

摘 要

描述复杂自适应系统中的作用者行为的动力学法则必须容纳类似生命系统特有的预测、选择、学习、进化等特征。作用者和它们的行为各不相同，作用者构成的系统绝不等同于它们的简单组合，不能用还原论的思路来处理，也不能用服务于确定论、还原论的传统数学来描述，必须探讨全新的描述工具。复杂网络研究的发展为我们提供了进行这种探讨的某种可能性。

复杂网络可以描述广泛多样的实际系统。我们进行了相当数量的实际系统的实证研究，希望能够根据实证研究得到的统计结果把实际系统分类，找出每一类系统的实证统计性质共性，进而试图建立模型，理解它们的共同动力学特征和机制。这样一步一步地达到对复杂系统的更普遍一点的理解。在当前复杂网络研究的广阔范围中，我们的兴趣集中在所谓的“广义合作网络”上。这里网络的节点（参与者）在参与许多合作项目。在研究中只考虑参与者在项目中的合作关系，不讨论它们之间的竞争、对抗或其他类型的关系，而且定义参与者在项目中的合作关系为它们之间的边。这样，一个项目就可以用一个它包含的参与者两两连边所构成的完全图来表示。这样的网络不一定是社会网络，也可以包含其他许多类网络，但是由于拓扑结构具有共同特征，它们的统计性质和动力学特性也具有共同特征。在我们实证地研究过的实际网络中，我们发现好莱坞演员合作网、中国旅游线路网、城市公交线路网、中药方剂网、以及淮扬菜肴网都可以用广义合作网络来描述。我们对这些实证数据分析之后，得到了一些共同的统计特征：首先，项目是这类网络中的主导因素，项目如何组成是网络演化的关键问题。参与者项目度分布（即参与者参与的项目数目的分布）很可能是这类网络最主要的统计性质，在很大程度上左右顶点的点强度分布和度分布。其次，参与者的同类性系数也随着这样的分布规律的变化而呈现单调依赖关系的变化。我们建议了一个简明的、体现上述思想的网络演化模型，来说明这两项共同统计特征的动力学机理，由此模型的解析分析和数值模拟产生的统计结果与实证结果很好地符合。最后，根据这类被称为广义合作网络的共同拓扑特征，我们建议群落和层次的一种定量定义和相应的划分群落和层次的方法。由此所建议的一个新的统计参量——网络

交连度可能富有实际意义。

由于实际的重要性，城市公交系统一直是研究的热点之一。我们从复杂网络的角度，研究旨在说明城市公交系统演化和性质的网络操纵者博弈模型。从这个角度出发，我们陆续提出三个模型，模型越来越简化，意思越来越集中，本论文中重点介绍最后一个最简网络操纵者博弈模型。我们建议最大简化地把公交公司和乘客看作网络的两个主要操纵者，他们的利益在某种意义上说是互相矛盾的，于是，城市公交网络系统的演化被最简化地看作网络的两个操纵者之间博弈的过程，博弈的均衡解决定网络最后的演化走势。对这样的模型进行了极端情况和均衡情况的解析，以及一般情况下的数值模拟研究，所得到的网络统计性质与我们对北京、上海、南京、杭州的实证统计结果相当好的符合，说明这个模型很可能抓住了最根本的特征。这种思想很可能推广到许多其它系统。

广义合作网络忽略了竞争因素。当在网络中加入竞争的因素时，网络会有什么样的特性呢？鉴于此，我们介绍了三个考虑到竞争因素的系统，这个工作是浅陋的。但期望这几个系统能够作为一个引子，能够激起研究者们对合作竞争网络作更深入的研究。

[关键词]：复杂网络，广义合作网络，项目度，同类性，层次，群落，操纵者博弈，竞争

Abstract

The dynamics principles, which describe the actors' behaviors in complex self-adaptive systems, should include some characters similar to what are shown by the living systems, such as forecasting, choice, learning, and evolution. Each actor and his behavior are different. Also, a system consisting of actors is not a simple combination of them. It cannot be treated by reduction theory, and cannot be described by the conventional mathematics, which serve for determinism and reductionism. We must develop some new methods and tools to describe self-adaptive systems. The latest development in the complex network gives us a possible way for searching the possibility.

The complex network can describe variety different practical systems. We have carried out empirical studies on quite some practical systems; wish to classify the systems based on the results and find out the common statistical properties of every kind of them. The next step could be setting up suitable models so that we could understand their common dynamical characteristics and mechanism. In this way we can reach, step by step, to a more universal understanding on complex systems. Among the very wide scope of complex network studying, our interests are concentrated on the so-called "generalized collaboration networks". In such networks the nodes (actors) are participating a lot of collaborative acts. In the study we shall only consider actor's cooperation, and ignore the competition, confrontation or other type of relations between them. We propose define actor's cooperation in the acts as edges. In this way, one act can be described by a complete graph composed by the actors where every pair of them are connected with edges. Such a network may not be a social network, it can belong to many other kinds of networks too, but because the topological structure has common characteristics, their statistics properties have common characteristics. Among the practical networks we empirically studied, we found that the Chinese tourist line network, the urban public traffic network, the traditional Chinese medicine prescription

network, and Huai-Yang recipe network can be described by the generalized collaboration networks. After the analysis on the empirical data, we have obtained some common statistical properties: First of all, the act is a leading factor in this kind of networks. The act degree distribution (the distribution of the number of the acts, in which the actors are participating) should be, very possibly, the most important property in such kinds of networks. It, in a great extent, determines the node strength degree and degree distribution. Secondly, the assortativity of the actors shows a monotonic change with the change of such distribution law. We propose a simplified model based on these ideas so as to show the common dynamical mechanisms of the two common properties. The analytic and numerical investigation results of the model show very good agreement with the empirical results. Finally, according to the common topological characteristics of the so-called generalized collaboration networks, we propose a quantitative definition and corresponding division method for community and hierarchy. A new network statistical property, the degree of interweavement, has been proposed by this consideration that may have practical importance.

Since urban public traffic systems are practically important, they have been extensively and intensively studied. We have studied urban public traffic networks from the viewpoint of complex networks and game theory. From that, we have suggested three manipulator game models of urban public traffic networks continuously. We try to construct a simplest model with the idea that hits the nail on the head. In this paper, we mainly present the last model of the three. We suggest a very simplified viewpoint in which the evolution of urban public traffic network can be considered as a game process between the two network manipulators, and the equilibrium solution of the game determines the last evolution tendency of the network. We have performed analytical discussion on some extreme and the equilibrium situations and numerical discussion on the general evolution, the obtained statistical properties are in a good agreement with the empirical ones obtained by the

investigations on the urban public traffic systems in Beijing, Shanghai, Nanjing and Hangzhou. This shows that the model grasps the most fundamental characteristic of the system. The idea, very possibly, can be extended and used in many other systems.

Generalized collaboration networks ignore the factor of competition. If competition joins in network, what characteristic may appear? As for it, we introduce three systems which take competition into account. The research is rough. However, we expect as a beginning, the work could inspire people to research the collaboration and competition network deeply.

[Keywords]: Complex network, collaboration, generalized collaboration networks, act degree, assortativity, hierarchy, community, manipulator game, competition

第一章 引言

科学家们用了几百年的时间，几乎将所有的东西拆解成分子、原子和夸克，体现了一种强烈的还原论思想，相信这种思想，也就是相信宇宙是受自然法则支配的，相信宇宙最终可以被完全理解。迄今为止，由培根、牛顿、伽里略、笛卡儿等人开创的近代科学已深入人心，由培根首先倡导的分析和实验方法，经过伽里略等许多学者的实现和发展，深入到几乎所有的科学领域；牛顿归纳的物理学的定律以及相应的教学方法，则成为人类对于宇宙（包括社会）认识的基本框架；而笛卡儿则从思想方法上进行了深刻的概括与总结。近代科学是非常成功的，但如果把近代科学看作是人类文明发展的最终完成，那就与科学的本意背道而驰了。二十世纪一开始，相对论和量子力学的出现就对牛顿力学所代表的宇宙观提出了尖锐的挑战。大自然向人类展示出了丰富多彩、变化万千的景象。从上个世纪 30 年代到 40 年代，从贝诺朗菲的一般系统理论，到维纳的控制论、香农的信息论，一系列新学科出现了。它们冲破了还原论那种学科分割，只见树木不见森林的、目无全牛的框架，从全局、从发展的眼光去观察和认识世界，为人们打开了眼界，开拓了视野。在此期间，现代数字式电子计算机的诞生，对人类社会和科学打开了一种全新的认识方式。而哥德尔不完全性定理的证明，这从另一个角度告诉我们，现有的逻辑方法并不能穷尽人类的认识，它只是人类认识的一种方法，而决不能构成最终完成的、无法再前进的、终极的理论体系。所有这一切大大推动了 20 世纪后半叶的科学进步。

自上个世纪七八十年代开始，复杂性科学真正走入了国内外科学家的视野，成为一个新兴的研究领域。许多杰出的科学家为发展复杂性科学进行了不懈的努力。使科学家们感到困惑的一个重要原因是，复杂性科学正处在探索阶段，一些基础问题亟待解决，如：是否存在从简单到复杂的自然法则的问题以及什么是复杂性的根源问题等，这些问题极大的阻碍着复杂性科学研究的发展，如果不能从理论上得到解决，那么复杂性科学就难以建立起坚实的理论基础，从而影响和制约复杂性科学的发展。在这个探索过程中，应该看到，该领域的研究工作者已经

作出了许多里程碑式的进展。从物理学的角度也可列举以下重大事件，如原胞自动机、人工生命、耗散结构和自组织现象、复杂网络等。

在对复杂网络的研究中，随机图理论^[1]在将近 40 年的时间里一直是研究复杂网络结构的基本理论，但绝大多数实际的复杂网络结构并不是完全随机的。例如，两个人是否是朋友，WWW 上两个页面之间是否有超文本链接等都不会是完全靠抛硬币来决定的。在 20 世纪即将结束之际，对复杂网络的科学探索发生了重要的转变，复杂网络理论研究不再局限于数学领域。人们开始考虑节点数量众多、连接结构复杂的实际网络的整体特性，在从物理学到生物学的众多学科中掀起了研究复杂网络的热潮，甚至于被称为“网络的新科学^[2-3]”。

复杂网络的研究方法即把复杂系统简化为节点以及连接节点的边的集合，其中节点代表系统中各个不同的个体，而边则用来表示个体之间的联系。复杂网络中每个节点或每条边都可以加上称为“权”的不同的说明。这很可能是一种合适的、抓住本质的抽象描述。目前而言，复杂网络理论的主要研究内容可以归纳为：

(1) 揭示刻画网络系统结构的统计性质，以及度量这些性质的合适方法；(2) 建立合适的网络模型以帮助人们理解这些统计性质的意义和产生机理；(3) 基于单个节点的特性和整个网络的结构性质预测网络的行为；(4) 提出改善已有网络性能和设计新的网络的有效方法，特别是稳定性、同步和数据流通等方面。有两篇开创性的文章可以看作是复杂网络研究新纪元开始的标志：一篇是美国 Cornell 大学理论和应用力学系的博士生 Watts 及其导师、非线性动力学专家 Strogatz 教授于 1998 年 6 月在 Nature 杂志上发表的题为《“小世界”网络的集体动力学》(Collective Dynamics of ‘small-world’ Networks)^[4]；另一篇是美国 Notre Dame 大学物理系的 Barabási 教授及其博士生 Albert 于 1999 年 10 月在 Science 杂志上发表的题为《随机网络中标度的涌现》(Emergence of Scaling in Random Networks)^[5]的文章。这两篇文章分别揭示了复杂网络的小世界特征和无标度性质，并建立了相应的模型以阐述这些特性的产生机理。

社会学是研究人类社会及其中现象的科学。长期以来，社会学家们意识到社会学研究面临的主要问题是理论和模型的模糊性，从而导致模型和理论难于在实

证研究中被证实,甚至难于被“证伪”。20世纪60年代以来,H.White以及后继的Boorman,Brieger,Freeman等人基于数学的图论提出的社会网分析方法,在模型的实证测量可能性和许多可测量的有效解释上前进了一大步。此后,一大批社会学家运用社会网分析方法提出了社会网上的传染病、信息、时尚、新知识等的传播理论、劳动力的求职和转职的理论、组织结构的形成理论等等。这些理论的结论都是可以实证验证的。从20世纪90年代开始,社会网分析方法和理论在全球社会学界得到了公认,为许多社会学的分支学科所接受^[6-8]。

社会网络可以分为单模式网络(只有同一类节点,称为参与者(actor))和双模式网络(有两类不同的参与者),甚至更多模式的网络。双模式网络中的最重要一种称为隶属网(affiliation network),其中一类节点是某种活动、事件或者组织中的参与者,而另一类节点就是它们参与的活动、事件或者组织(称为项目(act))。这种网络可以用一个二分图(bipartite graph)来比较好的描述。近年来,隶属网中的合作网络(collaboration network)(即任意两个节点之间的边只表示节点之间的合作关系的网络,忽略节点之间的其他关系,例如竞争、支配等)引起了分外的关注^[9-16]。在合作网络的研究中,常常把二分图向一类节点(常常是参与者节点)投影,得到单模式网络。这时参与每个项目的所有参与者节点之间都连有表示在此项目中合作关系的边,每个项目就表示为一个完全图(complete graph),整个单模式网络成为完全图的集合。

从1998年以来,对于实际网络统计性质的实证研究一直受到重视。大家都意识到这是复杂网络研究的基础工作。再好的模型也要有实证数据的支持,否则就是无源之水、无本之木。在实证研究的基础上就可能把实际系统分类,找出每一类的实证统计性质的共性。再进一步,就可能建立模型,对导致这些共性的动力学机制加以阐述,以求达到对这一类复杂系统的更深理解。

我们(以及合作者们)实证地研究了好莱坞和中国的演员合作网、中国航空线路网、中国旅游线路网、北京和扬州的公交线路网、中药方剂网、淮扬菜系网、中国铁路轨线及车次、以及长江的地理及航运网^[17-22]等等,建立和研究了一些相应的网络演化模型^[17-24]。其中一些网和其他许多实际网络虽然不是社会网络,但

是，如果强调整节点在它们所属的项目中的合作关系（例如中药药材在每幅中药方剂中合作、食品在每道淮扬菜肴中合作、旅游景点在每条旅游线路中合作、公交站点在每条公交线路中合作等等），也可以类似地定义合作项目和参与者，使整个网络成为完全图的集合。接下来的问题可能是：这样的定义有意义吗？我们说：除了每个具体系统的合作网络定义各有其方便之处以外，最重要的是，要看这样定义之后，这些（具有与社会合作网络完全相同拓扑结构特征的）网络是否会具有（包括社会合作网络的）共同统计特征和导致这些共性的共同动力学机制（这很可能是以前的社会网络研究没有考虑过的）。如果有，那么这种定义就可能揭示许许多多看起来十分不同，甚至毫不相干的复杂系统的共同动力学机制与规律，从而向认识复杂世界迈进一步。我们得到的初步研究结果倾向于肯定的答案，因此，我们建议把合作网络的概念推广到这些非社会网络的描述中去，称为广义合作网络（generalized collaboration networks）^[22,24,25]。这个建议大大扩充了合作网络的范围。

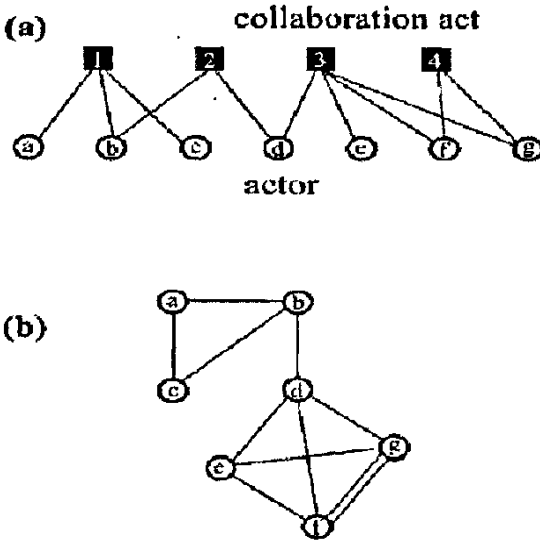
众所周知，节点之间的关系是有很多种的，合作只是其中比较重要的一种关系。和合作关系相对的另一重要关系就是竞争关系。我们对于广义合作网络的研究可以启发我们对竞争或合作竞争并存的复杂系统再作细致的描述和研究。

第二章 广义合作网络的项目度分布与点强度分布

2.1 实证统计调研

2.1.1 项目度、点强度概念

本文的广义合作网络的原始描述是由项目—参与者构成的二分图 (bipartite graph) 表示的。在广义合作网络的研究中, 把二分图向参与者节点投影, 得到单模式网络。这时参与每个项目的所有参与者节点之间都连有表示在此项目中合作关系的边 (忽略节点之间的其他关系、例如竞争、支配等), 每个项目就表示为一个项目完全图, 整个广义合作网络成为完全图的集合。如图 2.1.1 所示。



在这样的项目完全图的集合所描述的广义合作网络中, 节点 i 参与了多少个项目显然是描述节点 i 属性的一个重要物理量。于是我们定义这个量叫顶点项目度 h_i 。网络中顶点项目度的分布情况可用分布函数 $P(h)$ 来描述, 表示的是一个随机选定的节点的项目度恰好为 h 的概率。

在复杂网络研究中, 度 (degree) 一直都是单独节点的属性中简单又重要的概念。节点 i 的度 k_i 定义为与该节点连接的其他节点的数目。直观上看, 一个节点的度越大就意味着这个节点在某种意义上越“重要”。网络中节点的度的分布情况可用分布函数 $P(k)$ 来描述。如图 2.1.1(b) 中的 e 节点, 它和节点 f 、 d 、 g 相连, 所以它的度是 3; f 节点与节点 e 、 d 、 g 相连, 则 f 的度也是 3。如果从度的意义来看, 节点 c 和 f 是同样的重要地位。但是我们注意到节点 f 和节点 g 共同参与了两个项目, 它们之间连了两条边。显而易见, 这层意义上, 节点 f 比节点 e 的重要意义更强一些。为了描述这

(此图引自 Jun Ohkubo et al., PRE 72, 036120, 2005)

图 2.1.1 (a) 二分图的草图。(b) 图(a)的单模式投影。

样的区别，我们定义广义合作网络中的边权 ω_y 为两个节点共同参与的项目数，即计入重复连边时的边数。所以在此网络中，描述节点的连接属性时，不但要考虑到节点的近邻数，也该考虑到该节点与近邻节点的边的权重，于是定义关于节点的又一个量，即点强度 s_i 。其定义式为： $s_i = \sum_{j \in \Gamma_i} \omega_y$ ，其中 Γ_i 是节点 i 的邻近节点集合，同样的，网络中点强度分布情况可用分布函数 $P(s)$ 来描述。

2.1.2 一些例子的实证统计结果

近些年来的大量研究表明，许多实际网络的度分布主要有三种形式：幂律分布（也称为无标度(scale-free)分布）、指数分布、介于幂律分布和指数分布中间分布。表示介于幂律分布和指数分布中间分布有两种：广延指数分布^[30]（Stretched Exponential Distribution，简称 SED）、漂移幂律分布（Shifted Power Law，简称 SPL）。为了与后面介绍的模型解析结果相一致，本文中用 SPL 函数拟合所有的介于幂律和指数之间的分布函数。本组的许多统计结果也表明，不仅仅是度分布，还有项目度分布、点强度分布主要都是这三种分布形式。本组统计了很多的实际网络的性质，由于诸如数据量小、数据涨落比较大等等的原因，很多直接拟合的分布函数不是很直观，就用了另外一种方法同样的来判断分布拟合属于哪种函数，即绘制累积分布函数^[26]。累积分布函数和不累积分布函数定性的一致，下面以度为例作简要的推导说明。

如果度分布为幂律分布，不累积度分布为 $P(k) \propto k^{-r}$ ，累积度分布

$$P(k') = \sum_{k=k'}^{\infty} P(k), \text{ 则 } P(k') \propto \sum_{k=k'}^{\infty} k^{-r} \propto \int_{k'}^{\infty} k^{-r} dk \propto k'^{-(r-1)}, \text{ 即累积后的度分布仍然是幂律分布，幂指数比不累积分布少 1。}$$

如果度分布为指数分布，不累积度分布为 $P(k) \propto e^{-k/\kappa}$ ，其中 $\kappa > 0$ 是一常数，

$$\text{累积度分布 } P(k') = \sum_{k=k'}^{\infty} P(k), \text{ 则 } P(k') \propto \sum_{k=k'}^{\infty} e^{-k/\kappa} \propto \int_{k'}^{\infty} e^{-k/\kappa} dk \propto e^{-k'/\kappa}, \text{ 即累积后的度分布仍然是指数分布，指数不变。}$$

如果度分布为 SPL 分布, 不累计度分布为 $P(k) \propto (k + \alpha)^{-\eta}$, 累计度分布

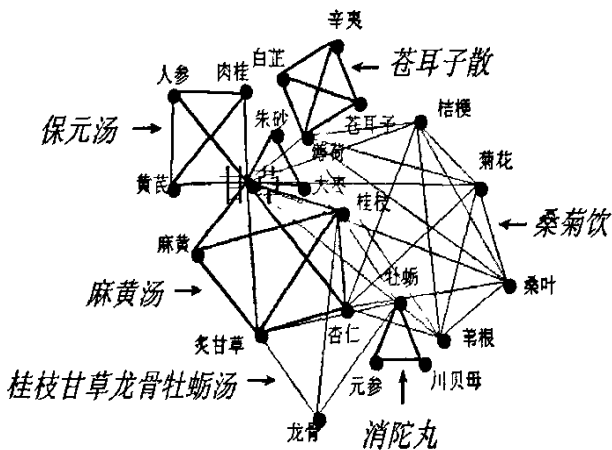
$$P(k') = \sum_{k=k'}^{\infty} P(k), \text{ 则 } P(k') \propto \sum_{k=k'}^{\infty} (k + \alpha)^{-\eta} \propto \int_{k'}^{\infty} (k + \alpha)^{-\eta} dk \propto (k' + \alpha)^{-(\eta-1)}, \text{ 即累计后}$$

的度分布仍然是 SPL 分布, 关键参数 α 不变, 幂指数少 1。

由上分析, 累计分布和不累计分布的拟合函数定性的一致, 而且关键参数有一定的关系, 那么用绘制累计分布图来分析实际数据的方法不但可以减少数据的涨落, 弥补实际数据的一些局限性, 也能更好的分析分布的规律。在我们的实际统计中, 我们更多的采取了这样的方法。

2.1.2.1 中药方剂系统的实证统计结果

我们把每一种中药方剂中的每一种药材定义为一个顶点, 处于同一道方剂中的两种药材之间的相互作用关系 (即在同一道中药方剂的制备过程中共同进行生化反应, 形成对于特定病人具有良好效果的药剂的合作关系) 定义为顶点之间的边, 则每道方剂成为其中药材互相作用构成的完全图项目。各道方剂中共同的药材可以看作是这些项目之间联系的桥梁, 这样就构成了一个中药方剂网络。如图



2.1.2 所示。

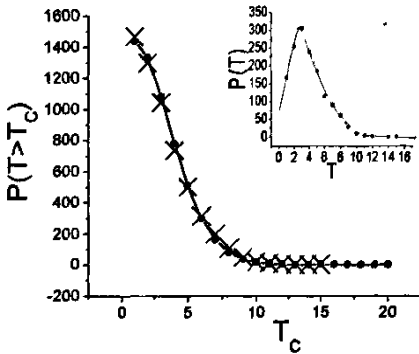
我们选用了文献^[27,28]中的 1536 道中药方剂 (完全图项目)、681 种药材 (参与者) 进行了统计。这些方剂是作为中华民族瑰宝的中药方剂中经过长期验证、经常使用的代表性部分。尽管数据量不太大, 但是仍然可以得到不错的统计规律, 并且和我们

图 2.1.2 中药方剂网络的完全图项目示意图

我们在其它广义合作网络中得到的结论以及我们后面将要讲述的模型模拟结果基本一致, 说明统计仍旧是合理可信的。

图 2.1.3 的插入图显示了统计得到的中药方剂网项目含顶点数 T 的分布, 主图显示了统计得到的中药方剂网项目含顶点数的累计分布。如图所示, 可以用一个近似于 $p(T) = \frac{1}{A} \frac{\lambda^{T+b}}{(T+b)!} \cdot e^{-\lambda}$ 的概率分布函数比较好的拟合这些数据。从插入图

中可见中药一幅方剂 (完全图项目) 最多包括 15 种药物 (参与者), 但是这样的完全图项目只有两个; 最少包括 1 种药物, 这样的完全图项目有 168 个。一道方剂由 3 种药材组成的可能性最大, 这样的完全图项目有 306 个, 而更少或更多药材组成的可能逐渐减少。平均项目含顶点数为: $\langle T \rangle = 3.90$ 。这体现中药方剂的特征。



图中实心圆圈表示相应的统计数据。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。主图中的大叉子符号表示用 $P(T)$ 函数拟合的结果 ($\lambda=6.5$, $b=6$)。

图 2.1.3 中药方剂网项目含顶点数 T 的累计分布 (插入图显示项目含顶点数 T 的分布)

图 2.1.4 显示了统计得到的中药方剂网项目度累计分布图, 图 2.1.5 显示了统计得到的点强度累计分布图。它们都显示了很好的 SPL 分布, 介于幂律和指数分布之间。但它们的关键参数并不相同。点强度分布实证数据中显示 $\alpha=20$, 项目度分布实证数据中显示 $\alpha=6$ 。图中的直线表示数据的最小二乘方拟合的结果。

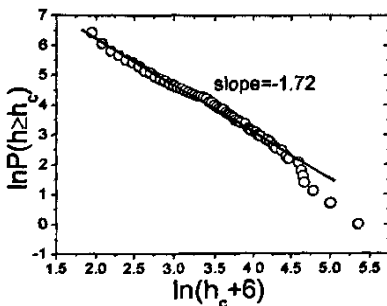


图 2.1.4 中药方剂网项目度累计分布图

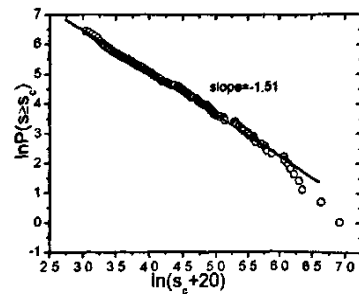


图 2.1.5 中药方剂网点强度的累计分布图

2.1.2.2 中国旅游线路系统的实证统计结果

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络。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)的两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。如图 2.1.6 所示。

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(参与者节点,它们之间共有 719 条边)进行了统计研究。如上所述,

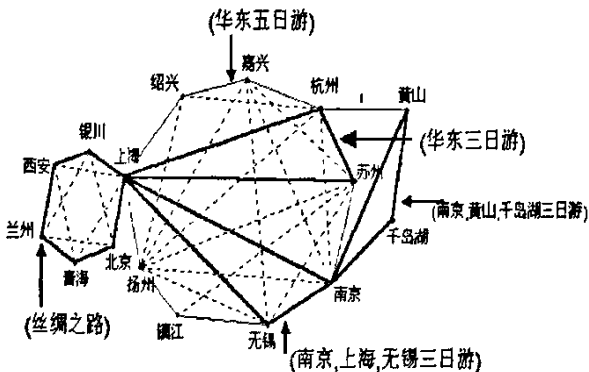


图 2.1.6 中国旅游线路网络示意图

各个完全图之间通过共用的节点架起了桥梁,从而构成了整个中国旅游线路网络。

图 2.1.7 显示了统计得到的旅游线路网的项目含顶点数 T 的分布。一条旅游线路最多包括 18 个旅游点,最少包含 1 个。

图 2.1.7 说明一条旅游线路由 2

个旅游点组成的可能性最大,而更少或多个旅游点组成的可能逐渐减少。图 2.1.8

显示了统计得到的旅游线路网项目含顶点数 T 的累计分布，同中药方剂网类似，

可以用一个近似半枝泊松分布函数 $p(T) = \frac{1}{A} \frac{\lambda^{T+b}}{(T+b)!} \cdot e^{-\lambda}$ 比较好的拟合这些数据。

图中大叉子符号表示用 $P(T)$ 函数拟合的结果 ($\lambda=7.5, b=6$)。

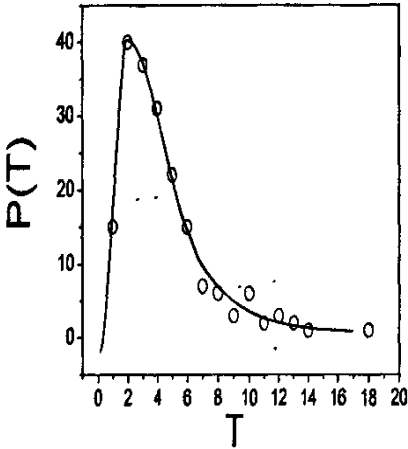


图 2.1.7 中国旅游线路网项目含顶点数分布图

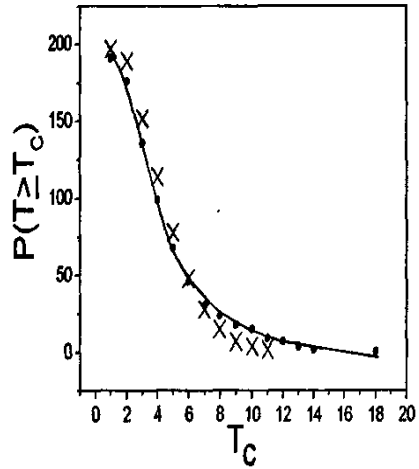


图 2.1.8 中国旅游线路网项目含顶点数累计分布图

图 2.1.9 显示了统计得到的中国旅游线路网顶点项目度累计分布，图 2.1.10 显示了统计得到的中国旅游线路网点强度累计分布，它们和上述中药方剂网结果一样，都显示了很好的 SPL 分布，并且关键参数有所不同，但相差不大。

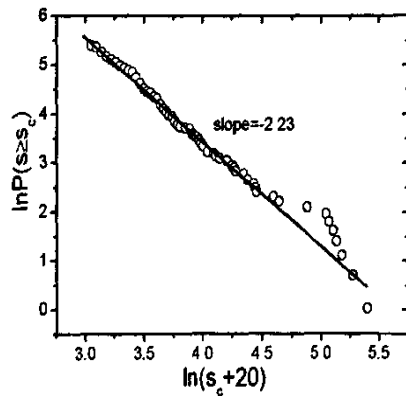
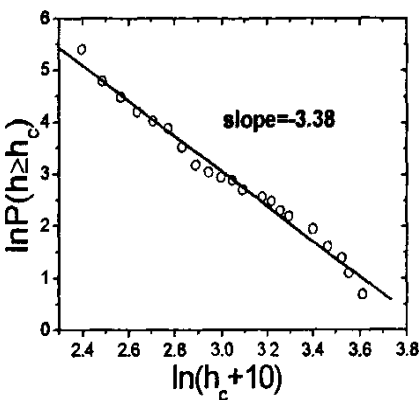


图 2.1.9 中国旅游线路网顶点项目度累计分布图 图 2.1.10 中国旅游线路网点强度累计分布图

2.1.2.3 中国淮扬菜系系统的实证统计结果

中国传统的烹饪菜肴系统的网络描述与中药方剂系统的描述接近。中菜的名作少说也有几万道，最著名的属于八大菜系。其中，淮扬菜系为淮安、扬州、镇江三地风味菜的总称，以扬州菜最为著名，有悠久的历史 and 独特的个性。我们类似地定义淮扬菜谱中每一种食品为顶点（参与者），处于同一道菜肴中的两种食品之间的相互作用关系为边，因此每道菜肴就是一个完全图项目，它们通过公共顶点的连接构成中国淮扬菜肴网。我们在文献^[29]中选取了 329 道菜肴，242 种食品（共 1713 条边）进行了统计。中国淮扬菜肴网的统计结果和上述两种网络统计结果定性的一致，项目含顶点数的分布仍可看作是个变形的泊松分布。而项目度分布和点强度分布都是 SPL 分布，如图 2.1.11 和图 2.1.12 所示。

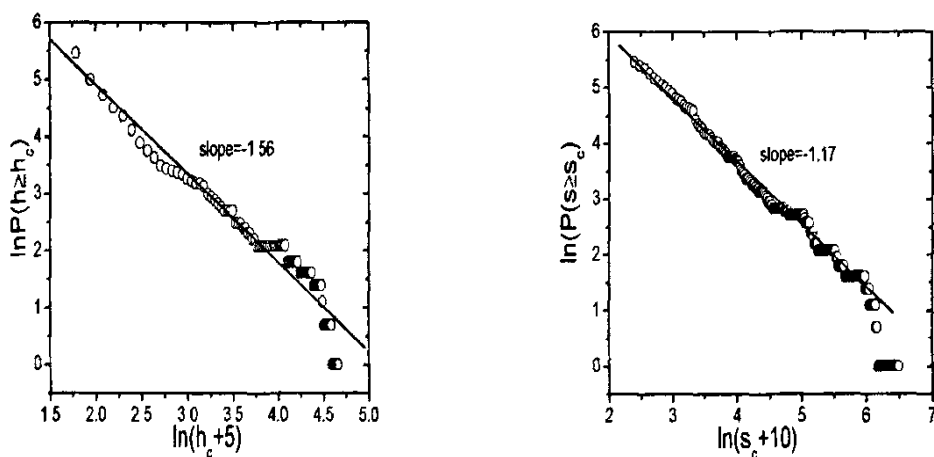


图 2.1.11 中国淮扬菜肴网顶点项目度累计分布图 2.1.12 中国淮扬菜肴网点强度累计分布图

2.1.2.4 好莱坞演员系统的实证统计结果

前面三种广义合作网络的数据量都不大，为了进一步说明这类网络实证统计的可靠性，我们调研了好莱坞演员系统。我们定义每一个演员为顶点（参与者），参加同一部影片合作演出的演员之间连边，这样每部影片就是一个完全图项目，这样完全图的集合构成了好莱坞演员合作网。我们下载的美国好莱坞演员合作网络的数据包括了 392,304 名演员和 181,455 部影片。它的项目含顶点数分布仍可看作是一个近似泊松分布。好莱坞演员合作网的项目度分布和点强度分布都可用

SPL 函数来很好的拟合。如图 2.1.13 和图 2.1.14 所示。

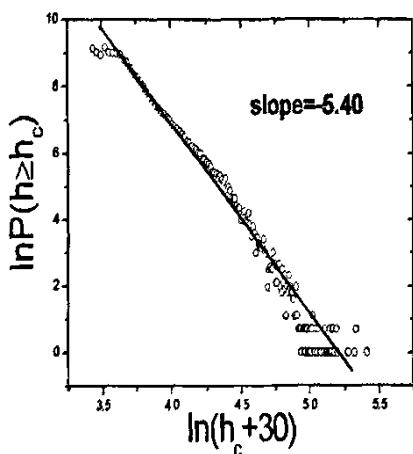


图 2.1.13 好莱坞演员合作网的顶点项目度
累计分布图

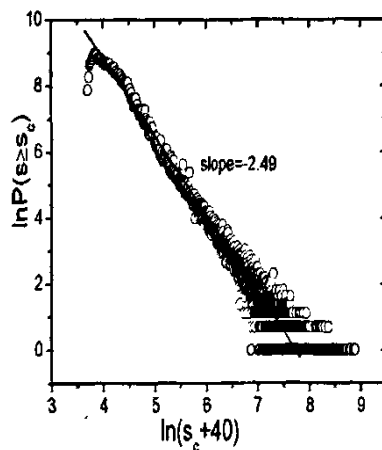


图 2.1.14 好莱坞演员合作网的点强度
累计分布图

2.1.3 小结

除了上述的四种广义合作网络以外，我们还实证统计了上海、北京、南京、杭州、扬州等地的公交线路合作网络，这些实证网络和上述四种广义合作网络一样，得出的项目度分布和点强度分布也都可以用 SPL 函数（趋向指数分布）来很好的描述，而 SPL 分布函数是可以根据其关键参数的连续变化连续的从幂律分布过渡到指数分布这样的一个范围很广的函数形式。所以，我们从这些实证研究中总结出的广义合作网络的第一个共同统计特征就是：广义合作网络中的项目度分布和点强度分布一般都具有处于幂律函数和指数函数这两种极端情况之间的分布函数，而且具有密切的一致关系，即如果网络的项目度分布趋于指数分布，则网络的点强度分布也是趋于指数分布的。那自然而然我们就会想了解，这样的广义合作网络是有什么样的物理机制来形成的呢，怎么会形成这样的一个共同的统计特征呢？于是，我们建议了一个网络演化模型来描述产生这种共性的机制。

2.2 一些模型的简介

ER 随机图和 WS 小世界模型的一个共同特征就是网络的连接度分布可近似用 Poisson 分布来表示，该分布在度平均值有一峰值，然后呈指数快速衰减。这

意味者度极大的节点几乎不存在。近年在复杂网络研究上的另一重大发现就是许多复杂网络,包括 Internet、WWW 以及新陈代谢网络等的连接度分布函数具有幂律形式。为了解释幂律分布的产生机理,Barabási 和 Albert 提出了一个无标度网络模型,现被称为 BA 模型^[5]。他们认为大多实际网络有如下两个重要特性:第一,增长(growth)特征:即网络规模是不断扩大的。例如 WWW 上每天都有大量新的网页产生;第二,优先连接(preferential attachment)特性,即新的节点更倾向与那些具有较高连接度的“大”节点相连接。这种现象也称为“富者更富(rich get richer)”或“马太效应(Matthew effect)”。例如,新的个人主页上的超文本链接更有可能指向著名的站点。

基于网络增长和优先连接特性,BA 模型的算法有两个要点:第一,网络顶点一个一个地增加;第二,新出现顶点与已经存在的顶点之间按照线性优选法则建立边的联系。由此模型得到的解析和模拟结果显示理想的幂律。Barabási 和 Albert 同时解析地说明了如果新出现顶点与已经存在的顶点之间完全随机地选择连接,那么顶点度的分布满足一个指数分布。在此后的许多旨在改进或修正 BA 模型的建议中,我们特别注意以下两个。第一是刘宗华、来颖诚等人建议的介于无标度网和随机网之间的网络模型^[31]。他们举出了科研合作网的数据统计结果介于指数函数和幂函数之间作为一个例证,说明网络的演化存在优选法则和随机选择这两种机制的竞争。类似的,他们假设网的顶点一个个地增加,但是设新顶点以概率 p ($0 \leq p \leq 1$) 随机连接旧顶点,以概率 $(1-p)$ 按照线性优选法则连接旧顶点。当 $p=0$ 时,刘来模型退化为 BA 模型,而 $p=1$ 时它退化为 BA 随机模型。在在 $0 < p < 1$ 时,模型具有介于幂律和指数之间的顶点度分布。第二个模型是李翔、陈关荣提出的,表现了关于优选和随机选择这两种机制竞争的更具体、更深刻的思想^[32]。他们举出世界贸易网作为例子,说明许多网络中存在“优先互相作用”的局域世界,例如内部的经济合作和贸易关系优先的各种区域经济合作组织、互联网中的域一路由器结构,以及蛋白质的族系和超族系。在李陈模型中网络的顶点也一个一个地增加,但是线性优选法则只适于随机选取一部分顶点构成的局域

世界内部。当局域世界线度 M （它包含的顶点数）等于 m （每个新顶点在它的局域世界中连接的新边的数目）时，新点和局域世界所有的顶点相连接，无优选可言，李陈模型退化为 BA 随机模型。当 $M = m_0 + t$ （ m_0 是初始顶点数， $m_0 + t$ 是 t 时刻的顶点数），局域世界就是整个网络，连接法则就是整个网络的线性优选，李陈模型退化为 BA 模型。当 $m < M < m_0 + t$ 时，模型具有介于无标度和指数之间的顶点度分布。

上节讨论了二分图对广义合作网络的描述。在社会网络研究中早就有关于 i 顶点的项目度（即参与者参加多少个项目） h_i 的定义以及关于 j 项目的“项目含顶点数” T_j 的定义。显然，在向参与者的投影单模式图中，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，从这个完全子图来说，它们中的每一个顶点的顶点度都等于 $T_j - 1$ 。当然，两个（甚至更多的）这样的完全图很可能共用一条甚至好几条边，那么从整个网络来看，对某个顶点 i 来说，它参与的所有的项目的 $T_j - 1$ 的叠加就是第一节中所定义的点强度 s_i 。

RDP (Ramasco, Dorogavtsev and Pastor-Satorras) 模型^[33]是我们所知道的唯一讨论过合作网络中项目度分布 (act degree distribution) 和度分布 (degree distribution) 之间关系的模型。这个模型设想每步时间演化过程中组建一个包含 T 个参与者的项目 (项目含顶点数为 T)。 T 可能是常数，也可能是遵从项目含顶点数概率分布 $P(T)$ 的随机数。其中 m 个是新参与点，其余 $T-m$ 个从旧顶点中按照正比于它的项目度 h 的概率优选，RDP 解析地证明了在参与者的投影单模式图中，度分布 $P(k)$ 和项目度分布 $P(h)$ 均为严格的幂函数，而且标度因子相同，都等于 $2 + m/(T - m)$ 。

我们在上述实际网络中得到的关于点强度分布 $P(s)$ 和项目度分布 $P(h)$ 的统计结果倾向于说明 RDP 模型得到的结论一定程度上是合理的，但是只代表了一中极

端情况。为此我们借鉴 BA 模型、刘来模型、李陈模型 RDP 模型，建议一个可能产生幂律、指数函数以及它们之间的各种点强度分布和项目度分布的简化模型，这种结果与我们的广义合作网络实际统计结果更加符合。以下我们将分两步来叙述我们建议的模型。

2.3 我们的最简化模型

设初始 $t=0$ 时有 m_0 个顶点，已经联接成若干个完全图项目，它们的项目度 h_{i0} 之和为 h_0 。每步时间演化过程增加一个新顶点，然后，按照一定法则选取 $T-1$ 个 (T 为常数) 旧顶点，把这 $T-1$ 个旧顶点和这个新顶点 (共 T 个顶点) 中两两之间尚未连接的边都连上，构成一个新的完全图项目。

2.3.1 最简化模型的优选连接情况

设上述选取 $T-1$ 个旧顶点的法则是：选取每个旧顶点 i 的概率正比例于它的项目度 h_i (项目度线性优选法则)，则仿照 BA 模型的准连续近似解法^[5]，我们列出项目度 h_i 的准连续近似演化方程：

$$\frac{\partial h_i}{\partial t} = (T-1) \frac{h_i}{\sum_j h_j} \quad (2.3.1)$$

由于在 t 够大时， $\sum_j h_j = T(t-1) + h_0 \approx Tt$ ，此方程可以简化为

$$\frac{\partial h_i}{\partial t} \cong (T-1) \frac{h_i}{Tt} \quad (2.3.2)$$

由此解得： $h_i(t) = Ct^\beta$ ，其中 $\beta = (T-1)/T$ ， $C = 1/t_i^\beta$ (注意当 $t=t_i, h_i(t_i)=1$)。由此可得 $P(h_i < h) = P(t_i > t/h^{1/\beta})$ 。考虑到 $P(t_i) = 1/(m_0 + t)$ ，可解得： $t_i < t/h^{1/\beta}$ 的概率为 $t/h^{1/\beta} - 1/(m_0 + t)$ 。因此，由 $P(h_i) = \partial P((t/t_i)^\beta < h) / \partial h$ ，可得到项目度分布为：

$$P(h) = \frac{t}{(m_0 + t)^\beta} h^{-\left(\frac{1}{\beta} + 1\right)} \quad (2.3.3)$$

即项目度分布为幂律，标度因子 $\gamma = 1/\beta + 1 = (2T-1)/(T-1)$ 。它随项目含顶点数 T 的变

化而变化, 若 $T=2, \gamma=3; T=3, \gamma=2.5, T=5, \gamma=2.25, T=7, \gamma=2.17; T=9, \gamma=2.12;$
 T 趋于无穷时, γ 趋于 2。如 2.2 节中所述, 总有点强度 $s_i = h_i(T-1)$, 由此可得
 点强度分布为:

$$P(s) \propto s^{-\nu} \quad (2.3.4)$$

且 $\nu = \gamma$, 即与 RDP 模型的结论相同, 点强度分布 $P(s)$ 和项目度分布 $P(h)$ 均为严格的
 幂函数, 而且标度因子相同。数值模拟结果与此解析结论很好地符合^[34]。

2.3.2 最简化模型的随机连接情况

若上述选取 $T-1$ 个旧顶点的法则是随机选取, 则类似地仿照 BA 模型的准
 连续近似解法^[5], 有:

$$\frac{\partial h_i}{\partial t} = (T-1) \frac{1}{m_0 + t - 1} \quad (2.3.5)$$

类似考虑当 $t = t_i, h_i = 1$, 解得:

$$h_i(t) = (T-1) \ln \frac{m_0 + t - 1}{m_0 + t_i - 1} + 1 \quad (2.3.6)$$

由此可得 $P(h_i < h) = P(t_i > (m_0 + t - 1) \exp[(1-h)/(T-1)] - m_0 - 1)$ 。考虑到 $P(t_i) = 1/(m_0 + t)$, 可解
 得:

$$P(h) = B e^{\frac{1-h}{T-1}} \quad (2.3.7)$$

$$\text{或} \quad \ln P(h) = (T-1) \ln B + 1 - \frac{h}{T-1} \quad (2.3.8)$$

其中 $B = (t + m_0 - 1) / [(T-1)(m_0 + t)]$ 。即项目度分布为指数规律, 斜率因子为 $\gamma = -1/(T-1)$ 。

它也随常数项目含顶点数 T 的变化而变化, T 趋于无穷时, γ 趋于 0。类似地, 点
 强度 $s_i = h_i(T-1)$, 由此可得点强度分布为:

$$\ln P(s) \propto -s/(T-1)^2 \quad (2.3.9)$$

这是 RDP 模型未讨论的另一个极端情况。这时点强度分布 $P(s)$ 和项目度分布 $P(h)$
 均为严格的指数函数, 但是斜率因子不同。 $P(s)$ 分布的斜率因子为 $P(h)$ 分布的斜

率因子的平方。数值模拟结果与上述解析结论也很好地符合。

2.3.3 最简化模型的部分优选连接、部分随机连接情况

如果上述选取 $T-1$ 个旧顶点的法则是以一定的概率 p 随机连接、以其余的概率 $1-p$ 优选连接, 则类似地仿照 BA 模型的准连续近似解法^[5], 有:

$$\frac{\partial h_i}{\partial t} = p(T-1)\frac{1}{t+m_0} + (1-p)\frac{(T-1)h_i}{T(m_0+t)} \quad (2.3.10)$$

类似可以解得:

$$h_i(t) = \left(1 + \frac{Tp}{1-p}\right) \left(\frac{m_0+t}{m_0+t_i}\right)^{(T-1)(1-p)/p} - \frac{Tp}{1-p} \quad (2.3.11)$$

由此可得:

$$P(h) = \frac{\beta}{(1+\alpha)} \left(\frac{h+\alpha}{1+\alpha}\right)^{-\eta} \quad (2.3.12)$$

其中 $a = \frac{Tp}{1-p}$, $\beta = \frac{T}{(T-1)(1-p)}$, $\eta = \beta + 1$ 。这时的项目度分布称为“移动幂律

(Shifted Power Law, 简称 SPL)”。容易得到, 当 $p=0$ 时, $\alpha=0$, (2.3.12) 式变

为典型的幂律; 而当 $p \rightarrow 1$ 时, $\alpha \rightarrow \infty$, (2.3.12) 式变为典型的指数函数。因此,

SPL 表示幂律和指数函数之间的分布。当 p 从 0 到 1 连续地变化时, 项目度分布连续地从幂律转变为指数函数。

类似地, 总有点强度 $s_i = h_i(T-1)$, 所以可以得到:

$$\frac{\partial s_i}{\partial \ln t} = p(T-1)^2 + (1-p)\frac{(T-1)S_i}{T} \quad (2.3.13)$$

令 $\alpha = \frac{T(T-1)p}{1-p}$, $\beta = \frac{T}{(T-1)(1-p)}$, $\eta = \beta + 1$, 可得:

$$p(s) = \frac{\beta}{T-1+\alpha} \left(\frac{s+\alpha}{T-1+\alpha}\right)^{-\eta} \quad (2.3.14)$$

类似地, 容易得到, 当 $p=0$ 时, $\alpha=0$, (2.3.14) 式变为典型的幂律; 而当 $p \rightarrow 1$ 时, $\alpha \rightarrow \infty$, (2.3.14) 式变为典型的指数函数。显然, 点强度分布的 SPL 分布中

的关键常数 α 是项目度分布的 SPL 分布中的关键常数 α 的 $T-1$ 倍。数值模拟结果与上述解析结论也很好符合。在实际中，极端的优选或者随机的情况总是罕见的，在它们之间的情况才是常见的，因此可以期望实际系统的点强度分布和项目度分布基本上都可以用 SPL 来描述。

2.4 我们模型的更近于实际情况

在 2.1 节中报道了实证统计结果说明了一般情况下 T 并不是常数，所述的实际系统中，项目含顶点数 T 的分布函数 $P(T)$ 都近于一个变形的泊松分布。这使得点强度分布 $P(s)$ 和项目度分布 $P(h)$ 之间的关系变的更为复杂。我们仅对此情况做数值模拟的研究，希望说明在此复杂情况下点强度分布 $P(s)$ 和项目度分布 $P(h)$ 之间仍定性地具有密切的一致关系，而项目含顶点数的分布 $P(T)$ 对其影响较小。

2.4.1 优选连接情况

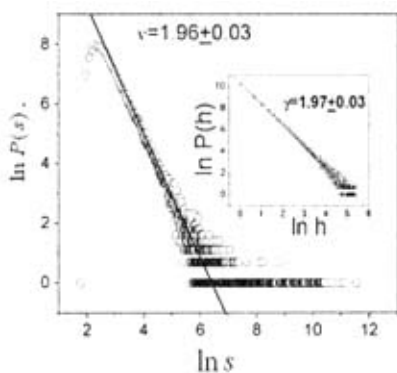


图 2.4.1 项目含顶点数为半枝泊松分布优选的点强度分布，插入图显示了项目度分布

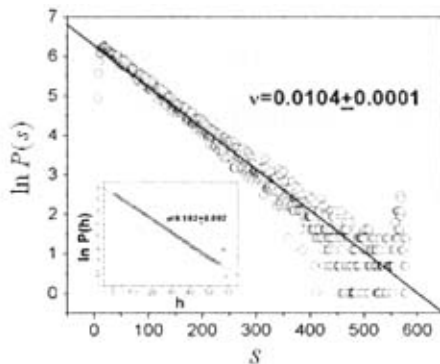


图 2.4.2 项目含顶点数为半枝泊松分布随机的点强度分布，插入图显示了项目度分布

我们根据大多数实际统计结果，令 T 为以半枝泊松分布函数 $p_\lambda(T) = \frac{1}{A} \frac{\lambda^T}{T!} \cdot e^{-\lambda}$ 的概率取值的随机数（其中 λ （这里取 λ 为 8.5）为函数的峰值， T 为大于或等于峰值 λ 的整数， $1/A$ 为归一化因子），然后在旧顶点中线性优选地（即与其顶点项目度正比例地）选取 $T-1$ 个构成一个完全图项目。这里取 λ 为 8.5 仅仅是为了做到使泊松分布函数峰值的数量级与我们大多数的实际统计结果接近。图 2.4.1 显示了项目度分布 $P(h)$ 和点强度分布 $P(s)$ 的数值模拟结果。结果显示这时项目度

分布 $P(h)$ 和点强度分布 $P(s)$ 都是幂律分布, 而且两种分布的幂律因子相同。取 λ 为其它值时结果相同。

2.4.2 随机连接情况

图 2.4.2 显示了我们对于 T 为上述半枝泊松分布, 随机选取 $T-1$ 个旧顶点时的数值模拟结果。结果显示这时项目度分布 $P(h)$ 和点强度分布 $P(s)$ 都是指数函数分布, 但是两种分布的斜率因子不相同(仍旧满足 $\nu \approx \gamma^2$)。

2.4.3 部分优选、部分随机连接情况

我们对 T 为上述半枝泊松分布, 以概率 p 随机选取, 以概率 $1-p$ 线性优选 $T-1$ 个旧顶点时的数值模拟结果显示项目度分布 $P(h)$ 和点强度分布 $P(s)$ 都是很好的 SPL 分布, 但是两种分布的关键参数 α 不相同。尽管如此, 这些数值模拟结果仍可以说明在实际情况下项目度分布 $P(h)$ 和点强度分布 $P(s)$ 之间仍定性地具有密切的一致关系, 而项目含顶点数的分布 $P(T)$ 对这两种分布的影响较小, 因此可以期待上述模型的解析结果被实证结果定性地或者接近的证实, 即实际系统的度分布和项目度分布基本上都可以用 SPL 来描述。

2.5 讨论

在广义合作网这类网络中, 由于拓扑结构具有共同特征, 它们的统计性质也具有共同规律。项目是这类网络中的主导因素, 项目如何组成是网络演化的关键问题。参与者的项目度分布很可能是这类网络的最主要的统计性质, 很大程度上左右了点强度分布。由于我们注重远离平衡的复杂系统的研究, 这类网络中各个参与者的信息往往是严重缺乏的, 通常它们只在各自的局域世界中比较密切的关联, 而越出这些局域世界, 就不具备这样的信息, 只能随机地参与项目的构成。

我们的实证结果和模型的结果定性上非常的吻合, 在一定程度上说明了我们模型的正确性和合理性。在中药方剂网中项目度最大的是甘草, 其次是茯苓、人参、当归等。这都是药性平和的辅助药物, 或者是协调、加强主、辅药材作用的协助药材。它们在许多不同方剂中广泛地应用, 成为网络中完全图项目间的最大桥梁。反之, 项目度很小的药材都是很少用的、具有某些独特功能的主、辅药,

例如刺猬皮、白果等。中药方剂网的平均项目度为 9.21，也就是平均看来每种药物参与 9 个方剂左右。统计的结果说明了中药方剂网的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，即构造方剂时既首选最常用的辅药，又考虑各种不同的具体因素（病的种类、病情、病人类型、季节等）来选各种药材，由于具体因素太多，互相又缺乏关联，相当于随机选取。类似地，中国旅游线路网在形成过程中的动力学机制既有优选的因素（首选知名度高的旅游点），又有随机因素（考虑各种不同的具体因素来选旅游点，因素太多，相当于随机选取）。中国淮扬菜系网和中药方剂网类似。好莱坞演员合作网的统计结果也表明在演化过程中，每部影片的导演挑选演员的原则也是既有优选因素又有随机因素，要考虑演员的知名度，也会在有限的情况下选择没有知名度的新人。这些系统的统计结果，即 SPL 分布的关键参数也可以定性的值明在网络的形成过程中，是哪种因素占主导地位。这应该是大部分广义合作网络的共同特征。

第三章 广义合作网络的同类性

3.1 同类性简介

3.1.1 同类性定义

在上一章中,我们介绍了广义合作网以及项目含顶点数 T 、项目度 h 、度 k 、点强度 s 的概念。顶点 i 的度 k_i 就是与 i 相连的顶点数目,即 $k_i = \sum_{j \in \Gamma_i} a_{ij}$, 其中 a_{ij} 为无权邻接矩阵元(顶点间有连边 a_{ij} 为 1, 没有连边为 0), Γ_i 为顶点 i 的邻点集合。顶点 i 的点强度 $s_i = \sum_{j \in \Gamma_i} \omega_{ij}$, 其中 ω_{ij} 为边权, 即顶点 i 和顶点 j 的合作强度(广义合作网中,指顶点 i 和顶点 j 在几个项目中共同合作出现), Γ_i 为顶点 i 的邻点集合。

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 (assortativity)”^[35]。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 (assortative mixing)”的;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 (disassortative mixing)”的。如果特别地研究具有不同度的节点之间的连接, 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 建议在无向网中计算度之间同类性的公式为^[35]:

$$r = \frac{M^{-1} \sum_{\Phi} (\prod_{i \in F(\Phi)} k_i) - \left(\frac{M^{-1}}{2} \sum_{\Phi} (\sum_{i \in F(\Phi)} k_i) \right)^2}{\frac{M^{-1}}{2} \sum_{\Phi} (\sum_{i \in F(\Phi)} k_i^2) - \left(\frac{M^{-1}}{2} \sum_{\Phi} (\sum_{i \in F(\Phi)} k_i) \right)^2} \quad (3.1.1)$$

其中 $F(\Phi)$ 表示第 Φ 个有连线的相邻节点对, M 表示网络中总的边数。

上一章中我们指出了当节点只考虑到邻点节点数而不考虑与邻节点连边合作的强度的话,就不能准确地反映实际网络的细致结构和功能。以交通网络为

例, 我们调研的许多的交通系统都可以在不同的空间来定义交通网络, 如 P 空间、L 空间和 K 空间^[36-43]。比如, 在 P 空间, 定义每一个公交站点为顶点, 至少有一条线路提供直接输送关系 (不需要换车) 的两个站点建立边的联系。也就是说, 如果两个站点必须转换路线才能互相输送乘客, 那么它们之间肯定没有直接连边。这样被定义的拓扑结构的交通网络也可被看作广义合作网。线路就是合作项目, 而站点就是参与者。一条线路上的站点两两互相连边形成一个项目子完全图。如在 L 空间中, 站点还是被定义为顶点, 如果两个站点至少在一条线路上是连续的相邻站点, 那么这两个站点才连边。这样被定义的拓扑结构的交通网络不能被看作广义合作网。事实上, P 空间交通网络比其它广义合作网络的边的含义更丰富一点, 因为交通网络的边刻画了交通能力或交通量的问题。在普遍的无权的 P 空间的交通网中, 边仅仅表示节点对直接运输关系。交通网中线路的信息反应在边上就是给边赋上叫“合作强度”的边权 ω_j 。显然, 这种讨论是适合所有的广义合作网的^[44-52]。那么也就有了上述的点强度 s_i 的定义。于是, 定义这类网络的加权同类性的公式修正如下^[35]:

$$r^\omega = \frac{H^{-1} \sum_{\Phi} (\omega_{\Phi} \prod_{i \in F(\Phi)} k_i) - \left(\frac{H^{-1}}{2} \sum_{\Phi} (\omega_{\Phi} \sum_{i \in F(\Phi)} k_i) \right)^2}{\frac{H^{-1}}{2} \sum_{\Phi} (\omega_{\Phi} \sum_{i \in F(\Phi)} k_i^2) - \left(\frac{H^{-1}}{2} \sum_{\Phi} (\omega_{\Phi} \sum_{i \in F(\Phi)} k_i) \right)^2} \quad (3.1.2)$$

其中 ω_{Φ} 表示第 Φ 个边的边权, H 表示网络中总边权数。如果网络中所有的边权都为 1 的话, 则加权后的定义 s 和 r^ω 就退化为无权情况下的定义 k 和 r ^[35]。

3.1.2 Newman 等人对同类性的讨论

Newman 实证地研究了科研合著网、演员合作网、公司董事会及其成员之间的合作网、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^[35]。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^[58]讨论了这个问题。他们通过一个模型的解析说明了一般情况下如果不允许网络中出现“多重边（边权大于 1）”，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，这个结论符合人们的直觉分析。如果网络中的节点归属于若干群落（即网络中内部联接密集，而对外联接稀疏的子图，对这个概念将在后面仔细讨论），那么度高的节点一定归属于更多的群落，度高的节点和度高的节点建立边联接的可能性就更大，但是这当然很可能导致它们之间的多重边，所以对多重边的限制一定大大消除了正同类性。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^[35]？这就是说，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^[56]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导致明显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这两个特性（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 这两篇论文的结论，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。更具体地说，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，而所有非社会网络的 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^[57]。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。

3.2 实证统计结果

3.2.1 项目度分布、点强度分布、度分布的实证结果

第二章已经显示了中药方剂网、中国旅游线路网、中国淮扬菜肴网、好莱坞演员合作网等四种广义合作网络的统计结果，结果表明了这些实际的广义合作网络的项目度分布和点强度分布是趋于一致的，并且都可以用 SPL 函数来很好的描述，SPL 函数的两个关键参数 α 和 η 可以传递系统的更详细的信息；第二章也详细的阐述了我们提出的简化和近于实际的模型，其解析和数值模拟的结果和实证统计结果定性很好的吻合。

度分布是自复杂网络研究以来非常重要的一个统计性质，度同类性的提出也是和度紧密相关的。我们实证统计了上述四种网络以及五个城市^[53]的 P 空间公交线路网络的度分布，结果显示了广义合作网络的度分布也是广泛的介于幂律和指数之间的，可以用 SPL 函数很好的描述，即 $P(k) \propto (k + \alpha)^{-\eta}$ (α 和 η 都为常数)。图 3.2.1 和图 3.2.2 分别显示了中药方剂网和中国旅游线路网的度累计分布统计结果，都是很好的 SPL 分布，图中直线表示数据的最小二乘方拟合结果。

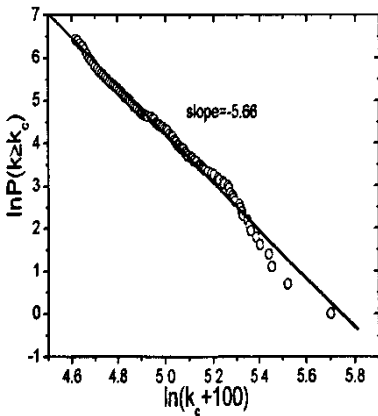


图 3.2.1 中药方剂网的度累计分布图

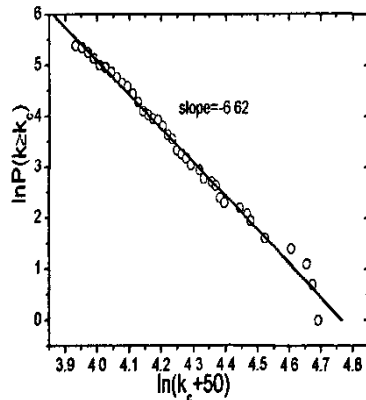


图 3.2.2 中国旅游线路网的度累计分布图

3.2.2 实证网络的同类性

我们实证计算了九种广义合作网络的度一度无权同类性以及加权同类性。表 I 除了显示了无权同类性以及加权同类性以外，还显示了九种广义合作网络的项

目度分布、点强度分布和度分布的 SPL 拟合函数的两个关键常数 α 和 η 。这些系统包括扬州公交线路网 (Bus Route Network of Yangzhou, 简称 BY)、杭州公交线路网 (Bus Route Network of Hangzhou, 简称 BH)、南京公交线路网 (Bus Route Network of Nanjing, 简称 BN)、北京公交线路网 (Bus Route Network of Beijing, 简称 BB)、上海公交线路网 (Bus Route Network of Shanghai, 简称 BS)、中药方剂网 (Traditional Chinese Herb Prescription Formulation Network, 简称 TC)、中国旅游线路网 (Travel Route Network of China, 简称 TR)、好莱坞演员合作网 (Collaboration Network of Hollywood Actors, 简称 HA) 和中国淮扬菜肴网 (Huai-Yang recipes of Chinese cooked food, 简称 HY)。

表 I 中的结果显示了广义合作网络中不计边权时的同类性系数有正有负, 但计入边权后的同类性系数全为正。因为如果不考虑边权, 就会忽略了影响系统中的重要因素, 所以我们有理由相信计入边权后的同类性系数更正确的说明了广义合作网络中度相关性; 五个城市公交线路网络的结果还可能说明了另外一个结论, 即网络规模越大, 加权同类性系数越大, 而不加权同类性却没有此规律。有人调研了 22 个波兰城市公交线路网^[39], 他们得出的结论是小规模 (节点数少于 500)

表 I: 九种实际广义合作网络的调研结果。表中, r 和 r^w 分别表示无权同类性系数和加权同类性系数; 项目度分布、度分布和点强度分布的 SPL 函数拟合参数 α 和 η 分别用 $\alpha-h, \alpha-k, \alpha-s, \eta-h, \eta-k, \eta-s$ 来表示; M 表示总项目数; N 表示总顶点数。

svstem	M	N	r	r^w	$\alpha-s$	$\eta-s$	$\alpha-h$	$\eta-h$	$\alpha-k$	$\eta-k$
BY	36	352	-0.11	0.04	infinite	0.015	infinite	0.38	100	5.01
BH	150	827	0.019	0.114	1000	18.5	30	9.9	1000	30.3
BN	252	1764	0.047	0.160	100	3.1	10	4.6	100	4.3
BB	572	4199	0.034	0.171	10000	37.8	1000	127.4	10000	77.9
BS	968	4374	0.05	0.202	200	3.3	30	6.6	200	5.3
TC	1536	681	-0.12	0.17	20	1.5	6	1.7	100	5.7
TR	240	171	0.15	0.36	20	2.2	10	3.4	50	6.6
HA	80000	210448	0.2653	0.00046	40	2.1	30	5.4	0	2.3
HY	329	242	-0.296	—	10	1.2	5	1.6	20	2.4

的城市公交线路网的无权同类性系数是负值，而大规模（节点数大于 500）的城市公交线路网的无权同类性系数是正值。表 I 中我们计算的中国四个大规模（节点数都是大于 500）的城市公交线路网也是显示正的无权同类性，扬州公交线路网的节点数小于 500，它显示负的无权同类性。我们的这个结果和上述提及的波兰城市公交线路网的结论是吻合的。这个表格中我们并没有对 SPL 函数的两个参数 α 和 η 作任何的结论，对此的讨论将在下两节中说明。

3.3 我们最简模型的同类性

在第二章中，已经详细的报道了我们提出的关于广义合作网络的最简化模型。我们数值计算了这个最简模型在不同的随机连接旧顶点和项目度线性优选旧顶点概率下所生成的网络的无权同类性系数和加权同类性系数。模拟计算所用的网络规模增长到 5000 个节点，项目含顶点数是固定值 6，进一步的计算表明，在定性的意义上计算结果与这两个参数的选择关系不大，结果如表 II 所示。

表中数据显示如下结论：（1）最简演化模型生成的网络在计入权重时显示正的同类性；（2）最简演化模型生成的网络在不计入权重时可能显示正的同类性或负的同类性；（3）加权同类性系数和无权同类性系数的值都随着随机选择旧顶点的概率 p 的增加而单调增加。结论（1）和（2）与表 I 的实证结果吻合的非常好。结论（3）可能能够更好的帮助我们理解表 I 中复杂的实证结果，这将在下面一节讨论。

p	r^w	r
0	0.0008 ± 0.003	-0.27 ± 0.0019
0.2	0.050 ± 0.003	-0.10 ± 0.0019
0.4	0.13 ± 0.004	-0.014 ± 0.0027
0.5	0.18 ± 0.005	0.060 ± 0.003
0.6	0.20 ± 0.005	0.10 ± 0.0035
0.8	0.28 ± 0.005	0.22 ± 0.004
1.0	0.31 ± 0.0045	0.29 ± 0.004

表 II：最简模型生成网络的无权同类性系数和加权同类性系数。表中， p 代表随机选择旧顶点的概率（项目度线性优选旧顶点的概率为 $1-p$ ）， r 代表是无权同类性系数， r^w 代表是加权同类性系数。

3.4 讨论

我们实证调研了九种广义合作网络并报道了部分结果。这些实证系统包括五种交通网络、三种技术网络（中药方剂网、中国淮扬菜肴网、中国旅游线路网）和一个社会网络（好莱坞演员合作网），这些网络都有相同的拓扑特征，它们都有两个基本的元素，一个被称为“项目”，另一个被称为“参与者”。在参与者投影网络中，项目就被表示为一个个项目子完全图，项目中的每两个参与者节点两两连边表示在项目中的合作关系。

实证结果有许多显见的结论。首先，项目度分布、度分布和点强度分布都一般的可以用 SPL 函数来描述。SPL 函数可以连续地从一个理想的幂律函数变化到一个理想的指数函数，当 $\alpha = 0$ 时，SPL 函数就是理想的幂律函数。当 $\alpha \rightarrow \infty$ 时，SPL 函数就趋向于理想的指数函数。另一个参数 η 是 SPL 函数的幂指数，表示函数衰减的快慢程度。其次，在我们的所有的实证网络中，加权同类性系数都是正值；而无权同类性系数可能显示正值或负值；第三个结论是从五个实证交通网络中得来的，即加权同类性系数和网络规模有单调的依赖关系。以上这些结论都是重要的，但是我们很难直接从这些实证数据中抽取出 SPL 函数的两个关键参数（ α 和 η ）与同类性系数的关系。然而，这两者之间的关系却是令研究者们十分感兴趣的。

我们的最简化模型，已经解析的得到了第一个结论。模型的最基本思想是广义合作网络是通过项目的逐步组织而演化生成的，其动力学演化机制可以分为两个趋向：随机选择和线性优先选择原则。基于这个思想，在随机选择旧顶点的概率 p ，项目度优选旧顶点的概率 $1-p$ 的情况下，我们写出了准连续演化方程，解析地得到了项目度分布和点强度分布是 SPL 分布，并且得出了随机选择旧顶点的概率 p 和 SPL 函数的两个关键参数 α 和 η 的清晰的关系式（见第二章）。我们不能从模型中解析得到同类性系数的清晰的表达式，但可以数值计算得到同类性系数。模型数值结果证实了实证调研的第二个结论，并且无论是无权同类性系数还是加权同类性系数都与随机选择旧顶点概率 p 有单调的依赖关系。那么可以得出

一个结论：同类性系数(r 和 r^o)和 SPL 函数关键参数(α 和 η)都与随机选择旧顶点概率 p 有单调的依赖关系。基于这个结论，就有可能间接的了解到 SPL 函数的关键参数是否和同类性系数有一定的依赖关系。由于 α 、 η 、 r 和 r^o 的数值都和项目含顶点数 (T) 有关系，我们就在表III中分别列出九种实际网络的处于项目含顶点数分布中峰值的 T_{max} 以作互相比较。

表III：九种实际网络的项目含顶点数分布峰值 T_{max} 。

	BY	BH	BN	BB	BS	TC	TR	HA	HY
T_{max}	14	16	21	20	15	3	2	4	4

据表III，实证网络可以被分为两组。项目含顶点数分布峰值较大的为一组（五个城市公交线路网络），项目含顶点数分布峰值较小的为一组。由于我们最简模型模拟的项目含顶点数是个小数 ($T=6$)，并且对于后一组广义合作网络，其实证数据显示了同类性系数 r 和 r^o 和 SPL 函数关键参数 α 和 η 有单调的依赖关系，所以为了用模型结果定量的对比实证结果，我们画了一个例图，即模型和后一组网络实证得到的无权同类性系数 r 与项目度分布 $P(h)$ 的关键参数 α 和 η 的关系图（分别用 $\alpha-h$ 和 $\eta-h$ 显示），如图 3.3.1 所示。

图 3.3.1 表明了尽管实证数据只有四个，但模型数据结果和实证结果在一定可容忍范围里还是定性吻合的，这个可容忍的误差是可以理解的，因为模型所用的项目含顶点数是个常数，而实际网络中的项目含顶点数都不可能是常数，其分布都是变形的泊松分布。这个图可能表明模型可以用来解释真实网络的两种拓扑性质（同类性系数和 SPL 函数关键参数）之间的关系。而且，模型中所表达的广义合作网络的演化机制也可能被更广泛的接受。广义合作网络中，同类性系数与 SPL 函数关键参数都与随机选择概率 p 的单调依赖关系，而随机选择概率 p 在实证网络很难被量化。但在清楚了图 3.3.1 所表达的意思后，我们就可能定性的猜想四个

广义合作网络中，中国淮扬菜肴网演化中的随机因素是最小的（也表明网络中有相对比较多的枢纽节点）；中药方剂网演化的随机因素是次小的；接下来就是中国旅游线路网；最后是好莱坞演员合作网（网络中枢纽节点相对少），其随机因素是最大的。

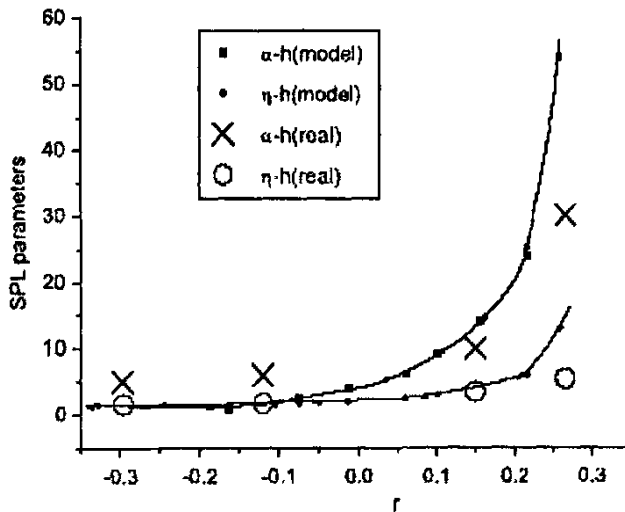


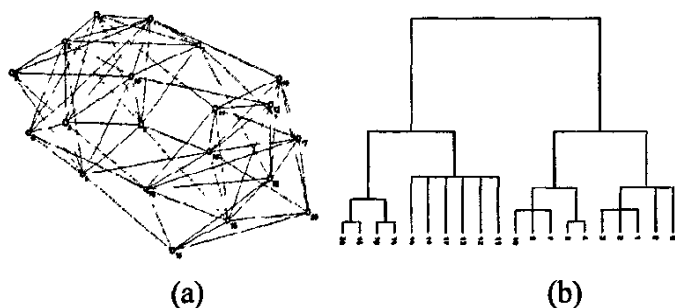
图 3.3.1: 模型和实证系统中的无权同类性 r 与项目度分布 $P(h)$ 的关键参数 α 和 η 的关系图。方块点代表的是模型得到的 $\alpha-h$ 与 r 的坐标点；实心圆圈点代表的是模型得到的 $\eta-h$ 与 r 的坐标点；叉子点代表的是实际系统得到的 $\alpha-h$ 与 r 的坐标点；空心圆圈点代表的是实际系统得到的 $\eta-h$ 与 r 的坐标点。图中实证得到数据点的实际系统从左至右依次是中国淮扬菜肴网、中药方剂网、中国旅游线路网、好莱坞演员合作网。

显然类似地，我们对五个城市公交线路网也可以画出与模型（比较大的 T ）对比的图形。图形表明模型和实证也能够定性的吻合，但比起其它广义合作网和模型的对比结果，其误差就大了很多。这可能说明了对于项目含顶点数比较大的网络，尤其是这类交通网络，模型需要进一步修正或有更好的描述这类交通系统的模型。

第四章 广义合作网络的群落、层次与交连度

4.1 简介

在社会网络研究中从来重视群落（也翻译为社团^[59]、凝聚子群^[8]等等）。与此相关的层次、派系也是重要的概念，这些概念在目前的复杂网络（包括大量的非社会网络）研究中同样受到重视。但是，迄今为止，仍缺乏得到普遍承认和采用的关于群落和层次的定量定义。社会网络中通用的定义是“群落是满足如下条件的一个参与者子集，其中参与者之间具有相对较强、较直接、较紧密、较经常、较积极的关系^[6,8]”。解绍、汪小帆最近的综述文章^[59]比较全面地总结了这个方向的研究动向和成就。其中所引用的论文中，影响最大的可能要算 Newman^[60-62]和 Palla 等人^[63]的论文。Newman 的一组论文中提出了划分群落的一系列重要算法，但是关于群落的定义仍旧沿用“在同一群落中节点的连接紧密，而群落之间的连接稀疏”这样的定性定义。在 Newman 之后，有一些科研小组提出了一些群落的定量定义以及相应的划分群落的方法，文献^[64]可能是一个代表性的例子。然而，这些方法和解绍、汪小帆的综述文章^[59]中介绍的不少其它方法在根本上类似，都是设法把网络“映射”为一个社会网络研究中所谓的“柱状图(dendrogram)”。这种图是一种树图，树叶就是节点，而树枝连接这些节点，或者在更高的层次上，连接节点的群落。如果能成功地、符合实际地完成这种映射，就能成功地划分出网络的各个群落，同时清晰地显示网络的层次结构，如图 4.1.1 所示。



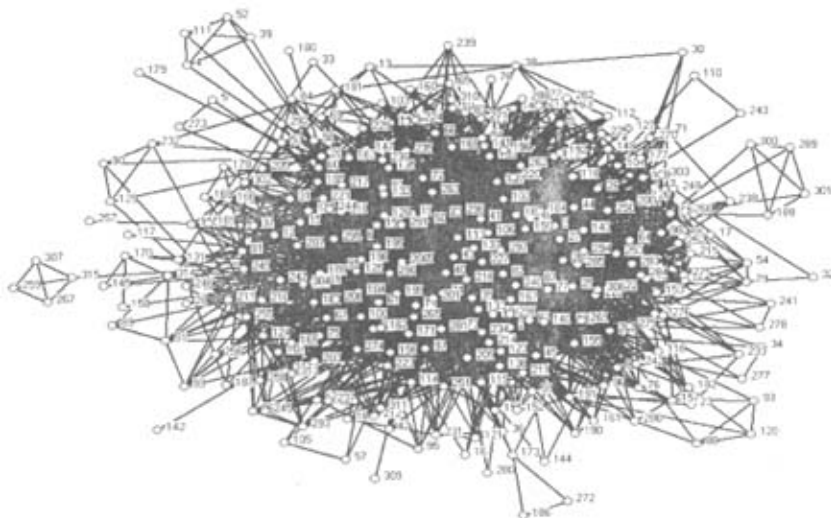
(引自 F. Radicchi 等, PNAS 101 (9) (2004) 2658)。

图 4.1.1 一个简单的网络(a), 和相应的柱状图(b)

这样，讨论就自然地涉及上述的层次概念。有关层次的研究论文相当不少^[65-68]，但是大部分论文都把层次作为一个大家可以直觉地接受的定性概念来处理，没有给出定量的定义。少数比较定量地定义层次例子有文献^[67]和文献^[68]。前者由网络的边权关系定义了网络的“层次单元”、“层次度”，并且画出了一个实际网络的层次度分布图。后者由实际网络的内容出发，从“命令”或“需求”、“决定”传递经过的（子）树图的短路边出发，定义了“层次路径”，并且计算了一些实际网络的层次路径。这些定义可以在一定程度上表述网络的层次结构，以及层次化的程度，但是缺乏上述为大家所希望的与划分群落的联系，而且，以我们的浅见，似乎缺乏作为一个揭示事物本质的定义，以便表述的直接、透彻、易懂。同时，也不符合我们在广义合作网中清晰划分群落和层次的需要。

我们曾经试用过好几种文献中建议的方法画出柱状图，以便得到划分群落和层次的算法，也试过用文献中提及的软件来直接的对在上两章中提及的几种实际广义合作网络划分群落和层次，但是效果都很不理想。经过仔细的分析，我们发现这些广义合作网络都是高度交连的，也就是说其中的许多完全图项目都共用许多个节点。对于本身就具有比较清晰的群落和层次结构的网络（例如存在各个层次的枢纽节点，基本上是树图结构的网络，以及完全图项目很少共用节点，交连很稀疏的广义合作网络。如果用 Netdraw 画出网络图容易用肉眼辨认出群落和层次结构。），则各种算法和软件常常都可以画出接近实际的柱状图，完成群落和层次的划分；而对于我们这里高度交连的广义合作网络，如果用 Netdraw 画出网络图，完全不可能用肉眼辨认出群落和层次结构（图 4.1.2 显示了根据全部实证数据用 Netdraw 软件画的中药方剂网络图。与图 2.1.2 显示的示意图相比，图 4.1.2 几乎完全看不出完全图项目，更谈不到直观地判断完全图项目交连所构成的层次和群落了）。在这种情况下，各种算法和软件画出的柱状图常常差别很大，而且都与实际吻合不好。这使我们怀疑把网络映射为一个柱状图是否能够在网络研究中普遍适用，也就是说，是否任何网络都一定能够映射为一个柱状图？因此，我们感觉必须寻找另外的出路。

Palla 等人的论文^[63]可能是迄今为止唯一讨论群落强烈地相互交连，以至于难



4.1.2 根据全部实证数据用 Netdraw 画的中药方剂网络图

于分割它们的情况的论文。他们建议这时运用社会网络研究中的另一个传统概念——派系^[63]。派系指“至少包括三个节点的最大完备子图”^[8]，意义是“建立在互惠基础上的群落”^[8]，也就是说派系中的任意两个成员之间都是“相关（即边联接）”的，而且派系之外不存在另外的同样连接派系中所有参与者的节点。显然，这就是我们一直在讨论的“完全图项目”，但是要清楚任何节点数大于2的完全图都可以分解成若干个更小的完全图，而派系是指其中最大的一个。我们所说的完全图项目是从实际意义定义的，完全图项目之外肯定不会再存在属于这个项目的参与者，所以它就是派系。Palla 等人^[63]定义一个含 k 个节点的派系为一个 k -派系。如果两个 k -派系公用 $k-1$ 个节点，称它们“相邻”。如果一个 k -派系可以通过若干个相邻的 k -派系到达另一个 k -派系，就称前后这两个 k -派系为彼此连通的。Palla 等人建议把所有彼此连通的 k -派系构成的集合定义为一个 k -派系群落。在此定义的基础上，Palla 等人建议了寻找派系，然后确定 k -派系群落的算法，并且编写了相应的、命名为 C-Finder 的程序^[63]。在我们的知识范围中，这很可能是迄今为止第一个不涉及把网络映射为一个柱状图的划分群落的方法。

Palla 等人的思想^[63]给了我们强烈的启发，然而，直接使用他们的算法和软件，

在我们研究的上述几种实际广义合作网络中划分群落的效果并不能与实际很好地符合。我们认为这可能是由于 Palla 等人的群落定义过强。我们在实证研究中观察到的完全图项目（派系）共用节点的情况各式各样，但是共用 $k-1$ 个节点的 k -派系很少。这样，按照 Palla 等人的定义划分出的 k -派系群落只占整个广义合作网络中完全图项目的很少一部分，我们认为这不能被称为符合实际。我们认为，每个完全图项目（派系）都可以被看作内部连接最密集，而对外连接相对稀疏的子图，也就是一个前文已经提及的“最基本的（或者作为基本单元的）”群落。在一个连通的广义合作网络中，任意两个完全图项目之间至少可以通过若干共用一个节点的相邻项目互相连通，所以可以建议把整个网络定义为一层次群落；相应地，由于共用至少两个节点而彼此连通的完全图项目的集合可以被定义为二层次群落；以此类推。一个广义合作网络的 h 层次群落被定义为至少共用 h 个节点而彼此连通的完全图项目的集合。也就是说，广义合作网络中的层次由完全图项目交连的程度差别而定义，交连的程度越大（共用的节点越多），层次越高；每一个层次上的完全图项目的集合就定义为这个层次的群落。显然，这种定义只适合于广义合作网络，它们完全由完全图项目（派系）的集合构成，而且实证数据已经包括哪些节点归属于哪个完全图项目。这种定义的适用范围虽然比较窄，但是根据这种定义可以编写相对简单的程序，从实证数据方便地划分和画出各个层次的群落，而且其不涉及把网络映射为一个柱状图的划分群落方法也可能具有更广泛的启发意义。

4.2 群落和层次的定义及其简单算法

在第二章开始，我们就阐述了广义合作网络的原始二分图描述以及其投影到参与者节点上的项目子完全图(act complete subgraph, 简称 ACSG)集合的定义。之前，我们在单粒子完全图集合的广义合作网络定义上给出了详细的关于广义合作网络的群落和层次定义^[70]。为了更简明扼要的描述群落和层次的定义，这里报道的定义是基于广义合作网络的二分图描述之上的。

4.2.1 层次和群落定义

我们建议的基于二分图的群落和层次的定量定义如下：

- (1) 定义： $X = \{x_1, x_2, \dots, x_n\}$ 表示参与者(actor)集合， x_i 表示第 i 个参与者；
 $T = \{T_1, T_2, \dots, T_m\}$ 表示项目子完全图集合， T_j 表示第 j 个项目子完全图； $x_i \in T_j$ 表示第 i 个参与者是第 j 个项目子完全图中的节点； $x_i \notin T_j$ 表示第 i 个参与者不是第 j 个项目子完全图中的节点。
- (2) 定义 $A = (a_{ij})_{n \times m}$ 为二分图隶属关系矩阵，其中 $a_{ij} = \begin{cases} 1, & x_i \in T_j \\ 0, & x_i \notin T_j \end{cases}$ 。矩阵性质如下：
- 1) 若 $\sum_{i=1}^n a_{i\alpha} a_{i\beta} \neq 0$ ，则 T_α 与 T_β 相邻，即 $T_\alpha \cap T_\beta \neq \Phi$ 。
 - 2) 若 $\sum_{i=1}^n a_{i\alpha} a_{i\beta} = h > 0$ ，则 T_α 与 T_β 恰好共用 h 个参与者。
 - 3) 若存在一组正整数 t_1, t_2, \dots, t_s ，使得 $(\sum_{i=1}^n a_{ip} a_{iu_1}) \cdot (\sum_{i=1}^n a_{iu_1} a_{iu_2}) \cdots (\sum_{i=1}^n a_{iu_s} a_{iq}) \neq 0$ ，
 则 T_p 与 T_q 连通，即 T_p 中的参与者可以通过若干路径与 T_q 中的参与者连接。
- (3) 1) 定义 $\Gamma_n = \left\{ (j, k) \mid \sum_{i=1}^n a_{ij} a_{ik} = h > 0, k > j, j = 1, 2, \dots, m \right\}$ 为 h 层指标集(indexing set)， h 称为层次数。
- 2) 定义 $T_h^* = T_\alpha \cup T_\beta$ 为 h 层次核心，其中 $(\alpha, \beta) \in \Gamma_n$ 。
 - 3) 对任意两个层次核心 $T_{\lambda_1}^*, T_{\lambda_2}^*$ (可以是不同层次的核心)，如果存在 $T_p \in T_{\lambda_1}^*$ ，
 $T_q \in T_{\lambda_2}^*$ ，使得 T_p 与 T_q 连通。则定义两个层次核心 $T_{\lambda_1}^*, T_{\lambda_2}^*$ 是连通的。
- (4) 1) 连通的且层次数不小于 h 的所有 λ 层次核心的并集 $\bigcup_{\lambda=h}^{h_{\max}} T_\lambda^*$ 称为 h 层次群落，记作 $C_h = \bigcup_{\lambda=h}^{h_{\max}} T_\lambda^*$ ，其中 $h_{\max} = \max_{j,k} \left\{ \sum_{i=1}^n a_{ij} a_{ik} \right\}$ 。
- 2) 对任意的 $T_k^* \in C_h$ ，如果存在某一 $T_h^* \in C_h$ ，使得 T_k^* 与 T_h^* 连通。则定义 T_k^* 与

h 层次群落 C_h 是连通的。

4.2.2 划分层次和群落的算法

基于我们的实证网络数据, 根据上述定义对实际系统划分层次和群落。其简要算法叙述如下:

Step1: 调用 act、actor 隶属关系的数据文件, 得到二分图隶属关系矩阵 $A = (a_{ij})_{n \times m}$ 。

Step2: 对 $k > j, j = 1, 2, \dots, m$, 记 $h = \sum_{i=1}^n a_{ij} a_{ik}$, 计算 h 。从计算结果中写出各层指

标集: $\Gamma_{h_{\min}}, \Gamma_{h_{\min \uparrow}}, \dots, \Gamma_{h_{\max \downarrow}}, \Gamma_{h_{\max}}$, 其中 $h_{\min} = \max \left\{ \min \left\{ \sum_{i=1}^n a_{ij} a_{ik} \right\}, 1 \right\}$, 即

最小层次数。 $h_{\min \uparrow}$ 表示次小的层次数。 $h_{\max \downarrow}$ 表示次大的层次数。

Step3: 取 $(\alpha, \beta) \in \Gamma_{h_{\max}}$, 遍历 $\Gamma_{h_{\max}}$ 中的所有元素, 得到全部最高层次核心 $T_{h_{\max}}^*$ 。

根据两个层次核心连通的定义, 列出所有连通的最高层次核心 $T_{h_{\max}}^*$ 并作并集, 得到最高层次的群落 $C_{h_{\max}}$ 。

Step4: 取 $(\alpha, \beta) \in \Gamma_{h_{\max \downarrow}}$ (次高层指标集), 遍历次高层指标集中的所有元素, 得

到次高层核心 $T_{h_{\max \downarrow}}^*$ 。列出所有连通的次高层核心, 根据层次核心与群落连通的定义, 从中筛选出与 $C_{h_{\max}}$ 连通的若干次高层次核心, 将它们与 $C_{h_{\max}}$ 作并集, 得到次高层次的群落 $C_{h_{\max \downarrow}}$ 。

Step5: 重复 step4。最后取 $(\alpha, \beta) \in \Gamma_{h_{\min}}$, 遍历 $\Gamma_{h_{\min}}$ 中的所有元素。得到最低层次

核心 $T_{h_{\min}}^*$ 。列出所有连通的最低层次核心, 从中筛选出与上层次群落 $C_{h_{\min \uparrow}}$ 连通的若干最低层次核心。将它们与上层次群落作并集, 得到最低层次的群落 $C_{h_{\min}}$ 。

用二分图图示 4.2.1 为例来说明上述的算法过程。

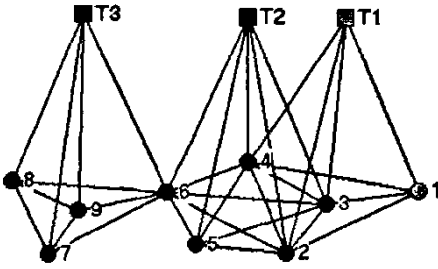


图 4.2.1 二分图及其投影的示意图。其中方块节点表示的是项目 T1、T2 和 T3, 1 到 9 的圆形节点表示参与者节点。

Step1: 调用 act、actor 隶属关系的数据文件, 得到二分图隶属关系矩阵 $A = (a_{ij})_{n \times m}$,

$$\text{即: } A_{9 \times 3} = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{pmatrix} \begin{matrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \\ x_6 \\ x_7 \\ x_8 \\ x_9 \end{matrix}$$

Step2: 当 $j=1, k=2$, $\sum_{i=1}^9 a_{i1}a_{i2} = 3$; 当 $j=1, k=3$, $\sum_{i=1}^9 a_{i1}a_{i3} = 0$; 当 $j=2, k=3$,

$$\sum_{i=1}^9 a_{i2}a_{i3} = 1。得到各层次指标集为: $\Gamma_3 = \{(1,2)\}$, $\Gamma_2 = \emptyset$, $\Gamma_1 = \{(2,3)\}$ 。$$

Step3: 取 $(\alpha, \beta) \in \Gamma_3 = \{(1,2)\}$, 得到 3 层次核心 $T_3^* = T_1 \cup T_2$, 也是 3 层次群落 C_3 。

Step4: 因为 $\Gamma_2 = \emptyset$, 所以 2 层次核心是空集。从而 $T_1 \cup T_2$ 也是 2 层次群落 C_2 。

Step5: 取 $(\alpha, \beta) \in \Gamma_1 = \{(2,3)\}$, 得到 1 层次核心 $T_1^* = T_2 \cup T_3$ 。显然, 1 层次核心是和 2 层次群落 C_2 连通的, 所以 $(T_2 \cup T_3) \cup (T_1 \cup T_2) = T_1 \cup T_2 \cup T_3$ 即是 1 层次群落 C_1 。

4.3 一些实际广义合作网络的群落和层次

我们对所研究的实际广义合作网络都进行了群落和层次的划分，下面分别以中药方剂网、扬州公交线路网和好莱坞演员合作网为例作具体的说明。

4.3.1 中药方剂网的群落和层次

我们的中药方剂网中，中药一幅方剂（完全图项目）最多包括 15 种药物（参与者），但是这样的完全图项目只有两个；最少包括一种药物，这样的完全图项目有 168 个。一道方剂由 3 种药材组成的可能性最大，这样的完全图项目有 306 个，而更少或更多种药材组成的可能逐渐减少。表 4.3.1 列出了各个项目含顶点数对应的完全图项目数。

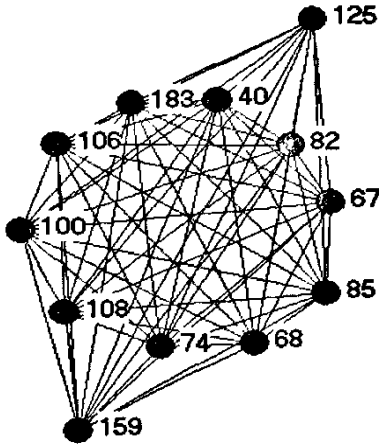
表 4.3.1 中药方剂网各个项目含顶点数对应的完全图项目数

项目含顶点数	1	2	3	4	5	6	7	8	9	10	11	12	14	15
项目数	168	254	306	238	186	115	90	61	29	10	5	2	1	2

由于中药方剂网是一个连通图，所以最大项目含顶点数为 15 就意味着这两个项目最多可能共用 14 个节点（在特殊情况也可能共用 15 个节点，如下面例子所讨论），也就是说最高层次数只可能是 14 或 15，而最高层次的群落只可能包括两个完全图项目。当然，这样的情况很难出现。实际上，这两个最大完全图项目只共用 1 个节点。在中药方剂网中，完全图项目共用节点数最多的是共用 10 个，也就是说实际的最高层次数是 10，实际的最高层次的群落只包括 2 个完全图项目（也就是说含顶点数大于或等于 10，且共用 10 个顶点的完全图项目只有 2 个）。它们之中一个是 10 完全图，另一个是 12 完全图，把 10 完全图的所有顶点完全包含在其中。图 4.3.1 显示了我们按照上面所述定义和思想画出的这个 10 层次群落图。最于最高层次它也就是 10 层次核心。

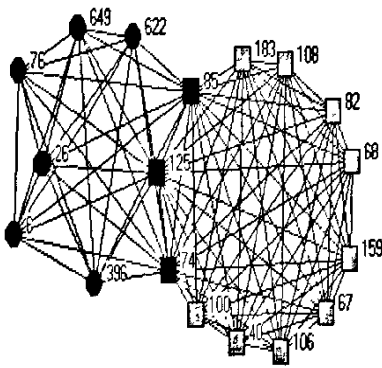
图 4.3.2 显示了中药方剂网的 9 层次群落图。它当然地必须包含图 4.3.1 所示的 10 层次核心。此外一个 9 层次核心由两个包含 9 种完全相同药材，但是由不同其比构成的两幅不同方剂组成，应该看作两个 9 完全图项目的共用所中 9 个节

点的 9 层次并，这是很特殊的情况。



10 层次群落图，包括 2 个完全图项目，一个完全图包含 10 个节点（图中央的 10 个灰色实心圆圈），另一个完全图包含 12 个节点。它们共用的 10 个节点用图中央的 10 个灰色圆圈表示。

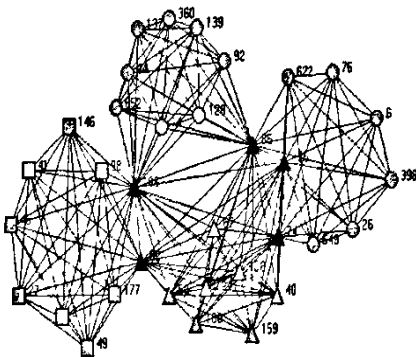
图 4.3.1 中药方剂网的 10 层次群落图（对于最高层次也就是 10 层次核心）。



9 层次群落包含图中黑色节点所示的 9 层次核心和方块节点所示的 10 层次核心。

图 4.3.2 中药方剂网的 9 层次群落

图 4.3.3 显示了中药方剂网的 8 层次群落图。它当然地必须包含 10 层次核心和 9 层次核心。此外存在两个 8 层次核心。其中不个由不个完全包含另不个 8 完全图的 10 完全图构成。另不个由公用 8 个节点的两个 9 完全图构成。



8 层次群落图包含图中三角节点所示的 10 层次核心、右边的圆形节点所示的九层次核心、左边的方形节点所示的一个 8 层次核心、以及上边的一个圆形节点所示的 8 层次核心。

图 4.3.3 中药方剂网的 8 层次群落图。

显然, 由于中药方剂网是一个连通图, 1 层次群落一定包括了所有的完全图项目, 一共包含 1536 道中药方剂(项目)和 681 种药材(参与者)。可以说图 4.1.2 就显示了这个 1 层次群落。其中每个完全图可以被看作前文已经提及的最基本的群落, 或者构成群落的基本单元。

4.3.2 扬州公交线路网的群落和层次

在第三章中, 已经说明了在 P 空间上的交通网络的定义, 将每个公交车站定义为顶点, 它们之间的连边定义为两个车站在同一条公交线路中的运输合作关系(不转车直达地运输乘客的合作关系), 则每条线路也表示为一个完全图项目。各个完全图项目之间通过共用的顶点架起了桥梁, 从而构成了整个公交网络。首先, 我们根据文献^[53]进行了扬州在 2003 年的公交线路网统计。表 4.3.1、表 4.3.2 分别列出了一些统计数据。

表 4.3.1 扬州 2003 公交线路网的一些统计结果

扬州 2003 公交线路网	顶点数	项目数	项目含顶点数范围	顶点项目范围
	289	26	6—40	1—13

表 4.3.2 扬州 2003 公交线路网每个项目含顶点数对应的完全图项目数

项目含 顶点数	6	11	12	14	15	18	19	20	21	22	24	25	26	27	29	34	40
项目数	1	1	2	4	1	2	2	1	2	3	1	1	1	1	1	1	1

图 4.3.1 显示的是扬州 2003 公交线路网整个网络图。从图中可以看出, 虽然扬州 2003 公交线路网的网络规模比较小, 但项目完全图也是十分稠密的。由于这个网络是个连通图, 而且最大的项目含顶点数是 40, 这意味着两个最大项目(一个有 40 个顶点, 另一个有 34 个顶点)最多可能共用 34 个顶点, 即最高层次数不可能高于 34, 而最高层次的群落只可能包括两个完全图。事实上, 这两个最大完全图共用了 24 个顶点, 也是网络中交连最多的顶点, 所以网络的最高层次数是 24。图 4.3.2 显示的是 24 层次群落图, 对于最高层次也就是 24 层次了。它由两个完全图项目(公交线路)合并构成。这两条线路都是通向郊区的长程公

交车，它们途径共同的主要市区部分。与中药方剂网相比，扬州公交线路网项目

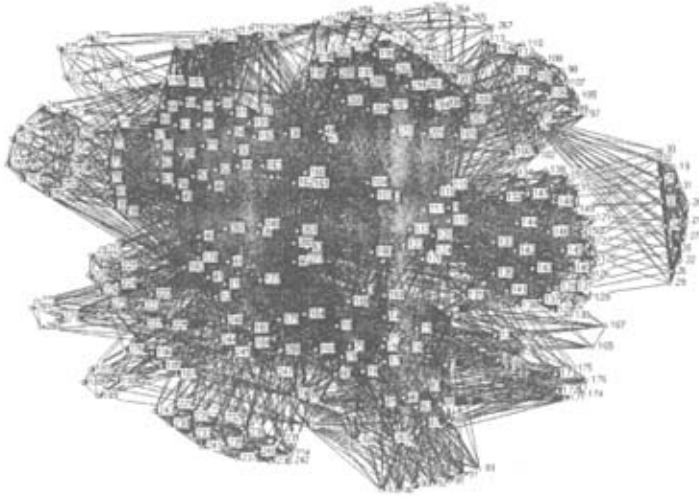
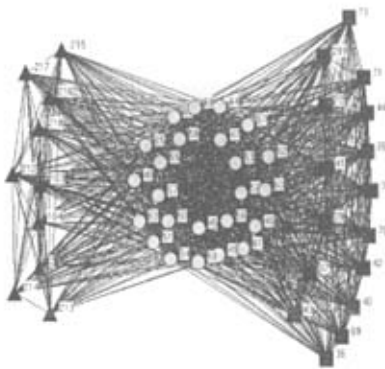


图 4.3.1 根据全部实证数据用 Netdraw 软件画的扬州公交线路网络图

完全图数目少得多，但是最高层次数大得多，说明交连的程度大得多。这既是方便乘客换乘的需要，也包含了大量的区间车和全线车、同一运行路线上不同快慢的公交车等不同完全图项目高度交连的实际情况。与中药方剂网的交连程度相当不同。

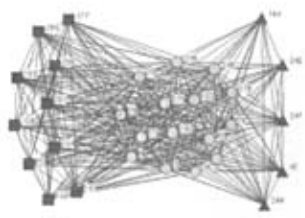


24 层次群落图（对于最高层次也就是 24 层次核心），包括 2 个完全图项目，一个完全图包含 40 个节点(图中央的 24 个灰色实心圆圈加上右侧的 16 个黑色方块)，另一个完全图(图中央的 24 个灰色实心圆圈加上左侧的 10 个黑色三角)包含 34 个节点。

图 4.3.2 扬州公交线路网的 24 层次群落图（对于最高层次也就是 24 层次核心）

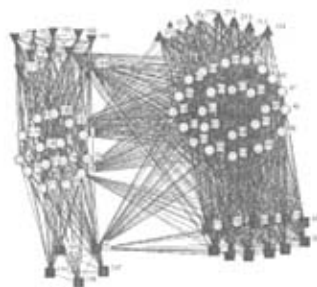
图 4.3.3 显示了扬州公交线路网的 15 层次核心图。这是扬州公交线路网的次

高层次的核心。它由两个分别包含 20 个节点和 25 个节点的完全图项目构成。图 4.3.4 显示了扬州公交线路网的 15 层次群落图。它当然地包括了图 10 显示的 24 层次核心和图 11 显示的 15 层次核心。



15 层次核心包括 2 个完全图项目，一个完全图包含 20 个节点(图中央的 15 个灰色实心圆圈加上右侧的 5 个黑色三角)，另一个完全图(图中央的 15 个灰色实心圆圈加上左侧的 10 个黑色方块)包含 25 个节点。

图 4.3.3 扬州公交线路网的 15 层次核心图。



15 层次群落图包含图中左半部所示的 15 层次核心和右半部所示的 24 层次核心。右半部所示的 24 层次核心共用左半部所示的 15 层次核心中的 6 个节点。

图 4.3.4 扬州公交线路网的 15 层次群落图。

图 4.3.5 显示了扬州公交线路网过图中所示的中心节点的 6 层次核心图。它由两个分别共用 6 个节点（包括中心节点）的 6 层次并构成。每个 6 层次并由两个共用 6 个节点（包括中心节点）的完全图构成。由于层次比较低，这样的核心有若干个。网络的 6 层次群落是高于等于 6 层次核心的并集，如图 4.3.6 所示。由图可见网络的 6 层次群落已经显示出高度交连网络难于用肉眼分辨群落结构的特征。

显然，由于扬州公交线路网是一个连通图，1 层次群落一定包括论文所有的完全图项目，一共包含 26 个完全图项目和 289 个参与者。可以说图 4.3.1 就显示了这个 1 层次群落。

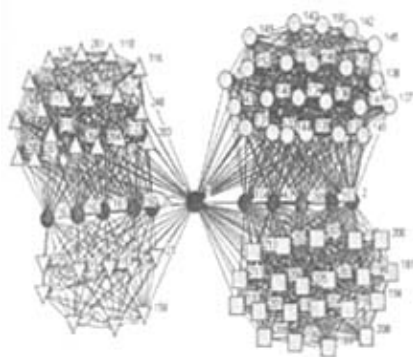


图 4.3.5 扬州公交线路网过图中中心节点的
6 层次核心图



图 4.3.6 扬州公交线路网的 6 层
次群落图

4.3.2 好莱坞演员合作网的群落和层次

在本节的讨论中,我们在下载的 392,304 个演员和 181,455 部影片的好莱坞演员合作网的数据中取出前 50 部或前 1500 部影片的数据,以方便我们的简单程序处理,方便不同目的的图示。

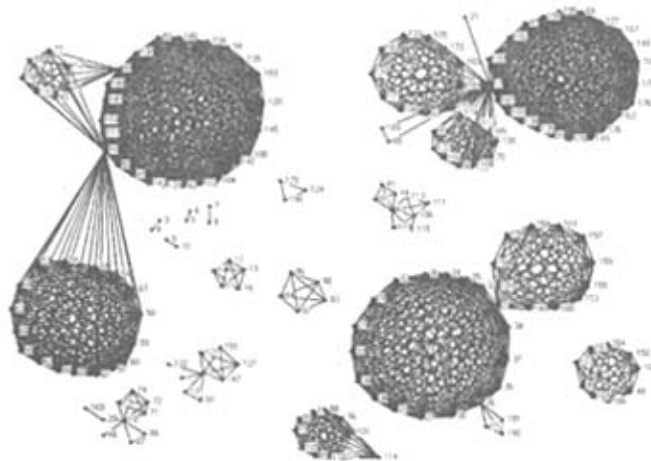


图 4.3.7 好莱坞演员合作网数据中前 50 部影片的群落图

我们用图 4.3.7 显示了根据好莱坞演员合作网数据中前 50 部影片的实证数据用 Netdraw 软件画的网络图。显然,此前 50 部影片的好莱坞演员合作网不是一个连通图,各个完全图项目交连很稀疏,是个弱交连的网络。这和前面图 4.3.1 和图

4.1.2 显示的扬州公交网和中药方剂网的情况呈现鲜明的对比。但是如果把演员网的数据增加到几百部或上千部影片（和前面讨论的三种实际网络数据的数量级相当），用 Netdraw 软件画出显示的图说明此网络的交连程度比图 4.3.7 显示的网络的交连程度密集，说明了前 50 部影片的好莱坞演员合作网的交连稀疏只是整个好莱坞演员合作网的早期特征。

当然，对于截取部分影片的好莱坞演员合作网，也可以用本章建议的方法划分群落和层次。划分的情况和上述两种网络类似，不再赘述。

4.4 交连度

上述几种实际网络的核心群落图使我们对于网络中核心、群落和项目子完全图间的重叠情况有了直观的印象。但是，这些图只是显示了网络中少部分的项目子完全图的重叠情况，人们自然就会想到，从整个网络来看，哪种网络的重叠嵌套最厉害呢？因此，有必要对网络的这种特征用一个量化的统计量来描述，我们称之为“交连度(interweavement)”，用 I 来表示。我们最初的想法^[70]是要比较不同实际广义合作网络中相同数量、同样大小的完全图项目交连的最大层次数，即定义交连度 $I = h_{\max} / MT_{\max}$ ，其中 M 表示总的项目数， T_{\max} 表示网络的项目子完全图所含的最大节点数目(上述 4.2.1 章节定义中的 T 是指的项目子完全图，以提醒区别)， h_{\max} 表示网络中的最高层次数(本章以外的章节中， h 指的是顶点项目度，以提醒区别)。但在后来的计算中，我们发现这个定义在两个方面上显得不合道理。首先，如果网络的 M 值很大(项目数很多)，即便网络交连很密集，依然会得到一个很小的 I 值；再者， h_{\max} 和 T_{\max} 并不能准确地描述网络的整体特征或平均特征。因此，我们建议了一个新的交连度的定义来比较不同实际网络的项目子完全图间的平均交连程度。

首先，定义 h' 为一个项目子完全图(ACSG)的层次数，即一个项目子完全图所参与的群落核心的最高层次数(与其它项目子完全图共用最多的节点数目)，这定义类似于文献^[63]中的“重叠大小(overlapping size)”的定义。则交连度定义如下：

$$I = \frac{E(h')}{E(T)} = \frac{(m_{h'_{\min}} h'_{\min} + m_{h'_{\min \uparrow}} h'_{\min \uparrow} + \cdots + m_{h'_{\max \downarrow}} h'_{\max \downarrow} + m_{h'_{\max}} h'_{\max})/M}{(m_{h'_{\min}} T_{\min} + m_{h'_{\min \uparrow}} T_{\min \uparrow} + \cdots + m_{h'_{\max \downarrow}} T_{\max \downarrow} + m_{h'_{\max}} T_{\max})/M} =$$

$$\frac{\rho_{h'_{\min}} h'_{\min} + \rho_{h'_{\min \uparrow}} h'_{\min \uparrow} + \cdots + \rho_{h'_{\max \downarrow}} h'_{\max \downarrow} + \rho_{h'_{\max}} h'_{\max}}{\rho_{h'_{\min}} T_{\min} + \rho_{h'_{\min \uparrow}} T_{\min \uparrow} + \cdots + \rho_{h'_{\max \downarrow}} T_{\max \downarrow} + \rho_{h'_{\max}} T_{\max}}.$$

定义式中, $E(h')$ 和 $E(T)$ 分别表示 h' 和 T (项目含顶点数) 的期望值; h'_{\min} 表示最小的项目子完全图层次数; $m_{h'_{\min}}$ 表示最小层次数的项目个数; $\rho_{h'_{\min}}$ 表示最小层次数的项目个数占总项目的比例; $h'_{\min \uparrow}$ 表示次大的项目子完全图层次数; $h'_{\max \downarrow}$ 表示次小的项目完全图层次数; h'_{\max} 表示最大的项目子完全图层次数。其它量类推之。

考虑这样的理想网络, 它包含 M 个项目子完全图, 每个项目子完全图都含有同样的节点数 $T = T_{\text{const}}$, 并且所有的项目子完全图都与别的项目子完全图共享了所有 T_{const} 个节点(每个节点参与了全部 M 个项目), 则有 $m_{h'} = m_{T_{\text{const}}} = M$, $I = (MT_{\text{const}})/(MT_{\text{const}}) = 1$ 。这是重叠最强的一种极端情况。

假设网络包含 M 个项目子完全图, 每个项目子完全图同样含有一样的节点数 $T = T_{\text{const}}$, 但与别的项目子完全图都不共享节点(每个节点只参加一个项目), 则有 $m_{h'} = m_0 = M$, $I = (M \times 0)/(MT_{\text{const}}) = 0$ 。这是重叠最弱的一种极端情况。

假设网络包含 M 个项目子完全图, 每个项目子完全图同样含有一样的节点数 $T = T_{\text{const}}$, 每个项目子完全图都和别的项目子完全图共享一个节点(即每个项目中只有一个节点参与了不止一个项目), 则 $m_{h'} = m_1 = M$, $I = (M \times 1)/(MT_{\text{const}}) = 1/T_{\text{const}}$ 。这是介于最强重叠和最弱重叠之间的一种可能情况。

综合上述, 显然有 $0 \leq I \leq 1$ 。

我们计算了一些实际网络的交连度, 并且就不同网络解释了其可能的实际含义。见表格 4.4.1。中国旅游线路网的交连度最大, 通常而言, 景点和景致、服务、娱乐、购物等联系在一定的, 其利益吸引了大批组团的旅游线路, 所以那些利益

大的景点就成了各条旅游线路的最佳选择,网络交连度大也就在情理之中了。北京市区公交线路网选取的是2003年北京市区65条线路,760个站点来统计的。

表格 4.4.1 实际网络的交连度以及可能的实际含义

实际网络	交连度	含义
中国旅游线路网	0.746	线路间共享重要景点趋势最强烈
中药方剂网	0.640	药方共用药材趋势比较强烈
扬州公交线路网 2003	0.435	线路间共用站点趋势一般
北京市区公交线路网 2003	0.401	线路间共用站点趋势稍弱
好莱坞演员合作网(前1500部影片)	0.444	影片间共用演员趋势一般
好莱坞演员合作网(前50部影片)	0.204	影片间共用演员趋势最弱

与扬州相比,北京是个大的多的城市,包含了许多条线路,但它的交连度比扬州还要略小些,这说明即便大的城市其线路之间的交连未必就随之密集。上一节给出了好莱坞演员合作网络前50部影片的全局网络图,直观看来交连稀疏,而计算得到的交连度相比较其它网络来看也是最小的。

4.5 小结及讨论

作为结论,我们建议了划分广义合作网络层次和群落的一种新的定量定义。基于实证数据,很容易就能检验不同层次上的群落划分的合理程度,而层次是根据网络中项目子完全图的重叠度^[61]来定义的。不同层次之上的群落都是交连的,但是线层次的群落是由更强重叠的项目子完公图组成的,所以比低层次的群落来的更密集。

一般的实际网络中,不通过恰当的计算,从实际数据中是很难直接的发现群落的存在,这和广义合作网络的情形是不同的,因此,很难直接回答如何最好的划分一个网络的群落。为了解决这个问题,Newman和Girvan建议了一个衡量网络划分质量的标准—模块度(modularity)^[61],用 Q 来表示。其定义式为:

$$Q = \sum [e_{ii} - (\sum_j e_{ij})^2] = TrE - \|E^2\|$$
 其中 e_{ij} 是矩阵 E 的矩阵元,表示网络中连接两个不同社团 i 和 j 的节点的边在所有边中所占的比例。 Q 表示的是群落内部的

边数占公部边数的比例。对每步划分, Newman 和 Girvan 都建议计算模块度并观察其变化, 最好的群落划分对应于模块度值变化最大的那步划分。为了说明我们划分广义合作网络层次群落的定义和算法是否恰当, 我们根据一个简单网络模型来做类似的判断。

假设一个网络包含 M 个项目子完公图, 每个项目子完公图都含有相同的节点数 $T = T_{const}$, 并且每个层次群落都包含一系列至多和其它两个核心相连接的核心 (即一个群落就是核心串起的链条)。考虑到两个项目子完公图重叠 h 个节点而组成的核心, 它含有 $2 \times T_{const} - h$ 个节点和 $T_{const}(T_{const} - 1)$ 条边 (计入重复边)。在广义合作网络中, 重复边的考虑是非常合理的, 例如两个电影演员在两部影片中合作, 他们之间连有两条边, 换言之, 如果忽略重复边, 那就要考虑边权。核心的模块度就是 $T_{const}(T_{const} - 1)/N$, N 表示网络的总边数。如果考虑到两个项目子完公图重叠 $h-1$ 个节点而组成的核心, 它包含 $2 \times T_{const} - h + 1$ 个节点和 $T_{const}(T_{const} - 1)$ 条边 (计入重复边), 虽然重叠情况变弱了, 但这个核心的模块度仍然是 $T_{const}(T_{const} - 1)/N$ 。这可能表明模块度在项目高度交连并且计算重复边的情况下并不适合衡量网络划分群落的标准。

于是, 我们建议了另外一个量, 称之为“景观参与者平均边密度”。其定义为:

$$D = N_e(T) / N_v(T)$$

其中, $N_e(T)$ 表示核心或群落内部的边数目, $N_v(T)$ 表示核心或群落内部的节点数目。上述中提及的 h 层次核心参与者平均边密度为 $D_h = T_{const}(T_{const} - 1) / (2T_{const} - h)$, 而 $h-1$ 层次核心参与者平均边密度为 $D_{h-1} = T_{const}(T_{const} - 1) / (2T_{const} - h + 1)$, 显然, 前者大于后者, 即重叠程度减弱时 D 值变小。

假设 h 层次群落以前面所述的链条式核心连接方式构成, 并且包含 m_h 个 h

层次核心, m_{h+1} 个 $h+1$ 层次核心, \dots , m_{h+n} 个 $h+n$ 层次核心。类似地, $h+1$ 层次群落就包含了 m_{h+1} 个 $h+1$ 层次核心, \dots , m_{h+n} 个 $h+n$ 层次核心。于是, 有

$$D_h = \frac{T_{const}(T_{const} - 1)(\sum_{i=h}^{h+n} m_i)}{(2T_{const} - h)(\sum_{i=h}^{h+n} m_i) - (\sum_{i=h+1}^{h+n} (i-h)m_i) - \delta(\sum_{i=h+1}^{h+n} m_i - 1 + m_h)},$$

其中 δ 表示两个相连的核心所共有的节点数。为了简便计算, 我们姑且假设 δ 为一常数。类似地, 有

$$D_{h+1} = \frac{T_{const}(T_{const} - 1)(\sum_{i=h+1}^{h+n} m_i)}{(2T_{const} - h)(\sum_{i=h+1}^{h+n} m_i) - (\sum_{i=h+1}^{h+n} (i-h)m_i) - \delta(\sum_{i=h+1}^{h+n} m_i - 1)},$$

令 $A = \sum_{i=h+1}^{h+n} m_i$, $B = \sum_{i=h+1}^{h+n} (i-h)m_i$, 得到:

$$\begin{aligned} & \frac{D_{h+1} - D_h}{T_{const}(T_{const} - 1)} = \\ & = \frac{A}{(2T_{const} - h)A - B - \delta(A-1)} - \frac{(A+m_h)}{(2T_{const} - h)(A+m_h) - B - \delta(A-1+m_h)} = \\ & = \frac{m_h(B - \delta)}{[(2T_{const} - h)A - B - \delta(A-1)][(2T_{const} - h)(A+m_h) - B - \delta(A-1+m_h)]} \end{aligned}$$

上式右式分子中, 因为一般情况下显然有 $B > \delta$, 则分子大于 0; 又因为分母是两个核心的节点数的乘积, 也是大于 0; 而左式分母显然也是大于 0 的; 所以可以得到 $D_{h+1} - D_h > 0$ 。这证明了普遍情况下, $h+1$ 层次群落都比 h 层次群落内部密集, 而我们关于层次群落的定义和算法在普遍情况下也是合理的。

第五章 城市公交线路网络的最简操纵者博弈模型

5.1 简介

复杂自适应系统的基本单元具有总结经验、积累知识、采取和改进对策等自适应功能，它们各自不同，各自之间的作用也不同，不可能使用传统的物理学描述工具。复杂网络很可能发展成为这类系统的合适描述工具，因而引起了广泛的兴趣。通常复杂自适应系统的每个基本单元都在致力于改进自身和环境，以便使自己的利益最大化。因此系统的演化过程可能看作各个基本单元之间进行博弈的过程。自从1998年复杂网络的研究热潮兴起以来，网络上的博弈一直是研究兴趣之一。就我们所知，这些研究集中于放置在一些著名网络模型（小世界、无标度等）上的少数博弈论的传统简化模型（囚徒困境(prisoner dilemma)和铲雪模型(snowdrift)等)的行为上^[71-77]。它们具有重要的理论意义和开创价值，然而，这些网络模型和博弈模型虽然具有广泛的实际背景，但是正因为“泛”，就难于直接与实际情况对应，更不能期望直接有益于解决实际问题，而且大量基本单元之间的繁杂博弈行为很难得到简明、一针见血的理解和相应的解析讨论结果。因此，有必要探索更集中、更简化、更结合实际的网上博弈模型。我们注意到在某些实际系统中，大量基本单元的利益可能有几个“代表者”，或者虽然没有成形的代表者，但是可能形成它们利益和意见的集中表示，因此可以说存在“无形”的代表者。在这种情况下，也许有可能仅仅考虑这些利益代表者之间的博弈，来决定整个网络的演化以及其他整体行为，从而得到远为简明的理解。如果可行，这些基本单元利益的代表者就可以称作“网络操纵者”。这个“网络操纵者博弈”的思想也许可以应用于许多实际系统。作为例子，本章讨论城市公交系统。

由于实际的重要性，城市公交系统一直是研究的热点之一^[78-80]。近年来，城市公交系统的网络描述以及由此得到的新特性也得到广泛的注意^[81-85]。除了对于大家都注意的一些统计性质（度分布、相关性等）的实证研究^[81-85]之外，城市公交的本行专家们长期以来提出和研究了许多旨在应用的模型^[78-80]，这些模型通常非常详尽、近于实际，但是可能缺乏整体观和简明性。我们还没有见到从复杂网

络的研究角度,旨在说明城市公交系统演化和性质的博弈模型。这样的模型也许可以具有整体观和简明性,这是物理学工作者感兴趣的。城市公交网络的节点一般定义为公交车站^[78-80],它们之间的博弈不好理解。如果能够找到上述的利益代表者,建议一个网络操纵者的博弈模型,并且能够做到由此模型产生的网络统计性质符合于实证结果,将是引人注意的新思想、新方法,而且这种思想和方法完全可能应用于其它系统。

城市公共交通系统的描述方法主要有三种:一种是被称为 L 空间的方法,即交通站点视为节点,若两个站点是某一交通线路上相邻的,它们就有连边;另一种是 P 空间的方法,即交通站点视为节点,若两个站点有直达交通线路,那么它们就有连边;还有一种是 K 空间方法,与第一种描述方法类似,即交通站点视为节点,若两站点间在实际中是相连的,它们就有连边。K 空间方法构造的网络是 L 空间方法构造网络的子网络,而 L 空间方法构造的网络又是 P 空间方法构造网络的子网络。

我们组对城市公共交通系统的研究主要是在 P 空间的定义基础之上的,不作特殊说明,本论文中“城市公交线路网络”就是指 P 空间定义的公交网络。我们就城市公交线路网络一共提出过三个操纵者博弈演化模型,这三个模型一次比一次简化,意义也越来越集中。本章中我们将对前两个模型作简单的介绍,对最近一个最简化操纵者博弈模型作详细的报道。

5.2 前两个网络操纵者博弈模型介绍

5.2.1 三方网络操纵者博弈模型^[86]

在公交网络的演化过程中,如果忽略掉许多次要因素,可以认为公交公司、乘客、交通管理部门对于网络演化的意见是最重要的。他们之间存在矛盾,也就是竞争关系,公交公司可能最优先考虑公司的经济效益;乘客只考虑出行的方便性和费用负担;而交通管理部分则优先考虑如何避免道路堵塞和事故。同时,他们也都了解三方面是缺一不可的,如果有一方不可忍受,自己必须做出让步,这就是互相合作的关系。如果假设三方面都在维持合作的前提下追求自己利益的最

大化，那么公交线路网络的演化过程可能可以看作一个三方博弈，最空达到三方都可以接受的均衡状态，形成对三者都有利的公交网络的过程。其模型要点如下：

(1) 假设有 n 个节点， n 随时间线性增长。给每个节点 i 赋一个点权 $a(i)$ （表征本站周转的客流），点权越大，说明在该站点周转的客流越多，该站点周围线路负荷越大。

(2) 给每两个顶点 i, j 之间赋边权 $b(i)$ （表征道路的拥挤程度），边权越大，该道路越拥挤。

(3) 每步时间建立一条新公交线路，每次建线路在这 n 个顶点中以均匀概率选择一个顶点作为起始站点。以一定的法则选择下一站点，法则有三种：

1) 以公交公司利益出发的法则：每次在最近邻的 k 个站点中选择 $a(i)$ 最大的作为线路的下一个站点。因为公交公司总是希望上车的乘客越多越好，它可以获得更多的收益。

2) 以乘客利益出发的法则：每次在最近邻的 k 个站点中选择顶点度 $d(i)$ 最大的作为线路的下一个站点。因为乘客为了出行方便，总是希望更多的站点能够转乘更多的线路。

3) 以交管部门利益出发的法则：每次在最近邻的 k 个站点中选择 $b(i)$ 和项目度 $c(i)$ 的乘积最小的作为下一个站点。因为交管部门总是希望站点选择在道路最畅通的路线上，并且站点通过的线路最少，尽量避免拥挤堵塞。

(4) 每一条新线建立后，计算三方操纵者的效益。

1) 公交公司的效益：站点平均乘车人数 \times 平均换乘次数 / 平均线路长度。公交公司希望以尽可能短的线路承载尽可能多的乘客，并且希望乘客尽可能多的转车以获取更多的收益。

2) 乘客的效益：平均项目度 / (平均换乘次数)。乘客希望各站点都通过尽可能多的线路，因而不需要多次转车，节约时间和车票钱。

3) 交管部门的效益： $1 / (\text{平均项目度} \times \text{平均重复边})$ 。交管部门希望每条道

路通行的公交车线路尽可能少，并且各站点通过尽可能少的线路。

每建一条线路后，比较三方效益，下次建线采取与三方效益成反比的概率挑选上述站点选择法则，也就是说，这一轮谁的效益最差，下一轮按照谁的站点选择法则的概率越大。

(5) 每条线路建好后，对点权和边权演化，即 $a(i) = a(i) * (1.0 - d(i) / \sum_{i=1}^N d(i))$ 、

$$b(i, j) = b(i, j) * (1.0 + l(i, j) / \sum_{i=1, j=1}^{N, N} l(i, j))。$$

(6) 如此演化 t 步，建立公交线路网。

5.2.2 基于博弈论的网络操纵者博弈模型

博弈论是一个分析竞争和合作行为的基本工具。本小节主要就是以博弈论中的类似于“囚徒困境”中的效益函数以及“纳什均衡”为基础建议的简单模型。对于公交系统，在上一个模型基础上我们认为最具有矛盾特征的一对操纵者是公交公司和乘客。于是，建立的模型要点如下：

- (1) 初始 n_0 个站点，在 n 个站点中随机选择 T （可能是常数或符合某种分布）个组成一个线路，初始组成 m_0 条线路。
- (2) 网络操纵者，即博弈双方为：公交公司、乘客。设网络（L 空间定义）平均最短路径为 \bar{l} ，平均最小换乘次数为 \bar{c} ，单位路径耗费时间费用为 s_1 ，单位换乘次数耗费的金钱费用为 s_2 。

当乘客选择最短路径乘车，平均换乘次数为 $c(\bar{l})$ ，则乘客耗费为：

$$p_1 = s_1 \times \bar{l} + s_2 \times c(\bar{l})， 公交公司收益为： b_1 = s_2 \times c(\bar{l})。$$

当乘客选择最小换乘乘车，平均路径为 $l(\bar{c})$ ，则乘客耗费为：

$$p_2 = s_1 \times l(\bar{c}) + s_2 \times \bar{c}， 公交公司收益为： b_2 = s_2 \times \bar{c}。$$

则效益函数为:

乘客 / 公交公司	最短路径	最小换乘
最短路径	$b_1, 1/p_1$	$b_1, 1/p_2$
最小换乘	$b_2, 1/p_1$	$b_2, 1/p_2$

在建立下一条线路前, 算效益函数。

(3) 根据效益函数, 求得博弈纳什均衡最优解, 即

当 $b_1 > b_2, (1/p_1) > (1/p_2)$ 时, 最优解为 $(b_1, 1/p_1)$;

当 $b_1 < b_2, (1/p_1) < (1/p_2)$ 时, 最优解为 $(b_2, 1/p_2)$;

当 $b_1 > b_2, (1/p_1) < (1/p_2)$ 时, 最优解为 $(b_1, 1/p_2)$;

当 $b_1 < b_2, (1/p_1) > (1/p_2)$ 时, 最优解为 $(b_2, 1/p_1)$ 。

(4) 下一条建线选站点的法则为:

如果最优解是 $(b_1, 1/p_1)$, 就在已建线路站点中选择站点。

如果最优解是 $(b_2, 1/p_2)$, 就在孤立站点中选择站点。

如果最优解是 $(b_1, 1/p_2)$ 或者是 $(b_2, 1/p_1)$, 就以 $b_1/(b_1+b_2)$ 的概率在已建线路站点中选站点, 以 $b_2/(b_1+b_2)$ 概率在孤立站点中选择站点。

(5) 每条线路建完后, 算博弈的纳什均衡最优解, 再选择下一条建线的法则。如此重复 t 步, 网络建成。

我们对上述两种模型都进行了数值模拟, 结果和我们的实证统计在一定的容忍范围里部分吻合。说明这两种模型是有一些道理和意义的, 操纵者博弈的思想都贯穿于其中。但是这两个模型都有一些明显的缺陷和不足。第一个模型所报道的选择法则和效益计算并不是只考虑某一个最关键的因素, 也并非考虑到所有的因素, 而是各自考虑了不同的因素, 不够精练或不够全面; 并且它考虑了三个方面的操纵者等等, 都使得模型相对的纷繁复杂而并没有非常接近简要。第二个模型中的效益函数中有假设的不确定的未知量, 这对网络博弈双方博弈过程有一定的影响; 效益函数矩阵中公交公司的收益中没有考虑路径所带来的间接影响等等, 使得模型并没有得到预期的思想集中和模型简化。并且, 这两种模型都没有集中

的体现某一个或几个重要因素，在数学上也无法解析证明。于是，我们在这两个模型的基础上，提出了一个最简化网络操纵者博弈模型，数学解析、数值模拟并与实证网络结果对比。

5.3 最简网络操纵者博弈模型思想和要点

我们建议最大简化地把公交公司和乘客看作网络的两个主要操纵者。他们的利益在某种意义上说是互相矛盾的。公交公司从根本上说最关心自己的经济效益，即利润；而乘客只关心如何节约金钱和时间去达到目的地。在大多数国家和地区，换乘率 c （定义为网络中任意两个车站之间完成公交交通所需的平均最小换乘次数）可能是双方利益的一种最简洁、最集中表示。换乘一次就必须购票一次（票价通常是固定数值），因此换乘次数越大，公交公司的收入越多，而乘客所花费的金钱和时间越多，得到利益越小。然而，双方肯定又都知道如果缺乏对方，博弈是不能进行的，因此当一方的效益太小，以至于不能忍受时，另一方肯定会让步。这样就可以期望在相当时间后网络的演化达到双方都可以接受的均衡状态。这样简明的思想如果可以表示为一个可解析的模型，而且所得的结论又与实证结果符合，当然是有价值的。

模型要点可以表述如下（按照上述 P 空间定义的公交线路网络，节点定义为公交车站，一条公交线路连接的两个车站之间连接一条表示直通关系的边）：

- (1) 在初始时刻已经有按照下文所述的公司建线法则建成的公交线路 l_0 条。根据实证统计，每条线路的站点数 T_l 遵从泊松分布。在以后的每个时刻按一定法则建设一条新线，共建线 l 条， l 远大于 l_0 。
- (2) 定义两个网络操纵者的效益函数：公交公司为 $b_c(t) = c(t)/c_e(t)$ ；乘客为 $b_p(t) = c_e(t)/c(t)$ 。其中， $c(t)$ 为某时刻演化得到的公交网络的换乘率， $c_e(t)$ 为下文所述的网络均衡演化法则达到同等规模网络的换乘率。这个定义体现了博弈双方利益矛盾的基本思想。用 $c_e(t)$ 相除是为了得到双方无量纲的效益，以便互相比大小，同时可以期望在达到博弈均衡时，双方的效益都为 1。
- (3) 公交公司和乘客都具有自己最满意的建线法则，如果有完全的决定权，他们

就会按照自己的建线法则来建设新的公交线路。为了最大地增大自己的效益（换乘率），公交公司的建线法则是：每条新线只与上一条旧线相交一点（只可能在一个车站换乘一条旧线）。乘客的建线法则是：每条新线与每一条旧线都相交一点（可以选择换乘任意一条旧线）。

(4) 最大简化地认为每个时刻完全按照一方的建线法则建设新线。在初始时刻和以后的每个时刻计算双方的效益，规定谁的效益小，下次建线就以谁的法则建线。这体现两个网络操纵者之间有竞争又有合作的关系。一方的效益小，他的怨言就更大，另一方很可能让步，因此下次建线就以上次效益小的一方的法则建线。实际情况当然是在建设每一条新线时都同时考虑双方的利益，但是那样就很难解析讨论。可以期望在演化时间够长，建设线路够多时，这种简化的考虑仍旧能得到在统计意义上与实际情况一致的结果。

(5) 如此演化直到建成 l 条线。

5.4 最简网络操纵者博弈模型解析讨论

我们在以前的文章中已经说明对于公交系统，把实际上遵从泊松分布的每条线路的站点数 T_l 认为是常量不会引起统计规律的原则性改变[]，因此为了在两个极端情况和均衡情况下进行解析讨论，假设 $T_l = n$ （常数）。

首先我们试图解析讨论公交公司的最大可能效益和相应建成的公交网络换乘率。如果每一条新线路都按照公司的建线法则建设，也就是只与上一条旧线相交一个点（为了简化，规定这个点不能选择已经是两条旧线交点的车站点）。这种情况下可以解析地得到网络换乘率。首先，最小换乘次数为 0 次的站点数为 $C_l^1 C_n^2$ ；最小换乘次数为 x 次的站点数为 $C_{l-x}^1 C_{n-1}^1 C_{n-1}^1$ ($x=1,2,\dots,l-1$)，而总的站点数为： $l \times n - (l-1)$ ，所以，换乘率为（总换乘次数/总站点数）：

$$c = \frac{0 \times C_l^1 C_n^2 + \sum_{x=1}^{l-1} x C_{l-x}^1 C_{n-1}^1 C_{n-1}^1}{C_{n-(l-1)}^2} = \frac{\sum_{x=1}^{l-1} x(l-x)(n-1)^2}{C_{n-l+1}^2} = \frac{(l+1)(l-1)(n-1)}{3(l \times n - l + 1)}$$

的均衡换乘率就可以算出公交公司的最大可能效益。

其次我们试图解析讨论乘客的最大可能效益和相应建成的公交网络换乘率。如果每一条新线路都按照乘客的建线法则建设，也就是与每一条旧线都相交一个点（为了简化，规定这些点都不能重叠，即不能选择已经是两条旧线交点的车站点再作为交点）。这种情况下也可以解析地得到网络换乘率。因为任意两条线都相交一点且只相交一点，所以最小换乘次数只能取 0 和 1。定义 c_1 为一个非交点与所有距它最小换乘次数为 1 的站点组成的点对数， c_2 为一个交点与所有距它最小换乘次数为 1 的站点组成的点对数，则有：最小换乘次数为 0 次的站点对数为 $C_l^1 C_n^2$ ；最小换乘次数为 1 次的站点对数为 $\frac{c_1 + c_2}{2}$ ，其中

$$c_1 = C_{nl - c_l^2 - c_l^2}^1 C_{nl - c_l^2 - n}^1 = (nl - l^2 + l)(nl - \frac{l(l-1)}{2} - n),$$

$$c_2 = C_l^2 C_{nl - c_l^2 - (2n-1)}^1 = \frac{l(l-1)}{2}(nl - \frac{l(l-1)}{2} - 2n + 1)。总的站点数为：$$

$$nl - C_l^2 = nl - \frac{l(l-1)}{2},$$

$$\text{所以，换乘率为： } c = \frac{0 \times C_l^1 C_n^2 + 1 \times \frac{c_1 + c_2}{2}}{C_{nl - c_l^2}^2} = \frac{c_1 + c_2}{2C_{nl - \frac{l(l-1)}{2}}^2} = \frac{c_1 + c_2}{(nl - \frac{l(l-1)}{2})(nl - \frac{l(l-1)}{2} - 1)}。由$$

下文得到的均衡换乘率也可以算出乘客的最大可能效益。

可以猜想在演化够长时间以后，双方的博弈达到一个均衡，这时双方的无量纲效益相等（且都等于 1）。以后如果由于任何因素，一方的效益变小，下一次就会按照他的建线法则建设一条新线，从而又使他的效益增大。因此以后的演化过程中双方的效益一定围绕这个均衡值波动，双方轮换以自己的建线法则建设一条新线。可以不失一般性地假设在此以后第奇数条公交线路与每一条旧线都相交一点，第偶数条线只与上一条旧线相交一点，以方便解析讨论。

对于新建的第 $2k$ 条线 ($k = 1, 2, \dots$)，新增的换乘为 0 的站点对为 C_n^2 ；新增的换乘为 1 的站点对为 $C_{n-1}^1 C_{n-1}^1$ ；新增的换乘为 2 的站点对为 $C_{n-1}^1 C_{2kn - k^2 - (2n-1)}^1$ 。对于新建的第 $2k-1$ 条线 ($k = 1, 2, \dots$)，增加的换乘数为： $C_{n-(2k-1)}^1 C_{(2k-1)n - (k^2-1) - n}^1 (c=1)$ 。

减少的换乘数为： $c'_{2k-1} = c(1 \rightarrow 0) + c(2 \rightarrow 1) + 2c(2 \rightarrow 0)$ ，其中：

$$c(1 \rightarrow 0) = \frac{(2k-1-1)^2}{4} = (k-1)^2,$$

表示原来换乘一次，建设这条新线后变为换乘零次。

$$2c(2 \rightarrow 0) = (C_{2k-1}^2 - \frac{(2k-2)^2}{4}) \times 2 = 2(k-2)(k-1), \quad c(2 \rightarrow 1) = 4 \sum_{i=1}^{k-2} i(n-k+i) - \frac{1}{6}(k-2)(k-1)(2k-3),$$

意义类似。所以，总的增加的换乘数为： $\Delta c_{2k-1} = (n-2k+2)(2nk-2n-k^2+1) - c'_{2k-1}$ 。总

的换乘数为 $\begin{cases} \sum_{k=1}^{l/2} (c_{2k} + c_{2k-1}), & l \text{ 为偶数} \\ \sum_{k=1}^{l-1/2} c_{2k} + \sum_{k=1}^{l+1/2} c_{2k-1}, & l \text{ 为奇数} \end{cases}$ ，总的站点数 $\begin{cases} \frac{4nl-l^2}{4}, & l \text{ 为偶数} \\ \frac{4nl-2l-l^2+3}{4}, & l \text{ 为奇数} \end{cases}$ ，可以求

出换乘率 $c = \begin{cases} \frac{\sum_{k=1}^{l/2} (c_{2k} + c_{2k-1})}{C_{(4nl-l^2)/4}^2}, & l \text{ 为偶数} \\ \frac{\sum_{k=1}^{l-1/2} c_{2k} + \sum_{k=1}^{l+1/2} c_{2k-1}}{C_{(4nl-2l-l^2+3)/4}^2}, & l \text{ 为奇数} \end{cases}$ 。由此表达式就可以方便地对演化到某时刻

的网络（分别 l 为偶数或奇数的情况）求得网络的换乘率以及相应的博弈双方效益。

我们根据解析模型思想进行了数值计算，其结果和解析结果是一致的。证明了我们的解析是正确的。

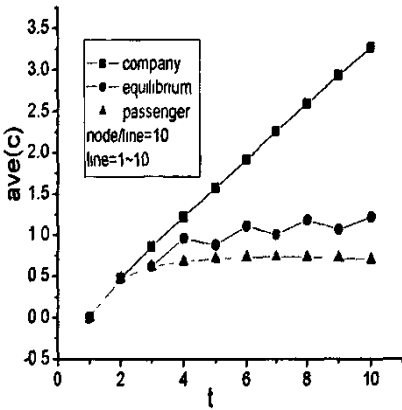


图 5.4.1 解析得到的三种建线方式下平均换乘率随线路数的变化图

图 5.4.1 显示了解析在三种建线规则下的平均换乘率的演化示意图。方块点代表的是按照公司利益建线得到的平均换乘率，三角形点代表的是按照乘客利益建线得到的平均换乘率，圆圈点代表的是均衡建线得到的平均换乘率。一共建了十条线，每条线 10 个站点。图中显示表明，按照公交公司利益建线，平均换乘率随着线路数越来越大，效益自然越来越高。按照乘客利益建线，

从第三条线开始，平均换乘率随着线路数越来越小，效益也是越来越高，但变化趋势没有公司的快。均衡建线的换乘率是增大和减小间隔而进的。解析示意图也

说明了我们关于换乘和操纵者效益的关系是可以理解的。

5.5 最简网络操纵者博弈模型数值模拟与实际对比

5.5.1 模型数值模拟结果

数值模拟中, 公交线路上的公交车站数 $T_f = n$ 为一个如前所述的泊松分布 $p_\lambda(T) = \frac{\lambda^T}{T!} \cdot e^{-\lambda}$ (其中 $\lambda = 20.1$, T 在 10 到 50 之间变化), 最后建成的公交线路数 $l = 150$ 。图 5.5.1 显示了按照模型数值模拟得到的效益演化曲线。由于初始的少量线路是按照公司建线法则建设的, 所以开始的若干时间内公司的效益明显大于乘客的。一定时间后, 双方的效益都围绕 1 作小幅度波动, 说明模型的演化确实导致均衡。

我们计算了如上演化到均衡状态的公交线路网的度分布 (一个车站可直达另一个车站的数目的分布)、点强度分布 (一个车站可直达另一个车站的直达关系边的数目的分布)、项目度分布 (一个车站停靠的公交线路数的分布)、项目含顶点数分布 (一个公交线路含车站数分布)、以及两个车站之间最小换乘数的分布。图 5.5.2 显示的是累计度分布函数曲线。图 5.5.3 显示的是累计点强度分布函数曲线。图 5.5.4 显示的是累计项目度分布函数曲线。图 5.5.5 显示的是累计项目含顶点数分布函数曲线。图 5.5.6 显示的是累计最小换乘次数分布函数曲线。

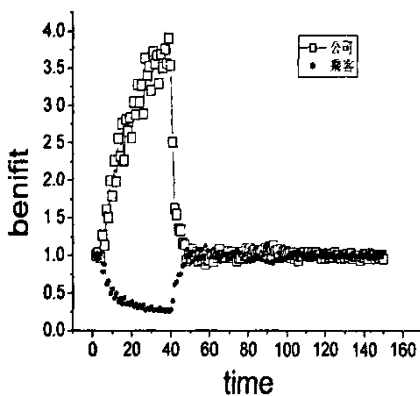


图 5.5.1 模型效益演化曲线

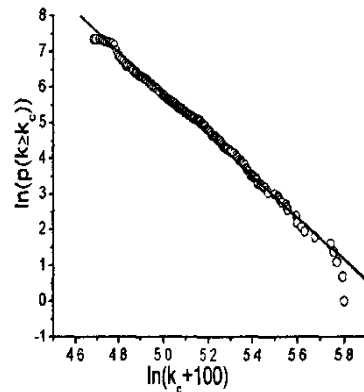


图 5.5.2 模型模拟得到的累计度分布曲线

模型模拟得到的度分布、点强度分布、项目度分布都是第二章项目报道的介

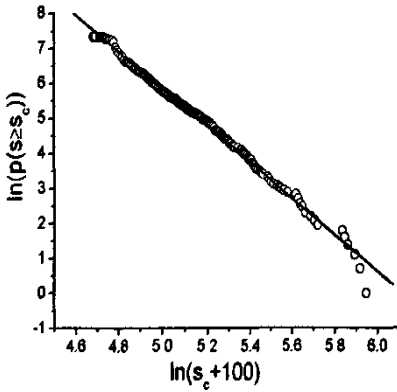


图 5.5.3 模型模拟得到的累计点强度分布曲线

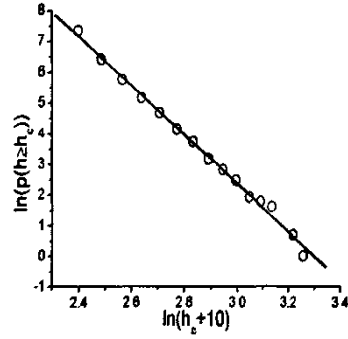


图 5.5.4 模型模拟得到的累计项目度分布曲线

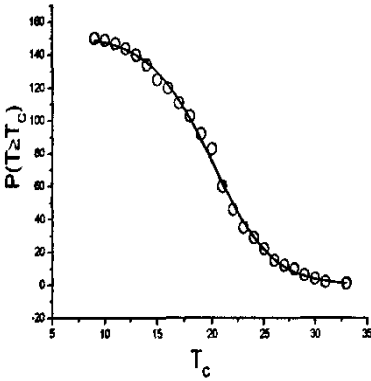


图 5.5.5 模型模拟得到的项目含顶点数
累计分布曲线

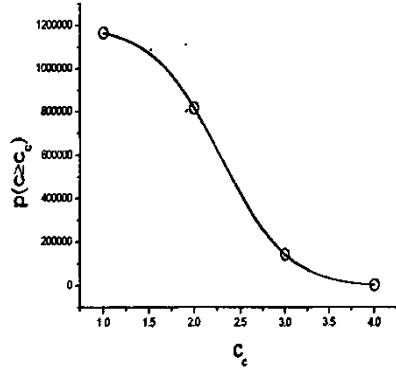


图 5.5.6 模型模拟得到的最小换乘次数
累计分布曲线

于幂律函数与指数函数之间的 SPL 函数分布，如 $P(k) \propto (k + \alpha)^{-\gamma}$ 。当关键参数 α 为 0，SPL 函数就转化为幂律函数；当 α 趋向于无穷大时，SPL 函数就转化为指数函数。上述是在数学近似解析下得到的结论，而针对于具体的数据，当 α 趋向于一个有限值时，SPL 分布曲线就近似于指数分布曲线了。我们得到的度分布、点强度分布和项目度分布都可用 SPL 函数很好的拟合，远离幂律函数分布，都更近似它们的指数分布。项目含顶点数分布是个泊松分布，这是模型根据实际情况预设的。最小换乘次数分布虽然只有四个数据点，但从累计的和 not 累计的曲线拟合来看，也可被认为是个泊松分布（或高斯分布）。这非常直观的符合实际情况。我们也计算了一些不同参数下模型模拟得到的结果，定性上不改变上述的结果。

下面一小节将报道我们一部分的实证结果以和模型模拟结果作比较。

5.5.2 实际城市公交线路网统计结果

我们实证统计了五个城市的公交线路网(包括中小城市和大城市),即扬州公交线路网、南京公交线路网、杭州公交线路网、上海公交线路网和北京公交线路网。已经在第三章中报道了五个城市的公交线路网的部分统计结果。下面以南京公交线路网为例显示其各个统计量的分布图以与上述模型模拟结果作比较。图 5.5.7 显示了累计度分布图,图 5.5.8 显示了累计点强度分布图,图 5.5.9 显示了累计项目度分布图,图 5.5.10 显示了累计项目含顶点数分布图,图 5.5.11 显示了累计最小换乘次数分布图。

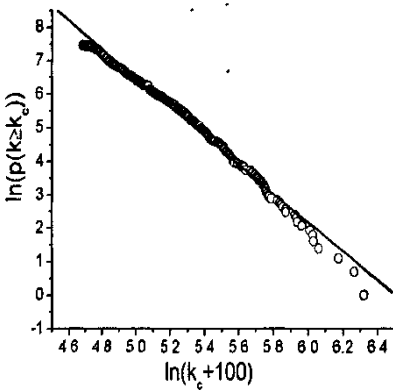


图 5.5.7 南京公交线路网的累计度分布

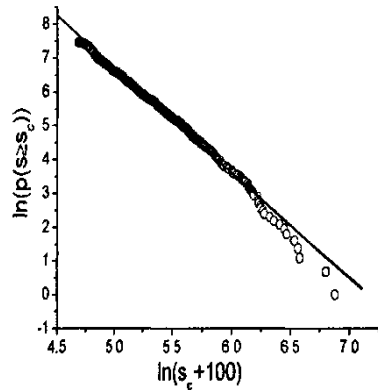


图 5.5.8 南京公交线路网的累计点强度分布

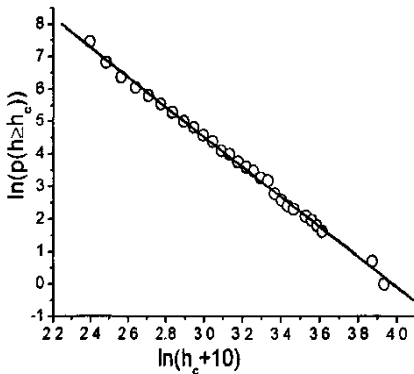


图 5.5.9 南京公交线路网的累计项目度分布
曲线

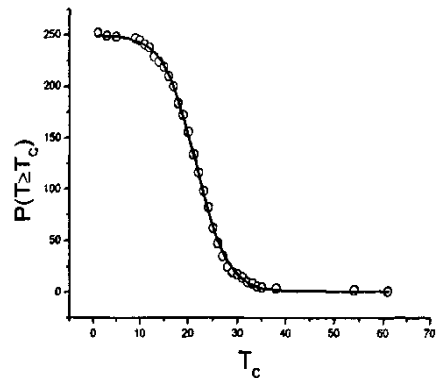


图 5.5.10 南京公交线路网的累计项目
含顶点数分布曲线

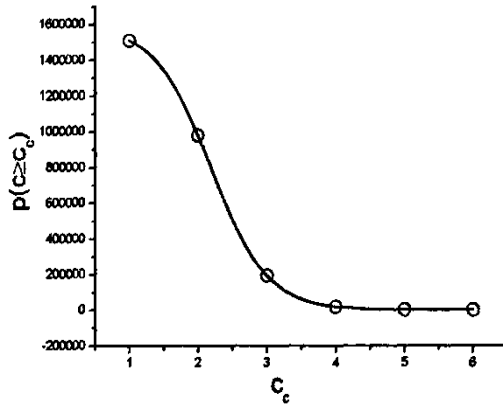


图 5.5.11 南京公交线路网的最小换乘
次数累计分布曲线

由上面显示的各分布图所示，南京公交线路网的度分布、点强度分布、项目度分布都可以用 SPL 函数很好的拟合，也都是远离幂律函数分布，接近指数函数分布。项目含顶点数分布和最小换乘次数分布也都可看作泊松分布(或高斯分布)。实证得到的结果和上一小节模型模拟得到的结果非常的吻合。其它四个城市公交线路网的实证结果和南京公交线路网的实证结果定性的吻合，都可以做为模型合理性的佐证。

5.6 小结

本章我们讨论了本组建议过的三种网络操纵者博弈模型。据我们知识所了解，本组首次建议了以网络操纵者博弈来演化网络的思想。传统的博弈模型，如囚徒困境和铲雪模型，都是有深远的研究意义，这些著名的模型用于网络中主要是讨论许多节点之间的博弈。但我们认为在一些复杂系统中，网络的演化并非基于节点之间的博弈，而是基于一些网络操纵者之间的博弈，这可能是很直接而易于理解的。本章我们重点报道了一个最简化公交网络操纵者博弈模型，所得到的解析和数值模拟结果与实证研究结论很好地符合。这样简明的思想是否能说明非常复杂的实际公交网络演化的本性呢？本章作为一种新的研究交通网络的试探，希望能够激发更多的相关研究。

第六章 三个实际系统的统计调研及其模型

6.1 简介

本章主要介绍三个实际系统的统计调研及其模型，这里只介绍我们做过的粗略工作，尚需要进一步深入研究。

首先介绍的是中国入境旅游系统。入境旅游指其他国家居民进入本国国境内所进行的旅行活动，是国际旅游事业中与出境旅游相对应的那部分旅游。考虑到旅游市场特征的一致性，我们选择中国大陆的 31 个省、自治区、直辖市为研究区域，涉及 13 个主要客源国，以中国入境旅游者中的外国人为研究对象^[87]，统计了 1995 年至 2004 年的入境旅游数据。从复杂网络的角度对其进行描述。

其次介绍的是奥运会竞技系统。奥林匹克运动是人类历史上不可分割的重要组成部分，是从现代奥林匹克主义中诞生的一个规模宏大的社会运动。奥运会竞技系统有很多诸如复杂性、动态性、随机性等特点，促使我们从复杂网络的角度对其进行深入的认识。我们统计了 1—28 届奥运会奖牌榜的数据^[88]，可以比较准确地反映出奥运会竞技的发展情况。

最后介绍的是图书借阅系统。图书馆是信息搜集、整理、存贮、加工与传播的中心，如何更广泛地了解图书的质量和受欢迎程度，以及怎样更好地提高图书馆的利用效率等问题都是各大图书馆的着重点。图书种类浩瀚繁多，我们选取了本校敬文图书馆的物理类图书借阅系统作为研究对象。经过筛选整理图书馆借阅数据，得到 3207 种被借阅过的物理图书，将其分类隶属于 227 类不同的物理学图书范围^[89-90]。

6.2 中国入境旅游系统

6.2.1 实证统计

在第二章中就已经详细的介绍了二分图的描述方法。此处也借鉴二分图的描述方法，将每个旅游目的地定义为项目(act)，每个客源国定义为参与者顶点(actor)，如果两个客源国选择同一个旅游目的地，则连边。此二分图的参与者投影图是个

特殊的全连通网络图。

图 6.2.1 显示了全国入境旅游人数的逐年变化情况。发现除了特殊的年份(2003 年“非典”的冲击)外, 总体趋势是持续增长的, 增长速度也是越来越快的。图 6.2.2 显示了双对数坐标下全国入境旅游人数逐年变化情况的累计图, t 表示第几个年份, k 表示人数。图 6.2.2 可以拟合很好的线性关系, 说明入境旅游人数是近于幂律增长的。入境旅游人数的高速增长, 主要应得益于中国的日益开放和发展。

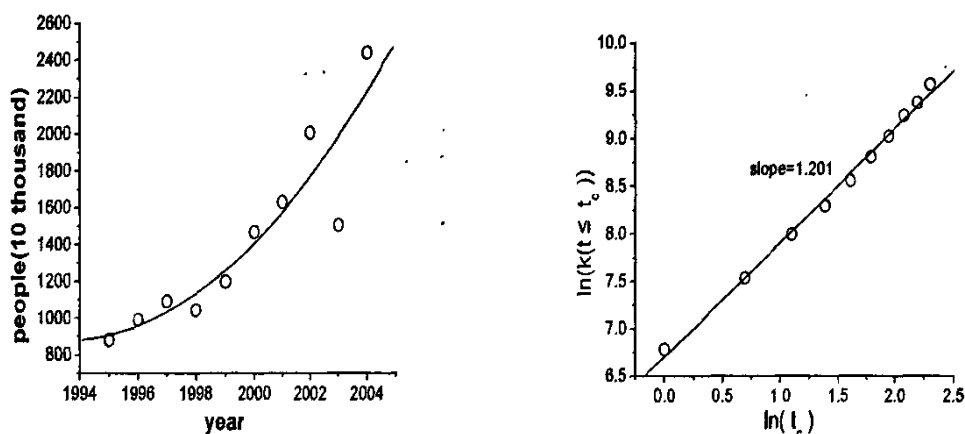


图 6.2.1 全国入境旅游人数的逐年变化图

图 6.2.2 全国入境旅游人数累计的逐年变化图

我们还统计了各省(区)的入境旅游人数的逐年变化和人数累计的逐年变化情况。统计结果显示, 在双对数坐标上, 均有很好的线性关系, 都是幂律增长, 但各省(区)的幂指数各不相同。幂指数范围在 0.96~1.97 之间。图 6.2.3 显示的是中国几个典型省(区)的入境旅游人数累计的逐年变化图。

从统计观点来看, 市场占有率 α 和市场增长率 β 是反映旅游市场竞争格局最直观、最简明的一组指标, 称为旅游市场竞争态, 记为 $Q(\alpha, \beta)$ 。给定合适的阈值划分标准 $Q_0(\alpha_0, \beta_0)$, 依据 $Q(\alpha, \beta)$ 可将区域旅游市场划分为明星市场、金牛市场、幼童市场和瘦狗市场四种类型, 它们分别处于由 $Q_0(\alpha_0, \beta_0)$ 所划分的四个象限中。我们统计了各个年份的旅游市场竞争态情况, 图 6.2.4 显示了 2000 年的入境旅游市场竞争态图。图 6.2.3 和图 6.2.4 都说明了不同特点的省(区)的入境旅游特点

是不同的，应该有选择的考虑措施来发展旅游市场。

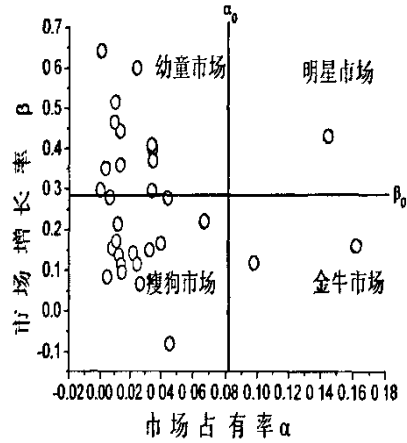
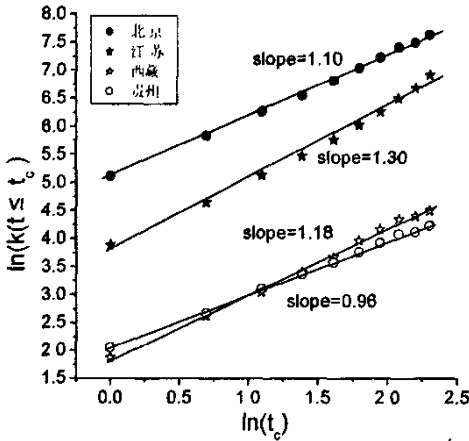


图 6.2.3 四个省(区)入境旅游人数累计随时间变化图 图 6.2.4 2000 年入境旅游市场竞争态

6.2.2 模型

模型要点如下：

- (1) N 个目的地， M 个客源地，演化 T 次；定义各个客源地大小（人数） B_i ，初值在 $0 \sim 10000$ 内随机赋值；定义旅客在目的地的消费因子 P_i ，初值在 $0 \sim 1$ 内随机赋值；定义客源地到目的地的距离因子 D_i ，在 $0 \sim 1$ 内随机赋值。
- (2) 定义目的地吸引因子 A_i ，表征各个目的地的吸引程度。吸引子分为长期因子 $A_i(l)$ 、短期因子 $A_i(s)$ 和稳定因子 $A_i(w)$ ，则 $A_i = c_1 \times A_i(l) + c_2 \times A_i(s) + c_3 \times A_i(w)$ ，其中 c_1, c_2, c_3 为可调参数并且 $c_1 + c_2 + c_3 = 1$ 。随机赋吸引因子初值 $0 \sim 1$ 。
- (3) 定义客源分配因子，即分配因子 $= \frac{\text{目的地吸引子}}{\text{距离因子}}$ ，将分配因子从大到小排序并

分为三个等级（对应于实际数据中的主要三个洲）。则客源选择目的地的分配

原则为：第一等级：客流量=客源大小 \times rand(0.6~1) \times 目的地吸引因子；

第二等级：客流量=客源大小 \times rand(0.4~0.6) \times 目的地吸引因子；

第三等级：客流量=客源大小 \times rand(0~0.4) \times 目的地吸引因子。

- (4) 定义目的地收益，收益=客流量 \times 消费因子。

(5) 客源大小、消费因子、吸引因子随时间演化，

客源大小： $B_i(t+1) = B_i(t) \times (1 + \xi_1)$ ，消费因子： $P_i(t+1) = P_i(t) \times (1 + \xi_2)$ ，吸引因子：以上步时间归一后的收益份额为长期因子；收益越少，改变策略的概率越大，以收益份额的倒数归一后的概率为短期因子；稳定因子重新随机赋值。

为了和实际对比，我们模拟了

30 个目的地市场，13 个客源国，10 年来的客流量变化情况。图 6.2.5 显示的是旅游总人数累计随时间的变化情况，图 6.2.6 显示的是四个典型目的地旅游人数累计随时间的变化情况，它们都是幂律增长的，并且幂指数各不相同。图 6.2.7 显示的是第 8 步时间演化的旅游目的地市场竞争态

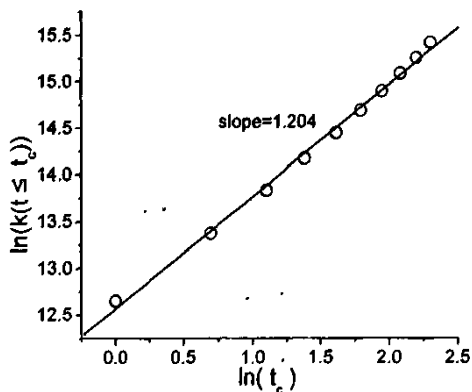


图 6.2.5 模拟旅游总人数累计随时间变化图

的目的地市场竞争态情况。从这些图中可以看出，模拟数值结果和实际统计结果定性的符合。

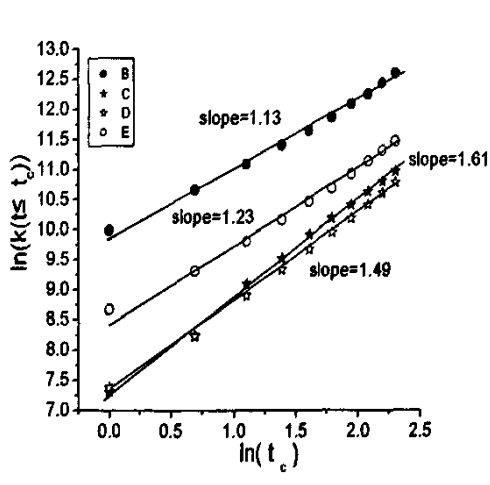


图 6.2.6 模拟得到四个目的地旅游人数累计随时间的变化图

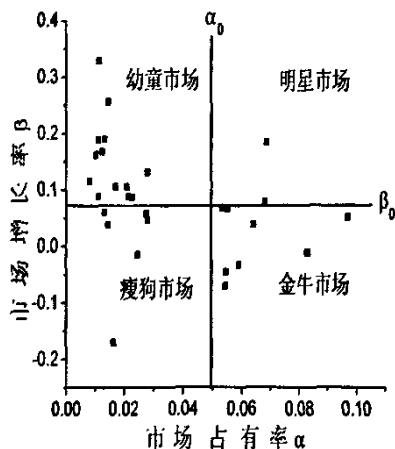


图 6.2.7 模拟得到第 8 步时间演化的旅游目的地市场竞争态

6.3 奥运会竞技系统

6.3.1 实证统计

奥运会竞技系统中,定义各个运动项目为项目(act),各个参赛国家为参与者顶点(actor),两个国家参与竞争同一个项目就连边。针对这个系统来说,比赛前,许多参与者参与一个项目,每两个参与者之间都有表示竞争的边;比赛后,许多边消亡(失败),只剩下三个参与者(金、银、铜牌得主)。在数据的统计上,我们将金、银、铜牌以一定的比例转换为成绩(总分值),以利于分析。

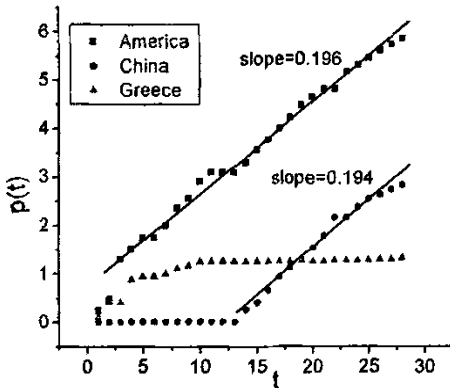


图 6.3.1 三个国家历届奥运会获奖成绩

从图中可以分析,美国属于竞争力一直很强的国家类型,中国属于一开始竞争力弱后来竞争力持续上升的国家类型,而希腊属于竞争力一直比较平稳的国家类型。

我们统计了各参赛国在历届奥运会中的获奖变化情况,发现有三种竞争类型的国家。图 6.3.1 显示的是代表三种类型的典型国家在历届奥运会的获奖变化情况,横坐标 t 表示奥运会届数,纵坐标 $p(t)$ 表示 t 届奥运会成绩。从图中可以分析,美国属于竞争力一直很强的国家类型,中

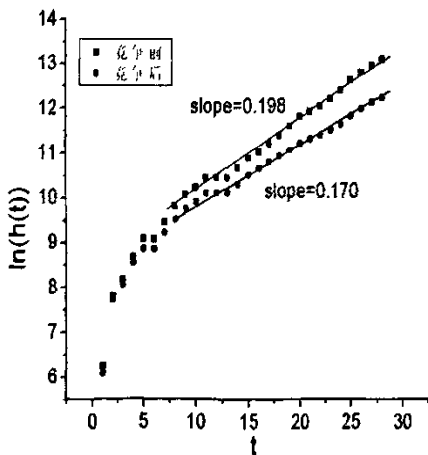


图 6.3.2 竞争前后总项目度随时间变化图

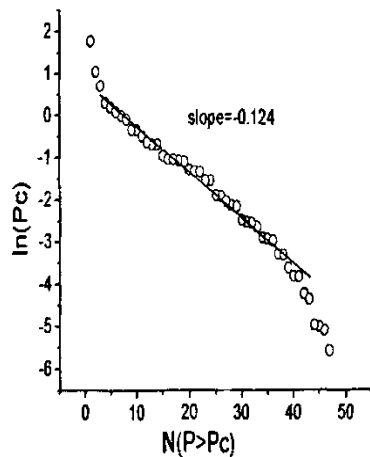


图 6.3.3 奖牌成绩的累计分布

我们对奥运会竞技网络比赛竞争前后总项目度进行了统计比较。竞争前的总项目度定义为总的项目数与参赛国家数的乘积, 竞争后的总项目度定义为总的项目数与获奖的国家数的乘积。图 6.3.2 显示了历届奥运会上网络竞争前后的总项目度对比, t 表示奥运会届数, $h(t)$ 表示 t 届奥运会的总项目度。由图可以看出, 竞争前后的总项目度随时间是趋于指数增长的, 并且随着时间的推移它们的差距逐渐拉开, 表明参赛的国家越来越多, 而获奖的国家相对来说反而越来越少了, 奖牌有被竞争力强的国家垄断瓜分的趋势。图 6.3.3 显示了奥运会奖牌成绩的累计概率分布, P_c 表示得奖成绩, $N(P_c)$ 表示得奖成绩为 P_c 的概率分布(国家数), 图在单对数坐标下线性拟合, 即趋向于指数分布。

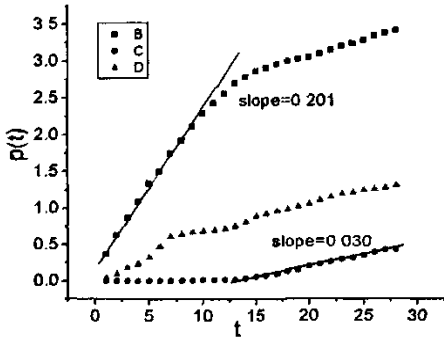
6.3.2 模型

模型要点如下:

- (1) 每届奥运会有 N 个国家参与, 有 M 个比赛项目。定义各比赛项目的难度因子 u , 在 $0 \sim 1$ 范围内随机取值。
- (2) 定义各参赛国的竞争因子 j , 其分为长期因子 $j(l)$ 、短期因子 $j(s)$ 和稳定因子 $j(w)$, 且 $j = c_1 \times j(l) + c_2 \times j(s) + c_3 \times j(w)$ 。其中 c_1, c_2, c_3 为可调参数并且 $c_1 + c_2 + c_3 = 1$ 。随机赋吸引因子初值 $0 \sim 1$ 。
- (3) 竞争因子从大到小排序。取前 n 个国家(可能得奖的国家), 在这 n 个国家中随机选取 3 个国家, 对这 3 个国家的竞争因子进行比较, 最大的得金牌, 其次的得银牌, 最小的得铜牌。将奖牌以一定的换算比例换算成总成绩。
- (4) 时间演化: 项目递增, $M(t+1) = M(t) \times (1 + \xi_1)$;

可能得奖国家数递增, $n(t+1) = n(t) \times (1 + \xi_2)$; 每届参赛者都以上一届的获奖概率为长期因子, 上一届的获奖概率越小, 本届的策略改变越大, 即以上届的获奖概率的倒数归一后所占比例为短期因子, 稳定因子随机赋值。

数值模拟中, 我们取时间 t 为 30 步。图 6.3.4 显示了代表三种类型的参与者获奖成绩随时间变化情况。图 6.3.5 显示了竞争后的总项目度随时间变化情况。图



6.3.6 显示了奖牌成绩的累计概率分布。与实际统计图对比，可以看出，数值模拟结果和实证定性的一致。

图 6.3.4 模拟参与者获奖成绩随时间变化

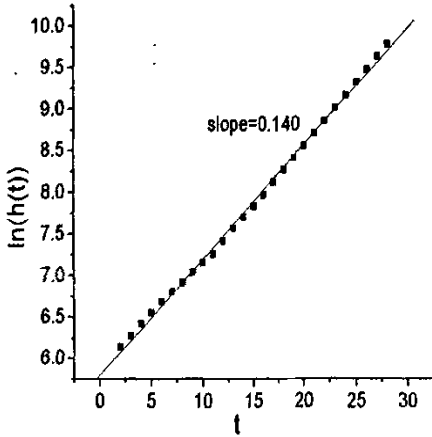


图 6.3.5 模拟竞争后总项目度随时间变化图

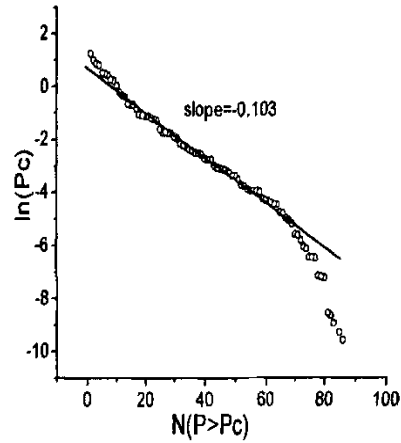


图 6.3.6 模拟奖牌成绩的累计概率分布

6.4 物理图书借阅系统

6.4.1 实证统计

物理图书借阅网中，定义每一类物理学图书范围为每一个项目(act)，每一种物理图书为参与者顶点(actor)，如果两种物理图书都属于同一类的物理学图书范围，则连边。则属于同一个项目中的顶点两两连边，构成一个项目子完全图。点权定义为每一种图书的被借阅量。

我们统计了物理图书借阅网的顶点度分布、点强度分布、项目度分布、项目含顶点数分布、二方组项目度分布(二方组指两个节点以及连接它们的边的集合，其项目度即为二方组参与了多少个项目)、集群系数(节点的邻点之间连边的概率)随点强度或项目度的变化、平均集群系数、最短道路长(连接两个节点的最短路径上

的边数)分布、平均最短道路长、同类性(度、点强度)、群落和层次、点权分布、点权随点强度或项目度的变化等。其它没说明的统计量的定义参见第二、三、四章, 这里不复赘述。

图 6.4.1 显示了项目度累计分布, 其分布是指数分布。顶点度分布、点强度分布、项目含顶点数分布、二方组项目度分布同样也都是指数分布。平均集群系数为 0.92, 平均最短道路长为 3.24, 说明它是个小世界网。图 6.4.2 显示了集群系数随项目度的变化情况, 是个指数变化。集群系数随点强度变化情况与随项目度变化情况类似。最短道路长分布是个泊松(高斯)分布。

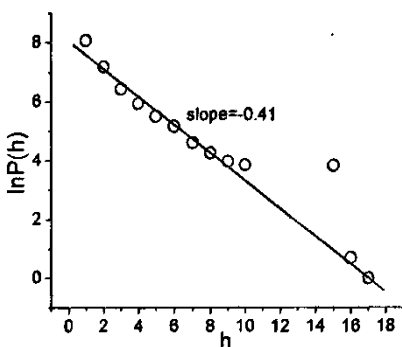


图 6.4.1 物理图书借阅网顶点项目度累计分布

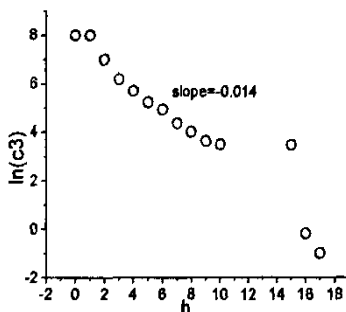


图 6.4.2 集群系数随顶点项目度的变化

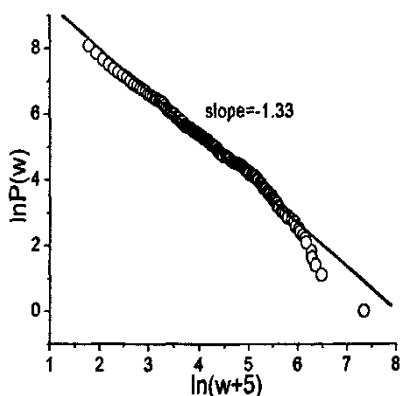


图 6.4.3 物理图书借阅网点权累计分布

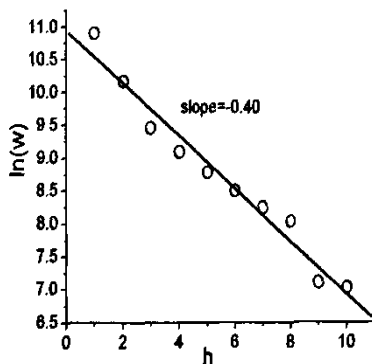


图 6.4.4 点权随顶点项目度的变化图

我们统计得到的度同类性为 0.83, 点强度同类性为 0.73。结果表示此图书借阅网无论是度还是点强度都是同类匹配的。此图书借阅网的最高层次是 93, 最低

层次是 1, 表明项目子完全图间最多重叠 93 个节点, 最少重叠 1 个节点。图 6.4.3 显示了此物理图书借阅网的点权累计分布, 是个很好的 SPL 分布(接近幂律分布)。图 6.4.4 显示了点权随顶点项目度的变化情况, 是按照指数变化的。点权随点强度变化和随顶点项目度的变化情况是一致的。

6.4.2 模型

模型要点如下(分两步):

(1) 借鉴第二章提出的模型, 建立网络项目、点和边的关系。

- 1) 初始 $t=0$ 时, 有 $N_0=50$ 个顶点, 构成 10 个项目。设项目含顶点数是指数分布(与实证一致)。
- 2) 网络生长, 组织项目。每增加一个顶点, 就在旧的顶点中随机选择 $T-1$ 个顶点, 和新顶点一起构成新的项目。项目中每两个顶点都连边。
- 3) 演化了 $t=500$ 步。网络项目建成, 总顶点数为 550, 总项目数为 510。

(2) 以建好的网络为基础模拟点权的演化

- 1) 计算各个顶点的项目度, 将 $(1/\text{项目度})$ 作为各个顶点的初始竞争因子(实证中点权与项目度的关系呈负相关, 因此如此定义)。
- 2) 每个项目中, 顶点间有竞争关系, 即谁会被更多的选择借阅。假设每个项目中顶点总共可以被 T (项目含顶点数)个人借阅, 即项目中的每个顶点都有 T 个竞争机会。竞争得到的份额跟相对竞争因子成正比。
- 3) 每轮项目竞争后, 每个顶点在其参与的项目中都得到份额(顶点项目权), 计算每个顶点的平均项目权作为每个顶点竞争得到的份额。
- 4) 时间演化, 每轮竞争都以上一轮竞争得到的份额作为竞争因子。重复上两步。竞争 t 步后, 演化结束。将 t 步演化顶点得到的份额累加作为最终顶点得到的点权(与实际数据一致)。

我们数值模拟了 $t=1500$ 步时的情形, 得到的各个统计量都与实际统计的结果大致定性的吻合。图 6.4.5 显示了点权累计分布, 是个幂律分布。图 6.4.6 显示了权重随项目度变化情况, 由图看出是指数关系。

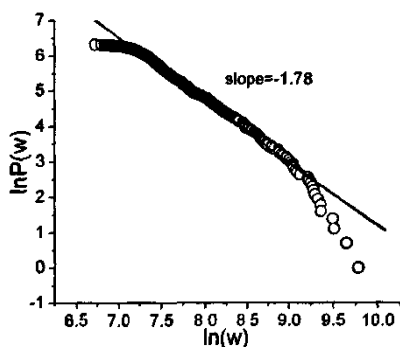


图 6.4.5 模拟得到的点权累计分布

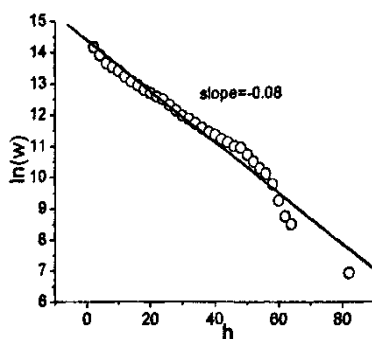


图 6.4.6 模拟得到的权重随项目度变化图

6.5 小结

以上简单介绍了三种实际系统从调研到建立模型的过程。这些工作都是比较浅陋的。在前面二、三、四章节中，我们主要研究的是广义合作网络，包括实际调研、模型建立、拓扑认识等。广义合作网络中我们忽略了竞争因素，自然而然，我们会联想到如果网络中加入竞争的因素，又会有怎样独特的性质和特点呢。第五章中提到了博弈，众所周知，博弈论是研究合作竞争系统的一个很好的方法。但就我们的粗浅认识，博弈论不是全能的。我们是否可以完全从复杂网络的角度来认识合作竞争系统呢？因为这个疑问，这也就有了上述三个系统的调查研究。除了这三个系统，我们的合作者们还调研了十几种涉及合作和竞争因素的系统，大家不断的有了一些新的发现、新的想法和新的疑问。期望这三个系统的粗陋研究能够作为一个引子，激发研究者们对诸如合作竞争并存的网络作更深入的研究。

参 考 文 献

- [1] Erdős P, Rényi A. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 1960,5:17~60
- [2] Barabási A L. *Linked: The New Science of Networks*. Massachusetts: Persus Publishing, 2002
- [3] Watts D J. The 'new' science of network. *Annual Review of Sociology*, 2004, 30:243~270
- [4] Watts D J, Strogatz S H. Collective dynamics of 'small-world' networks. *Nature*, 1998, 393(6684): 440~442
- [5] Barabási A L, Albert R. Emergence of scaling in random networks. *Science*, 1999, 286(5439): 509~512
- [6] Wasserman S. and Faust K., *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge 1994
- [7] 罗家德, 社会网分析讲义, 社会科学文献出版社, 北京 2005
- [8] 刘军, 社会网分析导论, 社会科学文献出版社, 北京 2004
- [9] Newman M. E. J., Scientific collaboration networks. I. Network construction and fundamental results, *Phys. Rev. E* 64, 2001, 016131; Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality, *Phys. Rev. E* 64, 2001, 016132
- [10] Barabási A. L., Jeong H., Neda Z. et al., Evolution of the social network of scientific collaboration, *Physica A* 311, 2002, 590
- [11] Bilke S. and Peterson C., Topological properties of citation and metabolic networks, *Phys. Rev. E* 64, 2001, 036106
- [12] Battiston S. and Catanzaro M., Statistical properties of corporate board and director networks, *Eur. Phys. J. B* 38, 2004, 345; Caldarella G and Catanzaro M., The corporate boards networks, *Physica A* 338, 2004, 98

- [13] Kou Z. and Zhang C., Reply networks on a bulletin board system, *Phys. Rev. E* 67, 2003, 036117
- [14] Newman M. E. J., Forrest S. and Balthrop J., Email networks and the spread of computer viruses, *Phys. Rev. E* 66, 2002, 035101(R); Ebel H., Mielsch L.-I. and Bornholdt S., Scale-free topology of e-mail networks, *Phys. Rev. E* 66, 2002, 035103(R)
- [15] Fan Y., Li M., Chen J., Gao L., Di Z., Wu J., NETWORK OF ECONOPHYSICISTS: A WEIGHTED NETWORK TO INVESTIGATE THE DEVELOPMENT OF ECONOPHYSICS, *Int. J. Mod. Phys. B* 18 17–19, 2004, 2505; Li M., Fan Y., Chen J., Gao L., Di Z., Wu J., Weighted networks of scientific communication: the measurement and topological role of weight, *Physica A* 350, 2005, 643; Li M., Wu J., Wang D., Zhou T., Di Z. and Fan Y., Evolving Model of Weighted Networks Inspired by Scientific Collaboration Networks, arXiv: cond-mat/0501655
- [16] Han D.-D., Liu J.-G., Ma Y.-G., Cai X.-Z. and Shen W.-Q., Scale-free download network for publication, *Chin. Phys. Lett.* 21, 2004, 1855
- [17] Xu T., Chen J., He Y. and He D.-R., Complex network properties of Chinese power grid, *Inter. J. Modern Physics B*, 18, 17-19, 2004, 2599; 陈洁, 许田, 何大韧, 中国电力网的复杂网络共性, *科技导报*, 4, 2004, 11-14
- [18] He Y., Zhu X., He D.-R., STATISTICS AND DEVELOPING MODEL OF CHINESE SKYWAY NETWORK, *Int. J. Mod. Phys. B* 18, 17–19, 2004, 2595
- [19] Zhu X., He Y., He D.-R., An investigation on Chinese skyway network, *Bulletin of APS*, 49, 1, 2004, 1007; He Y., Zhang P., Su B., He D., Search for function coefficient distribution in traditional Chinese medicine network, *Bulletin of APS*, 49, 1, 2004, 1006; Zhang Y., He Y., He D.-R., A network description on touristy in China *Bulletin of APS*, 49, 1, 2004, 1007; Sun A., Zhang P., He Y., Su B., He D., A complex network description on traditional Chinese medicine system, *Bulletin of APS*, 49, 1, 2004, 1006-1007; Hu J., Zhang P., Qiu R., Tang J., He D., Railway network in China, *Bulletin of APS*, 49, 1, 2004, 1008; Chen J., Xu T., He Y., He D.-R., Network description on Chinese power grid, *Bulletin of APS*, 49, 1, 2004,

- 1008; Wang B., Xu T., He D., A network description on geometry and economics of Yangtze drainage area, *Bulletin of APS*, 49, 1, 2004, 1007.
- [20] 许田, 张培培, 姜玉梅, 苏蓓蓓, 何大韧, 流行病传播模型与 SARS, *自然杂志*, 26, 2004, 20-25
- [21] 张培培, 侯威, 何阅, 何大韧, 淮扬菜系的网络描述, *复杂系统与复杂性科学*, 2, 2005, 49; 何阅, 张培培, 唐继英, 韩雪芳, 邱荣, 陈其娟, 周月平, 常慧, 何大韧, 中药方剂的合作网络描述, *科技导报*, 23, 2005, 36
- [22] Zhang P., Chen K., He Y., Zhou T., Su B., Jin Y., Chang H., Zhou Y.-P., Sun L.-C., Wang B.-H., He D.-R., Model and empirical study on some collaboration networks, *Physica A* 360, 2006, 599-616
- [23] Jiang Y., Xu T. and He D.-R., A self-adaptive model of actor-collaboration network, *Inter. J. Modern Physics B*, 18, 17-19, 2004, 2604; 何阅, 张培培, 许田, 姜玉梅, 何大韧, 一个科研合作网的双粒子图自适应演化模型, *物理学报*, 53, 2004, 1710-1715
- [24] 张培培, 何阅, 周涛, 苏蓓蓓, 常慧, 周月平, 汪秉宏, 何大韧, 一个描述合作网络顶点度分布的模型, *物理学报*, 55, 2006, 60-67
- [25] 何大韧, 关于合作网络研究进展与前景的浅见, 中国高等科学技术中心, *CCAST——WL workshop series: Vol. 170 (I), Second National Forum on Complex Dynamical Networks*, 205-212
- [26] 汪小帆, 李翔, 陈关荣. 复杂网络理论及其应用. 北京: 清华大学出版社, 2006
- [27] 朱玉祥, 中医处方指南, 金盾出版社(第二版), 北京, 1996
- [28] 刘东亮等, 中医常用方剂手册, 人民军医出版社, 北京, 1996
- [29] 北京民族饭店菜谱编写组, 淮扬菜谱, 中国旅游出版社
- [30] J. Laherrere and D. Sornette, *Eur. Phys. J. B*, 1998, 2: 525.
- [31] Liu Z., Lai Y.-C. et al., Connectivity distribution and attack tolerance of general networks with both preferential and random attachments, *Phys. Lett. A* 303, 2002, 337-344
- [32] Li X. and Chen G., A local-world evolving network model, *Physica A* 328, 2003,

274-286

- [33] Ramasco J. J., Dorogavtsev S. N. and Pastor-Satorras R., Self-organization of collaboration networks, *Phys. Rev. E* 70, 2004, 036106
- [34] 张培培, 何阅, 周涛, 苏蓓蓓, 常慧, 周月平, 汪秉宏, 何大韧, 一个描述合作网络顶点度分布的模型, *物理学报*, 55, 2006, 60-67
- [35] Newman M. E. J., Assortative mixing in networks, *Phys. Rev. Lett.* 89, 2002, 208701; Mixing patterns in networks, *Phys. Rev. E* 67, 2003, 0261126
- [36] Kurant M. and Thiran P., Extracting the traffic flows and the physical graphs from time tables. Arxiv. org. preprint physics/0510151.2005a
- [37] Kurant M. and Thiran P., Layered complex networks. arxiv. org preprint physics/0510194. 2006b
- [38] Sienkiewicz J. and Holyst J.A., Public transport systems in Poland: from Bialystok to Zielona Gora by bus and ream using universal statistics of complex networks. physics/0503099, 2005a
- [39] Sienkiewicz J. and Holyst J.A., Statistical analysis of 22 public transport networks in Poland. arxiv. org preprint physics/0506074.2005b
- [40] C.V.Ferber, T. Holovatch, Yu. Holovatch and V.Palchykov. arXiv: physics/0608125
- [41] X.Xu, J. Hu, F. Liu and L. Liu, *Physica A* 374(2007)441
- [42] Y-Z. Chen, N.Li and D-R. He, *Physica A* 376(2007)747
- [43] B-B.Su, HChang, Y-Z.Chen and D-R.He, *Physica A*(2007)doi:10.1016/j.physa.2006.12.049
- [44] W.Li and X.Cai, *Phy. Rev. E.* 69(2004)046106
- [45] M. E. J. Newman, *Phys. Rev. E.* 70(2004)056131
- [46] Z.Pan, X.Li and X. Wang, *Phys.Rev. E.* 73(2006)056109
- [47] M.Bathélemy, A.Barrat, R.Pastor-Satorras, and A. Vespignani, *Physica A* 346(2005)34
- [48] M.Li, Y.Fan, J.Chen, L.Gan, Z.Di and J. Wu, *Physica A* 350(2005)643
- [49] W-X.Wang, B-H. Wang, B.Hu, G. Yan and Q.Ou, *Phys. Rev. Lett.* 94(2005)188702

- [50] A. DeMontis, M. Barthélemy, A. Chessa. And A. Vespignani, arXiv:physics/0507106
- [51] C.C. Leung and H.F. Chau, arXiv:physics/0607134
- [52] J. Laherrere and D. Sornette, Eur. Phys. J. B2(1998)525
- [53] Hangzhou: <http://www.hzbus.com.cn/>; Nanjing: <http://www.njkgc.cn/>; Beijing: <http://www.bjbus.com/>; Shanghai: <http://www.jt.sh.cn/>. Yangzhou: <http://www.yzren.com/li-fe-bussetation.asp>
- [54] Z. Liu, Y-C Lai et al., Phys. Lett. A 303(2002)337
- [55] X. Li and G. Chen, Physica A 328(2003)274
- [56] Newman M. E. J. and Park J., Why social networks are different from other types of networks, *Phys. Rev. E* 68, 2003, 036122
- [57] Eguíluz V. M., Chialvo D. R., Cecchi G. A., Baliki M. and Apkarian A. V., Scale-free brain functional networks, *Phys. Rev. Lett.* 94, 2005, 018102
- [58] Park J. and Newman M. E. J., Origin of degree correlation in the Internet and other networks, *Phys. Rev. E* 68, 2003, 026112
- [59] 解绍, 汪小帆, 复杂网络中的社团结构, *复杂系统与复杂性科学*, 2, 2005, 1-12
- [60] Newman M. E. J., Girvan M., Finding and evaluating community structure in networks, *Phys. Rev. E*, 69, 2004, 026113
- [61] Newman M. E. J., Fast algorithm for detecting community structure in networks, *Phys. Rev. E*, 69, 2004, 066133
- [62] Clauset A., Newman M. E. J., Moore C., Finding community structure in very large networks, *Phys. Rev. E*, 70, 2004, 066111
- [63] Palla G., Derényi I., Farkas I., Vicsek T., Uncovering the overlapping community structure of complex networks in nature and society, *Nature*, 435, 2005, 814-818
- [64] Radicchi F., Castellano C., Cecconi F., Loreto V., and Parisi D., Defining and identifying communities in networks, *PNAS*, 101, 2004, 2658-2663
- [65] Ravasz E. and Barabási A-L, Hierarchical organization in complex networks, *Phys. Rev. E*, 67, 2003, 026112

- [66] Holme P., Huss M. and Jeong H., Subnetwork hierarchies of biochemical pathways, *Bioinformatics*, 19, 2003, 532-538
- [67] Costa L. F., The hierarchical backbone of complex networks, *Phys. Rev. Lett.*, 93, 2004, 098702
- [68] Trusina A., Maslov S., Minnhagen P. and Sneppen K., Hierarchy measures in complex networks, *Phys. Rev. Lett.*, 92, 2004, 178702
- [69] Guimera R., Uzzi B., Spiro J., Amaral L. A. N., Team assembly mechanisms determine collaboration network structure and team performance, *SCIENCE*, 308, 2005, 697
- [70] 郭雷等, 复杂网络, 上海科技教育出版社, 上海 2006
- [71] S-Y. Huang, X-W. Zou, Z-J. Tan and Z-Z. Jin, Network-induced non-equilibrium phase transition in the “game of life”, *Phys. Rev. E* 67, 026107 (2003)
- [72] M. Anghel, Z. Torozkai, K. E. Bassler and G. Korniss, Competition-driven network dynamics: emergence of a scale-free leadership structure and collective efficiency, *Phys. Rev. Lett.* 92, 058701 (2004)
- [73] H. Fort and N. Perez, Economic demography in fuzzy spatial dilemmas and power laws, *Eur. Phys. J. B* 44, 109-113 (2005)
- [74] M. G. Zimmermann and V. M. Eguiluz, Cooperation, social networks, and the emergence of leadership in a prisoner’s dilemma with adaptive local interactions, *Phys. Rev. E* 72, 056118 (2005)
- [75] K. Gao, W-X. Wang and B-H. Wang, Self-questioning games and ping-pong effect in the BA network, to appear in *Physica A*.
- [76] W-X. Wang, J. Ren, G-R. Chen and B-H. Wang, Memory-based snowdrift game on networks, to appear in *Phys. Rev. E*
- [77] C-L Tan, W-X. Wang, X. Wu and B-H. Wang, Effects of average degree on cooperation in networked evolution games, to appear in *Eur. Phys. J. B*
- [78] L. Zubieta, A network equilibrium model for oligopolistic competition in city bus

- services, *Transpn. Res. -B* 32, 413-422 (1998)
- [79] J. S. Chang and R. L. Mackett, A bi-level model of the relationship between transport and residential location, *Transpn. Res. -B* (2005)
- [80] Y. Kidokoro, Benefit estimation of transport projects—a representative consumer approach, *Transpn. Res. -B* 40, 521-542 (2006)
- [81] M. Rosvall, A. Trusina, P. Minnhagen, K. Sneppen. Networks and Cities: An Information Perspective. *Phys. Rev. Lett.* 94, 028701 (2005)
- [82] C. von Ferber, Yu. Holovatch, V. Palchykov, Scaling in public transport networks, arXiv:cond-mat/0501296 v1 12 Jan 2005
- [83] J. Sienkiewicz, J. A. Hoyst, Statistical analysis of 22 public transport networks in Poland, *Phys. Rev. E* 72, 046127 (2005)
- [84] P. P. Zhang, K. Chen. Y. He et al. Model and empirical study on some collaboration networks. *Physica A* 360, 599-616 (2006)
- [85] Y. Chen, N. Li and D-R. He, A study on some urban bus transport networks. *Physica A* 376(2007)747
- [86] B.B. Su, H. Chang, Y-Z. Chen, D.R. He, A game theory model of urban public traffic networks. *Physica A* 379(2007)291-297
- [87] 中国旅游年鉴 1996, 1997, 1999, 2000, 2002, 中国旅游年鉴编辑委员会主编; 中国旅游统计年鉴 1998, 2001, 2003, 国家旅游局; 中国旅游出版社
- [88] 奥林匹克百科全书网 <http://olympicnets.db66.com/list/jay.asp>
- [89] 中国图书馆分类法, 《中国图书馆分类法》编委会, 北京图书馆出版社, 1999
- [90] 《中图法》4 版的多重列类及分类方法, 姚红, 图书馆学刊, 2002(1)

攻读学位期间发表的学术论文目录

- 1、广义合作网络, 常慧, 何大韧, 复杂网络第八章, 上海科技教育出版社, 2006.11
- 2、Assortativity and act degree distribution of some collaboration networks, Hui Chang, Bei-Bei Su, Yue-Ping Zhou, and Da-Ren He, PhysicaA(2007)doi: 10.1016/j.physa.2007.04.045
- 3、中国旅游线路的合作网络描述, 常慧, 何阅, 张义勇等, 科技导报, 24(9)(2006)84-87
- 4、A game theory model of urban public traffic networks, B.B.Su, H.Chang, Y-Z.Chen, D.R.He, Physica A, 379(2007)291-297
- 5、Model and empirical study on some collaboration networks, Pei-pei Zhang, Kan Chen, Yue He, Tao Zhou, Bei-bei Su, Yingdi Jin, Hui Chang, Yue-Ping Zhou, Li-Cheng Sun, Bing-Hong Wang, Da-Ren He, Physica A, 360(2006)599-616
- 6、一个描述合作网络顶点度分布的模型, 张培培, 何阅, 周涛, 苏蓓蓓, 常慧, 周月平, 汪秉宏, 何大韧, 物理学报, 55(1)(2006)60-67
- 7、中药方剂的合作网络描述, 何阅, 张培培, 唐继英, 韩雪芳, 邱荣, 陈其娟, 周月平, 常慧, 何大韧, 科技导报, 23(11)(2005)36-39
- 8、Empirical study on dyad act-degree distribution in some collaboration networks, Hui Chang, Pei-Pei Zhang, YueHe, Da-RenHe, Bulletin of APS,(2006)
- 9、Generalized collaboration networks, H.Chang, B.B.Su and D.R.He, RIUPEEEC 2006(澳门第四届大学电工电子工程会议)文集
- 10、中药方剂网络的格子系数及其标度率, 常慧, 何阅, 何大韧, 第二届全国复杂动态网络学术论坛论文集(2005)
- 11、城市公共交通网络的一个博弈模型, 常慧, 全国复杂网络学术会议论文(2006)
- 12、Community and hierarchy in generalized collaboration networks, Hui Chang, Bei-bei Su, Da-Ren He, Bulletin of APS(2007)

- 13、A game theory model of urban public traffic networks, B.B.Su, H.Chang and D.R.He, RIUPEEEC 2006(澳门第四届大学电工电子工程会议文集)
- 14、A model study on dyad act-degree distribution in some collaboration networks, Yue-Ping Zhou, Hui Chang, Da-Ren He, Bulletin of APS (2006)
- 15、Empirical study on assortativity in some collaboration networks, Bei-Bei Su, Hui Chang, Da-Ren He, Bulletin of APS(2006)
- 16、Empirical study on a publisher questionnaire network, Lei Yuan, Hui Chang, Da-Ren He, Bulletin of APS(2007)
- 17、Empirical study on course selection network of Yangzhou University, Chun-Hua Fu, Hui Chang, Da-Ren He, Bulletin of APS(2007)
- 18、Empirical study on Yangzhou University library borrowing network, Jian-ru Tao, Hui Chang, Da-ren He, Bulletin of APS(2007)
- 19、Empirical study on competition in audience rating of China Central Television programe, Lingling Wu, Yue-ping Zhou, Hui Chang, Da-ren He, Bulletin of APS(2007)
- 20、Empirical study on tourism network in China, Yu Wang, Dan Shen, Hui Chang, Da-ren He, Bulletin of APS(2007)
- 21、Empirical study on human acupuncture point network, Jian Li, Dan Shen, Hui Chang, Da-ren He, Bulletin of APS(2007)
- 22、Empirical and model study on Travel-entering China, Xue-Fang Han, Qi-Juan Chen, Hui Chang, Da-Ren He, Bulletin of APS(2006)
- 23、Sale network of the top electronic companies in China, Yue-ping Zhou, Hui Chang and Da-ren He, Dynamics of continuous,Discrete and Impulsive systems (估计将在 2007 年 7 月份发表)

- 24、Bipartite network description on collaboration-competition systems, Yue-Ping Zhou, Hui Chang and Da-ren He, 管理科学国际会议论文集（科学出版社）估计将在 2007 年 7 月份发表

致 谢

首先，我要特别向我的导师何大韧教授表达由衷的尊敬和谢意。何老师以其渊博的学识和敏锐的洞察力，将我引入最前沿研究领域，而且以他的执著、求实、勤奋的工作作风影响和鼓励我在这个日新月异的领域中开展工作。何老师高洁的人品、循循善诱的话语、积极进取的精神将会一直鼓舞着我坚实的走好人生的路。本论文从选题一直到论文最后的撰写，始终得到了何老师认真细致的指导，因而才能够得以顺利完成。

感谢物理科学与技术学院的各位领导、老师在这几年中对于我的教导和帮助。

感谢陪伴我一起度过这几年美好岁月的同学和朋友们。

感谢我的好朋友杨振兴一直以来对我学业的支持和鼓励。

我还要特别感谢扬州大学复杂性科学研究中心的全体成员，感谢他们给予我方方面的关照和帮助。本论文报告的研究内容包含了非常大的工作量，因此经过了几届研究生的接力才得到今天的结果。本论文的绝大多数原始数据是由本组研究生们指导韩雪芳、陈其娟、陶建孺等本科生同学完成的，师姐苏蓓蓓和我等人对这些数据进行了整理和重新处理的工作。在此特别感谢南京航空航天大学陈永洲博士，他提供了四个城市公交系统的数据给我们做实证调研。本论文的工作中，包含了师姐张培培、师姐苏蓓蓓、周月平、师妹付春花、师妹冯爱霞等人的辛勤劳动和帮助。在我的工作中，众位师弟妹们都给予了有力的支持和鼓励。在此一并致以我深深的谢意。

特别感谢新加坡国立大学陈侃教授在第二章报道的工作中的辛勤劳动。

特别感谢刘春平老师在第四章报道的工作中的辛勤劳动。

特别感谢姜玉梅老师在日常生活中给予的帮助。

最后，谨以这篇论文作为一份微薄的礼物，

敬献给：

我的父母——是他们生养了我并送我进入学道；

并敬献给：

我的各位老师——是他们重塑了我，并启迪了我愚昧的心智。