中国旅游线路的合作网络描述

A Collaboration Network Description on the Chinese Travel Route System

- 常 慧/CHANG Hui1, 何 阅/HE Yue2, 张义勇/ZHANG Yi- yong1, 苏蓓蓓/SU Bei- bei1, 何大韧/HE Da- ren1
- 1. 扬州大学物理科学与技术学院, 江苏扬州 225002
- 2. 无锡市第一中学, 江苏无锡 214031
- 1. College of Science & Technology, Yangzhou University, Yangzhou 225002, Jiangsu Province, China
- 2. Wuxi No. 1 Middle School, Wuxi 214031, Jiangsu Province, China

[摘要] 从复杂网络角度研究了中国旅游线路的实证统计性质,包括平均距离、集群系数、度分布律、项目大小、项目度分布律以及同类性系数等;建议把中国旅游线路网归入"广义合作网络"一类;提出了一个中国旅游线路网的自组织演化建议模型,由此模型所得到的模拟结果与实证结果比较好地相符合。

[关键词] 复杂网络;旅游线路;合作网络

[中图分类号] O414.2, C811, O157.6, F590.1

[文献标识码] A

[文章编号] 1000-7857(2006)09-0084-04

Abstract: We have studied the empirical statistical properties of Chinese travel route system with a viewpoint of complex network. The properties include the average path length, clustering coefficient, degree distribution, act size, act degree distribution, assortativity and other. We suggest that Chinese travel route network can be included in the so-called "extended collaboration networks". We also propose a self-organizing developing model of Chinese travel route network. The simulation results based on the model are in quite good agreement with the empirical statistical results.

Key Words: complex network; travel route; collaboration network

CLC Numbers: O414.2, C811, O157.6, F590.1

Document Code: A

Article ID: 1000-7857(2006)09-0084-04

1 引言

复杂网络是 1998 年以来出现的一个研究新热点[1-6]。在许多 实际或理论问题的研究中,可能把复杂系统基本单元之间的相互 作用简化地表示为图论中赋权或不赋权的顶点(即上述单元)之间 的赋权或不赋权的"边"。这样,这些复杂系统就可以在一定规定 下表示为一个网络。这是对复杂系统建模的一种新尝试。在这些 研究中, 1998年 Watts 和 Strogats(WS) 间的论文起了开创作用。论 文报道了许多实际网络具有小的平均路径长</>
ا

定义为两顶点 间沿最短路径的边数的平均值)和大的集群系数 C (对每个有 kv 个连边邻点的顶点 v, C 定义为所有邻点之间的实际连边数被它 们的最大可能连边数来除,整个网络的C定义为各顶点集群系 数的平均值),建议了介于规则与随机之间的小世界网模型。此 后, Barabasi 和 Albert(BA) 1999 年著文建议了第二个著名的复杂 网络模型, 即无标度网的 BA 模型。无标度网的特征是度分布(描 述具有相同边数(度 k)的顶点的出现概率 P(k))显示的 "幂律"关 系, 即 $P(k) \sim k^{-\gamma}$, γ 称为特征标度因子^[4]。此后, 大量发表的复杂网 络论文涉及到通讯、电子、生物、地理、环境、交通等许多领域中的 诸多实际系统的实证研究[1-2]。

我们注意到社会网络中特殊的一类,即所谓"社会合作网络"⁷⁷。这类网络中存在两类节点。一类节点可称为"项目"(act,例

如电影影片或科研论文);另一类节点可以称为"参与者"(actor, 例如演员或科研人员)。两类节点之间的一条边代表一个参与者 参加一个合作项目[1,78];而同一类节点(例如参与者节点)之间的 一条边表示参加同一个项目的合作关系。可以定义一个量 h, 称 为节点 i 的 "顶点项目度",等于这个参与者节点参加的项目数 (如一个演员 i 曾经演过几部影片);也可以定义一个量 T_i, 称为 它的"项目大小",等于这个项目含有的参与者数(如一个影片 j 同时聘用几个演员)。如果只注意参与者之间的连边,参加同一个 项目 j 的参与者两两之间都连接一条表示在此项目中合作关系 的边,因此构成一个完全图。每个完全子图中的每一个节点的顶 点度都等于 T_i-1。当然, 2个(甚至更多的)这样的完全图很可能 共用 1 条甚至好几条边, 但是如果计算多重边(即把 2 个完全图 共用的 1 条边认为是 2 条重叠的边(例如 2 个演员在 2 部影片中 合作,应该算作合作2次)。以此类推,对每个完全图顶点度等于 T_i-1的结论仍旧成立,对网络整体顶点度等于所有 T_i-1的叠加。 如果我们把完全图的概念推广到这种计算多重边的情况(即认为 包含多重边,每2个顶点都连边的子图仍叫做完全图),仍可以说 向参与者的投影单模式网络是完全图的集合。

我们建议把合作网络的概念推广到非社会网络的描述中去, 称为广义合作网络^[9-10]。这个建议大大扩充了合作网络的范围。我

收稿日期: 2006-06-28

基金项目: 国家自然科学基金项目(70371071)

作者简介: 常 慧, 女, 江苏省扬州市扬州大学, 硕士生, 主要研究方向为复杂性科学; E- mail: changhui 330@gmail.com

何大韧(通讯作者), 男, 扬州大学物理科学与技术学院, 教授, 主要研究方向为非线性科学与复杂性科学;

E- mail: drhe@yzu.edu.cn



们已经报道过一个从广义合作网络的共性出发所建议的演化模型,以及由此模型预言的广义合作网络的第一个共同特征[^{5-10]},即顶点的度(即一个参与者连接多少个其他参与者)分布和项目度(即一个参与者参加多少个项目)分布的关系总是一致的,而且一般地采取所谓的 "广延指数分布"(Stretched Exponential Distribution,简称 SED 分布)的、介于幂函数和指数函数之间的分布函数[11]。

Newman 在 2002 年提出了标志网络的另外一个重要统计参数,即所谓 '同类性 '^[12]。同类性用一个在-1和1之间变化的参量 r 表征,表示网络中节点是否与类似自己的其他节点优先建立边连接。当 r 大于零时,节点优先与最类似自己的一类节点连接,网络称为 "同类匹配";当 r 小于零时,节点优先与最不类似自己的一类节点连接,网络称为 "非同类匹配"。如果特别地研究邻接节点的度, r 大于零表示邻接节点度的正相关,即度高的节点优先与度高的节点连接;而 r 小于零表示邻接节点度的负相关,即度高的节点优先与度低的节点连接。Newman 提出标志这种度之间同类性的计算公式为^[12]:

$$r = \frac{\sum_{jk} jk(e_{jk} - q_{j}q_{k})}{\sigma_{\alpha}^{2}}$$
 (1)

其中 e_k 表示连接 "剩余度"(即沿着一条边到达一个节点所看到的它的其他邻边数) 分别为 j 和 k 的节点的边所占的比例 (或者说随机选择一条边连接剩余度分别为 j 和 k 的节点的联合

概率分布),对一个无向图,
$$q = \sum_{k} e_{j_k}$$
, $\sigma_q^2 = \sum_{k} k^2 q_{k^-} [\sum_{k} k q_k]^2$ 。

Newman 实证地研究了科研合著网、演员合作网、公司董事会 成员之间的合作网、中学生的朋友关系网、e-mail 网等社会合作 网, 以及美国电力网、因特网、WWW 网、软件依赖关系网、蛋白质 相互作用网、代谢作用网、神经网络、海洋食物链网、洁净水食物 链网等非社会网络的 r 参数, 发现所有这些社会合作网的 r 参数 都为正, 而所有这些非社会网络的 r 参数都为负[12]。这些结果带来 的问题是: 这是不是普遍的规律? 如果是,原因(机理)是什么? Newman 和 Park 在 2003 年发表论文[13]讨论了这个问题。他们通 过一个模型的解析,说明了在一般情况下,如果不允许网络中出 现多重边(即一对顶点之间最多可能有一条边),则网络一定是度 负相关的。这就是说,网络度负相关几乎是自然的,除非允许多重 边。然而, 上述 Newman 实证研究的社会合作网也都不考虑多重 边, 为什么它们的 r 仍旧大于零[12]?由此推论, 影响同类性的, 除了 多重边之外,一定至少还有另外一个重要因素。Newman 和 Park 在 2003 年发表了另一篇论文[14]讨论这个问题。他们解析地讨论 了一个节点属于群落的模型, 说明了对于这样的网络拓扑结构, 即使不考虑多重边也会明显导致更大的集群系数和正的度同类 性。因此,如果社会网络都是分为群落的(这明显正确),而非社会 网络通常不分为群落(这有待进一步检验和说明),那么区别社会 网络和非社会网络的2个特性(特别大的集群系数和正的度同类 性) 就来自同一个拓扑结构原因, 即群落结构。 综合 Newman 和 Park这 2 篇论文的结论, 我们应该认为, 度同类性的正或负, 归根 结底还是取决于网络的拓扑结构。具体地说, 取决于网络是否具 有群落结构和多重边。这样,不一定所有社会网络的 r 参数都为 正, 所有非社会网络的r 参数都为负。但是, 由于社会网络都是分 为群落的, 而可能许多非社会网络不分为群落, 所以社会网络的r 参数为正的可能性比较大, 而非社会网络的r 参数为负的可能性 比较大而已。2003年之后,已经有人发现个别非社会网络的r参 数为正,大脑功能网络可能是这样的非社会网络的一个好例 子[15]。然而,这些实证报道缺乏与网络拓扑结构联系的进一步分 析,因此没有能够清晰地为上述的 Newman 和 Park 的分析结论 提供证据。

我们在上面介绍的模型[9-10]计算中是考虑多重边的,而且所 讨论的这些广义合作网络是由完全图集合构成的。显然,每个完 全图都可以被看作是内部连接最密集(因为所有可能连接的边都 已经连接,不可能再密集),而对外连接相对稀疏(至少不可能连 接所有可能的边)的子图,也就是一个"最基本的"(对此名词将另 文仔细讨论)群落。因此,广义合作网络肯定具有群落结构。按照 上述的分析结论,在计算多重边时,上面报道的模型和模型适用 的实证系统都应该显示正的r参数。这可能成为验证上述的 Newman 和 Park 建议机理的很好途径。我们已经证明,上述模型 的确显示正的 r 参数(将另文详细报道)。本文将报道中国旅游线 路网络的一些实证统计特征,包括节点度之间的同类性。其中,同 类性的结果由于可能验证上述的 Newman 和 Park 建议机理,应 该特别受到关注。我们还将建议一个中国旅游线路网的自组织演 化模型,并且将由此模型所得到的模拟结果与我们的实证研究结 果对比。我们发现的其他广义合作网络的共性将另文陆续报道。 近年来, 随着我国人民生活水平的迅速提高, 旅游业进入了一个 快速发展的时期。 "旅行社"、"旅游线路"这些概念已经为千家万 户所熟悉。"旅游线路"是指专为旅游者设计、能符合一类旅游者 需求、以一定交通方式将若干个旅游点和旅游活动项目贯穿起来 而形成的运行网路[16]。人们的兴趣和选择是千变万化的,实际推 出的旅游线路也就花样繁多。例如,按照空间范围来区分,有跨国 游、跨区(省、地、市)游、区内游、短途(一日、二日)游等;按照内容 来区分, 有观光游、休闲游、专题游、会(商) 务游等; 按照交通工具 来区分, 有泛海游、徒步游、自行车游、汽车越野游等。 通常, 每个 旅游线路有机地串联若干个有某种内在联系的,在观光、交通、食 宿、服务、购物、娱乐等方面具有特色和对一类旅游者具吸引力的 旅游点。每个旅游点都在整个线路的工作中和其他所有的旅游点 密切合作, 贡献自己的力量, 也分享一份收益。如果规定在一个旅 游线路中的各旅游点为顶点,处于同一旅游线路中(同一批游客、 同一次游览所经历)的2个旅游点之间的相互合作关系为边,则 一个旅游线路对应一个完全图描述的项目。这样,旅游线路体系 可以用完全图项目的集合来描述。旅游线路网也属于上述的广义 合作网络。

2 统计结果

我们经过调研和选择,选用了 2003 年全国的 240 个旅游线路(项目)、171 个旅游点(参与者顶点,它们之间共有 719 条边)进行了统计研究。如上所述,一个项目对应一个图论中的"完全图",各个完全图之间通过共用的顶点架起了桥梁,从而构成了整个中国旅游线路网络,如图 1 所示。

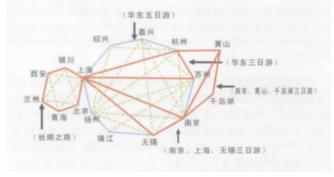


图 1 中国旅游线路网络的示意图

Fig. 1 A schematic map of the Chinese travel route network

我们的统计得到: 平均距离<l>=2.9, 集群系数 C=0.69。图 2 的主图显示了统计得到的旅游线路网大小T的累计分布,图 2 的

插入图显示了统计得到的旅游线路网的项目大小工的分布。一条 旅游线路最多包括 18 个旅游点, 最少包括 1 个。图 2 的插入图说 明一条旅游线路由2个旅游点组成的可能性最大,而更少或更多 个旅游点组成的可能逐渐减少。图 3 显示了相应的重复边累计度 分布和累计顶点项目度分布, 它们都属于很典型的 SED 分布, 而 且关键参数相同(μ=0.5), 说明旅游线路网络在形成过程中的动 力学机制既有优选的因素,又有随机的因素,即构成旅游线路时 既首选最常用的(顶点项目度最大的)旅游点,又考虑各种不同的 具体因素(客源情况、客人要求特点、可能的旅游点类型、距离、价 格、季节等)来选择各种不同的旅游点。由于具体因素太多,互相 又缺乏关联,相当于随机选取。这应该是大部分广义合作网络的 共同特征[9-10]。我们的统计结果表明,旅游线路网的平均顶点项目 度为: <h>=3.8. 即每个节点平均参加 3.8 个项目。在旅游线路网 中顶点项目度最大的是上海、北京这些特大城市,或者是张家界、 兵马俑这类特别著名的风景名胜或人类历史遗产。它们或者可以 提供特别多的客源,或者可以吸引、容纳特别多的客源,因此在许 多不同的旅游线路中广泛地出现,成为网络中完全图项目间的最 大桥梁。这些顶点在整个网络中起了关键性的作用。反之,顶点项 目度很小的顶点一般都是小城市,或者是一般的旅游景点。它们 只可能在少数线路(如短距离、近邻景点游)中起一般的作用。

旅游线路网分成许多不连通的分支。我们定义了一个描写网络不完全连通性的统计参数——不连通度,为 u=2k,/(N(N+1))。这里 k_u是网络中互不连通的顶点对的数目,N是网络中的总顶点数。我们统计得到的旅游线路网的不连通度数值是: u=0.141。

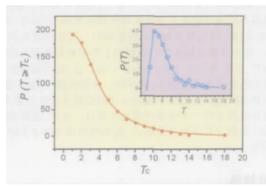


图 2 中国旅游线路网项目大小 T 的累计分布的实证统计结果注: 1.插入图显示项目大小 T 的分布; 2.图中实心或空心圆圈表示相应的统计数据; 3. 主图与插入图中的光滑曲线仅代表对数据点的一种可能的光滑连接。

Fig. 2 Empirical statistical results of the act size accumulative distribution of the Chinese travel route network (1. The inset shows the act size distribution; 2. In the figure the hollow or solid circles denote the corresponding statistical data. 3. The smooth lines in the figure and the inset only show a possible smooth connection between the data points.)

本文报道的同类性系数实证调研结果可能最应该受到注意。在计算重复边时, 旅游线路网的同类性系数 r=0.36, 标准偏差 $\sigma_{c}=0.02$; 而在不计算重复边时, 旅游线路网的同类性系数 r=0.15, 标准偏差 $\sigma_{c}=0.027$ 。其他可以用广义合作网络描述的实际系统的实证结果同样显示: 在不计算重复边时同类性系数 r 有正有负, 而在计算重复边时全部为正,与在引言一节所述结论很好地符合。这就是说, 度同类性的正或负取决于网络是否具有群落结构和多重边, 不一定所有社会网络的 r 参数都为正,所有非社会网络的 r 参数都为负。但是社会网络的 r 参数为正的可能性比较大,而非社会网络的 r 参数为负的可能性比较大。在计算多重边时,广义合作网络描述的实证系统都应该显示正的 r 参数。

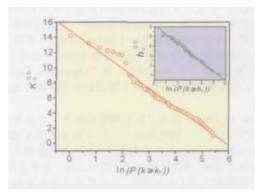


图 3 中国旅游线路网重复边的累计度分布的实证统计结果注: 1. 插入图显示顶点项目度累计分布;2. 图中空心圆圈表示相应的统计数据; 3. 主图与插入图中的直线表示数据的最小二乘方拟合的结果。

Fig. 3 Empirical statistical results of the degree accumulative distribution of the Chinese travel route network when multiple edges are counted

(1. The inset shows the act degree accumulative distribution; 2. In the figure the hollow circles denote the corresponding statistical data; 3. The linear lines in the figure and the inset show the least-square fitting results of the data points.)

3 模拟网络发展的自组织模型

文献[9,10]讨论的模型是所有广义合作网络的演化模型,普适性虽高,但不可能描述某一类具体的广义合作网络(例如旅游线路网)的演化。如果希望描述一类具体网络的动力学,还要给模型增加许多能够体现这一类系统特征的内容。我们在本节建议一个旅游线路网的自组织演化模型。在这个模型中引入旅游点节点的权(即它们的性质),其中之一定义为旅游点的"知名度"。在一些资料中[17-18]列出了这些资料的编者定义的具有"等级"的旅游点充。虽然这种等级的划分还不一定规范、有权威性,但作为大致成立的共识应该没有问题。我们建议规定一个旅游点节点的权(知名度)就等于它的地理范围内被文献[17,18]所列出的等级以上景区的个数。我们共计算了66个旅游点的知名度,最大知名度为1,发现总体的统计结果显示一个确定的趋势,即旅游点的知名度随其顶点项目度的变化显示单调的依赖关系,顶点项目度越大(即被列入的旅游线路越多),旅游点的知名度也越高。这一结论将被用于模型的建立。

3.1 模型的要点

其中各个常数是根据实证数据估计选取的,数值模拟说明结果对它们的依赖并不敏感。

- 1) 设初始时刻有 $y_0=30$ 个顶点(旅游点),这时的旅游线路数 (完全图项目数)为 0。当时间 t 离散地发展时,顶点(旅游点)的数量随时间增加的规律为: $y=y_0+t_2$,代表旅游点的加速开发和建设。
- 2) 给各个旅游点(i 顶点) 以均匀概率各赋一个在 0~1 之间 (为实数)的知名度 F; 在 0~1 之间(为实数)的位置因子; 在 1~10 (为整数) 之间的类型因子。位置因子简化地表示旅游点之间的地理距离; 类型因子简化地表示旅游点可能参加哪个类型的旅游线路(例如现代化大都市、民族风情、宗教文化、历史文化、风景名胜、自然保护区、丝绸之路等类型)。
- 3) 每个离散时间 $t_n(不为零)增加一定的旅游线路(完全图项目)。下面要说明的长、短 2 种旅游线路增加的数目各是目前已有旅游点总数的 <math>1/10($ 即长、短 2 种旅游线路(完全图项目) 数量相等, 随时间的变化关系均为: $N_{12}(t) = \sum_{i=1}^{n} (\frac{1}{2} \frac{1}{2} \frac{1}{$
- 4) 长旅游线路: 在上述 10 种类型中以均匀概率随机选取一种, 在这种类型的旅游点中随机选取一定数量(在 10 与 2 之间随机决定数量)的、知名度最高的点建立旅游线路, 不考虑位置因



- 子。 短旅游线路: 随机选取一个旅游点, 选取在它左右 0.01 范围内所有位置因子, 并且其知名度在它知名度的上下 10%之间($0.80F_i$ F_j $1.10F_i$) 的旅游点, 建立一条以它为中心的近程旅游线路。
- 5) 每步离散时间后旅游点(顶点)知名度的进化。 对知名度进行修正: 如上述,顶点项目度越大的旅游点知名度越大,而且可以预期它们的知名度会更大,由此令: F(j, t+1)=F(j, t)1, F(j, t)2, F(j, t)3, F(j, t)5, F(j, t)6, F(j, t)7, F(j, t)8, F(j, t)9, F(j, t)9,

如此不断演化。 我们的计算证明 t 时间步达到一定程度(> 10 000)之后,各个统计性质基本上达到稳恒。下面的计算结果是按照这个原则计算得到的。

图 4 显示了模拟得到的度分布和顶点项目度分布规律, 它们同样是典型的 SED 分布, 而且关键参数都是 μ =0.5, 与图 3 显示的统计结果很好地符合。

表 1 列出了模型模拟结果与前述实证统计结果的比较, 说明 我们的模型基本上是成功的,反映了旅游线路网演化的关键机 制。

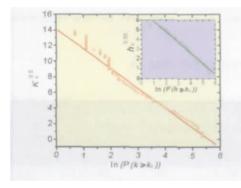


图 4 中国旅游线路网重复边的累计度分布的模型模拟结果注: 1. 插入图显示顶点项目度累计分布; 2.图中空心圆圈表示相应的统计数据; 3.主图与插入图中的直线表示数据的最小二乘方拟合的结果。

Fig. 4 Model simulation results of the degree accumulative distribution of the Chinese travel route

network when multiple edges are counted

(1. The inset shows the act degree accumulative distribution; 2. In the figure the hollow circles denote the corresponding statistical data. 3. The linear lines in the figure and the inset show the least-square fitting results of the data points.)

4 结论

进行了旅游线路统计性质的研究,并且通过分析,认为旅游线路网适合用广义合作网络来描述。在各种统计性质中,同类性系数和顶点项目度最值得注意,这是由于它们分别说明了一方面的网络重要拓扑性质和动力学特性。另外,我们建议一种旅游线

表 1 由模型进行模拟的结果与 2003 年 全国旅游网统计结果的比较

Tbl. 1 The comparison between the model simulation results and the statistical results of the Chinese travel route network in 2003

	平均	分支	平均	顶点度 SED	顶点项目度 SED
	道路长	系数	顶点度	分布 μ 因子	分布 μ 因子
2003 年统计	2.9	0.69	18.54	0.5	0.5
模型	2.7	0.67	19.32	0.5	0.5

路网络自组织演化的模型,由此模型的数值模拟能展示与统计参数很好符合的结果,说明这种模型是合理的。在统计意义上,旅游线路体系的迅速发展过程是可能如此简洁地描述的。

参考文献(References)

- [1] NEWMAN M E J. The structure and function of complex networks [J]. SIAM Review, 2003, 45: 167.
- [2] ALBERT R , BARABASI A L. Statistical mechanics of complex networks[J]. Reviews of Modern Physics, 2002, 74: 47.
- [3] WATTS D J, STROGATZ S H. Collective dynamics of "small-world" networks [J]. Nature, 1998, 393: 440.
- [4] BARABASI A L, ALBERT R. Emergence of scaling in random networks[J]. Science, 1999, 286: 509.
- [5]戴汝为, 操龙兵. 一个开放的复杂巨系统[J]. 系统工程学报, 2001, 16 (5): 376.
- [6] CHI L P, CAI X et al. Structural properties of US flight network[J]. Chin. Phys. Lett., 2003, 20(8): 1393.
- [7] WASSERMAN S ,FAUST K. Social Network Analysis: Methods and Applications [M]. Cambridge: Cambridge University Press, 1994.
- [8] J.J RAMASCO, S N DOROGAVTSEV, R P- SATORRAS. Self- organization of collaboration networks [J]. Phys. Rev. E, 2004, 70: 036106
- [9] ZHANG P, CHEN K, et al. Model and empirical study on some collaboration networks [J]. Physica A, 2006, 360: 599.
- [10] ZHANG P, HE Y, et al. 一个描述合作网络顶点度分布的模型[J]. 物理学报, 2006, 55: 60.
- [11] LAHERRERE J ,SORNETTE D. Stretched exponential distributions in nature and economy: fat tails with characteristic scales[J]. Eur. Phys. J. B, 1998, 2: 525.
- [12] NEWMAN M E J. Assortative mixing in networks [J]. Phys Rev Lett. 2002, 89: 208701; Mixing patterns in networks, Phys Rev E, 2003, 67: 0261126.
- [13] PARK J, NEWMAN M E J. Origin of degree correlation in the Internet and other networks[J]. Phys Rev E, 2003: 68026112.
- [14] NEWMAN M E J, PARK J. Why social networks are different from other types of networks [J]. Phys Rev E, 2003, 68: 036122.
- [15] EGUI LUZ V M, CHIALVO D R, CECCHI G A, et al. Scale-free brain functional networks[J]. Phys Rev Lett, 2005, 94: 018102.
- [16] 孙文昌. 现代旅游开发学[M]. 青岛: 青岛出版社, 1999.
- [17] 实用交通旅游图典编辑组. 实用交通旅游图典[M]. 福州: 福建省地图出版社, 2002.
- [18] 等级旅游区点[OL]. http://www.cnta.com/8- ssls/lyqd.asp

(责任编辑 李慧政)

科技导报》编辑部更正启事

我刊 2006 年第 6 期发表的 《从体内新的功能系统——支持储备及自体监控系统新学说》一文参考文献第 8 条应改为: [8] 原 林.筋膜学和经络穴位的有关研究[M]//钟世镇,主编.数字人和数字解剖学.济南: 山东科技出版社, 2005. 特此更正。