互联网新模型和网络地理学

李　强

（南京大学信息管理系　南京　210093）

摘　要　主要介绍了互联网最新的无尺度网络和蝴蝶结模型及网络地理学起源和成果，也论证网络计量学的发展需要专业计量工具。

关键词　网络模型　无尺度空间　网络地理学　网络计量

从1969年仅有4个节点的分组交换网ARPAnet到遍及全球的Internet，30多年间互联网以几何级数扩张，已被公认为第四媒体，改变全球亿万人的生活方式。通常我们所说的Internet有两层含义：一个是由众多计算机、路由器、通信线路组成的有形的开放物理网，通过一定的协议可实现多种信息交流，如电子邮件、Telnet、Gopher等；另一个则专指它最主要的用途——网页通过链接形成的虚拟的万维网(WWW)。互联网如此宽广无序，我们不禁想知道在它杂乱无章的背后究竟是怎样的呢？

1　互联网的新模型

1.1　*互联网是无尺度网络*　20世纪60年代描述大型通信网络和神经网络的典型模型是匈牙利爱尔特希(Paul Erdos)等人建立的随机模型。该模型中的节点遵从泊松分布，大部分节点的链接数接近平均值。IBM阿尔马登研究中心、AltaVista公司研究小组的布罗德尔(Andrei Broder)等人运用"网络蜘蛛人"调查了2亿个网页和15亿个超链接后，认为万维网具有非同构性或非均匀性，不是随机网络。万维网节点的链接数符合帕雷托分布，而不是通常认为的泊松分布，万维网是无尺度网络。

什么是无尺度网络？空中交通网就是无尺度网络。其特点是连接大部分中小城市机场的航线只有几条，最大的航空港如北京、巴黎、纽约的航线却可直达全国乃至全球。万维网中绝大部分节点的链接数远远小于平均链接数，极少数的门户网站、政府网站、企业网站的网页被众多网页链接。美国加州大学的法劳特索斯(Michalis Faloutsos)等人在路由器和域的层次上分析认为，物理节点的连接数也遵从帕雷托分布，也是无尺度网络。

无尺度网络中，设节点链接数为k的概率分布为P(k)，则P(k)与k^{-G}成正比(G为常数)，对入链接数$k_{入}$，G=2.1；对出链接数$k_{出}$，G=2.45；对路由器，G=2.5；对域，G=2.2。

图1是小型无尺度网络的拓扑图，通过10个连接最多的中心节点，可访问该网络70%以上的节点。新加入的节点与中心节点的连接概率远超过其它节点，这和随机网络是不同的。

非均匀性的无尺度网络具有连通性好、坚韧性和容错性较强的特点，不会因一些节点被随机去掉而导致崩溃。所以互联网无时无刻不遭遇病毒和黑客的攻击以及人为的通信障碍，但却始终正常运转。但若一部分中心节点被有目的地破坏，网络将受重创。比如位于北美的全球连通度最高的15个中心节点如果同时瘫痪，很多国家将成为网络孤岛。幸亏这种可能性微乎其微。

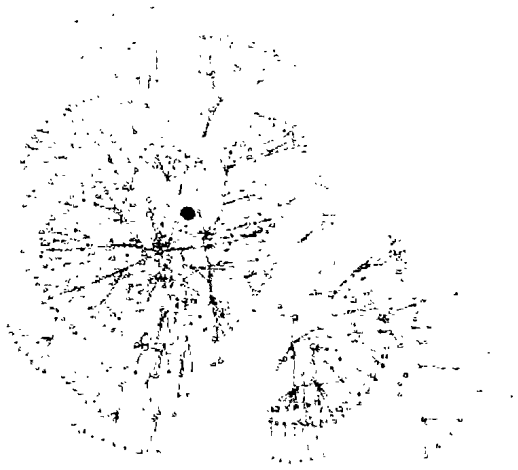

图1　小型无尺度网络拓扑图

1.2　*万维网的蝴蝶结模型和链接分析*　美国圣母大学巴拉巴希(Albert L Barabasi)的研究表明：万维网表现出高度的成群聚集现象。任意两个节点的平均距离(D)和网络总节点N之间满足：$D=0.35+2.06\lg N$

布罗德尔绘出了万维网的蝴蝶结模型图。他认为万维网的节点可分为五类：强连接部分、起点部分、终点部分、须和管道以及非连接部分。

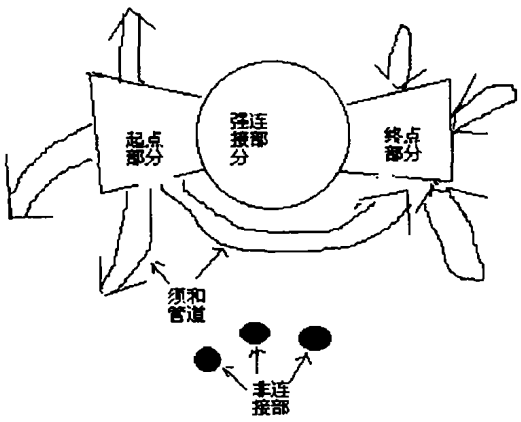

图2　万维网的蝴蝶结模型

强连接部分占28%，是万维网的心脏，其中的所有节点可以通过超链接相互访问。起点部分占22%，可以从其中的节点访问强连接部分和终点部分，但不能反向回访。一般是个人网站和尚无人知道的网站。终点部分占22%，只包含内部链接，不含任何外部链接，如一些企业网站和大学网站，强连接部分有指向此部分的链接。须和管道占22%，与强连接部分不能相互访问的节点。非连接部分占10%，不包含任何内外部链接，也不能从其它节点访问的孤岛节点，只能在浏览器地址栏直接输入访问。

这个模型提示，万维网中的相当一部分信息是单向流动的。以目前搜索引擎的搜索覆盖率和搜索机制，起点、须和管道以及非连接部分弱连接部分总计万维网的54% 基本不能被搜索。即便是强连接部分，从任意网页访问最远网页的平均点击数为28次，整个万维网中任意起始页和目标页能链接上的概率为24%，所以人们常常在网络中迷路。

黄奇等人在对南京高校网站学术资源的链接分析后认为，地域观念也是链接方向的影响因素，链接和访问量不一定能公正反映网站的学术地位。夏旭等人对10个医学搜索引擎的比较测试显示，学术网站被专业搜索引擎收录也具有很强的随机性，并没有权威标准。现在很多门户网站、搜索引擎对网站往往实行收费排序，也影响了小网站的知名度。语言也是影响因素之一。

假定学术网站分布也符合布拉德福定理，核心学术网站可从多个搜索引擎确定，但相关网站就未必。通常万维网中的垃圾信息很多，但不可否认，如同矿藏通常埋在荒无人烟的地方，有价值的网络资源也很可能隐藏在弱连接部分，比如诸多的学术性网站和专业网站或者某些FTP、BBS网站，由于诸多原因不为人知，这不代表这些网站没有学术或者情报价值。因为专业网站访问量本来就少，网站建设者的推广渠道比较窄；网站知名度积累需要一定时间，好网站也可能只被同样不为人所知的同行推荐链接。当然一般说来寻找弱连接部分相关学术网站的时间与其价值相比可能偏小。

1.3 *万维网的演变* 互联网和万维网成为无尺度网络并非偶然。依据优先连接原则，新网站总倾向于链接人气较高的网站；新终端也总倾向于连接拥有较多带宽资源的ISP。竞争的无尺度网络由于节点竞争指数不同，可能有两种演变趋势：竞争力强的节点连接数增加速度超过竞争力弱的节点，但不能取代后者，适者变富；或者竞争力最强的节点将所有连接吸引过来，取代同类节点成为超级节点，赢家通吃。究竟属于哪种，尚不得而知。

科学家发现细胞网络、物种网络和互联网一样也是无尺度网络，这可能是一种适者生存的自然选择。而互联网中的信息流动和细胞的反应动力学规律有相似之处。

2 网络地理学

2.1 *起源* 20世纪60年代以来，随着信息技术特别是网络技术的发展，传统的时间和空间概念受到挑战。计算机、数字通讯和媒体技术相结合所创造的赛伯空间(Cyberspace)以全新的数字空间逻辑改变着人们的生活。

赛伯空间是计算机库中抽象数据的图解表达，是信息空间的宇宙，它可能比物质空间更重要。以尼尔·史蒂文森的小说《雪崩》(Snow Crash)和基诺·里维斯的电影《黑客帝国》(The Matrix)为代表的一系列幻想作品表现了有智慧的人是如何驰骋三维时空而外的网络空间。它的神奇魅力吸引了无数研究者运用数论和统计学等各种手段试图从各个角度描绘电子空间的版图，使虚拟的网络形象化，由此诞生了网络地理学(Cybergeography)这一边缘学科。探寻计算机、路由器、网站、网页之后的秩序和美感，将枯燥的统计数据和自己对网络的理解变成易于欣赏和理解的"网络地图"。其研究对象为互联网的基础构造(如通信线路、节点分布等)、信息流、网络用户统计、信息结构或者关于赛伯空间的人文视点和艺术作品。各个时期的网络地图记载了网络的变迁和网络资源的分布，结合地理、经济、政治因素，我们可了解、评估某地区网络发展状况。

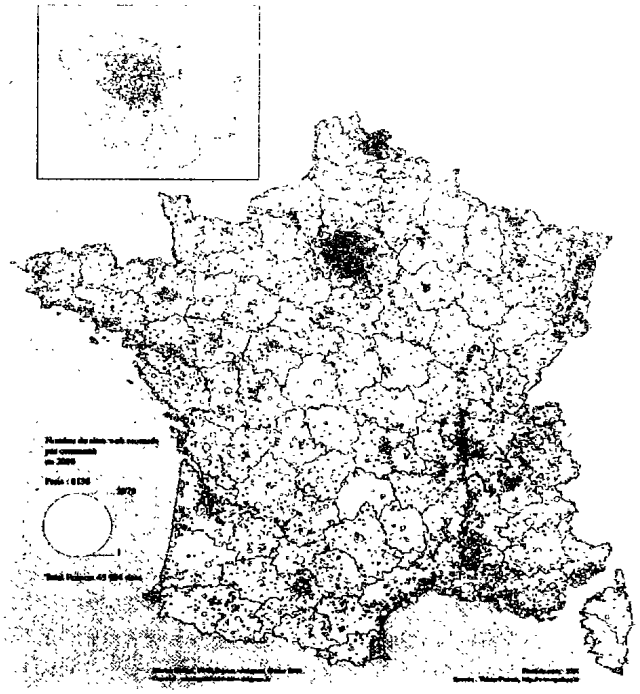

图3 网络地理学实例：法国的网站分布图

2.2 *主要成果和工具* 1964年兰德公司的保罗·巴兰(Paul Baran)在便笺上画了三种网络模型以研究它们的生存能力，被视为网络地理学的先驱。拉里·罗伯茨(Larry Roberts)绘制了最早的ARPAnet拓扑图。此后随着ARPAnet和其它网络的扩张，更多的基础构造图、拓扑图产生了。不仅有计算机绘制的网络拓扑图、3D赛伯空间模拟图，还有诸如伦敦IP密度分布图，精确到街区的圣·弗兰西斯科城域名所有者分布图、美国ISP分布图等源于统计资料的网络地图。绘图方式也由早期手工绘制变为计算机在软件控制下自动绘制。软件工具一般通过网络蜘蛛人对网站及其链接的搜索绘制网站结构图或网络拓扑图。如Visual Web、Microsoft's Site Analyst、Astra SiteManger、CLEARweb、Surfserf Internet，Cartographer等。比较著名的有：Nicheworks贝尔实验室的格拉汉姆·威尔斯(Graham Wills)设计，可描绘几十万个节点的大型网络拓扑图；Ptolo-

maeus 意大利法比奥·维纳科托拉(Fabio Vernacotola)设计，可绘制网站的等级结构图；Mapping the Web Infome 利莎·杰夫布赖特(Lisa Jevbratt)的网络艺术项目，采用专用软件绘制独特的有色环、点阵图、文本三种形式模拟网络蜘蛛所爬过的网站，比较抽象。

研究网络发展并逐步确立网络地理学的主要著作有：Katie Hafner 和 Matthew Lyon 的 Where Wizards Stay Up Late: The Origins of the Internet；

Peter H. Salus 的 Casting the Net: From ARPANET to Internet and beyond…；Martin Dodget 和 Rob Kitchin 的《赛伯空间的地图》(the Atlas of Cyberspace)一书用了 300 多张图片，比较全面地展示了网络地理学的成果，堪称集大成者。

网络地理学的图形比较注重美学效果，可视为艺术和技术结合的典范。

3 借鉴与启示

3.1 *中国的网络地理学* 每半年中国互联网络信息中心(CNNIC)对中国互联网发展状况作比较全面的统计，涉及网民数量、构成、网络行为、域名分布、ISP 及带宽等方面。赛迪资讯顾问公司(CCIDNET)每年推出 IT 行业软硬件评估报告。这些统计数据基本反映了我国互联网的基础构造，如果加以综合，完全可绘制成中国的网络地图。

3.2 *他山之石* 文献计量学的延伸学科—网络计量学(Webometrics)自诞生以来一直试图探索有效的评价网络信息的工具和定理，迄今似乎未有很大突破，大部分只是对传统文献计量学的重复和验证。主要原因在于研究者缺少专业网络信息计量工具，而依靠搜索引擎难以收集较准确可信的统计数据，更不用说提出新假说和数学模型加以验证。

如金岩对 19501 个 COM 类网站拥有的 34371 个网页分析后，认为 COM 类网站数量的国家分布不符合洛特卡定律。这一判断与 Roussean 使用 AltaVista 的研究结果相反。孰是孰非姑且不论，金岩所用统计数据中平均每个 COM 类网站平均拥有的网页少于 2 个，这似乎与事实相去甚远。图书馆学专家往往习惯用传统的文献计量学的方法手段研究网络计量学，而有意无意忽略网络链接的特性，原因也在于缺少科学计量工具。即便是网络计量学家常用的 AltaVista 的网站链接搜索功能，其稳定性、可靠性也不能令人满意。

网络地理学和网络计量学的研究对象有比较多的交叉部分，目的不尽相同，但都基于网络链接分析法。前者丰富的研究工具和统计分析方法可以移植到后者中，为后者注入活力。

3.3 *蝴蝶结模型的启示* 网络信息资源关系到社会影响力和话语权。对网站建设者来说，主动使自己的网站成为强连接部分节点，可极大提高访问量和知名度。对检索者和导航者来说，google 之类的 Pagerank 机制可帮助确定专业核心网站，而对非网络途径(如会议、报刊、论文)获知的相关学术网站也应适当关注追踪。

参 考 文 献

1 张兆晋. 互联网的新发现. Newton 科学世界，2002；(3)
2 黄 奇，李 伟. 基于链接分析的学术性—WWW 网络资源评价与分类方法. 情报学报，2001；(2)
3 夏 旭，李健康，葛 驰. 网络计量学在医学图书馆的应用探讨. 医学信息，2001；(12)
4 http://www.cybergeography.org
5 金 岩. 网络信息计量方法研究. 中科院硕士学位论文

(责编：晨王京)

(上接第 3 页)网络经济的快速发展，企业根据自身及经济变化适时转型也是非常重要的。早在 1999 年，B2C 正搞的如火如荼的时候，杨致远已意识到雅虎发展中的问题，果断将网站交给以经营见长的蒂姆·库格尔，以期"挽救"雅虎。蒂姆成功地将网站性质由"因特网搜索门户网站"转为"ICP + B2C + 个性化服务"，将商业模式由"注意力经济模式"转为"注意力 + 现金流"模式。随后，雅虎网上商店的营业额以 40% 速率上升。正是这次转型帮助雅虎度过了 2000 年席卷整个电子商务界的灾难。著名的 8848 也有过一次成功从"单打一"到"三合一"模式的转型。对企业的 CEO 来说，跟得上网络产业的变化速度和善于运用经济策略来维护企业的运行是一样重要的。

c. 选择适合网络经济的商品。企业在选择何种商品进行网络交易时，应遵循网络经济的特点选择合适的商品。一般来说，适于电子商务的产品有三个基本要求：无差异性；交割可以通过银行进行；交易场所受到一定的限制。如股票、电脑销售、旅游服务、金融服务、图书、拍卖等。

d. 电子商务与传统商务的整合。从本质上看，电子商务企业与传统企业不存在根本差异，成本与收益是两个具有决定作用的因素。纳斯达克的狂跌使包括电子商务网站在内的所有网络公司深刻认识到：网络公司的生存特质其实与传统企业并无本质区别；不赚钱，一样会倒闭；传统企业的管理经验及经济规律在电子商务中一样发挥着作用。新兴的商业模式必须与传统行业进行整合，充分利用传统企业的人才、资金、管理以及经济策略，才能更好地促进自身发展。

参 考 文 献

1 吴叔平. 电子商务的价值链与赢利模式. 上海：上海远东出版社，2000
2 刘明晶，熊婧汐. 增值商业模型. http://www.cnw.com.cn/cnw/2000/32/3216.asp
3 电子商务：点石成金，抑或点金成石. http://www.sinobnet.comc.cn/ec/jj-33.htm
4 http://j-ry.myrice.com/75unsuccess,.htm
5 正确把握电子商务的本质. http://www.real-eatate.tj.cn/cjpt/rdht
6 胡 超. 触摸电子商务的本质——专业化、智能化、个性化. http://www.amteam.org/a-chainstore/cast/leyou-ac-0404.htm
7 谭 智. 电子商务的"第三次浪潮". http://cnw.com.cn/cnw/2000/39/3917.asp
8 杨建民. 电子商务的成本分析. http://cnw.com.cn/cnw/2000/39/3917.asp

(责编：晨阳)