

从统计物理学看复杂网络研究

吴金闪^{1,2}, 狄增如¹

(1. 北京师范大学管理学院系统科学系, 北京 100875

2. 北京师范大学物理系, 北京 100875)

摘 要: 从统计物理学来看, 网络是一个包含了大量个体及个体之间相互作用的系统。本文从统计物理学的角度整理与总结了复杂网络目前的主要研究结果, 并对将来的研究工作做了一个展望。文章把网络分为三个层次——无向网络、有向网络与加权网络, 对不同网络的静态几何量研究的现状分别做了综述, 并结合网络机制模型设计与评价的需要, 提出了新的有待研究的静态几何量; 对网络机制模型做了总结与分析, 提出了有待解决的关于双向幂律网络的机制模型的问题; 部分地概括了网络演化性质, 网络的结构稳定性以及网络上的动力学模型的研究。然后, 以我们目前正在进行的两个方面的工作——科学家网络和产品生产关系网络——为例, 粗略地介绍了网络研究在一些实际问题中的应用。最后, 作为一个简单的补充和索引, 我们整理了复杂网络研究中部分常用的解析与数值计算的方法。

关键词: 统计物理学; 复杂网络; 综述; 随机图; 幂律; 无标度网络

中图分类号: O414.2

文献标识码: A

0 引 言

近年来, 关于复杂网络的研究正处于蓬勃发展的阶段^[1~3]。其研究者来自图论、统计物理学、计算机网络、生态学、社会学以及经济学等各个不同领域。网络研究的文章主要发表于 Phys. Rev. Lett., Phys. Rev. E, Physica A, PNAS 等物理类期刊, Nature, Science 等综合期刊, 以及 Ecology Letter, ACM 等专业期刊。2002 年 Rev. Mod. Phys. 的综述文章《复杂网络的统计物理学》^[1]在历史、基础与前沿等各个方面都写的非常之好, 得到了非常高的引用率, 已经在 SIS 被评为由突出影响的文章之一^[4]。但是这一年多以来, 网络研究的飞速发展, 新的理论研究、新的应用领域的发展和开辟, 使得我们有必要重新整理与总结这一领域的研究, 以促进复杂网络研究的发展。同样是网络的统计物理学, 本文更多的从复杂网络研究的不同方向进行总结, 在保证一定的广度的基础上突出深度, 阐述不同方向之间的联系, 并据此提出新的研究问题。

网络可以用来描述人与人之间的社会关系,物种之间的捕食关系,词与词之间的语义联系,计算机之间的网络联接,网页之间的超链接,科研文章之间的引用关系,以及科学家之间的合作关系,甚至产品的生产与被生产关系。网络还可以作为现象的背景舞台,例如在社会关系网络上讨论舆论的传播,接触关系网络上讨论传染病的传播,计算机病毒在 Internet 网络或邮件网络上的传播,在引文网络上研究新思想的提出与传播,在科学家网络上研究科学家之间的相互影响等。网络与现象结合还可以用来讨论网络的稳定性等结构与功能关系,例如在食物链网络上讨论个别或部分物种灭绝对整体生态系统的影响,在不同的网络上讨论传染病传播的控制,在科学家网络中讨论某个领域中不同的科学家的影响力对网络演化的影响。此外,网络本身的演化过程也是一个有趣的问题,例如 Internet 网络的形成被认为是无限定原则的,但是它却展现了一些重要而普适的结构特征与稳定性,再比如,对于某一个学科内的引文网络与科学家网络的演化机制的研究,有可能给出促进科学发展的新的方案与模式。

每一个系统中的网络都有其自身的特殊性质,有其紧密联系在一起的独特现象,有其自身的演化机制,但是由于都可以使用网络分析的方法,所以有其共性。例如关于顶点度值、介数的分析方法以及大量不同网络中存在的相同的统计特征,再如随机去点与选择性攻击对网络结构的影响及其分析方法。研究网络的几何性质,网络的形成机制,网络演化的统计规律,网络上的模型性质,以及网络的结构稳定性,并把它与具体系统结合起来是复杂网络研究的中心内容。

统计物理与图论都是研究这种共性的有力工具。网络 $G = (V, E)$ 作为图论的概念是指由一个点集 $V(G)$ 和一个边集 $E(G)$ 组成的一个图,且 $E(G)$ 中的每条边 e_i 有 $V(G)$ 的一对点 (u, v) 与之对应。记顶点数为 $N = |V|$, 边数为 $L = |E|$ 。如果任意 (u, v) 与 (v, u) 对应同一条边,则称为无向网络,否则为有向网络;如果任意 $|e_i| = 1$, 则称为无权网络,否则为加权网络。从统计物理学的角度来看,网络是一个包含了大量个体以及个体之间相互作用的系统,是把某种现象或某类关系抽象为个体(顶点)以及个体之间相互作用(边)而形成的用来描述这一现象或关系的图。统计物理学是从微观到宏观的桥梁。研究网络中顶点与边的度值与权值等微观性质与网络的几何性质、效率与稳定性等宏观性质之间的关系正是复杂网络研究的核心内容。因而,与图论的研究有所不同,统计物理学研究网络更侧重于从各种实际网络的现象之上抽象出一般的网络几何量,并用这些一般性质指导更多实际网络的研究,进而通过讨论实际网络上的具体现象发展网络上模型的一般方法,最后讨论网络本身的形成机制。统计物理学在模型研究、演化机制与结构稳定性方面的丰富的研究经验是统计物理学在复杂网络研究领域得到广泛应用的原因;而图论^[5]与社会网络分析^[6]提供的网络静态几何量及其分析方法是复杂网络研究的基础,并得到了充分的发展。

我们把个体与相互作用直接抽象为顶点与边的系统称为网络,例如 www 网络^[25], 合作者网络^[66,67], 并把网络的统计性质称为网络静态几何量;把关于实际网络演化的统计规律的分析称为网络演化性质的研究,例如 www 网络中网页数量的时间演化规律,而把关于具有特定几何性质的网络的形成机制的探索称为网络演化机制模型,例如探讨 Scale Free 网络的形成;把建立在网络上的其他模型,例如传染病^[12]、渗流模型^[13]的动态

过程称为网络上的动力学性质;把网络的各种攻击方式与响应称为**网络的结构稳定性**,例如选择性隔离对于传染病流行的控制作用。

本文是按照前面所述的逻辑来组织的,首先我们对三种网络的静态几何量研究的现状做了总结,并提出了一些新的研究对象;接着我们同样对网络机制模型做了回顾和展望,提出了有待解决的问题;随后部分地概括了网络演化性质与网络上的动力学性质的研究,并探讨了它们与网络研究的关系;接着总结了网络的结构稳定性研究,尤其是基于网络的传染病控制与食物链网络的鲁棒性的研究。最后,由于其特殊的重要地位,我们用科学家网络为例,做了一个包含以上几个方面的研究思路,并提出了产品生产关系网络(PION)的研究思路与计划。研究产品生产关系网络可以不仅把握经济学领域大量现象与关系,并且,由于这种网络结构有其特殊性,还可望促进网络研究的发展。最后一节,我们整理了在解析和数值计算中经常使用的一些技术,便于大家参考以及查找相应的文献。

1 网络上的静态几何量

在本节中我们将对网络的静态几何性质做一小结,并按照为网络机制模型的研究服务的思想提出了新的静态几何量。静态几何量指的是给定网络 $G = (V, E)$ 的微观量的统计分布或者宏观统计平均值。由于有向网络与加权网络有其特有的几何量,我们将分开讨论无向网络,有向网络与加权网络。

1.1 无向网络

目前已得到研究的典型无向网络包括:Internet 网络,电影演员合作网络,科学家合作网络,人类性关系网络,蛋白质相互作用网络,语言学网络,蛋白质折叠关系网络。

无向网络的基本几何量^[1,2]有:度及其分布特征,度的相关性,集聚程度及其分布特征,最短距离及其分布特征,介数(Betweenness)及其分布特征,连通集团的规模分布。

一个顶点的度是指与此顶点连接的边的数量,即

$$d_v = \sum_{l \in E} \nu_l^v \quad (1)$$

其中 ν_l^v 记号取值为 1 当路径 l 包含顶点 q ,否则为零,即

$$\nu_l^v = \begin{cases} 1 & \text{if vertex } q \text{ included in edges } l; \\ 0 & \text{if not} \end{cases}; \quad (2)$$

同时为了以后表述方便,我们定义其他两种类型的记号, uv 与 $_{lm}$ 分别表示顶点相邻与边相连,对于有向网络, uv 推广为 $_{lm}$, $_{lm}$ 推广为 $_{lm}$, $_{lm}$ 推广为 $_{lm}$ 。对于加权网络用记号 W_{uv} 代替 uv 就可以完整地描述。

$\forall v \in V$ 我们都可以得到其度 d_v 。度值的分布特征是网络的重要几何性质。规则网络各顶点度值相同,因而符合 δ 分布,随机网络符合泊松分布(见(2.1)节),大量实际网络存在幂律形式的度分布^[1,2],称为无标度网络(Scale Free Networks),包括 Internet 网络^[43,44],电影与电视剧演员合作网络^[17,31,45],科学家合作网络^[66,67],人类性关系网络^[46],蛋白质相互作用网络^[47],语言学网络^[48]等,同时还存在高斯型,如蛋白质折叠网

络^[45],和指数衰减型的幂率分布,如电视剧演员合作网络^[45]。

进一步的问题是,是否度分布可以完备地描述网络的特征呢?如果度分布与网络一一对应,我们就说度分布可以完整地描述网络,网络也只需要用度分布来描述。试想对于一给定的度分布,通过抽样,我们可以给对应网络的每一个顶点一个度值,保持度值的任意一种连接方式都构成一种网络。如果任意一种连接对应的其他几何量都一样,则我们认为实际上为同一网络。但是显然不能保证其他几何量也都一样。那么,保持度值不变的随机连接,即度与度之间没有相关性,是否可以成为大多数网络的代表呢?所以度分布之间的相关性是另一个重要的几何量。Newman 把它称为“匹配模式”^[15],意思是考察度值大的点倾向于和度值大的点连接,还是倾向于和度值小的点连接。具体的方法是,通过任意一条边都可以找到两个顶点,进而得到两个度值,这样通过所有的边我们就得到了两个序列,分析这两个序列的相关性即可。研究表明实际网络存在一定程度上的匹配模式,有的网络正向匹配,也有的网络反向匹配。但是,由于是无向网络,把哪一个顶点的度放入序列 a ,把哪一个顶点的度放入序列 b 是任意的,这一点对于相关性分析的影响并没有得到研究。实际网络的分析表明,不同的网络存在不同的匹配模式,有正相关也有负相关。

集聚程度的意义是网络集团化的程度,即考察连接在一起的集团各自的近邻之中有多少是共同的近邻。一种定义是对于每一个顶点 v ,找到其近邻集合 N_v ,记 $n = |N_v|$, N_v 中存在的边的数量为

$$M = \sum_{l \in E: x, y \in N_v} \frac{x \cdot y}{l} \quad (3)$$

则

$$C_v = \frac{M}{C_n^2} \quad (4)$$

于是我们可以得到所有顶点的集聚程度,它的统计分布是刻画网络的一个重要几何量,其平均值称为平均集聚程度 C 。

两点的最短路径 l_{ij} 定义为所有连通 (i, j) 的通路中,所经过的其他顶点最少的一条或几条路径。记 (i, j) 之间最短路径的集合为 S_{ij} ,相应的路径长度为 $d_{ij} = |l_{ij}|$ 。如果 (i, j) 之间不存在通路,那么记 $d_{ij} = N$ 。于是我们可以得到一个 $N \times N$ 的矩阵 $(d_{ij})_{N \times N}$ 。其分布特征是一个重要的全局几何量,其平均值称为平均最短路径 d 。

另一个重要的全局几何量是介数(Betweenness)^[16]。顶点 u 的介数含义为网络中所有的最短路径之中,经过 u 的数量。它反映了顶点 u 的影响力。记 (i, j) 之间最短路径的集合为 S_{ij} ,顶点 u 的介数定义为

$$B_u = \sum_{i, j} \frac{l_{i, j} \cdot S_{ij}^u}{|S_{ij}|} \quad (5)$$

由此可以得到每一个顶点的介数 B_u 。实证研究表明,大量实际网络的介数分布也拥有共同的统计特征^[16]。类似地,可以定义边的介数,Newman 等人发现,边的介数可以用于分析顶点的聚类^[23]。其基本思想是在包含不同集团的网络中所有最短路径经过次数最多的边,也就是介数最大的边,必然是联接两个集团之间的边。当然,建立在联接赋权基础

上的 Hierarchical Clustering 聚类算法也能给出网络上顶点的分类。

连通集团是指 G 的一个子图,在这个子图内,任意两点之间都存在通路。一个网络可能存在多个相互独立的连通集团。在渗流模型中,当系统处于临界状态时,连通集团的规模呈现出幂律分布。实证研究表明,对于大量的 Scale-free 网络,连通集团的规模也存在幂律分布^[1]。

关于无向网络的度分布还存在另外一种形式。计算近邻顶点的数目得到度值,然后把所有顶点的度值归一化后再把近邻顶点的度值相加得到顶点的二次度值,重复这一过程,直到得到稳定的度值^[6]。这样得到的每一个顶点的度值实际上是邻接矩阵的最大本征值对应的本征向量,上述过程实际上就是矩阵本征值的最速下降法。但是这样的度分布具有的几何意义目前没有得到研究。

1.2 有向网络

目前已得到研究的典型有向网络包括:WWW 网络,细胞内化学反应网络,食物链网络,引文网络,电力网络,神经网络。

当我们忽略边的方向的时候,或者反过来看认为任何一条边都是双向的时候,有向网络就成为无向网络。因此,关于无向网络的所有几何量都可以在有向网络中研究。有向网络的特殊静态几何量包括:In 度和 Out 度的分布特征,基于顶点的 In-Out 度关联性,基于边的(In-Out, In-In, Out-In, Out-Out)度关联性,双向比,In 集团和 Out 集团的集聚程度。

把无向网络的度推广为 In 度和 Out 度,

$$d_u^{\text{out}} = \sum_v \sum_v a_{uv} \quad (6)$$

以及

$$d_u^{\text{in}} = \sum_v \sum_v a_{vu}, \quad (7)$$

可分别研究 In 度和 Out 度的分布特征。实证研究表明,有向网络中存在 In 度和 Out 度的双向幂律分布,如 WWW 网络^[25, 49, 26],细胞内化学反应网络^[50, 51];但是也存在只有 In 度幂律分布的网络,如引文网络^[65];以及符合指数衰减的网络,如电力网络^[7]与神经网络^[45];此外,由于受顶点数规模限制,食物链网络的度分布特征仍然没有确定^[52~55]。

由于每一个有向网络的顶点都存在 In 与 Out 两个度值,研究这两个度值之间的关联将是有向网络的另一个重要特征。与无向网络的度相关类似,我们可以研究有向网络基于边的度相关。任意选取一条边,在边的两端存在 4 个度值,起点的 In 度和 Out 度,终点的 In 度和 Out 度,所以存在 4 种关联性,分别是 In-In, In-Out, Out-In, Out-Out 关联。尽管关联性复杂了,可是同时去掉了上一小节所提到的无向网络度相关分析的任意性。那么这样的度度相关的研究是否仅仅是无向网络中度关联性分析的平庸推广呢?不是。我们提到,所有新提出的几何量都是为了网络机制模型服务的,在下一节中大家可以看到在机制模型中,反馈机制是一个重要的因素,而有向网络的度相关性正是建立反馈机制的重要参考。

边的双向比,即双向的边占有所有边的比例,也是一个重要的概念,可以为设计双向幂

律分布的机制模型提供参考。

1.3 加权网络

目前,关于加权网络的实证研究不多,只有在 Newman 的工作^[66]中提到了一点关于科学家加权合作网络的研究。

首先,我们需要讨论加权网络的加权的必要性与方式。把一个实际问题抽象为加权网络的过程并不都是平庸的。当然,对于邮递员问题等包含距离关系的网络,直接把距离作为权就可以了,但是对于其他包含相似关系、亲密程度等社会关系的网络,如何加权就值得讨论了。尤其是,当系统当中包含多个层次的同一属性的关系的时候,就必须仔细研究其加权方式了。以下,我们以某一学术领域内的科学家网络^[66,68]为例,介绍加权网络。

为了研究某一学术领域的发展变化,某一个新的思想在此领域内的产生、传播,我们构造了一个科学家之间通过文献相互联系影响的网络。我们认为,对于研究此问题而言,科学家之间的合作关系^[66],引文关系^[65,68]以及讨论或书信交往(经常体现在致谢中)属于同一属性的关系,但是所起作用的程度有所不同。对于交流思想而言,合作是最直接的影响,其次是引文,再次为致谢。而且,在同为合作的关系内,合作的次数的不同,对于交流思想的影响也是不一样的。因此,在这个科学家网络中,每一条边代表了综合考虑了以上三个方面两个层次的相互作用以后的影响力程度。并且由于引文和致谢的有向性,此网络是一个加权有向网络。

加权网络的静态几何量包括:度及其分布特征,权及其分布特征,权的相关性,权与度的相关性,最短距离及其分布特征,介数及其分布特征与隧道现象,与相应无权网络的对比,距离关系与类聚分析,以及在加权网络上集聚程度的定义及其统计性质。

首先加权并不改变度与集聚程度等局域几何量,所以无权网络的局域几何量分析都可以在加权网络上实现。权的分布反映的是本领域内学术交流的活性,如果我们认为所有学术研究都将表现为文献发表与学术交流的话,那么它反映的是此领域内学术研究的活跃程度。

类似于无向网络的度相关性,权的相关性考察的内容是是否活跃的科学家倾向于和其他活跃的科学家合作,还是倾向于提携后进。一种可行的方法是把某一个点的权定义为从该点出发的所有的边的权重之和,即

$$W_u^{\text{out}} = \sum_v W_{uv}^{\text{out}} \quad (8)$$

以及所有到达该点的边的权重之和,即

$$W_u^{\text{in}} = \sum_v W_{vu}^{\text{in}} \quad (9)$$

其中,合作关系作为无向部分同时记入 W^{in} 与 W^{out} 。

权与度的相关性研究,考察的是科学家合作交流的广泛性与深入性的关系。用上一小节有向网络的度的定义,每一格点的有两个度值 d_u^{out} , d_u^{in} , 两个权值 W_u^{out} , W_u^{in} , 存在四种相关性 — $W^{\text{out}} d^{\text{out}}$, $W^{\text{out}} d^{\text{in}}$, $W^{\text{in}} d^{\text{out}}$, $W^{\text{in}} d^{\text{in}}$ 。忽略有向性,可以定义

$$\begin{cases} W_u = W_u^{\text{out}} + W_u^{\text{in}} \\ d_u = d_u^{\text{out}} + d_u^{\text{in}} \end{cases} \quad (10)$$

也可以讨论无向化网络的权与度相关性。

加权网络另一个有意思的特征量是单位权,即

$$U_v = \frac{W_v}{d_v} \quad (11)$$

它表示顶点每一个连接的平均权重。有关单位权的统计分布以及关联性的研究将是一个有趣的问题。

加权网络中的最短路径是指在两点之间所有连通的路径中,权数之和最小的一条或几条路径。在科学家网络中,它和新的思想在网络上的传播密切相关。与无权网络的顶点数最小相比,加权会改变最短路径。而且由于最短路径的改变,介数也将发生变化。在科学家网络中,介数反映了在本领域内,某位科学家影响力的大小。某一顶点的近邻顶点介数分布的两极分化性质称为隧道现象^[66],全部顶点的介数分布反映的是科学家影响力的层次。边的介数反映的是不同科学家之间的交流对学科发展的影响力的不同。同时,利用边的介数也可以对科学家做聚类分析。所有以上局域或全局几何量都需要与相应的无向网络或无权网络做对比研究。

最后,我们可以用两两顶点之间的距离来度量由于加权带来的顶点之间的亲密程度的不同。并且有了距离关系以后,我们就可以做聚类分析,给出科学家之间的交流集团。这种聚类分析还可以和基于边的介数的聚类分析对比。加权网络目前仍然是一块有待开发的沃土。

2 网络机制模型

上一节我们主要总结了对于实际网络的静态性质的统计研究,这一节我们去看一看什么样的网络模型会展现这些特定的统计性质。这是研究网络上的动力学模型的基础。如考虑不同网络上的传染病模型,用来描述接触性传染病的传播,谣言在人群中的传播等现象,或者不同网络上的渗流模型。对于这些问题,我们可以选择用微分方程来描述。例如以下 Logistic 模型,

$$\frac{dx}{dt} = \mu(1 - x)x \quad (12)$$

我们把 μ 看成是健康人与受感染人之间发生接触并获得传染的几率。所以在此模型中,所有个体是均匀混合的,没有局域近邻的概念,就好像是溶液中发生的化学反应一样。或者我们这样来理解 μ ,把疾病看作只要接触就传染,那么 μ 的意义是平均来看每一个体的近邻数目占所有人口的比例。因此,可见 Logistic 模型的背景是完全随机网络。作为补充,我们可以在规则网络上用模拟或者重整化群的方法来研究同一问题。而作为模型真实背景的是人的社会交往的网络,这样一些网络可以通过实证研究获得对其静态性质一定的认识。但是对于研究这些网络上的动力学模型来说,更加广泛的研究还需要网络模型,而不仅仅是实际网络数据。例如,人类相互认识的网络是 Small World 网络,那么当我们研究这样的网络上的动力学模型的时候,我们就需要一个相应特征的网络来作为我们的背景。当然,我们可以收集这个网络的实际信息,然后在这个实际网络上进行分析。

但是这有很大的局限性,尤其对于理论研究。如果我们能够设计一个具有这些性质的网络演化机制模型,那么我们进一步对网络上动力学模型的研究就可以以这些机制模型为基础。当然,如果发现现有的机制模型不能展现已有的几何性质,我们就需要修改模型。本节中,我们将分别介绍规则网络和完全随机网络^[5](以下简称随机网络),Small World 网络^[7]与 Scale Free 网络^[17]的机制模型。

2.1 规则网络与随机网络

我们把一维链,二维正方晶格等称为规则网络。规则网络是指平移对称性晶格,任何一个格点的近邻数目都相同。当然这只是一个习惯用法,不是下定义,比如 Carley Tree 显然不是随机网络,但是也没有规定说它属于规则网络。随机网络是另一个极端,由 N 个顶点构成的图中,可以存在 C_N^2 条边,我们从中随机连接 M 条边所构成的网络就叫随机网络。还有一种生成随机网络的方法是,给一个概率 p ,对于 C_N^2 中任何一个可能连接,我们都尝试一遍以概率 p 的连接。如果我们选择 $M = pC_N^2$,这两种随机网络模型就可以联系起来。对于如此简单的随机网络模型,其几何性质的研究却不是同样的简单。随机网络几何性质的研究是由 Paul Erdős, Alfréd Rényi 和 Béla Bollobás 在五十年代到六十年代之间完成的。为了强调随机网络研究与统计物理学的联系,我们从系综理论的角度重新表述了随机网络统计性质的研究结果。但是,为了在这一节中突出模型的统计性质而不是处理方法,我们把这一表述放在最后专门介绍方法与技术的一节中。

规则网络与随机网络的典型几何性质包括:度分布,平均集聚程度与平均最短距离。规则网络所有顶点都相同,因此其度值相同,度分布为 $(k - k_0)$,其平均集聚程度也只需要在一个点计算 $C = C_v = C_v$,其最短距离可以也只从某一个顶点开始计算从它到所有其他顶点之间的距离之和 $L \sim N^2$,然后计算其平均值 $d = \frac{L \times N}{2 \times C_N^2} \sim N$ 。对于随机网络 $G(N, p)$,包含了从空图到完全图的所有可能情况,因此随机图的几何性质需要对每一种可能图做平均,例如我们计算每一种可能图的最短距离,然后按照各自出现的几率做平均。一般公式与详细的计算见第 7 节。研究结果表明随机网络顶点的度值符合平均值为 Np 的泊松分布,其集聚程度约等于 p ,最短距离 $d \sim \ln(N)$ 。

对比规则网络与随机网络,我们发现,平均集聚程度与平均最短距离,这两个静态几何量能够很好地反映规则网络与随机网络的性质及其差异。规则网络的特征是平均集聚程度高而平均最短距离长,随机网络的特征是平均集聚程度低而平均最短距离小。规则网络的平均最短距离 $d \sim N$,而其集聚程度可以通过改变近邻数目 k_0 来调整,例如在如图 1 所示的规则网络中, $k_0 = 4$,集聚程度分别为 $\frac{1}{2}$ 。而在随机网络中,平均集聚程度非常的小,在图 1 所示的随机网络中,顶点数与边数都与规则网络相同,但集聚程度为 0.02。然而正是由于其集聚程度非常地小,所以其平均最短距离小。考察一个顶点 u 的近邻,假设其近邻数为 a ,那么在 a 个近邻的近邻之中相互重复的个数非常少,所以从 u 出发经过两次近邻关系我们可以找到正比于 a^2 的新顶点,最多经过 $\log_a N$ 个近邻关系,我们就可以穷尽整个网络。所以,其最短距离满足 $d \sim \ln N$ 。可见,对于规则网络,也正是由

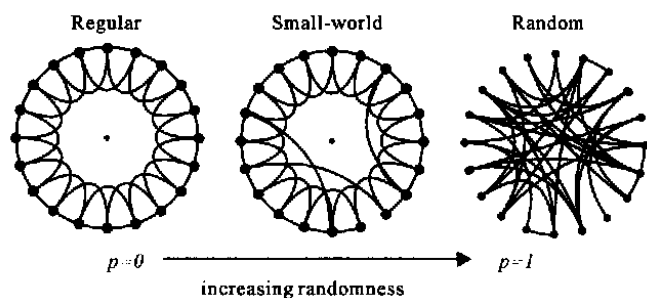


图1 Small World 网络模型。图中所示的 Small World 网络是在左图的规则网络基础上通过边的重连得到的。当 $p = 0$ 时,成为规则网络, $p = 1.0$ 时成为随机网络。本图取自文献[7]

于其集聚程度高,重复率很大,所以平均最短距离大。如此看来好像这是一对相互矛盾的几何量。

那么,是否存在一个同时具有高集聚程度,小最短路径的网络呢? 对于传染病模型,平均集聚程度对应于传播的广度,平均最短距离代表的是传播的深度。因此,如果实际网络同时存在宽的广度和大的深度的话,在这样的网络上的传染病传播显然将大大高于规则网络与随机网络。Watt 和 Strogatz 为我们找到了这样的网络模型—Small World 网络^[3,7]。

2.2 Small World 网络

Watt 和 Strogatz 发现,只需要在规则网络上稍作随机改动就可以同时具备以上两个性质。改动的方法是,对于规则网络的每一个顶点的所有边,以概率 p 断开一个端点,并重新连接,连接的新的端点从网络中的其他顶点里随机选择,如果所选的顶点已经与此顶点相连,则再随机选择别的顶点来重连。当 $p = 0$ 时就是规则网络, $p = 1$ 则为随机网络,对于 $0 < p < 1$ 的情况,存在一个很大的 p 的区域,同时拥有较大的集聚程度和较小的最短距离。一个典型的 Small World 网络见图 1 中间的示意图,其几何性质如图 2 所示。

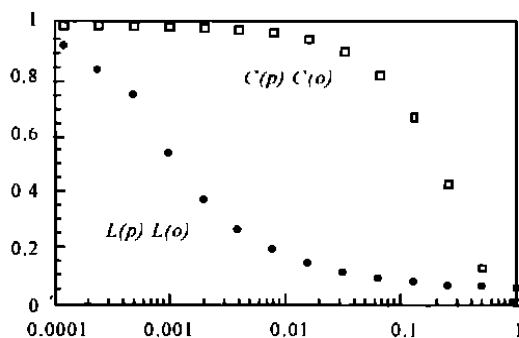


图2 Small World 网络的几何性质。同时有大集聚程度而小最短距离是 Small World 网络的重要特征,而且此性质在 p 略大于 0 到小于 1 的很大范围内存在。本图取自文献[7]

实证研究发现,大量的实际网络存在在这种 Small World 现象,见表 1。在 Watts 和 Strogatz 的工作之后,不同的作者做了许多 Small World 网络上的动力学模型的研究^[14, 36~38, 60~63],体现了平均集聚程度和平均最短距离的深刻的表现能力。我们将在网络的动力学性质一节对此做一小结。

表 1 实际网络的 Small World 现象

Network	Size	(k)	l	l_{rand}	C	C_{rand}
WWW[30],site level,undir	153,127	35.21	3.1	3.35	0.1078	0.00023
Internet[21],domain level	3015 - 6209	3.52 - 4.11	3.7 - 3.76	6.36 - 6.18	0.18 - 0.3	0.001
Movie actors[7]	225,226	61	3.65	2.99	0.79	0.00027
LANL co-authorship[66]	52,909	9.7	5.9	4.79	0.43	1.8×10^{-4}
MEDLINE co-authorship[66]	1,520,251	18.1	4.6	4.91	0.066	1.1×10^{-5}
SPIRES co-authorship[66]	56,627	173	4.0	2.12	0.726	0.003
NCSTRL co-authorship[66]	11,994	3.59	9.7	7.34	0.496	3×10^{-4}
Math. co-authorship[67]	70,975	3.9	9.5	8.2	0.59	5.4×10^{-5}
Neurosci. co-authorship[67]	209,293	11.5	6	5.01	0.76	5.5×10^{-5}
E. coli[51],substrate graph	282	7.35	2.9	3.04	0.32	0.026
E. coli[51],reaction graph	315	28.3	2.62	1.98	0.59	0.09
Words[48],co-occurrence	460,902	70.13	2.67	3.03	0.437	0.0001
Power grid[7]	4,941	2.67	18.7	12.4	0.08	0.005
C. Elegans[7]	282	14	2.65	2.25	0.28	0.05

列表中包括所研究网络的实际对象(Network),大小(Size),平均度值($\langle k \rangle$),平均最短距离(l),按随机网络计算的最短距离(l_{rand}),平均集聚程度(C),按随机网络计算的集聚程度(C_{rand})。通过对比实际网络与相应(相同顶点数和边数)随机网络的性质,可以发现网络的 Small World 特征。此表在文献[1]的基础上收集确认其他文献编辑而成,感谢作者 R. Albert 提供。

2.3 Scale Free 网络

在 Small World 网络的研究兴起之后,越来越多的科学家投入到复杂网络的研究中去。大家发现其实更多的其他几何量的特征也具有很大程度上的普适性和特定的结构功能关系。Scale Free 网络就是其中的一个重要方面。Scale Free 网络指的是网络的度分布符合幂律分布,由于其缺乏一个描述问题的特征尺度而被称为无标度网络。我们都知道幂律在统计物理学相变与临界现象,以及在自组织临界性(SOC)中的特殊地位^[24]。这种度分布的无标度性是否也会带来网络上动力学模型的特殊性质呢?我们也将在下节网络的动力学性质加以总结。实证研究发现,大量的实际网络可以被认为是 Scale Free 网络^[1],见表 2。

表 2 实际网络的 Scale Free 现象

Network	Size	k	out	in	l_{real}	l_{rand}
WWW[49]	325,729	4.51	2.45	2.1	11.2	8.32
WWW[28]	43,107	7	2.38	2.1		
WWW[26]	23,108	7.5	2.72	2.1	16	8.85
WWW[29],site	260,000			1.94		
Internet[43],domain	3015 - 4389	3.42 - 3.76	2.1 - 2.2	2.1 - 2.24		6.3
Internet[43],router	3,888	2.57	2.48	2.48	12.15	8.75
Internet[44],router	150,000	2.66	2.4	2.4	11	12.8
Movie actors[17]	212,250	28.78	2.3	2.3	4.54	3.65
Co-authors[66],SPIRES	56,627	173	1.2	1.2	4	2.12
Co-authors[67],neuro.	209,293	11.54	2.1	2.1	6	5.01
Co-authors[67],math.	70,975	3.9	2.5	2.5	9.5	8.2
Sexual contacts[46]	2,810		3.4	3.4		
Metabolic[50],E.coli	778	7.4	2.2	2.2	3.2	3.32
Protein[47],S.cerev.*	1,870	2.39	2.4	2.4		
Citation[65]	783,339	8.57		3		
Phone call[27]	533,106	3.16	2.1	2.1		
Words[48],co-occurrence*	460,902	70.13	2.7	2.7		

列表中包括所研究网络的实际对象(Network),大小(Size),平均值($\langle k \rangle$),Out 度幂律分布指数($_{out}$),In 度幂律分布指数($_{in}$),平均最短距离(l_{real}),按随机网络计算的最短距离(l_{rand})。可见,Scale Free 网络一般具有 Small World 特征。此表在文献[1]的基础上收集确认其他文献编辑而成。感谢作者 R. Albert 提供。

现在来看看 Scale Free 网络的形成机制。目前对于无向 Scale Free 网络,普遍认为偏好依附(Preferential Attachment)^[17]是一个很好地形成 Scale Free 网络的机制。具体模型如下。取初始 m_0 个顶点任意连接或完全连接。每一步在原网络 $G(t-1)$ 的基础上加上一个新的顶点,同时加上从此顶点出发的 m 条边,形成新的网络 $G(t)$ 。其中新加边的另一个端点按照正比于顶点度数的分布

$$u = \frac{d_u}{\sum_v d_v}$$

(13)

随机选取。重复以上新加点的过程足够多步所形成的网络的各顶点的度满足幂律分布 $p(k) \sim k^{-\gamma}$,见图 3。而且,指数 $\gamma=3$ 与模型的参数 m_0, m 无关。进一步的数值模拟表明,当 m 取某一范围内的随机数时,指数也不变。

鉴于实际网络的幂指数并不都是 $\gamma=3$,有的作者发展了 Barabási 和 Albert 的原始偏好依附模型,使得幂指数可以用模型参数来调整。在偏好依附模型中,顶点 v 的度值 k_v 可以认为是其吸引力的度量。那么,随着时间的推移,除去少数精品之外,大多数的顶点的吸引力会随着时间的减弱。一个考虑了顶点的历史模型^[22]可以改变幂指数 γ 。更多的可调参数模型甚至是破坏幂律分布的模型可以通过考虑更一般的网络演化过程得到。

另外一个要改动偏好依附模型的要求是偏好依附模型并不能复现所有的认为是 Scale Free 网络的静态特征。对于大多数实际 Scale Free 网络的 Small World 现象,模型给

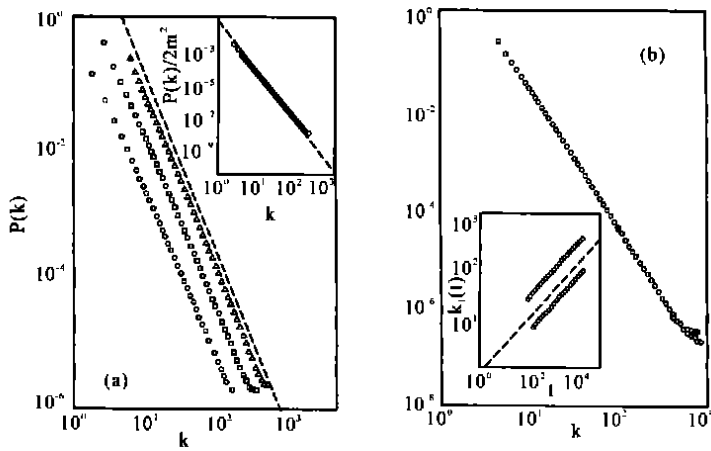


图3 偏好依附模型生成网络的度分布。(a) 度分布不随着 m 或 m_0 的改变而改变,从左到右 $m_0 = m$ 分别取值为 1, 3, 5, 7; (b) 在不同时刻, $N = 100000, 150000, 200000$, 取 $m_0 = m = 5$ 得到的度分布曲线完全重合,插入的小图是其中两个顶点的度值的演化曲线,与方程(29)相符。本图取自文献[20],感谢 Albert 提供

出了较好的结果,但是模型的度度相关性为零,这与实际网络中存在的匹配模式不相符合^[15]。更多的其他几何量的实际与模型的对比,以及在对比基础上对模型的修改工作还有很多值得做的地方。比如 Goh 和 Oh 等给出了模型上的介数的分布规律^[16],但是更广泛的实证分析仍然有待完成。这些相符的或者不相符的现象都给出了模型推广与修改的方向。

作为一个一般的网络演化的框架有哪些现象是可以进入模型来考虑的呢? 一个一般的网络演化包含五种现象: 加点、加边、重连、去边、去点。所谓加点就是 t 时刻在图 $G(t-1)$ 上加上新的顶点,并且加上若干从此顶点出发的边; 加边指的是 t 时刻在 $G(t-1)$ 原有顶点之间新加入若干连接; 去边与去点则是以上过程的逆; 而先去边后加边和起来就是重连,但是只有当加边和去边发生在同一顶点上的时候才刚好是重连,所以鉴于重连事件的几率有可能大大大于去边和加边发生几率之乘积,所以把重连独立出来。目前绝大多数网络演化模型都在这五种事件的范围内讨论。

在这样一个框架下来看,我们看到偏好依附模型只考虑了加点的行为,只是对于加点以后建立的从新加点出发的连接做了偏好性假定。从 Barabási 和 Albert 提出此模型以后,许多作者讨论了此模型的变例。Dorogovtsev, Mendes 和 Samukhin 的模型^[18]讨论了新加点加入之后, m 条新加边可以不从此点出发的模型,维持偏好性不变。模型还区分了 In 和 Out 边,并且基于 In 度提出了吸引力的概念,认为顶点 v 的吸引力就是 $A_v = A + k_v^{\text{in}}$ 。此模型也能呈现幂律的度分布,并且幂指数可调 $\gamma = 2 + \frac{A}{m}$ 。其实我们可以看到,实际上,此模型相当于考虑了加点和加边两件事情,只是由新加点引起的边不是一个确定数,内部原有点之间的新加边也不是确定数,但是仍然都维持偏好性。当我们只考虑 In 度分布的时候,可以忽略新加边 Out 端点的选择而只考虑 In 端点的偏好性。

有两个试验可以探讨加点和偏好性在模型中的不同地位。第一个试验是只考虑加

点,但是从新点出发的边的端点的选择不满足偏好性,而是随机选取。模型给出了指数衰减的分布^[17]。另外一个试验是只考虑加边。从一个非常稀疏的 N 顶点 M 边的随机网络开始,每一步就是随机选一个顶点,然后按照偏好性选择另一个顶点,如果不存在已连接的边,连之。或者是同时考虑活性偏好与连接偏好,即第一端点的选择和第二端点的选择都按照偏好依附的形式。当然,随着重复的次数增加,模型将趋向于完全网络,所以这样的模型的稳态显然不是幂律分布的。但是,是否存在着一个演化区间,在此范围内,网络呈现幂律分布呢?我们知道如果两个端点都随机选择就是随机网络模型,顶点的度符合泊松分布,那么按照偏好方式选择的随机网络模型呢?如果在一定时间段出现了幂律分布的话,那说明偏好性是其决定性因素,所谓引入新加点只是使得网络取得某种平衡,不至于发展为完全网络而已。Barabási 等人就研究了这样一个模型^[20],发现在演化的初期确实存在幂律形式的度值分布,但是随着时间的演化逐渐过渡到高斯分布。这说明偏好依附是 Scale Free 网络的核心机制,但是新加点也不可或缺。

Albert 和 Barabási 的第二个关于 Scale Free 网络的机制模型^[31]考虑了加点、加边、重连三种事件。每一时刻这三个操作分别以某一概率 $(1 - p - q, p, q)$ 发生,任何一种事件发生都遵循偏好性。理论分析与模拟的结果表明幂律分布与指数分布都可以在网络中出现,取决于 p, q 的值: $q < q_{\max}$ 为幂律形式的度分布,否则为指数形式。当然,这样的结果肯定是不完全的,因为当 $p \gg 0.5$ 的时候,加边占统治地位,必然使网络趋向于完全图。所以关于 p 也存在某一个阈值。

无向网络模型的一个比较全面的探索是由 Cooper 和 Frieze 完成的^[33],他们考虑了加边和加点两件事情,并且探讨了有无偏好性对模型的影响。其演化过程如下,在任一时刻 t ,以 $\frac{1}{N}$ 的几率加边发生在图 $G(t-1)$ 的原顶点之间,以 $1 - \frac{1}{N}$ 的几率加入新顶点。如果加入新顶点,则按照某一分布 p_i 产生一个随机数 i 表示从此顶点出发的边数,然后以 $\frac{1}{i}$ 的几率在原顶点中随机选择另一个端点,以 $1 - \frac{1}{i}$ 的几率按照偏好性选择另一端点。如果是在原顶点中加边,则按照某一分布 q_i 产生一个随机数 i 表示新加的边数。然后,以 $\frac{1}{i}$ 的几率在原顶点中均匀选择第一端点,以 $1 - \frac{1}{i}$ 的几率按照偏好性选择第一端点,接着以 $\frac{1}{i}$ 的几率在原顶点中均匀选择第二端点,以 $1 - \frac{1}{i}$ 的几率按照偏好性选择第二端点。文章对于各个参数对于模型的影响都做了理论分析,包括分布形式的显式结果或上下界分析。

再用这个框架去看引入顶点年龄历史的模型^[22],我们发现可以认为考虑活性与年龄的关系相当于在一定程度上考虑去点或去边的行为。那么,一个关键性机制突出,又能最大程度上复现相关几何性质,包含一般的网络演化五种现象的机制模型,将是无向 Scale Free 网络模型的最终目标。

从有向网络的角度来看,无向网络的偏好依附模型可以认为是只存在 In 度幂律分布,Out 度为 $(d^{\text{out}} - m)$ 分布或某种平庸随机分布的演化模型。既然实证研究表明了双向幂律网络的普遍性,那么机制模型的研究工作就需要对此做出回答。对于双向幂律分

除了以偏好依附为基础的这一大类模型之外,还存在一类特殊的确定性 Scale Free 网络——层次结构网络,见 Barabási 等人的文章^[8]。对于实证研究而言,有一个静态统计量能够很好地区分这两类模型:顶点集聚程度的分布规律 $p(c_v)$ 。

布模型,如果在单向幂律分布模型的基础上把所有的单向边都换成双向边,那自然就有了双向幂律分布。所以双向比——双向边占有所有边的比例对于有向网络来说是一个重要静态几何量。

还有一种实现双向幂律模型的方式是独立地构造 In 和 Out 边的机制,例如类比于按照偏好性建立的 In 边的反馈机制,我们也可建立 Out 边的反馈机制。基于以上想法,Tadic 建立了以下模型^[32]。从 m_0 顶点 M 边的随机网络开始,在 t 时刻,以 $\frac{A_v^{\text{in}}}{A_v^{\text{in}} + k_v^{\text{in}}}$ 的几率加入一个新的顶点,并带来从此顶点出发的 m 条边,以 $1 - \frac{A_v^{\text{in}}}{A_v^{\text{in}} + k_v^{\text{in}}}$ 的几率从原有网络的顶点内产生 m 条边。从新加顶点出发的边的终点按照偏好性选择,即

$$p^{\text{in}}(u) = \frac{A_v^{\text{in}} + k_u^{\text{in}}}{\sum_v (A_v^{\text{in}} + k_v^{\text{in}})} \quad (14)$$

而从原有顶点出发的边按照 Out 度的偏好性选择起点,即

$$p^{\text{out}}(u) = \frac{A_u^{\text{out}} + k_v^{\text{out}}}{\sum_v (A_v^{\text{out}} + k_v^{\text{out}})}, \quad (15)$$

然后再按照 $p^{\text{in}}(u)$ 来选择其终点。当然这样的模型肯定会展现双向幂律分布。

还有一个因素也可以在单向幂律的基础上产生双向幂律——In 度与 Out 度的高度相关性。考察每一个顶点的 $(d^{\text{in}}, d^{\text{out}})$ 之间的相关性,如果存在高度正相关说明,对外联系活跃(Out)的顶点也常常是获得外界联系多(In)的顶点。于是,自然也可以得到双向幂律分布了。

那么真实有向网络的双向幂律结构是否能够由以上这些因素来解释呢?这有待于实证分析来回答,尤其是关于度相关性的实证分析及其与理论模型的对比在这一点上就有重要价值。如果以上因素都不能够代表实际网络双向幂律结构的形成机制的话,那么研究者就不得不面对一个包含混合反馈机制的有向网络模型了,比如,In 边可以同时受 In 度和 Out 度的反馈来影响,以及 Out 边。

最后,我们谈谈加权网络的机制模型。由于加权网络的静态几何量与动力学性质没有得到广泛的研究,所以也就很难对其形成机制进行深入探讨了。从机制模型的角度来看,加权网络的演化既包含连接的演化又包含权重的演化,而且两者可能相互影响,看来要复杂得多了。

3 网络的演化性质

在设计网络机制模型的时候,我们主要关心静态几何量统计性质的再现,而忽略网络演化过程的统计特征,尽管机制模型设计的就是演化行为。然而,实际上,真实网络的演化性质也是可做实证分析的,尽管网络演化的时间序列包含大量的数据而且难以获得相关纪录。我们把关于实际网络演化的统计性质的研究单独列出来做为动态网络的研究,以区别于静态统计性质的研究。而两者都将为网络机制模型的研究提供启发和标准。目前,动态演化性质的研究比较多的网络是合作者网络^[67]与 WWW 网络^[25]。

5.1 时间演化性质

在 Barabási 等人关于合作者网络的动态演化性质的研究工作^[67]中,他们选择了数学和神经科学领域的研究者网络作为研究对象。首先文章以年为时间单位讨论了顶点(即研究者)数的时间演化,总文章数的时间演化。然后讨论了度分布的时间演化稳定性以及平均度值随时间的变化。接着统计了不同时间点的网络的平均最短距离和平均集聚程度,做了演化曲线。随后对网络最大集团的相对大小也做了随时演化行为的分析。例如图 4-a 所示为每一年的新加入研究者随时间变化曲线,图 4-b 为每一年发表的文章数量随时间变化曲线。

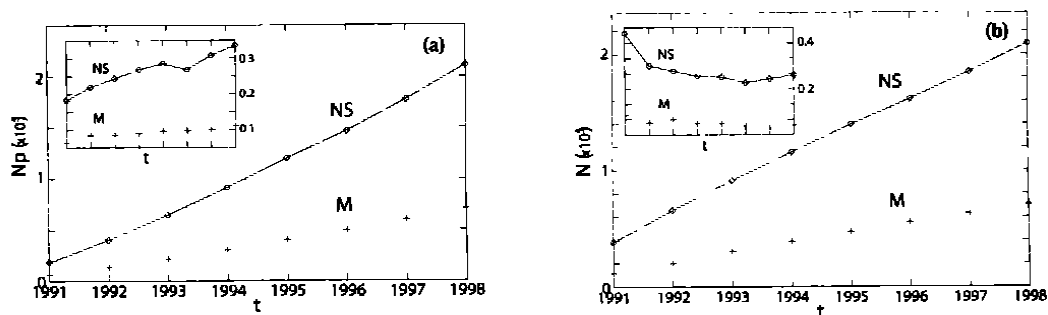


图 4 发表文章数量与科学家数量随时间的变化。(a) 数学与神经科学发表的文章数量,插图为每一年新发表的文章;(b) 两领域内研究者的数量,插图为每一年新增加的研究者。本图取自文章^[67]

除了顶点数和联接数的时间演化性质以外,网络上特定动力学过程的演化性质也应该是复杂网络研究的重要组成部分。那些与网络本身的演化密切相关的动力学过程,例如人与人之间接触网络上的传染病传播,人类社会关系网络中舆论的传播与信任的建立,其演化过程是在网络这样一个舞台上发生的,但是也会在一定程度上改变网络。这种现象的时间演化性质的研究无论对与网络本身还是动力学模型的研究都是很有好处的。目前这样的研究并不多见。从这个角度来看, SARS 的传播是一次代价高昂的非常珍贵的研究机会。此外,我们将在科学家网络中讨论的新思想的传播,也属于这类研究,可以看作是科学家网络上新思想的随机行走或有偏随机行走。

5.2 偏好性的检验

演化性质的另一个重要方面是对于演化机制模型的检验,例如对偏好依附假设的直接检验^[34]。对于 $t-1$ 时刻已经存在于网络之中的顶点 u , 定义从 $t-1$ 到 t 时刻的度值改变量为

$$k_u = k_u(t) - k_u(t-1), \quad (16)$$

它由两部分构成,新加点引起的连接和新加边引起的连接,分别记为 k_u^{new} , k_u^{old} 。如果关于新加点的偏好依附成立,那么, k_u^{new} 应该与 k_u 成正比,所以可以通过分析 (k_u^{new}, k_u) 的相关性来检验偏好依附。定义其关系为 $\langle k_u^{\text{new}} \rangle = \langle k_u \rangle \langle k \rangle$, 实证分析表明, $\langle k_u^{\text{new}} \rangle \sim \langle k \rangle$, 其中 $\langle k \rangle$ 分别为 0.8 与 0.75 有别于偏好依附的 $\langle k_u^{\text{new}} \rangle = \langle k \rangle$ 。对于由新加边带来的度值的改变,同

样可以通过实证分析来检验活性偏好与连接偏好的假设。首先我们假设活性偏好与连接偏好形式一致,即

$$\begin{cases} p^{\text{out}}(u) = \frac{k_u^\mu}{\sum_v k_v^\mu} \\ p^{\text{in}}(u) = \frac{k_u^\mu}{\sum_v k_v^\mu} \end{cases} \quad (17)$$

那么一条新的边出现在 (k_1, k_2) 之间的几率为

$$(k_1, k_2) \sim k_1^\mu k_2^\mu, \quad (18)$$

因此,统计由于新加边引起的度值改变量 $(k_1, k_2) = k_1^{\text{old}} + k_2^{\text{old}}$, 然后分析其与 (k_1, k_2) 的相关性就可以检验偏好性了。对于活性无偏好,连接有偏好的检验类似。

对于有向网络,这样的动态过程相关性的分析还可以做得更细,用来检验活性偏好与 In 度和 Out 度的关系,及其连接偏好与 In 度和 Out 度的关系。具体方法是把以上分析中的 k_u 换成 $k_u^{\text{in}}, k_u^{\text{out}}$ 。这样的实证分析可以在 WWW 网络中实现。

关于 WWW 网络的演化性质,目前已有顶点数目的随时演化,边的增加与更新等统计结果^[25]。发现网页更新的平均周期大约是 50 天,而且,COM 网页的更新速度要比其他域名网页的更新速度快很多。但是,这些演化行为的统计都不是专门针对网络结构的。对于网络结构而言,最重要的更新是网页上的链接的改变。这样的实证分析还有待进一步的工作。至于直接检验偏好依附假设的实证工作则更要等到收集足够多的网络演化数据之后了。但是不管如何,以上网络演化统计性质的分析将大大推进网络机制模型的研究。

4 网络的动力学性质

网络动力学性质的基本研究对象是动力学模型在不同网络上的性质与相应网络的静态统计性质的联系,包括已知和未知的静态几何量。如果我们发现了某个模型在某一网络上有某种特殊的表现,那么可以认为是这一网络的某种特征影响了这个模型的表现。这种特征有可能是已经得到研究的这种网络的几何特性,也有可能是没有被发现的几何特征,那么前者将印证网络上这些几何量的重要性,而后者将会推动网络本身研究的发展。当然,大多数网络上的模型工作是体现了某种已有几何量对模型的影响。在一定程度上说,这是一种关于网络的结构与功能关系的研究。关于规则网络上物理模型的研究可参见综述文章^[35],在此我们主要关心其他几类网络上的物理模型及其特殊的动力学性质。

4.1 网络上的物理模型

在 Watts 和 Strogatz 的文章^[7]中,在介绍了几何量之后,研究了一个简单传染病模型。在此模型中,接触传染率为 1,感染的顶点在一个单位时间以后退出系统。对于任何网络,这样的传染病都将在整个网络扩散,研究其扩散时间,发现对于从规则网络到随机

网络的所有 $p \in (0, 1)$ 网络, 其扩散时间刚好与最短路径一致。也就是说, 在规则网络上传播所需时间长, 但是只要 p 略大于 0, 就会得到迅速传播。这很好地说明了最短路径这一几何量的作用。其实, 从这个传染病模型来看, 任一顶点都同时向所有近邻传播, 当然所需时间就是网络的平均最短距离了。在此工作之后, 许多耦合微分方程系统引入 Small World 网络, 以研究驰豫过程、同步等现象。文章^[14]研究了 Small World 网络上的耦合 Hodgkin-Huxley 方程, 发现模型在 Small World 网络上同时具有驰豫时间短、共振性好的特征, 分别来源于网络的小最短距离和高集聚程度。Newman 等人研究了 Small World 网络上的渗流与传染病模型^[36]。在这个传染病模型中, 作者引入了可感染者比例 q 与 [7] 中的模型相区别。在规则网络上的研究结果表明, q 存在一个阈值, 只有 $q > q_c$ 传染病才能大范围流行。在 Small World 网络上的研究表明, q_c 仍然存在, 但是比规则网络小得多。而对于 Scale Free 网络上传染病模型的研究结果是, 对于 SIS 或 SIR 模型, 不存在类似的阈值^[2, 37, 38], 也就是说只要传染病发生, 就将大范围流传开来。因此, 对于这种网络上的传染病的控制就不能完全依赖于医疗卫生水平的提高, 而需要从改变网络结构入手。破坏网络结构的典型方法是隔离, 即强行断开某些联接, 或去掉某些顶点。关于去边与去点的方式的研究称为对网络的攻击, 所以, 网络的不同攻击方式下的结构稳定性是与网络上的物理模型紧密联系在一起的。

除了这种与网络结构、网络演化密切相关的模型以外, 很多作者研究了网络上的传统统计物理模型, 并与规则网络和分形网络上的结果进行了对比。例如网络上的 Ising 模型^[60, 61], XY 模型^[62], 及其临界现象^[63]。在 Small World 网络上的 Ising 模型研究发现, 对于以一维链、二维面和三维立方体为基础通过重连得到的不同的 Small World 网络, 相变点、比热有不同的规律。也就是说, 维数仍然是网络上动力学模型的一个重要参数。但是, 对于一般的拓扑网络, 维数并不是一个很显然的几何量。那么一般网络上的动力学模型是否也会受与维数对应的网络的某种宏观性质影响呢? 如果存在这样的类似宏观性质, 那么这种一般网络上的维数的几何定义也是一个有待解决的问题。

4.2 网络的容错与抗攻击能力

从 Small World 网络模型可以发现, 之所以它具有独特的几何性质, 并由此带来 Small World 网络上动力学模型的特殊性质, 完全是因为加入了极少量的长程联接的缘故。因此, 对于 Small World 网络, 只要能够识别这些长程联接, 然后针对长程联接进行攻击就可以在很大程度上改变 Small World 网络的结构。所以对于 Small World 网络的攻击的关键问题是长程联接的识别。这并不是一个平庸的问题。对于在规则网络的基础上改造而得的 Small World 网络, 直观上我们就能够识别长程联接。但是如果把点的顺序打乱重新再做一次图, 那长程联接与短程联接就不好区分了。恰恰当我们面对实际网络的时候, 这种关于顶点的先验顺序实际上就是不存在的。Pandit 和 Amritkar^[39]最先研究了这一问题, 提出了一种基于各阶路径的边的阶数的定义, 然后由边的阶数来反映其长程还是短程联接。但是这种识别方法不具有很好的区分能力。有一个很好的几何量——边的介数——可以很好地区别长程与短程边。边的介数的含义是所有最短路径中必须经过此边的次数。长程联接一般联接着两个局域集团, 因此不同的集团的两点之间的最短路径就会经

过这条长程联接。而不同集团之间的顶点对的数目 $C_{n_1+n_2}^2$ 远远大于集团内部顶点对数目之和 $C_{n_1}^2 + C_{n_2}^2$, 其中 n_1, n_2 为两集团顶点数。因此可以把边的介数做为判断长程联接的一种方法。Motter 等人的文章^[40]就研究了基于边的介数的攻击方式,发现 Small World 网络对于这样的攻击方式非常敏感。对于 Small World 网络,可以预见,所谓结构发生很大改变就是平均最短距离变长了,平均集聚程度变小了。因此,网络结构稳定性的研究,一般地从以下四个方面随着攻击的变化来讨论:平均最短距离,平均集聚程度,最大集团的相对大小以及集团规模分布模式。有的文章还用特征路径长度代替平均最短距离来讨论,因为平均最短距离随着攻击幅度的增加而增加,与效率的通常意义相反。特征路径长度的定义如下,

$$l^{-1} = \frac{2}{N(N-1)} \sum_{ij} \frac{1}{d_{ij}} \quad (19)$$

对一般网络的攻击方式可以选择去点与去边两种方式,从选择的方式上分为随机攻击和选择性攻击两种类型^[41],分别称为网络的容错能力与抗攻击能力。研究表明^[41,42], Scale Free 网络具有很强的容错性,但是对于基于顶点的度值或介数的选择性攻击抗攻击能力较差,对于基于边的介数的攻击也非常敏感。那么,这些选择性攻击之间是否等价,有关联吗?如果按照顶点的度值所去掉的点与按照顶点的介数所去掉的点之间存在很强的共同性的话,我们就说其实这两种攻击等价。还可以进一步研究基于边的介数的去边攻击与前述去点攻击之间的关联性。当我们考虑这几个关联性的时候,其实,已经从网络的结构稳定性回到了网络的静态几何性质。我们可以按照度值对所有顶点排序,然后我们可以再按照顶点的介数排序,如果两个序列之间存在很强的关联性则认为两种方式等价。可见,网络的结构稳定性与网络的静态几何性质也是紧密联系在一起的。文献^[42]不仅讨论了网络的最短距离等几何性质在去边去点攻击下的改变,还讨论了度值与介数等几何量的相关性。

实际网络的结构稳定性分析的一个重要实例是食物链网络。网络结构稳定性的理论研究表明,对于规则网络与随机网络等同质网络(每个顶点都相同),随机攻击与选择性攻击的效果相当,Small World 网络对于长程联接攻击非常敏感,Scale Free 网络对于随机攻击具有很强的鲁棒性,但是基于度值或介数的选择性攻击可以造成很大的结构性破坏。那么对于食物链网络呢?由于规模限制,食物链网络的度分布没有得到最终的研究结果,存在幂律分布、均匀分布、以及指数分布等多种形式。在这样的网络上,如果发生个别物种的衰落或死亡会对整体生态系统产生什么影响呢?Dunne 等人的文章^[56]对这个问题进行了研究。文章讨论了 16 个不同的食物链网络对随机攻击和基于度值的选择性攻击的响应。定义引起 50% 以上物种灭绝的去点率做为网络鲁棒性的度量。发现尽管网络的结构有所不同,但是网络的鲁棒性与网络联接的紧密程度具有很强的正向关联。但是这样的网络是否比同样顶点与边数的随机网络或 Scale Free 网络具有更强的鲁棒性呢?还有,食物链网络是有向网络,以度值为基础的攻击方法不足以反映有向网络的特征,是否可以试一试基于顶点介数的攻击方式呢?

网络结构稳定性问题另一个有意义的实例是 SARS 传染病的控制。我们认为 SARS 传染病是在居民基本活动网络上传播的疾病。近距离接触传染是其主要传播方式。因此

我们提出了以楼为顶点,人员流动为联接的居民基本活动网络。例如我的办公楼与家庭住所是网络中的两个顶点,因为我的存在,这两个顶点之间建立了联接。这样一个网络应该包含居民的工作或学习、必需品与必要服务的获得等几个方面。有了这样一个网络,加上关于传染病传播的更详细的调查,我们就可以以网络攻击为基础来讨论对传染病的控制。

首先,我们需要研究网络的基本结构特征,比如平均最短距离与平均集聚程度的特征,度分布等静态几何性质。然后,针对网络的特点与传染病传播的特点制订预先切断某些联接的方案,隔离某些顶点,改变网络性质。我们认为,传染病的隔离控制包含两个方面,第一是对得病人员及其接触者的隔离,第二是对网络结构的预先隔离,也就是对健康人的工作、学习、生活进行调整,例如在发现病人之前就有重点地隔离某些区域,限定某些人在家办公等等。由于条件所限,目前我们没有开展关于这个网络的数据调查工作。Scale Free 网络的攻击与网络上传染病的控制相结合的模式研究正在进行中,我们的目标是找到 Scale Free 网络上的一种攻击方法,使得大规模迅速传播的传染病转为小规模缓慢传播。

鉴于目前 SARS 在全世界范围,尤其在中国部分省市的流行,研究这样一个问题就更有紧迫性和实际意义了。实际网络研究还发现,人类朋友关系网络^[7],人类性伙伴网络^[46]正是一个有 Small World 特征的 Scale Free 网络。在这样的网络上采用什么方式来攻击,才能使传染病的流行得到控制,目前还有一个有待研究的问题^[64]。

网络的效率是网络结构与功能关系的另一个重要组成部分。河流与血管的分支网络做为水流与血液的传输网络,其网络结构与传输效率的关系在分形几何的研究中已经得到了深入的研究^[58,59],发现了普适的分支行为标度律(Allometric Scaling)。数学家从理论上证明了,镶嵌在 d 维空间的树形传输网络,其最优效率下的分支系数为 $\frac{d+1}{c}$ 。食物链网络可以认为是能量与物质流的传输网络,Garlaschelli 等人^[57]研究了这一网络的结构与效率关系。首先从食物链网络中产生一个最小生成树。对于树中的每一个顶点,定义包括直接与间接的上级顶点数,即直接与间接捕食者数量 A_i ,定义顶点的耗散量 $C_i = \sum_{k>i} A_k$,其中 $k>i$ 的含义是物种 i 的上级顶点。研究 C_i 与 A_i 的关系发现,两者成幂律关系

$$C \sim A^{\frac{d+1}{d}} \quad (20)$$

而且,不同的食物链网络有相同的指数 $\frac{d+1}{d} = 1.13 \pm 0.03$ 。对于河流网络 $\frac{d+1}{2} = 1.5$, 血管网络 $\frac{3+1}{3} = 1.33$ 。这说明,不同的食物链网络遵循相同的优化规则。但是 1.13 不能表达为整数维空间中的分支优化条件 $\frac{d+1}{d}$ 。这是因为食物链网络是一个一般的拓扑网络,不能镶嵌在某一个整数维欧氏空间中。但是由此也引出一个问题,对于一般的网络,是否存在类似维数的宏观几何性质。这个问题在我们讨论 Small World 网络上的物理模型的时候出现过。如果有这种决定网络效率与模型动力学性质的宏观几何性质存在,那么它在一般的网络中如何定义呢?

网络结构与功能关系的另一个重要方面是利用网络独特的表现力发掘那些只能通过网络这个工具加以研究的对象。网络上交流集团的类聚分析就是其中之一。网络上的顶点之间是存在某些交流集团的。例如 WWW 网络中,存在某一个领域内的网页相互联合的现象,科学家网络中存在团队现象。这些联合或团队内部顶点之间的联接会比团队之间的联接更加频繁。那么,只要这样的几何结构存在,我们就可以通过结构上的分析来发现这些团队,而不需要依靠内容。关于团队所对应的几何结构有两种方式来描述。第一,团队内部成员之间的集聚程度会比团队之间的高。第二,不同团队的顶点之间的最短路径一般需要经过长程联接。联接赋权的层次类聚方法^[61](Hierarchical Clustering)基于第一种描述,而 Newman^[73]等人提出的基于边的介数的类聚方法的前提是第二种描述。联接赋权方法的基本思想是,找到任意两个顶点之间的所有路径,包括最短路径,以及比最短路径长的其他路径,有时候可以只取到比最短路径长 3 个顶点以内的路径,记下长度为 l 路径的条数 n_l 。对于属于同一个集团内的两个顶点,存在各阶路径的条数显然要比不同集团的两个顶点之间要多。因此定义亲密程度

$$S = n_l^{-1} \quad (21)$$

其中, S 为 $(0, 1)$ 之间某一常数。然后按照亲密程度做类聚分析。基于介数的类聚方法是计算每条边的介数,去掉介数最大的边,然后再计算剩下边的介数,再去掉介数最大的边,直到没有边,记下所有的去边的过程,反过来就是各个顶点组成集团的顺序。对于实际系统,有时候两种方法给出相同的结果,有时候不同。因为他们基于不同的假设,不能说哪个更好一些。除了以上包含所有路径的赋权方法以外,路径赋权方法还存在只考虑独立路径、或者只考虑自回避路径的赋权方法,这些方法与介数分类方法的区别与联系还有待研究。像这样的网络上独特的分析方法及其应用也是网络研究的重要组成部分。

5 广泛的前景:科学家网络和产品生产关系网络

在介绍了网络静态几何量,加点、加边、去点、去边的等复杂网络研究的一般问题之后,我们来看看网络几何学在实际系统中的应用——网络研究中最丰富的部分。统计物理学研究网络的最终目标就是从实际系统抽象出一般方法,然后应用于更加广泛的实际系统。因此,作为统计物理学研究对象的网络与实际网络的研究是相辅相成的。网络的顶点与边的关系可以在多种实际问题中加以具体化,目前已得到研究的网络包括语言学^[48]、生物学^[47, 50]、生态学^[52]、通信^[43, 45]、电力^[45]、社会学^[71]等各个领域的具体关系所形成的网络。在此我们准备对这些研究加以总结,仅以目前我们自己正在开展的工作为例,粗线条地介绍一下把网络应用于实际系统研究的思路与方法。

首先,我们介绍一下科学家网络。在第 1.3 节中我们提出,可以通过建立一个某一领域内的科学家网络来研究这一领域的发展变化,讨论某一个新的思想在此领域内的产生、传播。在实际工作中,我们选取经济物理学领域的科学家网络作为我们的研究对象。

经济物理学是近年迅速发展起来的一个交叉学科,它把统计物理学和复杂性研究的方法用于研究包括金融和企业规模等经济学问题^[69, 70]。在这之间涌现了大量新的思想与新的处理方法。对于这样一个新兴学科的发展,我们拥有比较完备的文献纪录,能够得

到整个网络的发展演化的图像。在这个网络中我们考虑科学家之间通过合作、引文、通信或讨论的方式形成的相互影响,讨论思想与创见在这个网络上的诞生与传播等问题。当然,除了对于这些关于具体学科的问题的研究之外,通过这个网络我们能够考察加权有向网络的一般问题,包括静态几何量、网络演化行为和网络机制模型三个部分。

第一个问题是数据的收集。SIS 等大多数数据库提供作者、引文、摘要等信息,但是有关致谢的内容只有通过阅读原文来获取,所以一个大规模的数据库的建立将包含很大的工作量。初步的收集工作我们已经完成。对于这一网络我们可以研究的静态统计量包括:度分布(包含 In 和 Out 度),权的分布,度的相关性,权的相关性,度与权的相关性,最短距离的统计性质及其与相应的无权网络的对比,权的加入引起的顶点与边的介数及其统计性质的改变,加权网络的集聚程度的定义与性质。能在此网络上加以研究的演化性质包括:顶点数目、顶点的权与度值的时间演化,加点与加边行为中偏好性假设的检验,此时偏好反馈的馈入量可以是度值也可以是权值。

此外,科学家网络上可以做一个独特的演化行为的统计分析工作——某一个思想(往往体现为某篇重要文章)在此网络上的传播行为的统计特征。以某篇文章的发表与被引用的时间序列作为研究对象,分析其影响力随着时间扩散的统计性质,就好像是对网络上的随机行走或传染病模型一样加以研究。并且,由于引文本身属于构成网络关系的一部分,这种既可以认为是网络上的模型又可以认为是网络本身演化过程的行为的统计性质的研究具有特殊的意义。

最后,是否可以发展一个加权网络的演化机制模型用来描述一个学术领域的发展变化呢?权的增加由网络中已经存在的边的重复出现而得到,而这种重复出现可能受到度的增加的影响。我们仍然可以用加边加点作为机制模型的研究对象,但是此时的加边加点将同时受到顶点的权与度的影响。

其次,我们介绍一下产品生产关系网络。正象把食物链中的捕食关系抽象为网络关系^[52~55]所形成的网络可以用来探讨生物圈中的能量物质流问题,了解网络的结构与抗干扰能力之间的关系,并促进环境保护与生态平衡的研究一样,把经济关系中的产品生产关系抽象为网络关系就可以研究经济学中的物流关系,从而技术的发展可以在网络上体现为新的顶点的创造或者新的边的形成。资源与物品在生产关系中的地位将体现为网络上顶点介数的分布特征,因而此网络对于关键资源的保护、储存与相应的策略的研究都将提供重要的信息。同时,边的介数体现的是某项技术或某种生产的重要程度,这对于发现并保障这种关键技术具有重要意义。把边的介数或者赋权路径用于分类所得到的生产部门的区分可以与实际生产部门对比。

从网络研究本身而言,生产关系网络包含了一些目前研究的网络所没有的新的特征,例如,多对一关系,即只有同时具有产品 a 和产品 b 才能生产产品 c ,所以将来最短路径就不足以描述生产链的相关问题了,那就要促使网络研究去提出新的几何量;再比如替代关系,即在 a 原料不足的情况下,我们可以换一套生产方案或者以 a 代替,这种行为在网络上会体现为什么结构都是有待研究的问题。对于这些新的结构我们已经有了初步的描述方法:引入每一条边的 Out 权,可以描述多对一关系;引入副本,即顶点 u , u 表示一种产品,所有的 Out 关系都相同,但是可以接受两种方式的 In 关系。对于实际网络这样的

描述是否可行,甚至是否有新的结构需要描述都有待于实际数据的研究。

我们很高兴地发现,在我们认识到这一重要问题并提出理论方案的同时,这种产品投入产出关系(PIOTs)的实际数据已经有研究者在统计和分析了^[72]。只是非常可惜的是在我们提出的一个网络中,顶点是一种产品而不是一个 POITs 中的生产部门,这可能需要改变一下产品投入产出表的统计方式。希望在不久的将来,网络的研究方法能够被引入到这一研究工作中去,也许会发现更多新的网络几何特征或者发现对于生产关系相关问题研究的新的思路。

6 解析与数值计算方法

网络本身是个多个体复杂系统,研究网络的工作也是个复杂系统(笑话),因此为了使读者有一个更好的全局概念,我们在前文中除了定义,基本上不列数学公式。但是有些结论没有经过解析推导或数值计算是不能让人信服的,所以在这一节中,我们补上这些技术工作。尽管这样分开处理的方式增加了我们整理此文的难度,但是我们希望这能使大部分只是想了解一下这个领域的读者更加轻松,能使有兴趣投身这个领域的研究者有更好的全局观。

尽管大部分工作可以通过计算机生成网络进行研究,但是通过系综理论、平均场近似、生成函数方法和重整化群等统计物理学的常用技术我们也可以解析地得到网络性质的近似表达式。这一节我们就从这两个角度介绍部分经常使用的技术,更加全面的了解请参看文献^[1,2]以及下文中每一个方面的具体文章。我们先介绍网络的计算机生成与存储,以及计算某些几何量的算法,然后用系综理论的框架整理了随机网络的研究方法,最后介绍偏好依附模型的解析处理。

6.1 图的表示与算法

图有两种表示方法:邻接矩阵与邻接表。对于无权图 $G(V, E)$ 邻接矩阵是 $N \times N$ 的矩阵,行号与列号代表顶点,如果顶点 i 与顶点 j 相连,则相应矩阵元为 1,否则为零。也就是矩阵 $(^{uv})_{N \times N}$ 。加权网络的邻接矩阵为 $(^{uv}W_{uv})_{N \times N}$ 。邻接表是 N 个元素构成的数组,每一个元素代表相应顶点的近邻。也就是把邻接矩阵分为一行一行表示,对于第 i 行,列出所有 $^{iv} > 0$ 的顶点。对于加权网络,列出顶点编号之后,还要带上权重。对于稀疏图,利用邻接表的方式可以节约大量的存储空间。

对于实际网络,只要把顶点与边的信息利用邻接矩阵或邻接表存储起来就可以供下一步研究使用了。对于模型网络,例如 Scale Free 网络,就必须先按照网络模型来生成了。生成 Scale Free 网络的方法有两种:机制演化与 Monte Carlo 模拟。机制演化就是按照偏好依附模型以 m_0 个顶点的完全图为初始值,按照偏好的方式进行加边加点,在足够长时间以后形成稳定度值分布的网络;Monte Carlo 模拟的方法是按照幂律分布选取 N 个随机数,并赋值给每一个顶点,然后按照某种方式在各自度值的约束下连接这些顶点,也可以得到度值符合幂律分布的网络。例如,度值大的顶点优先选择度值小的顶点来连接,直到用掉所有的度值,然后第二大的顶点重复这一过程;也可以让度值大的顶点优先

选择度值大的顶点来连结;也可以在所有顶点之间随机连接,如果出现联接数大于所要求的度值的情况就断开重新连接。不管这种连结优先方式如何,生成的网络其度分布都是幂律形式的,只会改变顶点之间的匹配模式。关于匹配模式的研究表明,偏好依附模型生成的网络其顶点的度值之间基本上没有关联,也就是与随机连结方式类似。

在网络的静态几何量之中,度值、集聚程度都是局域性质,可以在网络表示中简单地求得。而最短距离、边与顶点的介数都与网络中任意两点的最短路径相关。因此,解决最短路径问题是网络上几何量算法的关键。在无向网络中,这可以通过广度优先搜索实现。广度优先搜索是图的常用的一种算法,给每一个顶点指定一个标记,开始全为零,从某一个顶点开始,访问其所有的近邻,访问过之后,改变其标记值为1;对于被访问到的近邻,同样先访问各自的近邻,并改变未被访问过的顶点(其标记值为0)的标记值,直到所有的顶点都被访问。利用这一算法就可以研究从某一个顶点出发的所有最短距离。至于所有顶点之间的最短距离可以简单地重复上述单源最短距离的算法,也可以在此基础上略加修改^[74]。但是对于稀疏网络改进不大。

6.2 随机网络的几何性质

图论与概率论是网络研究的重要工具。在随机网络模型一节,我们已经了解,包含 n 个顶点的随机网络有两种 $G(n, M)$ 与 $G(n, p)$ 。现在我们用这两种网络分别构造两个概率空间。记 $N = C_n^2$, $G(n, M)$ 包含了 C_N^M 个可能的网络,假设所有的网络等几率,则 $G(n, M)$ 成为概率空间。给定网络 $H \in G(n, M)$ 的几率为

$$P(G_M = H) = \left(C_N^M \right)^{-1}. \quad (22)$$

对于随机网络 $G(n, p)$,实际上它包含了 2^N 个所有可能的网络,从空图到完全图。记 $e(H)$ 为网络 H 的边数,给定网络 $H \in G(n, p)$ 的几率为

$$P(G_p = H) = p^{e(H)} (1 - p)^{N - e(H)}. \quad (23)$$

从统计物理学系综理论的角度我们可以这样来看概率空间 $G(n, p)$ 。我们把 $e(H)$ 看作图 H 的宏观量,类比于物理系统微观状态的能量,概率空间 $G(n, p)$ 关于 e 的分布为

$$P(e) = C_N^e p^e (1 - p)^{N - e}. \quad (24)$$

而边为 e 的微观图也构成概率空间 $G(n, e)$,其中每一个图都等几率。所以随机图 $G(n, p)$ 是正则系综, $G(n, e)$ 则可以认为是微正则系综。

在组合系数 C_N^e 上,图论与统计物理学有不同的处理方式。图论定义了图的同态关系——如果图 $V(G_1)$ 与 $V(G_2)$ 之间存在可逆映射 $\phi: V(G_1) \rightarrow V(G_2)$,使得对于 $\forall x, y \in V(G_1)$,如果 $xy \in E(G_1)$ 则 $\phi(x)\phi(y) \in E(G_2)$ 。所以给定图 H 相当于给定了 H 以及 H 同态的所有图。从统计物理学来看,这相当于顶点不可区分的情况,这时候网络结构成为不同网络的唯一区别,同态图就如同微观状态的简并度。但是对于顶点可区分的网络,在网络结构相同的情况下,相同的连结方式连在不同顶点上仍然有值得区分的意义。由于目前研究的实际系统顶点都是可以区分的,我们不考虑网络的同态。

当我们说随机图 $G(n, p)$ 具有性质 Q 指的是在随机图的概率空间中,从平均意义上有性质 Q 。现在我们讨论随机网络顶点的度值分布。我们首先对于每一个微观图统计其顶点度值分布规律,然后按照不同微观图出现的几率做一个平均,就可以解决这个问题

题。但是,即使对于微正则系综 $G(n, e)$ 中的样本点也存在不同的度值分布形式,因而实际上实现这一方法并不简单。现在我们换一个近似的方法。由正则系综分布函数(24)式可见,分布函数的最大值点是边数的平均值 $\bar{e} = N_p$, 且二项分布衰减的非常快,因此,我们可以尝试在分布函数峰值附近的微观图上最近似的平均。这样,问题就变成微正则系综随机图 $G(n, N_p)$ 的度值分布形式。

重新表述问题如下:在 n 个顶点的网络中等几率地选取 $1/2 n(n-1)p$ 条边,求平均意义上顶点被选取次数的分布形式。首先,我们来看某一个顶点被选取的几率。在第一条边的选取的时候,每一个顶点具有 $(n-1)$ 条可以被选取的边,被选取的几率相同,又因为每选取一条边产生两个顶点,所以任一顶点被选取的几率 $= 2/n$ 。假设 $p \ll 1$, 则被选取的边并不影响与其关联的顶点再次被选取的几率。因此几率为 p 的顶点被选取的近独立事件构成概率空间,其分布符合平均值为 $\bar{k} = n \times 1/2 n(n-1)p = (n-1)p \simeq np$ 的泊松分布,即

$$P(k) = e^{-np} \frac{(np)^k}{k!}. \quad (25)$$

一个更有物理学风格的方法^[1]是计算在随机图 $G(n, p)$ 中,一个顶点可能有的所有联接数,尽管此联接数不一定是某个网络中它的度值的含义,

$$P(k_i = k) = C_{n-1}^k p^k (1-p)^{n-1-k}. \quad (26)$$

对所有的 k_i 作频数统计,当 n 很大时,上述二项分布可以用(25)式泊松分布代替。

当然,解决度分布问题更加直截了当的方式是产生足够多的随机图样本点,然后进行统计。按照二项分布的尖峰特征,只要少量的样本就可以得到比较准确的网络性质。见图5,图中只用了一个微观图来计算。当我们实际拿到一个网络的数据进行分析时,一般我们直接在这个网络上统计性质 Q 是否存在,很少需要大量的同质网络组成系综来讨论。这与统计物理中实际测量一个系统是一样的。我们的理论是建立在宏观状态与微观状态的关系的基础上的,但是对于实际测量,我们也只需要一个系统就可以了。

随机网络研究的另一个重要方面是与统计物理学联系在一起的相变现象,这也是随机网络得到充分研究的原因之一。随机网络的相变现象表现为几何性质随着参数 p 的突现性,存在很多的几何性质 Q 只有当 $p > p_c^Q$ 才会出现,展现出类似于二级相变的行。例如我们讨论网络最大集团的相对大小。图的一个连通子图称为图的集团,其中顶点数最多的称为最大集团。所研究的问题是 p 值对网络最大集团相对于整个网络的相对大小的影响。记最大集团大小为 S , 其相对大小为 $s = S/n$ 。可以看到,如果 $p \rightarrow 1$ 则,最大集团与整个网络大小相当,而 $p \rightarrow 0$ 则相反。那么, $0 < p < 1$ 的情况呢? 这在随机网络的研究中被称为网络的演化性质。

研究这些问题要用到随机图演化序列构成的概率空间 \tilde{G}^n , 定义为从空图到完全图的随机图序列 $\tilde{G} := G(n, 0), G(n, 1), \dots, G(n, N)$ 组成的等

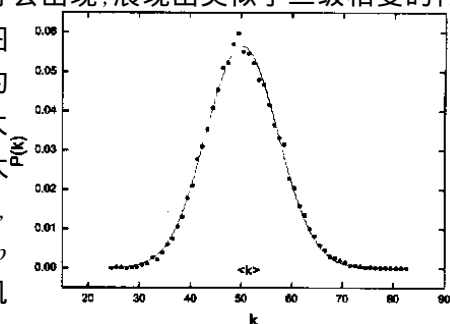


图5 随机网络的度分布。图上各点来自于随机生成的一个随机网络,其中 $N = 20000$, $p = 0.0025$; 连线为用于拟合的泊松分布

几率空间,其中 $G(n, t)$ 表示 t 时刻的图,也表示有 t 条边的图。当 $n \rightarrow \infty$ 时这个空间可以看成是由各个 $G(n, M)$ 空间中组合而成,或者看成由不同 p 值的 $G(n, p)$ 空间组合而成的空间。关于 s 的问题就可以在这样的概率空间中得到研究。结果表明对于最大集团,当 $p < p_c^s = 1/n$ 时不存在大规模集团,当 $p = p_c^s = 1/n$ 时,最大集团大小为 $S \sim n^{\frac{2}{3}}$,当 $p > p_c^s = 1/n$ 时最大集团大小 $S \sim n$ 。

熟悉统计物理渗流理论的话就会很容易发现这种现象与渗流理论的联系。实际上他们本来就是同一件事情,只不过从不同的角度进行研究。例如我们知道,在渗流问题中,在 $p = p_c$ 附近,各连通集团的规模展现幂律分布。同样,在随机网络中,也存在这种幂律形式的规模分布。

6.3 Scale Free 网络的偏好依附模型

对偏好依附模型的一般处理方法包含连续性方法^[17,20],速率方程^[19], Master 方程^[18]与生成函数方法^[9],这也是统计物理学经常使用的方法。这一节,我们围绕偏好依附模型的幂律分布形式介绍连续性方法,速率方程与 Master 方程在复杂网络研究中的应用。

考虑 Scale Free 网络的偏好依附模型,从顶点的度值变化的角度,假设其度值连续,我们有如下方程,

$$\frac{\partial k_i}{\partial t} = m(k_i) = m \frac{k_i}{k_j} \quad (27)$$

每一步我们加入 m 条边,即增加 $2m$ 个度值,于是 $k_j = 2mt$, 则

$$\frac{\partial k_i}{\partial t} = \frac{k_i}{2t} \quad (28)$$

顶点 k_i 是在 t_i 时刻进入系统的,且满足初始条件 $k_i(t_i) = m$, 于是

$$k_i(t) = m \left(\frac{t}{t_i} \right)^{\frac{1}{2}} \quad (29)$$

利用上式,结合 t_i 的分布,就可以知道任意 t 时刻 k_i 的分布,即

$$P(k_i(t) < k) = P\left(t_i > \frac{m^2 t}{k^2}\right). \quad (30)$$

在 t 时刻, $t_i \in [0, t]$ 的分布函数为

$$P(t_i) = \frac{1}{m_0 + t}. \quad (31)$$

结合以上两式就可以得到 k_i 的分布,

$$P(k) = \frac{\partial P(k_i(t) < k)}{\partial k} = \frac{2m^2 t}{m_0 + t} \frac{1}{k^3} \sim 2m^2 k^{-3} (t \rightarrow \infty). \quad (32)$$

方程(27)是一种平均意义上的描述,实际随机过程中所加入的边只是按照 (k_i) 的分布来选择,并不能保证确实让每一个顶点的度值增加那么大,也正是因此称为连续性方法,所以对实际随机过程分布函数演化的描述需要用到其他方法。引入分布函数 $p(k; t_i, t)$,表示 t_i 时刻引入系统的顶点在 t 时刻度值为 k 的几率,其初始条件为

$p(k; t_i, t_i) = (k - m)$, 满足如下 Master 方程,

$$p(k; t_i, t + 1) = p(k; t_i, t) + W^{get}(k - 1) p(k - 1; t_i, t) - W^{get}(k) p(k; t_i, t) \quad (33)$$

其中 $W^{get}(k) = \frac{k}{2t}$ 。所有顶点的度值分布定义为所有不同时刻加入系统的顶点的平均行为, 即

$$P(k) = \lim_{t \rightarrow \infty} \frac{t_i p(k; t_i, t)}{t}. \quad (34)$$

由以上两式可以求得度分布

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \sim k^{-3}. \quad (35)$$

速率方程从另一个分布空间描述。 t 时刻度值为 k 的顶点组成一个集团, 集团大小为 $N_k(t)$ 。当新顶点进入系统时, N_k 可能发生改变, 其变化率满足如下方程,

$$\frac{dN_k(t)}{dt} = m W^{get}(k - 1) - m W^{get}(k) + k, m. \quad (36)$$

其中 $W^{get}(k)$ 为所有度值为 k 的顶点获得新的联接的几率,

$$W^{get}(k) = \frac{kN_k(t)}{dN_d(t)}. \quad (37)$$

N_k 同样满足方程 $kN_k(t) = 2mt$, 且

$$P(k) = \lim_{t \rightarrow \infty} \frac{N_k(t)}{dN_d(t)}. \quad (38)$$

结合一上方程可以解得度分布 $P(k)$, 具体计算见文献[19]。

7 结束语

本文介绍了复杂网络研究的现状与前景, 尤其突出了复杂网络研究与统计物理学的关系。2002 年 Rev. Mod. Phys. 的综述文章^[1]在历史、基础与前沿等方面都写的非常之好 (已经在 SIS 被评为有突出影响的文章之一), 涵盖了比较全面的研究工作。本文在更多的程度上强调复杂网络研究的方向, 以及各个方向之间的联系。基于这种联系, 在总结整理目前复杂网络研究工作的基础上, 我们提出了以下有待研究的新问题: 无向网络邻接矩阵最大本征值所对应的度分布, 有向网络双向比、顶点的 In-Out 度相关、顶点的四种匹配模式, 加权网络权的关联、度权相关、单位权、类聚分析, 双向幂律网络机制模型, 一般网络的维数, 以及网络上的传染病控制与针对 SARS 的居民基本活动网络研究、经济物理学领域的科学家网络、产品生产关系网络等研究计划。希望我们这些新的问题与研究计划的提出有助于推进网络研究的发展。但是, 由于我们自身工作仅限于其中的某些方面, 我们不可能全面地了解所有的研究状况, 行文当中的详略取舍也受我们自身兴趣的影响。而且, 现在几乎每一期的 PRE 或者 PRL 都会刊登新的研究结果, 应该说很多其他方面的工作, 其他作者的贡献没有包含在本文当中。对此我们的处理方式是我们没有提及的或没

有研读过的文章一概不列在参考文献当中,由此引起的疏忽以及不方便也只能作为文章的缺陷的一部分了。

致谢:作者对杨展如老师的鼓励与支持深表感谢,没有杨老师的鼓励我们就不可能把这么多个头绪整理成文。作者吴对系统科学系诸位老师的热烈讨论与所营造的和谐向上的氛围表示感谢,尤其感谢王有贵老师多方面的支持。

参考文献

- [1] Albert R, Barabási A-L. Rev. Mod. Phys., 2002, 74: 47-97.
- [2] Dorogovtsev S N, Mendes J F F. Adv. Phys., 2002, 51: 1079-1187.
- [3] Strogatz S H. Nature, 2001, 410: 268-276.
- [4] SIS: Faster Breaking Papers, <http://www.esi-topics.com/fbp/fbp-december2002.html>.
- [5] Bollobás B. Modern Graph Theory, New York: Springer-Verlag, 1998; Random Graph, London: Academic, 1985.
- [6] Scott J. Social Network Analysis: A Handbook, London: Sage Publications, 1991.
- [7] Watts D J, Strogatz S H. Nature, 1998, 393: 440-442.
- [8] Barabási A-L, Dezső Z, Ravasz E, et al. Scale-free and Hierarchical structures in complex networks, unpublished, <http://www.nd.edu/~networks/papers.htm>.
- [9] Newman M E J, Strogatz S H, Watts D J. Phys. Rev. E, 2001, 64: 026118.
- [10] Ramezanpour A, Karimipour V. Phys. Rev. E, 2002, 66: 036128.
- [11] Yook S H, Jeong H, Barabási A-L. Phys. Rev. Lett., 2001, 86: 5835-5838.
- [12] Moore C, Newman M E J. Phys. Rev. E, 2000, 61: 5678-5682.
- [13] Schwartz N, Cohen R, ben-Avraham D, et al. Phys. Rev. E, 2002, 66: 015104(R).
- [14] Lago-Fernandez L F, Huerta R, Corbacho F, et al. Phys. Rev. Lett., 2000, 84: 2758-2761.
- [15] Newman M E J, Phys. Rev. Lett., 2002, 89: 208701; Phys. Rev. E, 2003, 67: 026126.
- [16] Goh K-I, Oh E, Kahn B, et al. Phys. Rev. E, 2003, 67: 017101.
- [17] Barabási A-L, Albert R. Science, 1999, 286: 509-512.
- [18] Dorogovtsev S N, Mendes J F F, Samukhin A N. Phys. Rev. Lett., 2000, 85: 4633-4636.
- [19] Krapivsky P L, Redner S, Leyvraz F. Phys. Rev. Lett., 2000, 85: 4629-4632.
- [20] Barabási A-L, Albert R, Jeong H. Physica A, 1999, 272: 173-187.
- [21] Yook S H, Jeong H, Barabási A-L. Modeling the Internet's Large-Scale Topology, PNAS 2002, 99: 13382-13386.
- [22] Krapivsky P L, Redner S, Leyvraz F. Phys. Rev. Lett., 2000, 85: 4629-4632.
- [23] Girvan M, Newman M E J. Community structure in social and biological networks, PNAS, 2002, 99: 7821-7826.
- [24] Per Bak, How Nature Works: The Science of Self-Organised Criticality, New York: Copernicus Press, 1996.
- [25] Guillaume J-L, Latapy M. The web graph: an overview, Proceedings of Algotel 2002, <http://citeseer.nj.nec.com/guillaume02web.html>
- [26] Broader A Z, Kumar S R, Maghoul F, et al. WWW9/Computer Networks, 2000, 33(1-6): 309-320.
- [27] Aiello W, Chung F, Lu L. Proceedings of the 32nd ACM Symposium on the Theory of Computing, New York: ACM, 2000, 171-180.
- [28] Kumar R, Raghavan P, Rajalopagan S, et al. Proceedings of the 9th ACM Symposium on Principles of Database Systems, 1999, p. 1.
- [29] Adamic L A, Huberman B A. Science, 2000, 287: 2115.

- [30] Adamic L A, Huberman B A. *Nature* 1999, 401: 131.
- [31] Albert R, Barabási A-L. *Phys. Rev. Lett.*, 2000, 85: 5234-5237.
- [32] Tadic B. *Physica A*, 2001, 293: 273-284.
- [33] Cooper C, Frieze A. A general model of web graph, In *ESA*, 2001, 500-511.
- [34] Jeong H, Neda Z, Barabási A-L. *Euro. phys. Lett.* 2003, 61: 567-572.
- [35] 伍法岳,杨展如. *物理学进展*, 1981, 1 (4): 525-541.
- [36] Moore C, Newman M E J. *Phys. Rev. E*, 2000, 62: 7059-7064.
- [37] Lloyd A L, May R M. *Science*, 2001, 292: 1316-1317.
- [38] Paster-Satorras R, Vespignani A. *Phys. Rev. Lett.* 2001, 86: 3200-3203.
- [39] Pandit S A, Amritkar R E. *Phys. Rev. E*, 1999, 60: R1119-1122.
- [40] Motter A E, Nishikawa T, Lai Y-C. *Phys. Rev. E*, (2002), 66: 065103(R).
- [41] Albert R, Jeong H, Barabási A-L. *Nature*, (2000), 406: 378-382.
- [42] Holme P, Kim B J, Yoon C N, et al. *Phys. Rev. E*, 2002, 65: 056109.
- [43] Faloutsos M, Faloutsos P, Faloutsos C. *Comput. Commun. Rev.* 1999, 29: 251.
- [44] Govindan R, Tangmunarunkit H. *Proceedings of IEEE INFOCOM 2000*, Israel: Tel Aviv (IEEE, Piscataway, N.J.), 2000, 3: 1371.
- [45] Amaral L A N, Scala A, Barthélemy M, et al. *Proc. Natl. Acad. Sci. U. S. A.*, 2000, 97: 11149-11152.
- [46] Liljeros F, Edling C R, L. Amaral A N, et al. *Nature*, 2001, 411: 907-908.
- [47] Jeong H, Mason S P, Oltvai Z N, et al. *Nature*, 2001, 411: 41-42.
- [48] Ferrer i Cancho R, Solé R V. *Proceedings of the Royal Society of London*, 2001, B 268: 2261-2266.
- [49] Albert R, Jeong H, Barabási A-L. *Nature*, 1999, 401: 130-131.
- [50] Jeong H, Tombor B, Albert R, et al. *Nature* 2000, 407: 651-654.
- [51] Fell D A, Wagner A. *Nat Biotechnol*, 2000, 18: 1121-1122.
- [52] Williams R J, Martinez N D, Berlow E L et al. *Santa Fe Institute Working Papers*, 01-07-036.
- [53] Dunne J A, Williams R J, Martinez N D. *PNAS* 2002, 99(20): 12917-12922.
- [54] Camacho J, Guimerà R, Amaral L A N. *Phys. Rev. E* 2002, 65: 030901R.
- [55] Camacho J, Guimerà R, Amaral L A N. *Phys. Rev. Lett.*, 2002, 88: 228102.
- [56] Dunne J A, Williams R J, Martinez N D. *Ecol. Lett.*, 2002, 5: 558-567.
- [57] Carlascelli D, Caldarelli G, Pietronero L. *Nature*, 2003, 423: 165-168.
- [58] Banavar J R, Maritan A, Rinaldo A. *Nature*, 1999, 399: 130-132.
- [59] West J B, Brown J H, Enquist, B J. *Science*, 1999, 284: 1677-1679.
- [60] Forogovtsev S N, Goltsev A V, Mendes J F F. *Phys. Rev. E*, 2002, 66: 016104.
- [61] Herrero C P. *Phys. Rev. E*, 2002, 65: 066110.
- [62] Medvedyeva F, Holme P, Minnhagen P, et al. *Phys. Rev. E*, 2003, 67: 036118.
- [63] Goltsev A V, Forogovtsev S N, Mendes J F F. *Phys. Rev. E*, 2003, 67: 026123.
- [64] Dezső Z, Barabási A-L. *Phys. Rev. E*, 2002, 65: 055103(R).
- [65] Redner S. *Eur. Phys. J. B*, 1998, 4: 131-134.
- [66] Newman M E J, *Phys. Rev. E*, 2001, 64: 016131; Newman M E J. *Phys. Rev. E*, 2001, 64: 016132;
- [67] Barabási A-L, Jeong H, Neda Z, et al. *Physica A*, 2002, 311: 590-614.
- [68] Egghe L, Rousseau R. J. *Amer. Soci. Info. Sci. Tech.*, 2003, 54: 193-202.
- [69] Mantegna R N, Stanley H E. *An Introduction to Econophysics: Correlations and Complexity in Finance*, Cambridge: Cambridge University Press, 1999.
- [70] Bouchaud J-P, Potters M. *Theory of Financial Risk*, Cambridge: Cambridge University Press, 1999.

- [71] Anderson R M, May R M. Infectious Disease of Humans: Dynamics and Control, Oxford: Oxford Univ. Press, 1991.
- [72] Hubacek K, Giljum S. Ecological Economics, 2003, 44: 137-151.
- [73] Girvan M, Newman M E J, Proc. Natl. Acad. Sci. USA, 99: 7821-7826.
- [74] Cormen T H, Leiseison C E, Rivest R L, et al. Introduction to Algorithm, MIT Press, 2001.

COMPLEX NETWORKS IN STATISTICAL PHYSICS

WU Jin-shan^{1,2}, DI Zeng-ru¹

(1. *Department of Systems Science, School of Management, Beijing Normal University, Beijing, 100875*

2. *Department of Physics, Beijing Normal University, Beijing, 100875*)

Abstract : From the point view of Statistical Physics, Networks are general many-body systems including all kinds of relations on physical and non-physical interactions. In this article, the authors review most of the main results and organize them by traditional logic of Statistical Physics. An outline for research works in the future is presented. According to the different property of interactions, Complex Networks are classified into three categories, undirected networks, directed networks and weighted networks. First, the definitions of some static geometrical quantities and their properties on such networks are collected, and some new static geometrical quantities are proposed on the basis of research works on construction and evaluation of mechanism model of networks. Second, we turn to the mechanism models. Most kinds of models are included there, while we point out that the mechanism of bi-scale-free networks is waiting for our more attention. Development of networks is related closely with the mechanism model. Some of the works in this aspect are put into the third part. Results in physical models on networks and the dynamical properties of networks are organized in the next section. In this part, the exploration on robustness, structure-function analysis and special properties of physical models such as random walk and Ising model should be extremely interested to physicists. Then we take two of our recent works network of scientists in Econophysics and networks of physical input and output tables as examples to demonstrate some initial application of networks analysis. At last, a brief introduction of the common analysis and numerical methods are listed for further reading. After all, Complex Networks, as an attempt to extend Statistical Physics into general systems, should be one of the central focuses of traditional researchers in Statistical Physics.

Key Words : Statistical Physics; Complex Networks; Review; Random Graph; Power Law; Scale-free networks