

复杂网络模型在传染病预防控制中的应用

黄淑琼 魏晟 岳丽 向浩 刘丽 刘莉 甘秀敏 许川 聂绍发

【导读】 为探讨复杂网络模型在传染病预防控制中的应用,根据复杂网络的基本属性,利用网络模型和常用软件分析传染病传播过程。相对于传统的流行病学方法,复杂网络的理论不仅能描述传染病动态传播过程,也能进行传染病预测。通过复杂网络的理论来研究疾病的传播,能深入理解到网络的拓扑结构对疾病传播有重要影响,从而找到控制疾病传播的更有效的方法。

【关键词】 流行病学; 复杂网络; 传播动力学; 传染病控制

Application of complex networks model in preventing and controlling communicable disease
HUANG Shu-qiong, WEI Sheng, YUE Li, XIANG Hao, LIU Li, LIU Li, GAN Xiu-min, XU Chuan, NIE Shao-fa. Department of Epidemiology and Biostatistics, School of Public Health, Tongji Medical College, Huazhong University of Science and Technology, Wuhan 430030, China

Corresponding author: NIE Shao-fa, Email: sf_nie@mails.tjmu.edu.cn

【Introduction】 To discuss the application of complex networks models in preventing and controlling communicable disease, analyze and control the spread of infectious diseases by using the models and the software of complex networks based on its basic properties. Compared with conventional epidemiological approach, the complex networks theory, as a new theory, not only can describe the dynamic process of infectious diseases spreading but also forecast the situation of infectious disease. The influence of the network's topology on the infectious diseases transmission can be deeply understood through the research on disease spreading by its theory, so to control the spread of diseases. Complex networks theory approach can be used in epidemiological research for having much advantage compared with those conventional epidemiological approaches.

【Key words】 Epidemiology; Complex networks; Transmission dynamics; Communicable disease control

复杂网络 (complex networks) 作为研究系统复杂性的一项重要工具^[1], 成为近半个世纪以来科学家们关注的新热点。复杂网络理论在流行病学领域的应用主要在于研究传染病的传播机制, 对传播动态过程的描述和预测, 以更好地预防控制传染病的发生。对传染病的描述和预测已有很久远的历史, 最早使用单纯总结经验型的统计方法, 通过对以往发病资料的收集, 用标准的统计预测模型进行疾病发病趋势研究^[2], 主要注重预测性和应用性。近年来随着数学和物理科学成果不断引入, 传染病传播机制模型为其预测理论研究提供了许多强有力的基础^[3], 这类模型强调解释性和研究性, 比前一类模型复杂得多。复杂网络理论在流行病学中应用尚处于萌芽期, 本文中探讨复杂网络理论在研

究传染病传播机制的应用, 为理论流行病学提供新方法和思路, 对研究传染病预测、预警和控制有借鉴价值。

基本原理

1. 基本属性: 复杂网络由大量的节点通过边的连接构成^[4], 是具有复杂拓扑结构和动力行为的大规模网络。复杂网络理论溯源于 1736 年德国伟大数学家 Euler 解决 Konigsberg 七桥问题^[5], 是以图形研究问题, 后来平面图论渐渐发展为用立体拓扑结构来分析空间网络结构, 但是直到 20 世纪 90 年代以来这一理论才蓬勃发展。

复杂网络理论研究可以简单概括为紧密相关且逐渐深入的以下四方面内容: 通过实证方法测定网络结构的统计性质; 构建相应的网络模型来展现这些统计性质; 在已知网络结构属性及其形成规则的基础上, 预测网络中的行为规范^[6]; 在预测网络行为规范的基础上探讨网络控制策略。前两方面是基础, 后两方面是应用。复杂网络简单图示见图 1。

DOI: 10.3760/cma.j.issn.0254-6450.2009.04.023

基金项目: 国家自然科学基金青年科学基金 (30700680)

作者单位: 430030 武汉, 华中科技大学同济医学院公共卫生学院流行病与卫生统计学系

通信作者: 聂绍发, Email: sf_nie@mails.tjmu.edu.cn

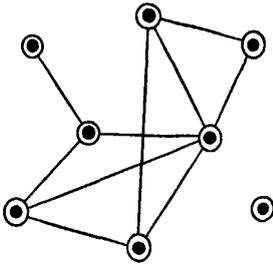


图1 8个顶点和10条边构成的简单网络

复杂网络的静态几何量主要有网络聚集系数、平均路径长度、度及其分布特征等^[7]。

(1) 网络聚集系数：聚集系数 (clustering coefficient, C) 又称聚类系数, 用来描述网络中节点的聚集情况, 即网络的紧密程度。其计算方法为: 假设节点 i 有 k_i 条边把它与其他 k_i 个节点相连, 这 k_i 个节点之间最多可能存在 $k_i(k_i - 1)/2$ 条边, 而这 k_i 个节点之间存在的边数实际只有 E_i 的话, 则它与 $k_i(k_i - 1)/2$ 之比定义为节点 i 的聚集系数 C_i , 用算式表示 $C_i = 2E_i / k_i(k_i - 1)$ ^[5]。网络的聚集系数就是整个网络中所有节点的聚集系数的平均。

(2) 平均路径长度：平均路径长度 (average path length, L) 是网络的分离程度指标。网络研究中, 两节点 i 和 j 间的距离 (d_{ij}) 定义为连接两者的最短路径的边的数目; 网络的平均路径长度则是所有节点对之间距离的平均值, 用算式表示 $L =$

$$\frac{1}{\frac{1}{2}N(N+1)} \sum_{i>j} d_{ij}$$

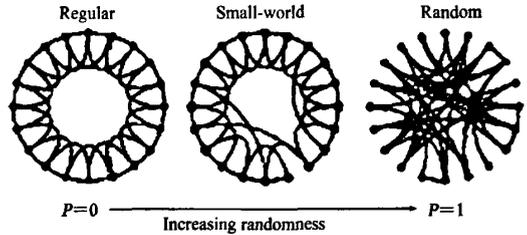
其中 N 为网络的节点数。

(3) 度及其分布：度 (degree) 是单独节点的属性中的重要概念。网络中节点 i 的度 k_i 为该节点连接的边的总数目, 所有节点 i 的度 k_i 的平均值称为网络的平均度, 记作 $\langle k \rangle$ 。网络中节点的度分布 (degree distribution) 用分布函数 $P(k)$ 来描述, 其含义为一个随机选定的节点恰好有 k 条边的概率, 也等于网络中度为 k 的结点数占网络结点总数的比值。度分布函数对于传染病传播机制研究具有重要的作用。

2. 复杂网络模型: 复杂网络按照拓扑结构特征可分为: 完全规则网络、完全随机网络、小世界网络和无标度网络^[8]。自从20世纪末小世界网络和无标度网络为代表的复杂网络理论兴起以来, 许多复杂网络都呈现出小世界和无标度的特性, 从而复杂网络上的疾病传播和控制问题成为新的关注点。

(1) 小世界网络模型: 小世界网络 (small-world network) 是 Watts 和 Strogatz 在 1998 年提出的基于人类社会网络的网络模型; 它是介于规则网络和随

机网络之间的网络 (WS 网络), 具有较短的平均路径长度和较高的聚类系数, 它通过调节一个连接参数从规则网络向随机网络过渡。小世界网络构造过程如图 2 所示: 将规则网络中的每条边以概率 P 随机连接到网络中的一个新节点上, 构造出一种介于规则网络和随机网络之间的网络。规则网络和随机网络则分别是小世界网络在 P 为 0 和 1 时的特例^[9]。

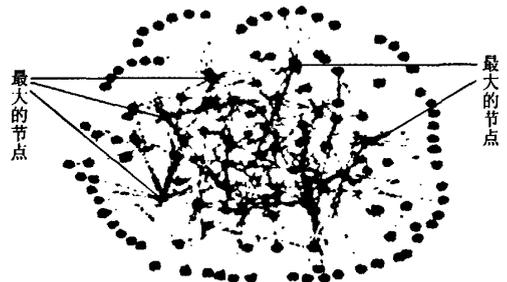


The figure shows the transition from a regular network to a random network via a small world network. $N=20, K=4$

图2 小世界网络的构造

小世界网络模型对于传染病免疫预防策略采用局域控制方法具有重要的启发意义, 通过免疫被感染节点一定路径范围内的节点, 可以很好控制疾病传播, 达到控制代价和提高效果之间的一种最优平衡。

(2) 无标度网络模型: 虽然小世界模型能体现现实世界的小世界性和高聚集性, 但小世界网络模型的连接度分布可近似于 Poisson 分布, 仍属于均匀网络。其实许多实际网络并不是均匀网络, 节点的连接度没有明显的特征长度, 即无标度网络 (scale-free network), 又称 BA 网络。1999 年 Barabási 和 Albert 发现许多实际网络的无特征尺度, 节点的度服从幂律 (power-law) 分布, 即 $P(k) \propto 2m^2 k^{-3}$, 这表明无标度网络的度分布函数可由幂指数为 3 的幂律函数近似描述。无标度网络结构如图 3 所示。



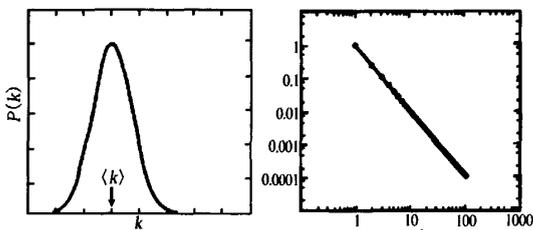
(展示 130 个节点的 BA 网络, 图中标注的 5 个节点是网络中最大的节点)

图3 无标度网络的拓扑结构示意图

无标度网络考虑了实际网络的两个重要特性:

增长特性和优先连接特性。增长特性表明网络大小随时间一步步生长,网络的规模总是不断扩大。优先连接特性体现出新的节点更倾向于与那些具有较高连接度的“大”节点连接,这种现象在社会学领域又被称为“马太效应”。

由于度分布呈现出幂律分布,具有非均匀性,绝大多数节点的度都相对较小,仅少量节点的度相对较大,无标度网络表现出很强的鲁棒性和脆弱性。鲁棒性指如果随机移走少量节点后网络中的绝大部分节点仍是连通的,该网络的连通性对节点故障具有鲁棒性。脆弱性是相对于蓄意攻击策略,即蓄意去除网络中度高的节点就可破坏无标度网络的连通性,非均匀性使无标度网络对蓄意攻击具有高度的脆弱性。两种不同拓扑结构网络的度分布如图 4 所示。



a. 均匀网络,指数分布 b. 无尺度网络,幂律分布 (log-log 曲线)

图 4 两种不同拓扑结构网络的度分布图

无标度网络特征对于分析传染病动态传播过程具有更高参考价值,其优先连接特性表明传染病传播过程中人群连接度较高的节点更易传播疾病,找出节点度高的个体,采取相应防控措施,减少传染病的传播范围和强度。

3. 复杂网络的应用步骤:

(1) 人群接触网络生成:利用复杂网络理论研究传染病传播过程的客观基础和前提是构建人群接触网络。选取假定封闭的特定人群作为研究主体,通过调查,了解人与人之间的接触情况,构建一个人群接触网络,进行计算机录入分析,得到一些网络特征参数。社会接触网络常用的软件是 Ucinet 网络分析集成软件^[10],其中嵌入了 NetDraw 和 Pajek 应用软件程序。

(2) 网络模型识别:依据构建的网络结构,进行特征分析,关键是三个基本网络特征参数,然后进行网络度分布的检验。应用 Matlab 程序分析网络节点度及度分布规律,从而进行网络模型识别。

(3) 预测网络系统行为:在进行模型识别后,应用相关网络模型分析传染病传播过程及特征,继而探讨合适的预防控制措施。利用网络模型预测网络系统行为,主要是找出关键性节点,这些节点是连接内部小团体的连通点,如果这些节点受到影响(感染),那么疾病传播将会很迅速。从另一角度来说,这些节点感染后,又能很容易被发现,也加速了患者就医的速度,缩短病程,减少单个传染源传播疾病的时间,从而控制疾病的传播规模。因此,关键性节点对控制传染病在人群中传播起到重要作用。

(4) 采取相应预防控制措施:度较大的节点被感染后将会快速传播疾病,因此,防止度较大的节点成为传染源是防止疾病快速传播的第一步。另外,利用被感染节点周围的局部信息,控制被感染节点周围的部分节点,从而达到控制传染病在整个网络大范围传播的目的。

实例分析

本例数据来源于一起重症肺炎暴发调查,研究范围是所有肺炎病例和接触者,共 35 人,以指示病例为线索起点,通过追踪和个案调查结果按网络生成规则构建一个小型人群网络,就是由许多节点与连接两个节点之间的一些边组成,其中每个节点代表真实系统中不同的个体,而边则代表个体间的关系,如果两个节点之间具有某种关联就在两节点之间连一条边,反之则不连边^[4]。先用 Ucinet 软件生成一个人群接触情况的矩阵,再通过 NetDraw 软件绘成网络结构图(图 5)。

对构建的人群接触网络的节点度和度分布的关系在双对数坐标下展示(图 6),近似于直线,可以考虑符合幂律分布,利用无标度网络模型拟合。本网络结构可看出有些关键性节点,如节点 1、2、3、4、5、

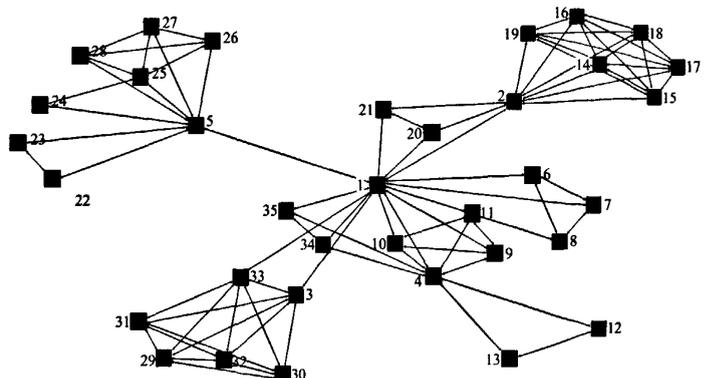


图 5 疾病传播网络拓扑结构(包含 35 个节点)

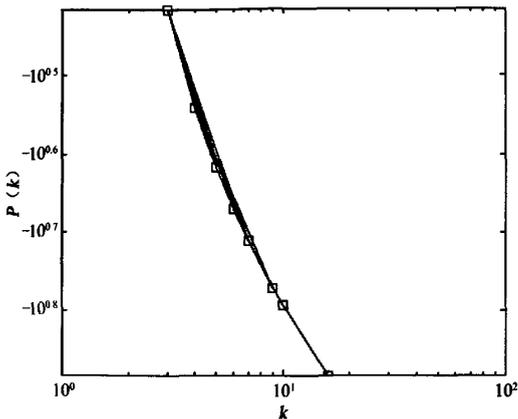


图6 双对数坐标下节点度的概率分布

33 节点度很高,这些关键性节点是疾病防控的重点对象,继而有针对性地采取控制措施。

讨 论

传染病的传播过程中充满了偶然因素的影响,它的传播显然是一种复杂现象,这是因为传染病传播的载体——人和人之间的交流、接触、联系所形成的系统是复杂的,易感染和有传染性的个体都具有主动性,其行为方式对于易感染性和病程长短有影响,也会随外界情况的变化产生适应性的改变,这些都造成了问题的复杂性。

复杂网络理论研究传染病传播机制,以宏观和微观相结合为重要特色^[5],其在流行病学应用的优势作用有以下四个方面。

1. 更新传统的传染病传播阈值理论。传染病传播数理模型是用来描述传染病的传播过程,分析受感染人数的变化规律,预测传染病在人群中传播的高潮。如基于均场理论模型的一个重要依据即疾病的传播存在着一个临界值,当有效传播率大于该临界值时,感染的个体将使疾病扩散,反之亦然。由于疾病波及的范围与传播强度呈正相关,如果人群中的疾病长期存在,则必然波及大量个体。但实证研究表明,有些传染病如流感仅波及少数个体而长期存在。因此,经典的传染病传播模型难以准确解释某些传染病在人群中的传播规律。换言之,传统的传染病传播模型由于其理论基础的缺陷,难以全面、系统地解释传染病在人群中的传播行为,这在很大程度上阻碍了传染病预测预警方面取得突破性进展。

网络的拓扑结构影响疾病的传播,复杂网络上的传播行为与均匀网络上的传播行为有着根本的不同,最近的研究结果表明无标度网络不存在传播阈

值。传染病在人群中的流行应看作是服从某种规律的在复杂网络上的传播行为,因此,建立传染病传播模型时,不仅要考虑疾病本身的传播特性,还要考虑网络的拓扑结构对疾病传播的影响^[8],才能全面掌握传染病的传播过程。

2. 强调传染源初始节点位置对传播的影响。传染病是通过人与人之间的接触进行传播的,如果人群中的个体用网络中的节点来表示^[11],可能发生传染和被传染的关系用网络中的边来表示,就构建出疾病传播所依赖的人群接触网络结构^[12]。疾病在网络中的传播过程是随机出现一个节点(病例),通过与其邻近节点(其他人)接触,把疾病传给邻近节点,然后每个邻近节点再各自进行接触传播,如此进行下去,就可知道在整个网络中疾病的传播情况^[13]。疾病在网络上的传播与初始感染节点的位置关系密切。防止度较大的节点成为传染源是防止疾病快速传播的第一步,可以有效地延缓疾病的扩散,为进一步采取控制措施、防止疾病大面积传播赢得时间。

3. 体现局部免疫控制策略优势。疾病传播网络中免疫节点是指当疾病传播到该节点时,该节点既不会被疾病感染,也不会将疾病向其他节点传播。所以,也可以看作是从网络中删除该节点。较好的免疫策略应该是在对较少的节点进行免疫的情况下,能够最大程度地延缓疾病的传播。

免疫控制策略有三大类:随机免疫(完全随机地选取网络中的部分节点进行免疫)、目标免疫(选取少量度最大的节点免疫)和“熟人”免疫(对随机选出的邻近节点进行免疫)。其中“熟人”免疫属局部免疫控制方法,这种策略回避了目标免疫中需要知道全局信息(每个节点的度)从而找出度大的节点进行免疫。局部免疫控制在对网络全局信息未知的情况下,利用被感染节点周围的局部信息,控制被感染节点周围的部分节点,从而达到控制疾病在整个网络大范围传播的目的(不需要考虑节点的度,仅考虑传染源周围一定路径长度内的节点^[14])。

4. 进行疾病传播和预防控制模拟仿真。利用复杂网络建立传染病传播模型模拟传染病传播这一复杂现象,通过改变参数数值,探讨哪些因素对于传染病暴发流行最重要,采取哪类措施能够更有效地控制疾病传播,何时采取效果最好,利用模拟仿真可以更好地为传染病预防控制决策提供理论支持^[15]。另外通过相关软件应用,进行网络构建和传播模拟仿真,可以对传染病的传播过程、控制措施的决策和效果有一个比较直观的认识。

复杂网络理论应用于传染病预防控制,具有很 多优势,但也存在一定的局限性。其局限性主要 在于其研究前提是针对具体人群,构建的模型 的普及性需进一步证实。另外,复杂网络理论 在传染病控制方面的应用主要限于人与人之间 直接接触传播的疾病,其他类型的传染病不太 适用。

参 考 文 献

[1] Auyang SY. Foundations of complex-system theories. 2版.上海:上海科技教育出版社,2002:419.

[2] 孙振球. 医学统计学.北京:人民卫生出版社,2002.

[3] 马知恩,周义仓,王稳地,等. 传染病动力学的数学建模与研究. 北京: 科学出版社,2004:8.

[4] Newman MEJ. The structure and function of complex networks. Siam Review, 2003, 45(2):167-256.

[5] 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,2006:260.

[6] 刘涛,陈忠,陈晓荣. 复杂网络理论及其应用研究概述. 系统工程,2005,23(6):1-7.

[7] Albert R, Barabasi AL. Statistical mechanics of complex networks. Reviews of Modern Physics,2002,74(1):47-97.

[8] 赵一帆.复杂网络中基于疾病传播模型的阻塞传播.科学技术与工程,2007,7(24):6397-6400.

[9] Watts DJ,Strogatz SH. Collective dynamics of 'small-world' networks. Nature, 1998,393(6684):440-442.

[10] 罗家德. 社会网分析讲义. 1版. 北京:社会科学文献出版社,2005:105-111.

[11] Martin Kilduff, Wenpin Tsai. Social networks and organizations. 北京:中国人民大学出版社,2007:208.

[12] 王旻,郑应平. 基于复杂网络的疾病传播. 科技导报(北京),2005,23(5):21-24.

[13] 孙胜秋. 用复杂网络理论研究疾病的传播. 重庆师范大学学报(自然科学版),2005,22(4):1-5.

[14] 许丹,李翔,汪小帆. 复杂网络病毒传播的局域控制研究. 物理学报,2007,56(3):1313-1317.

[15] 余雷,薛惠峰. 基于复杂网络的传染病传播模型研究. 陕西科技大学学报(自然科学版),2007,25(3):126-129.

(收稿日期:2008-10-22)

(本文编辑:张林东)

· 读者 · 作者 · 编者 ·

本刊对统计学方法的要求

统计学符号按 GB 3358-1982《统计学名词及符号》的有关规定一律采用斜体排印,常用:①样本的算术平均数用英文小写 \bar{x} (中位数用 M);②标准差用英文小写 s ;③标准误用英文小写 s_x ;④ t 检验用英文小写 t ;⑤ F 检验用英文大写 F ;⑥ 卡方检验用希文小写 χ^2 ;⑦ 相关系数用英文小写 r ;⑧ 自自由度用希文小写 v ;⑨ 概率用英文大写 P (P 值前应给出具体检验值,如 t 值、 χ^2 值、 q 值等), P 值应给出实际数值,不宜用大于或小于表示,而用等号表示,小数点后保留 3 位数。

研究设计:应告知研究设计的名称和主要方法。如调查设计(分为前瞻性、回顾性还是横断面调查研究),实验设计(应告知具体的设计类型,如自身配对设计、成组设计、交叉设计、析因设计、正交设计等),临床试验设计(应告知属于第几期临床试验,采用了何种盲法措施等);主要做法应围绕 4 个基本原则(重复、随机、对照、均衡)概要说明,尤其要告知如何控制重要非试验因素的干扰和影响。

资料的表达与描述:用 $\bar{x} \pm s$ 表达近似服从正态分布的定量资料,用 $M(Q_n)$ 表达呈偏态分布的定量资料,用统计表时,要合理安排纵横标目,并将数据的含义表达清楚;用统计图时,所用统计图的类型应与资料性质相匹配,并使数轴上刻度值的标法符合数学原则;用相对数时,分母不宜小于 20,要注意区分百分率与百分比。

统计学分析方法的选择:对于定量资料,应根据所采用的设计类型、资料具备的条件和分析目的,选用合适的统计学分析方法,不应盲目套用 t 检验和单因素方差分析;对于定性资料,应根据所采用的设计类型、定性变量的性质和频数所具备的条件及分析目的,选用合适的统计学分析方法,不应盲目套用 χ^2 检验。对于回归分析,应结合专业知识和散布图,选用合适的回归类型,不应盲目套用直线回归分析;对具有重复实验数据检验回归分析资料,不应简单化处理;对于多因素、多指标资料,要在一元分析的基础上,尽可能运用多元统计分析方法,以便对因素之间的交互作用和多指标之间的内在联系做出全面、合理的解释和评价。

统计结果的解释和表达:当 $P < 0.05$ (或 $P < 0.01$) 时,应说对比组之间的差异具有统计学意义,而不应说对比组之间具有显著性(或非显著性)差异;应写明所用统计分析方法的具体名称(如:成组设计资料的 t 检验、两因素析因设计资料的方差分析、多个均数之间两两比较的 q 检验等),统计量的具体值(如: $t=3.45, \chi^2=4.68, F=6.79$ 等);在用不等式表示 P 值的情况下,一般情况下选用 $P > 0.05, P < 0.05$ 和 $P < 0.01$ 三种表达方式即可满足需要,无须再细分为 $P < 0.001$ 或 $P < 0.0001$ 。当涉及总体参数(如总体均数、总体率等)时,在给出显著性检验结果的同时,再给出 95% 可信区间。