

复杂网络中的 病毒传播

FUZA WANGLUO ZHONG DE BINGDU CHUANBO

郭长睿 蔡绍洪 著



中国地质大学 出版社
ZHONGGUO DIZHI DAXUE CHUBANSHE

复杂网络中的病毒传播

FUZA WANGLUO ZHONG DE BINGDU CHUANBO

郭长睿 蔡绍洪 著

图书在版编目(CIP)数据

复杂网络中的病毒传播/郭长睿,蔡绍洪著. —武汉:中国地质大学出版社,
2017.8

ISBN 978—7—5625—4106—6

I. ①复…

II. ①郭… ②蔡…

III. ①计算机病毒—网络传播—研究

IV. ①TP309.5

中国版本图书馆 CIP 数据核字(2017)第 214820 号

复杂网络中的病毒传播

郭长睿 蔡绍洪 著

责任编辑:王 敏

责任校对:张咏梅

出版发行:中国地质大学出版社(武汉市洪山区鲁磨路 388 号) 邮编:430074

电 话:(027)67883511 传真:(027)67883580 E-mail:cbb@cug.edu.cn

经 销:全国新华书店

Http://cugp.cug.edu.cn

开本:880 毫米×1230 毫米 1/32

字数:112 千字 印张:3.875

版次:2017 年 8 月第 1 版

印次:2017 年 8 月第 1 次印刷

印刷:武汉三新大洋数字出版技术有限公司

ISBN 978—7—5625—4106—6

定价:22.00 元

如有印装质量问题请与印刷厂联系调换

前 言

自然界和人类社会中的很多复杂系统都可以抽象为各种复杂网络。近年来,复杂网络的研究处于高速发展的阶段,复杂网络的大量新特性不断被人们揭示出来。复杂网络的研究涉及到数理学科、工程学科、生命学科和社会学科。复杂网络理论的发展将对各种现实网络的安全性、稳定性、控制策略、网络同步和网络演化机制等研究和应用产生重大的影响,同时成为人们研究复杂系统的新途径。

各种病毒的流行给人类带来了巨大的威胁和灾难。本书所说的病毒是一个广义抽象的概念,它包括各种生物病毒、细菌、计算机病毒、谣言、故障、污染物,以及一些负面因素等。因此研究复杂网络中病毒的传播与控制的机理成为一个重要而紧迫的挑战性课题。对人类健康和财富产生危害的各种病毒,它们难以防治的一个根本原因就是病毒的变异能力非常强,短时间内会出现许多变种,使得原有的免疫系统对变异后的病毒没有免疫作用,造成病毒大流行。对于时常爆发的病毒大规模传播事件,人们需要有效且成本较低的病毒控制手段。病毒传播的局域控制策略具有成本低、效果好的优点。病毒传播的局域控制策略是指,如果发现网络中的某些节点已经感染病毒,那么就对与感染节点有链接的健康节点进行免疫,使这些有可能感染病毒的节点具有较高的免疫力。

虽然人们对网络病毒变异与控制的研究已经取得了一些进展,但就目前而言,这一领域仍有许多重要的问题没有得到深入研究,一些重要性质和机理还不够清晰,一些变化规律还需要进一步揭示。例如:病毒在多种网络模型中变异与传播的规律还没有完全揭示;网

络拓扑结构对病毒变异与传播的影响还不清楚;病毒变异理论模型需要继续发展,现有的病毒变异理论模型还不能完全解释病毒变异与传播的规律;还没有在理论上形成具有局域控制策略的病毒传播动力学;一些具有病毒局域控制策略网络模型中的病毒传播也需要进一步探索;很多实际因素在病毒局域控制策略下对病毒传播的影响也有待深入研究,等等。对以上问题进行深入系统的研究,将进一步丰富和完善复杂网络的理论体系,对各类病毒危害的有效控制具有理论价值和实践意义。

本书对复杂网络中病毒的变异特性和病毒传播的局域控制策略进行了深入系统的研究。首先,本书在 Bitstring 病毒变异模型的基础上进行了拓展,建立了无标度网络中的 Bitstring 病毒变异模型;使用计算机仿真方法系统地研究了无标度网络中病毒变异与传播的规律,并考查了各种因素对具有变异特性病毒传播的影响;得出了决定病毒感染比例振荡周期的是节点感染时间,病毒变异概率较小时也能达到较大的病毒稳态感染比例等有益的结论;较好地解释了真实网络中病毒传播的周期性和病毒变异概率值较小时病毒仍然能够大规模传播的原因。其次,本书将病毒传播的局域控制策略扩展到二维规则网络和二维稀疏网络中,建立了二维规则 SIS 网络、二维规则 SIRS 网络、二维稀疏 SIS 网络和二维稀疏 SIRS 网络中具有局域控制策略的病毒传播模型;在理论上分析了这 4 种模型的病毒传播动力学方程,对这 4 种病毒传播模型中的病毒传播过程进行了计算机仿真;研究结果表明局域控制策略对这些网络中的病毒传播具有较好的控制效果,并根据研究结果制定了控制病毒传播的方案。

本书的研究揭示了无标度网络中病毒感染比例随时间振荡的规律;决定病毒感染比例振荡周期的是节点感染时间 t_I ,病毒感染比例振荡周期 $T=t_I+1$;节点感染时间 t_I 值越大,则病毒稳态感染比例 I 也越大;病毒变异概率 u 越大,则病毒稳态感染比例 I 也越大,当 u 非常小时,病毒将不能够在网络中持续稳定传播;当变异概率 u 较小

时,病毒也能达到较大的稳态感染比例;当病毒在较大规模网络中能够持续传播时,在较小规模网络中可能会灭绝。具有局域控制策略的二维规则 SIS 网络、二维规则 SIRS 网络、二维稀疏 SIS 网络和二维稀疏 SIRS 网络中病毒的传播随时间的演化最终会达到一个稳定状态;只有当病毒传播效率 λ 大于一个临界值 λ_c 时,病毒才能在网络中持续稳定地传播;局域控制范围 d 越大,稳态感染比例 I 越小;稳态感染比例 I 与网络中的节点密度 ρ 也有关系。

本书的研究具有以下一些创新点:

(1)本书首次拓展和建立了无标度网络中病毒变异的 Bitstring 模型,以及具有局域控制策略的二维规则网络 SIS 模型、二维规则网络 SIRS 模型、二维稀疏网络 SIS 模型和二维稀疏网络 SIRS 模型,并进行了系统研究。

(2)本书采用动力学方程稳定性分析和计算机仿真的方法对上述病毒变异模型和控制策略模型进行研究,取得了较好的效果,拓展了对复杂系统的其他运行机理和发展规律进行深入研究渠道。

(3)本书的研究结果与真实网络的实际结果相吻合,较好地解释了真实网络中病毒变异、传播与控制的很多现象。

(4)通过本书的研究,探寻了复杂网络中病毒传播和病毒变异规律,初步揭示了病毒传播和变异的机理。

本书的研究结果,对进一步充实和完善现有复杂网络病毒传播控制与变异理论体系,对掌握病毒在复杂网络中传播与变异的规律和特性,从而更好地对之加以利用和控制,具有重要的理论价值和实际指导意义。

作 者
2017 年 7 月

目 录

§ 1 绪 论	(1)
1.1 研究的背景、目的及意义	(2)
1.1.1 研究的背景	(2)
1.1.2 研究的主要目的	(3)
1.1.3 研究的重要意义	(4)
1.2 国内外相关的研究进展及评述	(4)
1.2.1 复杂网络研究的历史回顾及研究进展	(4)
1.2.2 复杂网络病毒控制与变异的研究现状	(6)
1.2.3 复杂网络病毒变异与控制研究存在的问题	(7)
1.3 主要研究内容及方法	(8)
1.3.1 研究的主要内容	(8)
1.3.2 研究的主要思路和方法	(8)
1.3.3 研究的整体结构及框架	(9)
§ 2 复杂网络的基本概念及研究模型	(11)
2.1 复杂网络的基本概念与统计性质	(11)
2.1.1 网络的界定	(11)
2.1.2 度与度分布	(12)
2.1.3 网络直径与平均路径长度	(13)
2.1.4 聚类系数与聚一度相关性	(13)
2.1.5 介数	(14)
2.2 常见的复杂网络基本模型	(15)
2.2.1 规则网络模型	(15)

2.2.2	ER 随机网络模型	(17)
2.2.3	WS 小世界网络模型	(20)
2.2.4	BA 无标度网络模型	(23)
2.3	现实世界中的典型网络	(27)
2.4	本章小结	(28)
§ 3	复杂网络中的病毒传播与控制策略	(29)
3.1	传统病毒传播动力学模型	(29)
3.1.1	SIS 病毒传播模型	(29)
3.1.2	SIR 病毒传播模型	(30)
3.1.3	SIRS 病毒传播模型	(30)
3.2	复杂网络中的病毒传播动力学	(31)
3.2.1	ER 随机网络中的 SIS、SIRS 病毒传播动力学模型	(31)
3.2.2	小世界网络中的 SIS、SIRS 病毒传播动力学模型	(32)
3.2.3	无标度网络中的 SIS、SIRS 病毒传播动力学模型	(32)
3.3	复杂网络中的病毒传播控制策略	(33)
3.3.1	随机免疫策略	(33)
3.3.2	目标免疫策略	(34)
3.3.3	熟人免疫策略	(34)
3.3.4	局域控制策略	(35)
3.4	本章小结	(35)
§ 4	复杂网络中的病毒变异	(36)
4.1	病毒变异特性及病毒变异研究	(36)
4.1.1	病毒变异特性	(36)
4.1.2	病毒变异模型研究进展	(37)
4.2	病毒变异的 Bitstring 模型	(37)
4.3	小世界网络中的病毒变异	(39)
4.3.1	小世界网络中病毒变异的 Bitstring 模型	(39)
4.3.2	小世界网络病毒变异模型中的病毒感染比例	(39)

4.4	BA 无标度网络中的病毒变异	(39)
4.4.1	BA 无标度网络病毒变异模型中的病毒感染比例振荡	(39)
4.4.2	BA 无标度网络病毒变异模型中的病毒感染比例	(50)
4.5	本章小结	(55)
§ 5	二维规则网络中的病毒传播局域控制	(57)
5.1	ER 随机网络、WS 小世界网络和 BA 无标度网络中的局域控制 ...	(57)
5.2	二维规则网络中的 SIS 病毒传播局域控制	(58)
5.2.1	二维规则网络中的 SIS 病毒传播局域控制模型	(58)
5.2.2	模型动力学方程解的稳定性分析	(61)
5.2.3	计算机仿真	(62)
5.3	二维规则网络中的 SIRS 病毒传播局域控制	(66)
5.3.1	二维规则网络中的 SIRS 病毒传播局域控制模型	(66)
5.3.2	模型动力学方程解的稳定性分析	(69)
5.3.3	计算机仿真	(70)
5.4	本章小结	(74)
§ 6	二维规则稀疏网络中的病毒传播局域控制	(76)
6.1	二维规则稀疏网络中的 SIS 病毒传播局域控制	(76)
6.1.1	二维规则稀疏网络中的 SIS 病毒传播局域控制模型	(76)
6.1.2	模型动力学方程解的稳定性分析	(79)
6.1.3	计算机仿真	(80)
6.2	二维规则稀疏网络中的 SIRS 病毒传播局域控制	(84)
6.2.1	二维规则稀疏网络中的 SIRS 病毒传播局域控制模型	(84)
6.2.2	模型动力学方程解的稳定性分析	(87)
6.2.3	计算机仿真	(88)
6.3	本章小结	(94)
§ 7	总结与展望	(96)
7.1	总结	(96)

7.1.1	研究的核心内容	(96)
7.1.2	主要创新点	(96)
7.1.3	主要结论	(97)
7.2	存在的问题与展望	(99)
7.2.1	存在的问题	(99)
7.2.2	研究展望	(99)
参考文献		(101)
致 谢		(113)

§ 1 绪 论

网络存在于人类社会的方方面面。提起网络,人们首先想到的是国际互联网。自 1969 年出现的互联网发展至今已经深深地改变了人们的工作、学习、生活以及思维方式,使得人类社会进入到信息时代,成为人类历史上重要的科技成就之一。除了互联网,对人们工作和生活有着重大影响的网络还有人际关系网、大型输电网、公路铁路航空运输网、企业和金融关系网、病毒传播网、食物链网、生态网、大规模集成微电子电路网和各类通信网等,同时各种新的网络还在不断地被发现或被创造出来。人们的工作生活与各种网络息息相关,可以说,整个人类社会就是一个由各种网络所组成的极其复杂的超级网络。

人类社会中的很多网络给人们的工作和生活带来了巨大的便利,极大地促进了生产效率和生活质量的提高。但是一些重要的网络一旦出现异常,则会给人们的工作和生活带来极大的影响,甚至可能给经济带来巨大的损失。近年来多次发生的计算机病毒在互联网上的大规模爆发,造成了严重的影响,2006 年 7 月纽约输电网陆续崩溃造成的大规模停电事故,2008 年初中国南方凝冻导致的大范围电网瘫痪,2007 年美国次贷危机导致的全球金融风暴等。在自然界和人类社会中一些网络的存在,本身就会对人们的生命和财产造成灾难。例如:病毒传播网,2003 年爆发的 SARS 病毒,2009 年在全世界肆虐的甲型 H1N1 流感,等等。这些事例使得很多具有不同学术背景的研究人员将目光投向了复杂网络。

近年来,复杂网络的研究涉及到数理学科、工程学科、生命学科和社会学科。在这一新兴的研究领域内,物理学家、数学家、生物学家、社会学家和计算机专家等研究人员通力合作。他们考察了各种真实网络的特性,并建立了大量的网络模型,研究了发生在这些网络中的各种动力学过程,初步揭示了各种复杂网络的共性及特性,建立了研究各种真实网络的普适方法。

对复杂网络进行研究已经成为一个重要而紧迫的挑战性课题。复杂网络所展示的各种特性已经改变了人们对于网络的传统理解,甚至被称为“网络的新科学”(Barabási, 2002; Watts, 2004)。当前对复杂网络的研究正在起步,大量未知领域等待人们探索。复杂网络理论的发展将对各种现实网络的安全性、稳定性、控制策略、网络同步和网络演化机制等研究及应用产生重大的影响。

1.1 研究的背景、目的及意义

1.1.1 研究的背景

病毒是一个抽象的概念,包括各种生物病毒、细菌、计算机病毒、谣言、故障、污染物,以及一些负面因素,等等。

从古至今,病毒对人们生命和财产的危害从来没有停止过。病毒造成的生命死亡和经济损失无法估量,十分惊人。

黑死病是一种由细菌引起的传染病,于 1340 年传播到欧洲后,因此病死亡人数 7500 万人。天花病毒是人类历史上消灭的第一个传染病,18 世纪欧洲死于天花病毒的人数达 1.5 亿人以上。西班牙流感病毒爆发于 1918—1919 年,全世界共有 10 亿人感染(世界总人口为 17 亿人),有 2500 万~4000 万人因此病死亡。艾滋病毒死亡率近 100%,至 2005 年底,全世界共有 3860 万名艾滋病毒感染者。

甲型 H1N1 流感于 2009 年 3 月被发现,迅速在全世界蔓延,至 2009 年 12 月,全世界死亡人数超过 1 万人。

计算机病毒能对计算机系统造成破坏,甚至造成网络瘫痪。各种计算机病毒已经在全世界造成了数百亿美元以上的经济损失。如 2007 年 1 月初肆虐网络的熊猫烧香病毒,在短时间内出现 700 多个变种,上百万用户感染此病毒,造成的经济损失无法估计。

近年来爆发的世界金融危机,开始时只表现为少数几家公司的倒闭,但迅速发展成为席卷整个世界的金融风暴。当金融系统处于某个不稳定的临界状态时,金融网络中的某个环节一旦出现故障(如呆账坏账、不良贷款、有毒资产、流动性障碍等),那么故障将会在整个网络中传播扩散,不断放大,形成链式反应,甚至造成整个金融系统的崩溃和瘫痪。

因此,怎样利用控制手段限制病毒在网络中传播是学者们目前非常关注的一个问题(Pastor-Satorras, 2002; Cohen et al, 2003)。

对于危害人类健康和财富的各种病毒,难以防治的根本原因就是病毒容易产生变异,使得原有的免疫系统对变异后的病毒没有免疫作用。因此,研究病毒在网络中传播时发生变异的规律和应对策略,也是当前复杂网络研究中的一个热点问题(Andreasen et al, 1996; Andreasen et al, 1997; Girvan et al, 2002; Lin et al, 1999; Smith et al, 1999)。

1.1.2 研究的主要目的

(1)通过本书的研究,探寻复杂网络中病毒变异和病毒传播的规律,揭示病毒传播和变异的机理,初步建立起复杂网络中病毒传播与变异控制的理论框架,为进一步丰富和完善复杂系统理论研究做出贡献。

(2)研究复杂网络中病毒传播的动力学机制,分析局域控制策略

对网络中病毒传播的限制作用,揭示病毒在复杂网络中的传播机理,为人们控制复杂网络系统中病毒的传播提供理论指导。

(3)研究各种复杂网络结构中的病毒变异过程,建立各种复杂网络的病毒变异模型,探索病毒变异在各种复杂网络模型中的机制,为真实系统中控制病毒变异提供理论依据。

1.1.3 研究的重要意义

从认识论的角度看,通过对复杂网络中病毒传播控制和病毒变异的深入研究,人们能够深入地认识自然界和人类社会中的各种复杂网络,更加深刻地理解自然与社会。

通过对复杂网络中病毒传播控制和病毒变异的规律及机理进行深入的系统研究,可以深入了解复杂系统的内部结构和功能,丰富复杂系统的研究方法,推进复杂系统理论的研究发展,充实和完善复杂网络病毒传播控制和病毒变异的理论研究体系,因此本书的研究具有重要的理论意义和现实意义。

根据本领域研究的所得结果及其揭示的规律,人们可以有针对性地提出一些限制病毒在网络中传播与变异的控制手段,尽可能地将病毒造成的损失降低到最低限度。因此掌握病毒在复杂网络中传播与变异的特性并加以利用和控制,具有重要的社会价值和实践意义。

1.2 国内外相关的研究进展及评述

1.2.1 复杂网络研究的历史回顾及研究进展

最早有关网络的研究可以认为是大数学家欧拉(Euler)解决的哥尼斯堡(Konigsberg)七桥问题。这一问题开创了一门新的数学分

支——图论,现在认为这也是网络科学的起点。

20 世纪 60 年代,两位匈牙利数学家建立了随机图理论(Random Graph Theory)(Erdős & Rényi, 1959, 1960, 1961), ER 随机图理论在数学上开创了复杂网络理论的系统性研究。同时, Milgram (1967) 提出了“六度分离”(Six Degrees of Separation) 原则,意思是说世界上任意两人之间的平均距离是六。Milgram 做了邮件传递的社会实验,验证了这一论断。Watts 和 Strogatz (1998) 提出了小世界网络模型,揭示了复杂网络中的小世界性质。研究发现,在许多真实的网络中都存在小世界效应,这是复杂网络研究的一次突破。

Barabási 和 Albert (1999) 提出了无标度网络模型,揭示了复杂网络中的无标度性质。研究发现很多真实网络具有无标度的拓扑结构,这是复杂网络研究的又一次突破性进展。

复杂网络理论由此迅速发展,很多研究领域中的复杂系统都可以归结为复杂网络(Strogatz, 2001; Dorogovtsev & Mendes, 2002; Boccaletti et al, 2006; Albert & Barabási, 2002; Newman, 2003; 汪小帆等, 2006; 方锦清等, 2007)。现在复杂网络理论已经发展成为一门多学科交叉的新兴学科。

随着复杂网络研究的兴起,复杂网络中越来越多的特性不断被人们所发现和研究。例如:近年来研究者们非常关注的复杂网络的社团(Adamic & Adar, 2003; Milo et al, 2002; Gleiser & Danon, 2003)、模体(Milo et al, 2004)和分形自相似(Song et al, 2005)等性质。复杂网络中的各种新概念层出不穷,复杂网络理论研究正飞速发展,日新月异(Johnson et al, 2010; Qian et al, 2010)(Cafieri et al, 2010; Liu, 2010; Donner et al, 2010; Rozenfeld et al, 2010; Liu et al, 2010; Wilds & Glass, 2009)。

在复杂网络快速发展的同时,研究人员也在不断地推动复杂

网络的应用,例如:研究国际互联网的性质(Faloutsos et al, 1999; Zhou & Mondragon, 2004; Newman, 2002)、复杂网络上的病毒传播(Boguna & Pastor-Satorras, 2002; Dezsó & Barabási, 2002; Newman & Watts, 1999)、相继故障(Moreno et al, 2002; Crucitti et al, 2004; Goh et al, 2003)、同步(Belykh et al, 2004; Sorrentino & Ott, 2009; Yin et al, 2008)和动态网络控制(Wang & Chen, 2002)等。

1.2.2 复杂网络病毒控制与变异的研究现状

关于病毒传播的研究已经进行了很多年,提出了多种类型的病毒传播模型(Bailey, 1975; Anderson & May, 1992; Diekmann & Heesterbeek, 2000)。在较为简单的病毒传播模型中,参与病毒传播的个体(病毒宿主)一般分为3种状态: Susceptible(可被感染或健康状态)、Infected(已被感染状态)和 Removed(移除或免疫状态)。使用这3种状态的不同变化及组合就可以构造出不同的传播模型,例如: Susceptible $\rightarrow$ Infected $\rightarrow$ Susceptible(SIS模型)。在这一模型中,个体状态可以从健康状态变化到被感染状态,经过一段时间后又可以痊愈成为健康状态;同时有可能再一次被感染。除了SIS模型外,还有SIR模型和SIRS模型等。

传统的病毒传播理论认为,只有当病毒的传播率大于临界传播率时,病毒才能够在种群中长期存在,如果当病毒的传播率小于这个临界值时,病毒将从种群中自动消失。但Pastor-Satorras和Vespignani(2001)发现,大规模无标度网络中的病毒临界传播率趋近于零,即传播率很低的病毒也容易在大规模无标度网络中持续传播。

为了控制病毒在网络中大规模传播,学者们提出了很多控制策略,例如:随机免疫(Anderson & May, 1992)、目标免疫(Pastor-

Satorras & Vespignani, 2002) 和熟人免疫 (Cohen et al, 2003) 等控制方式。许丹等 (2007) 提出了“局域控制 (Local Area Control)”的病毒控制手段。这些控制策略都能够在一定程度上限制病毒在网络中传播。

前面提过, 病毒难以防治的根本原因就是病毒容易产生变异, 使得原有的免疫系统对变异后的病毒没有免疫作用。因此人们开始研究病毒在复杂网络中变异的规律。Cornell 大学的 Girvan 等 (2002) 提出了病毒变异的 Bitstring 模型, 并使用这一模型研究了均匀混合网络中的病毒变异。武汉大学的 Shao 等 (2006) 利用 Bitstring 模型研究了小世界网络中的病毒变异, 并得到了一些有意义的结论。

1.2.3 复杂网络病毒变异与控制研究存在的问题

虽然复杂网络中病毒变异与控制研究已经取得了一些良好的成果, 但仍然还存在许多急需解决的问题。

(1) 还未对多种网络模型中病毒变异与传播的规律进行深入研究, 网络拓扑结构对病毒变异与传播的影响还不清楚。

(2) 在网络中变异传播的病毒对网络的生成及演化具有重要影响, 但目前还不了解其中的机理。

(3) 病毒变异理论模型需要继续发展, 现有的病毒变异理论模型还不能完全解释病毒变异与传播的规律。

(4) 对于时常爆发的病毒大规模传播事件, 人们需要有效且成本较低的病毒控制手段, 目前病毒控制策略的效果并不完美。

(5) 局域控制策略具有成本低、效果好的优点, 但没有从理论上研究具有局域控制因素的病毒传播动力学。

(6) 规则网络和稀疏网络中的病毒传播局域控制也有待深入研究。