
- [159] Zhou T. Statistical mechanics of information systems: information filtering on complex networks [D]. Fribourg: University of Fribourg, 2010.