

文章编号: 1672 - 3813(2010) 02-03 - 0173 - 14

复杂网络观察

吕琳媛¹ 陆君安² 张子柯¹ 闫小勇³ 吴 晔^{4,5} 史定华⁶ 周海平⁷ 方锦清⁸ 周 涛^{9,10}

- (1. 弗里堡大学物理系, 弗里堡 CH-1700; 2. 武汉大学数学与统计学院, 武汉 430072;
3. 石家庄铁道大学交通运输学院, 石家庄 050043; 4. 波茨坦大学复杂系统动力学交叉中心, 波茨坦 D-14469;
5. 北京邮电大学理学院, 北京 100876; 6. 上海大学数学系, 上海 200444;
7. 贵阳学院计算机科学系, 贵阳 550005; 8. 中国原子能科学研究院, 北京 102413;
9. 电子科技大学互联网科学中心, 成都 610054; 10. 中国科学技术大学近代物理系, 合肥 230026)



摘要: 总结了几位作者对复杂网络研究中存在的重要问题和发展趋势的讨论, 其中既包括度分布和度指数的分析和计算, 各种不同动力学之间的内在一致性, 网络加速增长机制这样的基本问题, 也包括了网络中尺度这类细致深入的结构分析。还就复杂网络与其他重要研究方向深入结合的现状和未来展开了讨论, 包括复杂网络中的链路预测问题, 复杂网络在信息推荐系统中的应用, 复杂网络与信息物理系统的可能结合, 复杂网络和人类动力学的结合以及复杂网络在国家安全方面可能的重要战略地位。

关键词: 复杂网络; 度分布; 加速增长; 同步; 疾病传播; 网络交通; 流驱动的动力学; 中尺度; 链路预测; 推荐系统; 信息物理系统; 人类动力学

中图分类号: N94

文献标识码: A

Looking into Complex Networks

LÜ Lin-yuan¹, LU Jun-an², ZHANG Zi-ke¹, YAN Xiao-yong³, WU Ye^{4,5},
SHI Ding-hua⁶, ZHOU Hai-ping⁷, FANG Jin-qing⁸, ZHOU Tao^{9,10}

- (1. Department of Physics, University of Fribourg, Fribourg 1700, Switzerland;
2. School of Mathematics and Statistics, Wuhan University, Wuhan 430072, China;
3. Department of Transportation Engineering, Shijiazhuang Tiedao University, Shijiazhuang 050043, China;
4. Interdisciplinary Center for Dynamics of Complex Systems, University Potsdam, Potsdam D-14469, Germany;
5. School of Science, Beijing University of Posts and Telecommunications, Beijing 100876;
6. Department of Mathematics, Shanghai University, Shanghai 200444, China;
7. Department of Computer Science, Guiyang College, Guiyang 550005, China;
8. China Institute of Atomic Energy, Beijing 102413, China;
9. Web Sciences Center, University of Electronic Science and Technology of China, Chengdu 610054, China;
10. Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China)

Abstract: This article summaries the discussions about the open issues and research tendency of complex networks by several active scholars. The statement covers both fundamental problems like the understand-

收稿日期: 2010 - 06 - 21

作者简介: 吕琳媛(1984 -), 女, 北京人, 博士研究生, 主要研究方向为信息物理, 包括链路预测, 推荐算法和网络排序算法等。

注: 本文系第六届全国网络科学会议中“复杂网络观察”博文大赛优秀作品的集萃。

ing of power-law degree distributions , the underlying connections between different flow-driven dynamics and the mechanism leading to acceleratingly growing , and the in-depth analyses like the understanding of mesoscales in complex networks. Moreover , we introduce some typical interdisciplinary studies where complex networks play a major role , including the link prediction in complex networks , recommender systems for user-object bipartite networks , the integration of cyber physical systems and complex networks , the human dynamics in the online network space , and the possibly important role of the studies of complex networks in national security.

Key words: complex networks; degree distribution; acceleratingly growing; synchronization; epidemic spreading; network traffic; flow-driven dynamics; mesoscales; link prediction; recommender systems; cyber physical systems; human dynamics

1 引言

如果从 1998 年 Watts 和 Strogatz 提出小世界网络模型^[1] ,以及 1999 年 Barabási 和 Albertt 提出无标度网络^[2] 算起 ,复杂网络这个研究领域已经走过了它的第 1 个 10 年。即便去除掉开创时期的文章 ,仅 2000 ~ 2009 年发表的明确以“复杂网络”(complex networks) 为主题(topic) 的 SCI 论文就达到了 4 925 篇之多 ,引用高达 76 834 次 ,而且依旧呈现强劲的上漲势头(见图 1) 。得益于我国一些学者较早意识到复杂网络研究的重要性 ,并推动了我国网络科学的迅猛发展。如图 2 所示 ,复杂网络在我国呈现出了类似的迅猛发展的势头 ,在统计的 SCI 论文中 ,有中国地址的论文共有 1 318 篇 ,超过了四分之一。

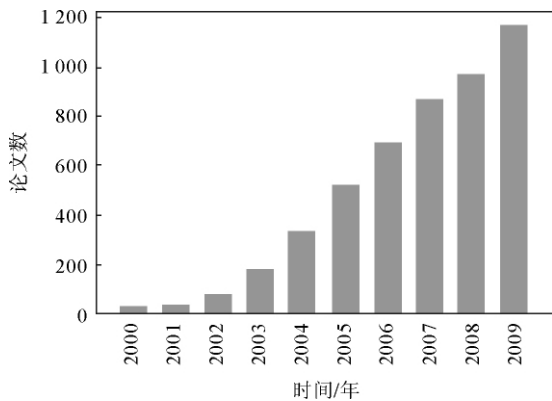


图 1 以复杂网络为主题的 SCI 论文数

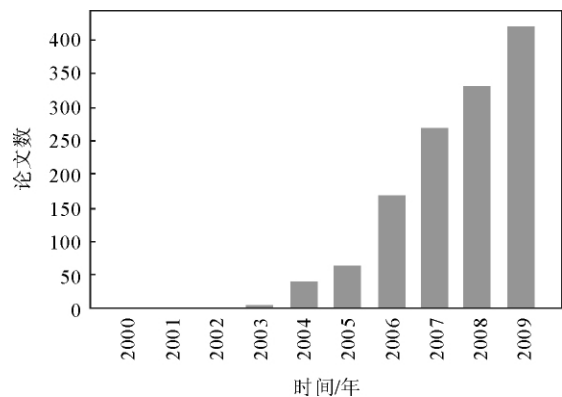


图 2 以复杂网络为主题的含中国地址 SCI 论文数

在欣喜之余 ,我们也需要警醒 ,这 1 318 篇论文获得的引用仅仅是 7 229 次 ,还不到总数的十分之一 ,说明我国学者在复杂网络方面的研究工作距离世界平均水平还有比较明显的差距。本文正是在这种冷静的兴奋、谨慎的乐观情境下对于复杂网络研究中存在的重要问题和发展趋势的一次集中讨论。从这些讨论中 ,可以看出 ,在复杂网络研究中作出真正重要的贡献 ,离不开“精”、“深”、“广”三字。

以前对于网络结构的刻画 ,要么很宏观 ,例如网络的连接密度和平均距离 ,要么很微观 ,从节点的局部性质入手 ,比如节点的度和簇系数。这些量都只能从比较粗糙的层面揭示网络的结构 ,而本文要述及的中尺度 ,包括模体、圈、核、群落等等 ,有望提供连接微观与宏观的桥梁——这类研究体现了从粗放到精细的趋势。以前对幂律度分布的拟合 ,往往是双对数坐标下简单的最小二乘法 ,拟合得好不好 ,也就是肉眼判断一下。如何判断一个分布是不是幂律的 ,幂指数如何确定 ,离散和连续分布的区别等等 ,最近又重新成为学术界争论的焦点——这类研究体现了从简浅到深刻的趋势。以前复杂网络的研究 ,大部分集中于网络结构分析和

建模。对于网络功能的研究,也主要集中在传播、同步、交通、博弈等少数几个方面,虽然号称交叉学科,但和其他学科关系松散,基本上可以说是自娱自乐。本文介绍了复杂网络和链路预测、信息推荐、人类动力学等研究方向业已存在的深入结合,展望了其与信息物理系统的可能联姻——这类研究体现了从狭促到广阔的趋势。

本文将尽最大可能忠实于原作品的风格,希望能对我国学者在复杂网络方面的研究提供借鉴。

2 几何增长网络的度分布和度指数

自从 Barabási 等人^[3]利用模块增长构造出第一个确定性的层次网络起,层次网络^[4-6]、伪分形图^[7-9]、阿波罗网^[10-13]等几何增长网络的度分布和度指数就一直是颇有争议的问题。他们在文[3]中首先得到了第一个确定性的层次网络度概率的尾部呈现幂律,度指数是3的对数与2的对数之比。后来觉得不妥,在文[4]中又给出了该层次网络度概率密度的尾部,也呈现幂律,度指数是1加上3的对数与2的对数之比。文[7]采用补分布方法,首先求出了伪分形图的补分布,然而表达式是一个离散的幂律,得到了伪分形图的度指数为1加上3的对数与2的对数之比。文[10]引入了阿波罗网络,用补分布方法求出了阿波罗网络度的补分布也是一个离散的幂律,度指数为1加上3的对数与2的对数之比。不知何故,然后文[10]的作者在文[13]中又将度指数更改为3的对数与2的对数之比。

由于几何增长网络不仅突破了每次只增加一个结点的限制,而且因其特殊的拓扑性质(自相似和高群集),以及便于精确解析广泛受到网络界的青睐。然而,几何增长网络的正确度分布(概率、密度、补分布)是什么?正确的度指数需不需要在两个对数之比基础上加1。文献上对这个问题的认识就像过山车一样,迄今没有一个令人信服的说法。因为这类网络度的非零概率一般都是幂律,但其中存在几何增长的度概率为零的区间,使得网络是否无标度和度指数难于决断。而且这种现象在实际网络统计和模型网络模拟中也经常会遇到。这就引发了如何准确判断一个网络的度分布是否为无标度,以及如何确定其度指数的重要问题。我觉得这或许会是“复杂网络圈”的一个好论题。

首先关于幂律和标度要有一个正确的定义,文献[14]对确定的和随机的两种情况都有详尽的论述。如果确定的单调递减序列或随机的补分布函数存在一个常数使得序列值或函数值能够表示为正比于幂律的形式,那么就称为无标度的幂律指数,即那个常数除掉负号就称为标度指数。对随机的情况,如果存在密度函数,密度函数也是幂律,但标度指数需要加1。不过要注意,在复杂网络里,无标度网络的度指数都是指概率密度的标度指数,因此需要加1。这可能是 Barabási 等人^[2]在用平均场方法推导 BA 模型网络度分布时,得到的是概率密度,现在已经约定俗成,既不便更改也无须争辩。

下面以伪分形图^[7]为例子来讨论。文[7]精确分析了该网络的结点度和结点数,结点只有度 $2, 2^2, 2^3, \dots, 2^{l-1}, 2^l, 2^{l+1}$, 对应的结点数分别是 $3^l, 3^{l-1}, 3^{l-2}, \dots, 3^2, 3, 3$, 结点总数为 $3(3^l + 1)/2$ 。于是,文[7]得到该网络的稳态度概率是在 $2, 2^2, 2^3, \dots$ 上取值的幂律,系数为2,幂指数为负3的对数与2的对数之比。因为当中穿插了许许多多概率为零的度,这不是单调递减序列,所以无法判断。文[7]采用补分布来绕开这个困难,给出了一个离散幂律的表达式,幂指数仍为负3的对数与2的对数之比,但指出度指数是1加上3的对数与2的对数之比。随后几何增长网络的文献几乎都采用这种方法确定无标度网和度指数。

我们现在来评论一下上述工作。文[7]得到的稳态度概率是精确的,度指数也是正确的。然而补分布表达式容易误解,对离散情况,众所周之补分布应该是一个递减的连续阶梯函数。正确的补分布可以借助 Heaviside 阶梯函数表示出来。另外,文[4]中给出的概率密度尾部也不准确,离散的概率密度也不是单调递减序列。因此,对几何增长网络,我认为度分布最好采用递减的连续阶梯函数,度指数是其幂指数除掉负号后加1。如果采用度概率,必须说明定义域,度指数仍需在幂指数除掉负号后加1。

事实上,只有当度间隔为1的概率值服从幂律,才能直接得出无标度网和正确的度指数;否则需要采用补分布方法。实际网络统计和模型网络模拟都会遇到在最小度和最大度之间出现度概率为零的情况。如果只有偶尔几个零问题还不至于太大,若度概率为零的点比较多,问题就严重了。所以一般建议采用补分布方

法,即统计大于等于度 k 的结点数与总结点数之比来画图,因为这样更准确。不过要注意,度指数是其斜率除掉负号后加 1。

(史定华)

3 网络同步、交通和传播动力学的内在一致性

2005年1月, Motter、昌松和 Kurths 在《美国物理评论 E》上发表了一篇不算太难的文章^[15], 提出网络同步的时候, 一个节点被耦合的强度总和应该归一。以前每一条边上的耦合强度都是 1, 那么一个度为 k 的节点, 被耦合的强度之和就是 k 现在要归一, 就是要除以一个 k ——这样得到一个归一化的拉普拉斯矩阵。这个方法虽然简单, 但是原来同步能力很差的无标度网络^[16], 一下子同步能力就变强了。所以说很快就成为了网络同步研究中耳熟能详的著名方法, 事实上, 一多半的提高网络同步能力的尝试, 都受到了这个方法的启发^[17], 而这篇论文在 2005 年《美国物理评论 E》上发表的 2 590 篇论文中引用排名第 3。

2006年2月和2006年4月, 中科大小组接连在《美国物理评论 E》上发表了两篇关于交通流的论文, 前一篇是仅知道网络局部信息的交通流问题^[18], 后一篇是知道网络整体结构的交通流问题^[19]。这两篇文章也是一发表就受到广泛关注, 目前在 2006 年《美国物理评论 E》上发表的 2 425 篇论文中引用排名第 6 和第 4。两篇文章都涉及到同一个问题, 就是交通选路时选择短路和选择流量小的路之间的矛盾: 如果总是走偏街背巷, 尽管不会堵, 但是太绕路, 如果总是走大路, 则有可能完全堵死。有趣的是, 尽管两篇论文所利用的信息和路由策略非常不同, 但是结论却很相似, 前者发现最优的策略是按照与邻居度 k 成反比传送信息包, 后者指出最优路由策略可以通过按照与节点度 k 成正比的数值进行路径加权。

假设我们考虑一个带有自由参数 α 的一个与度相关的量 k^α , 会发现这些最优值的位置简直太漂亮了, 要么是 1, 要么是 -1。尽管在一些假设的基础上, 似乎也能够对这样的最优参数给出似是而非的解释, 但是真正从头做起的完整的解析却很困难, 因为同步能力中涉及的特征值问题和网络交通流中涉及的路径问题, 都有巨大的复杂性。这里, 我想特别强调的一点是, 这些似乎很不一样的物理过程, 都可以看做一种“流驱动的动力学”(flow-driven dynamics): 同步中是耦合信号在传输, 交通中则是信息包在传输, 而更加直观的例子, 是传播动力学, 其中疾病在网络中传播。当然, 这其中也有本质的不同之处, 例如耦合信号之间互相可以干扰, 信息包和疾病则不会; 传染源可以复制, 而信息包则不会。

因为看到了这种隐隐约约的内在一致性, 我希望能够找到一种非常简单但是又能体现这种流驱动的动力学本质的物理过程, 并尝试一些解析。在所有流驱动的动力学中, 接触过程(contact process)无疑是最简单的之一, 所以我先以此为突破口。我们发现, 如果允许一个个体在选择接触对象的时候(只能在邻居中选择)按照与其度 k 的一个函数 $f(k)$ 成正比的概率选择, 那么可以利用泛函变分的方法证明最优的函数是 $f(k) = 1/k$ 。注意, 这里并不依赖于 $f(k)$ 的具体形式。这篇文章 2008 年 12 月也在《美国物理评论 E》上发表^[20]。

得到这样的解析结果固然可喜, 但是并不意味着我们真正理解了这些动力学之间的一致性, 甚至不能表明是否真正存在某种有价值的一致性。流驱动的动力学是否拥有一些性质, 在同步、交通和传播中都能表现出来呢? 对我来说, 依然是完全没有头绪的!

(周涛)

4 网络同步的中尺度分析——同步从度大的区域开始

最近读了 CHAOS 20 010202(2010) 的专辑征稿文章: Announcement: Focus Issue on Mesoscales in Complex Networks, 觉得复杂网络中尺度(Mesoscales)问题的确是值得十分重视的研究方向, 正如征稿文章指出的“专辑的目的是通过对复杂网络中尺度层次的研究来推进非线性科学特别是复杂系统的研究”。复杂网络中尺度并不是今天才提出来的, 至少在 2006 年文献[21]就明确提出网络的中尺度问题, 不过这方面的成果似乎不多, 而这个问题又十分重要, 因此值得我们深入研究。

复杂网络可以在不同层次用不同尺度去刻画。在小尺度(微观、局部)层次,就是点和边的层次,可以讨论它的度、介数、聚类系数等属性,但是要刻画点和边的动力学就不能孤立静态地分析了。在大尺度(宏观、全局)层次,就是从网络的整体分析其参数的统计规律。文献[21]中指出,在这两种尺度中间,仍然存在一个包含各种不同尺度的很大的过渡空间,这就是所谓的中尺度。这些尺度可以理解为子结构(子图),相对于整个网络而言,子结构有自己的拓扑实体,例如模体(motifs)、小圈子(cliques)、核(cores)、环(loops),或者更一般地说成社团。这里对中尺度的定义已经有初步的界定。目前对于大尺度和小尺度两头的属性,研究得比较清楚,可是中间层次的属性,很多都还没有搞清楚,就譬如社团发现问题至今仍然存在着争议。实际系统的社团的确定很难做到是唯一的,原因就在于不同的尺度会有不同的划分。

复杂动力网络的同步已经得到广泛的研究,目前主要集中在给定拓扑结构和动力学后判断网络能否最终达到同步,主要利用主稳定方法和Lyapunov方法(少数利用图方法),它是从大尺度来分析网络节点的集体行为。可以说它是一种只看结果(最终是否同步)不顾过程的研究。至于同步过程表现出来的属性很少研究。实际上同步过程是很重要的,因为同步本来就是一个渐进过程,在网络趋于全局同步的过程中有没有什么有趣的现象发生,不同拓扑结构的网络的同步过程有什么差异,同步过程如何揭示拓扑尺度和发现社团,等等,对于它的研究有利于揭示复杂系统的演化机理,有利于理解不同复杂系统的时空差异。

研究网络的同步过程离不开网络中尺度层次,网络中尺度有利于揭示同步过程,网络同步过程的中尺度研究可以展示实际系统丰富的时空复杂性。要分析清楚同步过程,需要引进能够刻画同步过程的中尺度物理量。本节介绍几篇文献在这方面的研究。文献[22]讨论的是KM振子模型,首先引入了全局序参数和局部序参数,前者刻画全局同步程度,后者刻画网络中同步块所占的比例。通过计算ER随机网络和SF无标度网络在耦合强度增加时这两个参数的变化,发现在开始耦合强度较小时,全局序参数虽然为0,可是局部序参数已经增加,说明虽然全局不同步,但是局部同步已经开始;接下来随着耦合强度的增加SF网络的全局序参数迅速增加;当耦合强度继续增加时,ER网络的全局序参数和局部序参数都急剧增加而超过了SF网络,说明ER网络的同步是在比较短的时间内发生,不像SF网络同步所需要的时间那么长。文献又引入两个很有意思的参数:GC(最大同步块所包含的节点数目)、NC(局部同步块的数目)。发现在开始耦合强度较小时,GC值是SF网络大于ER网络,而NC值是ER网络大于SF网络,当耦合强度继续增加时,ER网络的GC值急剧增加而超过了SF网络,而ER网络的NC值急剧下降而低于SF网络。说明SF网络由于存在hubs节点,它的同步是以hubs为中心凝聚方式趋于全局同步,而ER网络由于节点度较均匀,它的同步是以许多小社团分别同步开始,然后在比较短的时间内这些小社团相互趋于全局同步;SF网络由于节点的不均匀性,大多数节点的度值较小,这些度很小的节点达到同步所需要的时间更长,所以造成SF网络完全同步所需时间大于ER网络。可见度分布的不同造成同步过程和方式的差异。文献计算了网络中不同度的节点属于最大同步块的概率,发现度大的节点即使在耦合强度较小时属于最大同步块的概率仍然很大,而度小的节点即使在耦合强度较大时属于最大同步块的概率仍然很小。

在同行工作的基础上,我们最近在中尺度意义下,也研究了不同拓扑结构的复杂网络的同步以及广义同步过程^[23-24]。现在我们可以概括为以下几点:1)不同结构的网络其同步方式和过程是完全不同的,SF网络的同步是以hubs为中心凝聚方式趋于全局同步,边缘点(度小的节点)影响同步时间,而ER网络的同步是以许多小社团分别同步开始然后相互趋于全局同步。SF无标度网络完全同步所需时间比ER随机网络长。2)同步是从度大的节点开始的,然后再蔓延到网络的其他节点,对于SF无标度网络尤其明显。由于实际网络大多是无标度网络,是否可以说现实的复杂系统的集群行为也是从密集处开始。3)社团网络的同步过程是从完全不同步到部分同步,再到聚类同步最后实现全局的完全同步;根据同步过程可以发现网络的拓扑结构,网络社团结构的发现依赖于尺度;分层社团结构的网络的同步过程与特征值谱的分布有关。

(陆君安)

5 复杂网络的加速增长现象

许多实际的网络具有加速增长特征。通过对万维网、因特网、科学家合作网、引文网络等的实证研究发

现,在观测的生长期,这些网络的平均度是随时间增加的,即边的数量比节点的数量增长更快,这种现象一般被称为加速增长^[25]。过去十年间,已有不少文献对加速增长网络的拓扑性质进行了模拟、解析或数值分析^[26-32],但对网络为何会加速增长这一问题还较少受到关注。本节将对这一问题进行粗浅的探讨。

网络规模的增长通常会表现在两个方面:一是节点数量的增加,二是边数量的增加。当有新的个体进入网络时,网络的节点数量会增加,同时至少会有一条边加入网络中。此处将节点数量理解为网络的演化时刻,则每个时刻网络边增加的数量可以表示为 $m(t)$,如果网络的边是加速增长的,则 $m(t)$ 应该是一个增函数^[25]。需要注意的是,每个时刻增加的边并不一定都是连接新加入节点的,网络中的“老”节点之间也可能产生新的连接^[27]。例如对于因特网,一个新设备通常是用一条线路而不是多条线路接入网络的。网络边数量较节点数量增加的快,更多的原因是网络中既有设备之间为增加吞吐量或提高速度而增加新的线路。这种现象在合作网络中存在的更为普遍,例如科学家论文合作网络,相当规模的论文是由已发表过论文的作者完成的,这种情况下网络甚至没有增加节点,而只是增加了边(或提高了边的权值)。对于以此种方式演化的网络,我们可以把其加速增长现象理解为:在网络“年轻”时,边更多地是以连接新节点的形式增加;而随着网络年龄的增长,越来越多的边会在老节点之间产生。

那么,为什么很多实际的网络中会有这种“初期以连接新节点为主、后期以连接老节点为主”的边加速增长特性呢?本节尝试用经济学中的边际效用递减理论^[33]来对其进行初步解释。边际效用递减理论认为,在消费者连续消费某种消费品时,随着所消费的该消费品的数量增加,其总效用虽然相应增加,但消费品的边际效用(即每消费一个单位的消费品所带来的效用增加量)有递减的趋势。如果我们认为网络是自组织的,那么可以把网络自身理解为一个消费者,把网络演化过程中边的增加理解为一次消费行为,而把边连接新节点还是老节点理解为两种消费品。在网络增加边的一次“消费行为”中,是选择新节点连接还是选择老节点连接,取决于二者中的哪个会带给网络自身更大的“效用”。而根据边际效用递减理论,连接新节点带给网络的效用会越来越小,网络会越来越不倾向于连接新节点,网络会呈现出边加速增长的趋势。例如对于公交网络,在网络发展早期,开辟线路的主要目的是覆盖公交服务盲区,那么连接新站点对提高网络整体效率的效用显然比连接旧站点更大;而随着网络发展,服务盲区越来越少,连接新站点的效用会递减,而减少既有站点间的换乘次数对提高网络整体效率的效用会加大,因此线路更倾向于在既有站点间连接。这样,公交网络的节点增速降低,而边的增速提高,网络会加速增长。对于科研合作网络也可以进行类似的解释:当新出现一个研究方向时,发表论文的“门槛”可能较低,也就是说新节点进入网络的“成本”(可以理解为负的效用)较低;但随着这一方向研究的不断深入,新研究者发表论文的“成本”会越来越高,而“老”研究者由于有了更多的积累,发表论文的“成本”会越来越低。网络节点的增加会减速,网络呈现边加速生长趋势。

当然,以上解释只能说是一种类比,对于驱动网络加速增长的内在机制我们还并不十分了解,尚需结合各种实际网络的具体情况进行更深入的分析(最近琳媛、子柯和周涛提出了一种新的观点,即认为无标度网络要保持度指数的稳定,必须采取加速的增长模式^[34])。但本节从经济学角度所给出的这些解释,为开发更符合实际的网络演化模型提供一种新的视角,有助于更深入地理解复杂网络的演化机制。

(闫小勇)

6 复杂网络中的链路预测

复杂网络研究近几年发展迅速,可谓枝繁叶茂。六七年前,如果你问一个人的研究领域,他说“复杂网络”。你大概就知道他是干什么的了。如果现在有人问你,你做什么的,如果你只说“复杂网络”人家就会问“具体哪个方向”,然后你说“网络上的动力学”,人家又会继续问“同步,传播,交通,博弈还是什么呢”。可见,如今的复杂网络研究已经不仅仅局限于自身理论和方法的研究,而是成为多学科交叉的研究方向和多学科研究的有力工具。例如生物学中的新陈代谢网络、蛋白质相互作用网络、神经网络;社会学中的性关系网络、朋友关系网络、合作网络;交通系统中的航空网络、铁路及公路网;信息系统中的互联网、万维网等等。复杂网络研究已经与生物学、社会学、交通科学等产生了深刻的联系,而链路预测则是将复杂网络与信息科学

联系起来的桥梁。

网络中的链路预测是指通过已知的网络结构信息预测网络中尚未产生连边的两个节点之间产生链接的可能性^[35]。该研究在理论和应用两方面都具有重要意义。

首先, 链路预测的研究可以从理论上帮助我们认识复杂网络演化的机制。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网络演化模型提供一个简单统一且较为公平的比较平台。类似地, 如何刻画网络中节点的相似性也是一个重大的理论问题。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, 链路预测可以一展身手。链路预测本身也带来了有趣且有重要价值的理论问题, 也就是如何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行性分析。这方面的研究有望建立链路预测的统计力学基础, 迈出从统计力学的观点理解数据挖掘的第一步。

链路预测还具有重要的应用价值。以蛋白质相互作用网络和新陈代谢网络为例, 节点之间是否存在相互作用关系, 是需要通过大量实验结果进行推断的。我们已知的实验结果仅仅揭示了巨大网络的冰山一角。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。如果能够事先在已知网络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, 就有可能提高实验的成功率从而降低试验成本并加快揭开这类网络真实面目的步伐! 另外, 在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工具。链路预测算法还可以用于分析演化网络, 即对未来可能连边的预测, 如在线社交网络的朋友预测(推荐)问题。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或者判断一个手机用户是否产生了切换运营商(例如从移动到联通)的念头^[36]。此外, 关于错误链接的预测对于网络重组和结构功能优化也有重要的应用价值。例如在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, 我们就有可能应用链路预测的方法对其进行纠正^[37]。

链路预测作为数据挖掘领域的研究方向之一在计算机领域已有一些早期的研究。他们的研究思路和方法主要基于马尔科夫链和机器学习。这些方法往往要涉及节点的属性信息。虽然应用节点属性等外部信息的确可以得到很好的预测效果, 但是很多情况下这些信息的获得是非常困难的, 甚至是不可能的。比如很多在线系统的用户信息都是保密的。另外即使获得了节点的属性信息也很难保证信息的可靠性, 即这些属性是否反映了节点的真实情况。更进一步, 在能够得到节点属性的精确信息的情况下, 如何鉴别出哪些信息对网络的链路预测是有用的, 哪些信息是没用的仍然是个问题。因此与节点属性信息相比较, 已观察到的网络结构或者用户的历史信息更容易获得也是更可靠的。

基于节点相似性的链路预测方法最近受到了广泛的关注。此方法的一个重要前提假设就是两个节点之间相似性(或者相近性)越大, 它们之间存在链接的可能性就越大。因此如何定义节点的相似性就成为该方法的一个核心问题。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或者是基于图论的矩阵森林方法。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^[38]提出了基于网络拓扑结构的相似性定义方法, 并分析了若干指标在社会合作网络中链路预测的效果。周涛、吕琳媛和张翼成^[39]在6种不同网络中比较了9种已知的基于局部信息的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9种已知指标的预测能力。特别地, 最近的研究结果显示, 新提出来的相似性指标在进行群落划分^[40]和含权网络权重设置^[41]的时候也比原有指标好。吕琳媛、金慈航和周涛^[42]进一步在噪音强度以及网络密度可控的模型中细致分析了局部路径指标的性能, 发现这个指标在网络的平均最短路径较小的时候具有与全局结构指标可匹敌的预测能力, 甚至在噪声较大的情况下可以预测得更加准确。刘伟平和吕琳媛^[43]提出了两种基于网络局部随机游走的相似性指标, 通过与其他5种相似性指标的比较, 发现有限步的随机游走可以给出比全局收敛后的预测精度更好的结果。

链路预测另一类方法是基于最大似然估计的。Clauset, Moore 和 Newman^[44]认为很多网络的连接可以看作某种内在的层次结构的反映。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^[37]假设我们观察到的网络是一个随机分块模型(Stochastic Block Model)的一次实现,在该模型中节点被分为若干集合,两个节点间连接的概率只和相应的集合有关。他们所提出的基于随机分块模型的链路预测方法,可以得到比文献[44]更好的结果。与此同时,该方法不仅可以预测缺失边,还可以预测网络的错误链接。所有最大似然估计的方法都存在一个共同的问题,就是计算非常耗时,只能处理很小的网络。

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^[45]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”有深刻的关联。Leskovec、Huttenlocher 和 Kleinberg^[46]则注意到了近期“社交平衡理论”的定量化研究成果,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。

借助复杂网络的理论和方法从统计物理的角度来考虑链路预测问题给出了全新的视角。现有的工作只是起点,还有很多需要进一步研究和完善的地方,如对于大量算法在各种不同网络中的预测能力的系统分析尚欠,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(Multi-Partite Networks)的讨论虽然有,但非常少,也不系统。综上所述,我们认为链路预测在未来的一段时间里可以围绕以下几点进行研究:

1) 丰富和提高现有相似性预测的算法,特别是针对有向网络、含权网络、多部分网络、含异质边的网络等较复杂的情形,提出新的相似性指标。

2) 对已知算法的性能进行深入细致的分析,揭示算法性能和网络结构特征之间的关系,希望得到各种算法在不同网络中的可预测性极限。

3) 利用网络系综和最大似然估计的思想和技术,建立相似性框架下链路预测的理论基础。

4) 基于链路预测的思想,建立一个比较和评价不同演化模型优劣的平台。

5) 实现有代表性的链路预测的应用研究,并开展自适应性的快速算法研究以实现在巨大规模的实际系统中的应用。

(吕琳媛)

7 物理方法在推荐系统中的应用

但凡看过情景喜剧《天才也性感(The Big Bang Theory)》的人,我想没有不被剧中人物谢尔顿(Sheldon)的那种与生俱来的古灵精怪、单纯的性格和执着的信念所吸引的。理论物理学家这个职业的具体内容也随着谢尔顿的出色演出而逐渐被广大观众所熟知。该剧由浅入深的将我们日常生活中所存在的大量物理学现象娓娓道来,从而使得非物理专业人士也能快速理解各种生活现象背后看似深奥的原理。热传导(Heat Conduction)便是众多普遍存在的物理现象中的一种。所谓热传导,是热传递三种方式(即热传导、热对流和热辐射)中的一种。它的工作原理非常简单:当两个不同温度的物体相接触时,热量会从温度高的物体传递到温度低的物体上,直到两者的温度相同,所以热传导方法有助于提高系统中低温度物体的温度。而我们知道,在信息极为丰富的互联网中,对于用户来说,最为迫切问题就是如何帮助他们找到那些他们所感兴趣但不易找到的信息。而这也正是推荐系统所最为关注的问题之一。假设以精确性为衡量一个推荐系统好坏标准的话,那么只要将算法设计得更加容易推荐那些热门的物品即可。举个例子来说,对于一个电影网站,如果一味的倾向于向用户推荐《阿凡达》、《功夫熊猫》之类的热门大片,固然用户会喜欢,推荐的精度也会很高。但这样的推荐结果对于用户来讲是没有任何信息含量的(因为大家早就通过各种渠道了解了)。反之,如果能够推荐一些适合用户喜好的,但鲜有人关注而用户还不知道的影片。因为长尾效应的存在,推荐那些被收藏次数少、质量高的影片反而能起到“四两拨千斤”般的“惊艳”效果,从而提高用户对系统的信任

和黏着性。正是基于这些考虑,近来一些物理学家尝试将热传导的方法应用到推荐系统中来,期望可以利用热量传递的原理,合理的提高温度较低物体的温度,更有利用推荐算法来发现那些不易被用户所察觉的“冷点”信息^[47-48](即被收藏次数较少的物品)。

有了这个想法,在没有其他额外信息(如用户属性、物品属性、描述等)的情况下,利用网络结构(只有节点和连边)来实施基于热传导的推荐方法便成了第一选择。事实上,对于一个固定的网络结构来说,是比较容易应用热传导方法的。因为网络中的节点可以看做是物体,而是否有连边则可看做是两个物体是否有接触。能量只会在有连边的两个节点之间进行传递。那些被收藏次数多的物品可以看做是温度较高的“热点”,被收藏次数少的物品则可以看做是温度较低的“冷点”。能量根据连边的由温度高的节点流向温度低的节点。不难想象,只要给予足够长的时间,所有节点都会达到相同的温度。这种稳态将最大限度地发掘出所有隐蔽的“暗信息”。但很可惜,面对所有温度相同的物体,此时任何推荐算法都会一筹莫展:如何从中选取合适的物品给用户呢?于是,精确性和多样性便组成了一把双刃剑,综合起来衡量推荐结果的质量。具体来说,就是考虑温度传递的步数与推荐效果的关系。从目前的实验结果来看,在用户-物品组成的二部图中,两步传递会得到较好的推荐结果。多步传递由于涉及到了重复的全局信息,在没有考虑这些重复信息的负作用时,将会得到比较差的推荐效果。因此,从简单和便于实现的角度来说,两步传递是目前热传导方法所采用的主要方法。

另外,物理学中还有一种被称之为物质扩散(Mass Diffusion)的方法也被广泛地应用到推荐系统中了^[49-53]。从本质上来讲,物质扩散等同于推荐系统的常用的随机游走(Random Walk)方法,只是不同学科对相同方法的不同称呼而已。基于物质扩散和基于热传导的推荐算法的区别在于:基于物质扩散的方法在进行个性化推荐时,系统的总能量是保持不变即守恒的;而热传导在推荐过程中,目标用户(即被推荐用户)的收藏品将被视作恒温热源,源源不断的给系统提供能量,所以系统的总能量随着传递步骤的增加是在不断增加的。换言之,对于物质扩散,相当于有固定的初始能量在系统中传递,最后的系统稳态结果是和节点度(即物品被收藏数目)成正比的,所以它倾向于推荐那些度较大(较流行)的物品,相当于一个凸透镜,将用户的视野汇聚在那些较流行的节点上,从而也就不难理解这种方法会对提高推荐的精确性有很大帮助。而对于热传导,因为热源存在的缘故,从而保证系统中有足够的能量可以传递到那些“冷点”上。也正是这个热源的存在,导致系统的最终稳态结果是所有节点温度相同,所以相对于物质扩散来说,热传导倾向于推荐那些度较小(较不流行)的节点,相当于一个凹透镜,把用户的视野发散到了那些较不流行的物品上,从而提高了推荐的多样性^[48-53]。文献[48]将两者结合起来设计了一套行之有效的混合算法,发挥二者的优势,同时在精确性和多样性上提高了推荐算法的性能。

目前的科学研究越来越向交叉科学方向发展,各领域相互学习、借鉴和渗透的趋势也越来越明显。在为各学科提供新鲜血液的同时,也由此诞生了很多新兴学科和研究方向,如生物物理、信息物理、金融物理、经济地理等。物理学作为一门基础性学科,其解释日常生活现象背后原理的优势使得它更容易地被其他学科所认同和接受,并迅速在各学科得到广泛应用。因此我们有理由相信,除了热传导和物质扩散,会有其他的物理学方法和原理将被应用到推荐系统中来,为这一新兴领域的发展壮大添砖加瓦。

(张子柯)

8 信息物理系统与网络科学

信息物理系统(Cyber physical system, CPS)是集计算、通信与控制于一体的下一代智能系统,是计算进程和物理进程的统一体。该系统通过人机交互接口实现和物理进程的交互,使用网络化空间以远程的、可靠的、实时的、安全的、协作的方式操控一个物理实体。CPS包含了将来无处不在的环境感知、嵌入式计算、网络通信和网络控制等系统工程,使物理系统具有计算、通信、精确控制、远程协作和自治功能。它注重计算资源与物理资源的紧密结合与协调,其涉及应用领域非常广泛,包括智能交通系统、远程医疗、智能电网、航空航天等多个领域^[54-55]。如果说Internet的普及给人类社会带来了翻天覆地的变化,那么CPS的普及将给人

类带来另一次工业革命。目前,CPS的研究才刚刚起步,对研究者来说机遇与挑战并存。本节将简要介绍CPS的特点、CPS与物联网的关系、CPS面临的机遇与挑战、CPS对复杂网络研究的影响。

8.1 CPS的特点

从20世纪40年代麻省理工学院发明数控技术到如今基于嵌入式计算系统的工业控制系统遍地开花,工业自动化技术早已成熟,并且这种技术已经广泛地应用到了各行各业的生产线上。但是,这些控制系统基本是封闭的系统,即便其中一些工控应用网络也具有联网和通信的功能,但其工控网络内部总线大都使用当前的工业控制总线,网络内部各个独立的子系统或者说设备难以通过开放总线或者互联网进行连接,其通信距离很短,通信功能比较弱。而CPS则把通信放在与计算和控制同等地位上,这是因为CPS强调的分布式应用系统中物理设备之间的协调是离不开通信的。CPS对网络内部设备的远程协调能力、自治能力、控制对象的种类和数量,特别是网络规模上远远超过现有的工控网络。美国国家科学基金会(NSF)认为,CPS将让整个世界互联起来,如同互联网改变了人与人的互动一样,CPS将会改变我们与物理世界的互动。2007年7月,美国总统科学技术顾问委员会(PCAST)在题为《挑战下的领先——竞争世界中的信息技术研发》的报告中列出了八大关键的信息技术,其中CPS位列首位。

8.2 CPS与物联网的关系

物联网是在计算机互联网的基础上,利用RFID、无线数据通信等技术,把世界上万事万物连接起来的网络。在这个网络中,物品(商品)能够彼此进行“交流”,而无需人的干预。其实质是利用射频自动识别(RFID)技术,通过计算机互联网实现物品(商品)的自动识别和信息的互联与共享。当人们还陶醉在物联网能够把物与物连在一起时,没想到CPS又冒了出来,而物联网所擅长的基于RFID的连接,对于CPS来说太过简单。在很多应用中,CPS对接入网络的设备的计算能力的要求远非RFID能比。以基于CPS的智能交通系统为例,虽然目前人们使用的汽车都嵌入了各种具有计算功能的电子系统,但是这些嵌入式系统的计算能力还远未达到智能交通系统对汽车之间的协同能力的要求。事实上,满足CPS要求的汽车电子系统的计算通常都是海量运算^[56]。海量运算往往是很多CPS接入设备的特征,因此,接入设备通常具有强大的计算能力。如果从计算性能的角度出发,把一些高端的CPS应用比作胖客户机/服务器架构的话,那么物联网则可视作超级瘦客户机服务器^[57],由于物联网中的物品不具备控制和自治能力,通信也大都发生在物品与服务器之间,因此物品之间无法进行协同。

8.3 CPS面临的机遇与挑战

如果物联网的市场规模像人们所说的有上万亿元,那么,CPS的市场规模则难以计数,因为CPS涵盖了小到智能家庭网络大到工业控制系统乃至智能交通系统等国家级甚至世界级的应用。更为重要的是,这种涵盖并不仅仅是将现有的家电简单地连在一起,而是要催生众多具有计算、通信、控制、协同和自治性能的设备。下一代工业将建立在CPS之上,随着CPS技术的发展和普及,使用计算机和网络实现功能扩展的物理设备无处不在,并将推动工业产品和技术的升级换代,极大地提高汽车、航空航天、国防、工业自动化、健康/医疗设备、重大基础设施等主要工业领域的竞争力。CPS不仅会催生出新的工业,甚至会重新排列现有产业布局。尽管CPS前景无限,但其带来的挑战也是物联网所无法比拟的。CPS所面临的挑战主要来自以下几个方面:1) 计算速度,由于很多CPS需要海量运算,当前嵌入式系统的计算速度还无法满足这一要求,可以想象在智能交通系统中高速运动的汽车长时间等待一条指令的情况下时将会发生什么危险事情。2) 网络通信效率,就目前的互联网来说,在网络拥堵的情况下点击一个页面出现几秒钟延迟是常见的事情,然而在CPS中,如果一辆汽车需要立刻采取刹车措施时迟迟得不到指令是很危险的。3) 研究条件,从CPS的特点我们已经知道它是一个横跨多个学科领域,尽管人们已经意识到学科交叉的重要性,但在目前的教育和评价体制下,学科之间的融合显得缺乏动力。由于工业界和学术界对科研成果的评价指标不一致,使得高校和科研机构不愿投入时间和资源来研究和弄懂工业界所遇到的实际问题。

8.4 复杂网络在CPS中的应用

复杂网络在经历了10年的发展之后,目前正向着各个应用领域前进,CPS的出现给复杂网络领域的研

究者带来了无限机遇。CPS 最大的特点就在于它是由很多具有通信、计算和决策控制功能的设备组成的智能网络,这些设备可以通过相互作用使得整个系统处于最佳状态。例如在机器人足球比赛中,当某个机器人准备传球时,它会收集每个同伴的信息,然后通过计算得出一个最佳的传球方案,并且将该方案传给所有队员,让队员们配合这个传球过程,通过这种方式可以提高整个球队的水平。该思想同样可以应用于很多其他的系统,如交通系统中车辆之间通过通信和计算得出最佳行车路线,并避免各种交通事故;电力系统中各个站点通过信息传递从而动态调整负荷,避免大规模级联故障等。这些系统的运行其实就是复杂网络的动力学过程,只不过这些过程集合了复杂网络中的信息传播、同步、博弈等多种动力学过程,研究的内容也更加复杂。一方面,CPS 的出现给复杂网络提供了大量的研究课题;另一方面,复杂网络的研究必将极大地推进 CPS 的研发和应用进程。我们相信,CPS 和复杂网络的结合必将加速推动科技进步的步伐并给人类的生产 and 生活方式带来巨变。

(周海平)

9 人在互联网上的行为动力学

追寻人的行为规律是科学家们梦寐以求的理想,长期以来,由于人的行为的复杂性以及缺少记录,对人的行为规律的研究一直停留在定性分析阶段或者简单的认为人的行为就是一个没有规律的泊松随机过程。研究人的行为规律的重要性不言而喻,很多自然、经济、科技等行为是受人的行为驱动的。随着 2005 年,Barabási 关于人写信规律的 Nature 文章^[58-59]出来后,有一批科学家重新进入了这个领域,由于现代科技的发展,人的某些行为已经有被记录下来,可以用来研究人的行为规律。Barabási 主要关注的是等待时间以及 inter-event 时间,这两个物理量在人的行为特点里起很大的作用。关于这两个刻画人的行为的物理量,在很多领域已经被深入研究过了,比如经济系统里连续两次交易之间的时间间隔,地球系统里连续两次地震之间的时间间隔,气候系统里连续两次下雨或者极端气候之间的时间间隔等。在这些系统里,都已经发现了与泊松过程不同的性质,也就是说 inter-event 分布不是指数分布,而是 power-law 或者 Weill 等分布,但由于人的行为缺少记录,很少人来关注在人的行为里去寻找有没有规律,直到 Barabási 听说他们是人工把那几个人写的信的记录一条条找出来统计出来的规律。

因特网作为一个虚拟社会,人在因特网上的行为与现实生活的行为具有共性,研究其规律同样非常重要,同时因为因特网具有实时记录特点,研究人在因特网上的行为特点必将引起社会学家、统计学家、物理学家以及网络设计者的关注。目前关于人在网上的行为的定量研究还局限于网站访问的 inter-event time 等,定性研究倒是有一些,比如关于网络舆论的形成、传播等,但局限于定性或者半定量的统计性质分析,缺少理论以及背后机制分析。下面几个方向或许会成为热门的研究方向。

人在“BBS”上的行为规律分析,首先是抛弃一切内容的纯物理分析内在规律并建模,不局限于 inter-event 分析,而挖掘更多的刻画人的行为规律的物理量。其次结合部分内容分析,把人在“BBS”上的行为模式分类,可以分为自组织、外力驱动等模式,可用于识别真假舆论等。

由于人在“BBS”上的每一句话都有记录下来,所以结合语义分析,可以得到网络舆论形成的内在机制以及详细的过程特点。举个简单的例子,一个帖子后面的回复可以分成 3 类:正面、反面,以及中立,通过分析可以定量地得到后面的帖子是怎样受到前面帖子性质的影响,得到一个舆论形成的详细过程。

人在网络上参与的舆论传播行为,在宏观角度上,把一个站点,或者一个站点的分站定义为一个复杂网络的节点,以某个突发事件为研究对象,互相转载认为边,在这种网络上,一个事件的传播是一种复杂的行为,事件从 A 站点起源,传出到 B 站点,同时 B 站点的用户看到后,回到 A 站点追源,减慢了事件在 A 站点的影响衰减。这与生活中的流行病传播具有共性。我把病传给你,我好了,你没好,你会把病又传给我。所以定量分析这种相互作用过程,不但有助于理解网络舆论传播,而且对现实流行病传播过程的理解也有帮助。

最重要的一点,现在的复杂网络或者人的行为动力学的研究都是后知研究,通俗说叫事后诸葛亮,都是对过去记录的数据进行分析,得到其中的规律,并建模分析等。而对社会学家或者动物学家或心理学家来

说,还有一种研究方法,那就是为了验证某一个推测或者理论,人为设计一套合理的实验来验证。复杂网络的六度分离已经得到了实验的验证,但还有很多理论或推测没有得到验证。由于因特网的开放性、广泛性以及自组织特点,还有成本低,会是一块很好的实验田。

(吴 晔)

10 密切关注和应对“网络中心战”

网络科学分析为复杂系统研究提供了一个的重要平台和有力武器,有助于人类应对重大社会的和多方面的挑战,但是网络科学的研究成果与任何事物一样具有两面性:它既可以造福人类,又可以祸害人类。因此,需要我们密切关注和应对。

这让我联想到二战投在日本的二颗原子弹造成人类的空前悲剧。1945年8月6日美国B-29轰炸机“埃诺拉·盖伊”号在广岛投下一枚代号“小男孩”的原子弹,造成死亡14万人。1945年8月9日:第二颗代号“胖子”的原子弹在长崎爆炸,造成约8万人死亡。负责运送原子弹的“印第安纳波利斯”号巡洋舰的舰长战后得知自己运送了原子弹,同时也了解到原子弹的破坏力后,最终由于无法忍受自己的罪恶,而在家中开枪自尽!负责创建洛斯阿拉莫斯(武器)实验室和“曼哈顿工程”的“美国原子弹之父”奥本海默也因此类似的主因而自杀!

历史的教训值得注意。当年,曼哈顿工程虽然是一项人类伟大的创举,它开创了科学、军事和工业的三位一体的“大科学”典范,它是一个具有领导统一、组织严密、分工明确、高效协调的复杂网络,提供了原子弹成功研制保证。但是它同时也是一个生命、死亡和战争的疯狂竞赛。曼哈顿工程既揭开了原子能的神奇之谜,也第一次把最具破坏力的武器——原子弹赋予了人类。

2009年《科学》专辑上John Bohannon关于“调查网络科学:它的黑暗面”一文的最重要贡献是:大胆地揭开了网络分析的黑暗面,揭露了美国军方利用网络分析犯下的罪行,他直截了当地指出:美国“国家安全局对从e-mail和电话通话中获得的海量信息数据库进行挖掘。”“这种做法并不关心一个被拘留者是否是无辜的”。美国西北大学的网络科学家Brian Uzzi指出“所有强有力的(网络分析)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。”“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。”据推测:被关押在伊拉克和阿富汗的人数为5万人。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。事实却是如此,这是多么令人可怕的情形!事实充分说明网络科学的分析方法被美国军方不恰当地应用于“反恐”方面,并被不负责任地错误地扩大化到无辜的广大普通民众生活。这就是目前网络科学应用出现的一个黑暗面!这个反面教训是非常深刻的。

更应引起全世界人民的警惕和重视的是,美国正在大搞的未来“网络中心战”!它可能产生比原子弹更大的“爆炸”威力和更严重的不可预测后果。我认为,现代网络科学必须接受“曼哈顿(原子弹)工程”的历史深刻教训,不仅在“反恐”网络分析方面,而且特别是“网络中心战”方面,不应被各国军方不负责任地扩大化和效仿。任何负责的国家都要认真研究长远之计,研究应对这种可能发生的方法、策略和措施等。今后全世界科学工作者不可等闲视之,都应该联合起来,行动起来,绝不能让历史的悲剧重演!

(方锦清)

参考文献:

- [1] Watts D J, Strogatz S H. Collective dynamics of small-world networks [J]. Nature, 1998, 393: 440 - 442.
- [2] Barabási A-L, Albert R. Emergence of scaling in random networks [J]. Science, 1999, 286: 509 - 512.
- [3] Barabási A-L, Ravasz E, Vicsek T. Deterministic scale-free networks [J]. Physica A, 2001, 229: 559 - 564.
- [4] Farkas I, Derényi I, Jeong H, et al. Networks in life: scaling properties and eigenvalue spectra [J]. Physica A, 2002, 314: 25 - 34.
- [5] Ravasz E, Barabási A-L. Hierarchical organization in complex networks [J]. Phys Rev E, 2003, 67: 026112.

- [6] Ravasz E , Somera A L , Mongru D A , et al. Hierarchical organization of modularity in metabolic networks [J]. *Science* ,2002 , 297: 1551 – 1555.
- [7] Dorogovtsev S N , Gotsev A V , Mendes J F F. Pseudofractal scale-free web [J]. *Phys Rev E* ,2002 65 : 066122.
- [8] Jung S , Kim S , Kahng B. Geometric fractal growth model for scale free networks [J]. *Phys Rev E* ,2002 65: 056101.
- [9] Zhang Z Z , Rong L L , Zhou S G. A geometric fractal growth model for pseudofractal scale-free web [J]. *Physica A* ,2007 377: 329 – 339.
- [10] Andrade J S , Herrmann H J , Andrade R F S , et al. Apollonian networks: simultaneously scale-free , small world , euclidean , space filling , and with matching graphs [J]. *Phys Rev Lett* ,2005 94: 018702.
- [11] Zhang Z Z , Rong L L , Zhou S G. Evolving apollonian networks with small-world scale-free topologies [J]. *Phys Rev E* ,2006 , 74: 046105.
- [12] Guo J L. Scale-free networks with self-similarity degree exponents [J]. *Chin Phys Lett* ,2010 27 : 038901.
- [13] Andrade J S , Herrmann H J , Andrade R F S , et al. Erratum: apollonian networks: simultaneously scale-free , small world , euclidean , space filling , and with matching graphs [J]. *Phys Rev Lett* ,2009 , 102: 079901.
- [14] Li L , Alderson D , Doyle J C , et al. Towards a theory of scale-free graphs: definitions , properties , and implications [J]. *Internet Math* ,2005 2 : 431 – 523.
- [15] Motter A E , Zhou C S , Kurths J. Network synchronization , diffusion , and the paradox of heterogeneity [J]. *Phys Rev E* ,71 , 2005: 016116.
- [16] Nishikawa T , Motter A E , Lai Y C , et al. Heterogeneity in oscillator networks: are smaller worlds easier to synchronize [J]. *Phys Rev Lett* ,2003 91: 014101.
- [17] Zhao M , Zhou T , Chen G R , Wang B H. Enhancing the network synchronizability [J]. *Front Phys China* ,2007 2 : 460.
- [18] Wang W X , Wang B H , Yin C Y , et al. Traffic dynamics based on local routing protocol on scale-free networks [J]. *Phys Rev E* ,2006 73: 026111.
- [19] Yan G , Zhou T , Hu B , et al. Efficient routing on complex networks [J]. *Phys Rev E* ,2006 73 : 046108.
- [20] Yang R , Zhou T , Xie Y B , et al. Optimal contact process on complex networks [J]. *Phys Rev E* ,2008 78: 066109.
- [21] Arenas A , Diaz-Guilera A , Perez-Vicente C J. Synchronization processes in complex networks [J]. *Physica D* ,2006 224: 27 – 34.
- [22] Gomez-Gardenes J , Moreno Y , Arenas A. Paths to synchronization on complex networks [J]. *Phys Rev Lett* ,2007 98: 034101.
- [23] Chen J , Lu J A , Wu X , et al. Generalized synchronization of complex dynamical networks via impulsive control [J]. *Chaos* , 2009 ,19: 043119.
- [24] Liu H , Chen J , Lu J A , et al. Generalized synchronization in complex dynamical networks via adaptive couplings [J]. *Physica A* ,2010 389: 1759 – 1770.
- [25] Dorogovtsev S N , Mendes J F F. Evolution of networks [J]. *Adv Phys* ,2002 51: 1079.
- [26] Barabási A-L , Jeong H , Néda Z , et al. Evolution of the social network of scientific collaboration [J]. *Physica A* ,2002 311: 590.
- [27] Sen P. Accelerated growth in outgoing links in evolving networks: deterministic versus stochastic picture [J]. *Phys Rev E* , 2004 69 : 046107.
- [28] Shi D H , Chen Q H , Liu L M. Markov chain-based numerical method for degree distributions of growing networks [J]. *Phys Rev E* ,2005 71 : 036140.
- [29] Gagen M J , Mattick J S. Accelerating , hyperaccelerating , and decelerating networks [J]. *Phys Rev E* ,2005 72: 016123.
- [30] Yuan W J , Luo X S , Zhou J F , et al. Acceleratingly growing scale-free networks with tunable degree exponents [J]. *Physica A* , 2008 387 : 5311 – 5316.
- [31] Smith D M D , Onnela J P , Jones N S. Master-equation analysis of accelerating networks [J]. *Phys Rev E* ,2009 79 : 056101.
- [32] Zhang Z Z , Fang L J , Zhou S G , et al. Effects of accelerating growth on the evolution of weighted complex networks [J]. *Physica A* ,2009 388: 225 – 232.
- [33] Varian H R. *Intermediate Microeconomics: a Modern Approach* [M]. New York: W W Norton & Company ,2006: 65.
- [34] Lü L , Zhang Z K , Zhou T. Zipf’s law leads to heaps’ law : analyzing their relation in finite-size systems [DB/OL]. [2010 – 05 – 03]. <http://arxiv.org/abs/1002.3861>.

- [35] 吕琳媛. 复杂网络链路预测[J]. 电子科技大学学报: 自然科学版, 2010, 39(5): 651–661.
Lü Linyuan. Link prediction on complex networks [J]. Journal of University of Electronic Science and Technology of China: Natural Science Edition, 2010, 39(5): 651–661.
- [36] Zhang Q M, Shang M S, Lü L. Similarity-based classification in partial labeled networks [J]. Int J Mod Phys C, 2010, 21: 813.
- [37] Guimera R, Sales-Pardo M. Missing and spurious interactions and the reconstruction of complex networks [J]. PNAS, 2009, 106: 22073.
- [38] Liben-Nowell D, Kleinberg J. The link-prediction problem for social networks [J]. J Am Soc Inform Sci Technol, 2007, 58: 1019.
- [39] Zhou T, Lü L, Zhang Y C. Predicting missing links via local information [J]. Eur Phys J B, 2009, 71: 623.
- [40] Pan Y, Li D H, Liu J G, et al. Detecting community structure in complex networks via node similarity [J]. Physica A, 2010, 389: 2849.
- [41] 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: 2949.
- [42] Lü L, Jin C H, Zhou T. Similarity index based on local paths for link prediction of complex networks [J]. Phys Rev E, 2009, 80: 046122.
- [43] Liu W P, Lü L. Link prediction based on local random walk [J]. Europhys Lett, 2010, 89: 58007.
- [44] Clauset A, Moore C, Newman M E J. Hierarchical structure and the prediction of missing links in networks [J]. Nature, 2008, 453: 98.
- [45] Lü L, Zhou T. Link prediction in weighted networks: the role of weak ties [J]. Europhys Lett, 2010, 89: 18001.
- [46] Leskovec J, Huttenlocher D, Kleinberg J. Predicting positive and negative links in online social networks [C]. Proc WWW 2010. New York: ACM, 2010.
- [47] Zhang Y C, Blattner M, Yu Y K. Heat conduction process on community networks as a recommendation model [J]. Phys Rev Lett, 2007, 99: 154301.
- [48] Zhou T, Kuscsik Z, Liu J G, et al. Hybrid algorithms to customize and optimize diversity and accuracy of recommendations [J]. PNAS, 2010, 107: 4511.
- [49] Zhang Y C, Medo M, Ren J, et al. Recommendation model based on opinion diffusion [J]. EPL, 2007, 80: 68003.
- [50] Zhou T, Ren J, Medo M, et al. Bipartite network projection and personal recommendation [J]. Phys Rev E, 2007, 176: 046115.
- [51] Zhou T, Jiang L L, Su R Q, et al. Effect of initial configuration on network-based recommendation [J]. EPL, 2008, 81: 58004.
- [52] Zhang Z K, Zhou T, Zhang Y C. Personalized recommendation via integrated diffusion on user-item-tag tripartite graphs [J]. Physica A, 2010, 389: 179.
- [53] Zhou T, Su R Q, Liu R R, et al. Accurate and diverse recommendations via eliminating redundant correlations [J]. New J Phys, 2009, 11: 123008.
- [54] Stankovic J, Lee I, Mok A, et al. Opportunities and obligations for physical computing systems [J]. IEEE Computer, 2005, 38(11): 23–31.
- [55] Edward L. Cyber physical systems: design challenges [C]. Proc 11th IEEE international symposium on object oriented real-time distributed computing. USA: IEEE Press, 2008.
- [56] Mohan P, Padmanabhan V, Ramjee R. Nericell: rich monitoring of road and traffic conditions using mobile smartphones [C]. Proc 6th ACM Conference on Embedded Network Sensor Systems. New York: ACM Press, 2008.
- [57] Wang M, Cao J, Li J, et al. Middleware for wireless sensor networks: a survey [J]. J Comput Sci Technol, 2008, 23: 305–326.
- [58] Barabási A-L. The origin of bursts and heavy tails in human dynamics [J]. Nature, 2005, 435: 207–211.
- [59] Oliveira J G, Barabási A-L. Darwin and Einstein correspondence patterns [J]. Nature, 2005, 437: 1251.