

基于自治计算的流行病传播网络建模与推断*

杨博^{1,2}, 刘际明³, 杨建宁^{1,2}, 白媛^{1,2}, 刘大有^{1,2+}

¹(吉林大学 计算机科学与技术学院, 吉林 长春 130012)

²(符号计算与知识工程教育部重点实验室(吉林大学), 吉林 长春 130012)

³(香港浸会大学 计算机科学系, 香港)

Modeling and Inferring Epidemic Networks Based on Autonomy Oriented Computing

YANG Bo^{1,2}, LIU Ji-Ming³, YANG Jian-Ning^{1,2}, BAI Yuan^{1,2}, LIU Da-You^{1,2+}

¹(College of Computer Science and Technology, Jilin University, Changchun 130012, China)

²(Key Laboratory of Symbolic Computation and Knowledge Engineering of Ministry of Education (Jilin University), Changchun 130012, China)

³(Department of Computer Science, Hong Kong Baptist University, Hong Kong, China)

+ Corresponding author: E-mail: liudy@jlu.edu.cn

Yang B, Liu JM, Yang JN, Bai Y, Liu DY. Modeling and inferring epidemic networks based on autonomy oriented computing. *Journal of Software*, 2012, 23(11): 2955–2970 (in Chinese). <http://www.jos.org.cn/1000-9825/4300.htm>

Abstract: In previous research, most related works on inferring the structures of diffusion networks are designed for recovering the process of information propagation. The learning data adopted by these works is distinct in terms of both format and features from the available surveillance data of epidemics. Therefore, the existing methods are not competent when dealing with epidemic surveillance data with some intractable properties such as coarse granularity, spatial and temporal multi-scale, and incompleteness. To address this issue, an AOC (autonomy oriented computing) based method is proposed to model epidemic networks, as well as to infer their structures from epidemic surveillance data. In this method, the structure of an epidemic network and the process of disease spread are modeled by an autonomous multi-agent system named D-AOC, and the parameters of the system are automatically estimated by a self-discovery process. During this process, the parameters are adjusted and thereafter, the behaviors of agents are updated by a feedback mechanism which combines the Monte Carlo simulation and swarm intelligence. The objective is to reduce the difference between emergent behavior of the D-AOC and observed surveillance data. Regulated by the feedback mechanism, it is expected that the D-AOC will keep evolving toward the real system to be simulated. In this way, the structure of epidemic network and main biological features related to the epidemic will finally be recovered. The effectiveness and applicability of the proposed method have been validated and discussed by analyzing the real surveillance data of the H1N1 swine-flu in Hong Kong during 2009. Moreover, one scenario of applying epidemic network inference is also demonstrated by a case study of epidemic risk assessment in Hong Kong.

* 基金项目: 国家自然科学基金(60873149, 60973088, 61133011, 61170092); 新世纪优秀人才支持计划(NCET-11-0204)

收稿时间: 2012-06-09; 定稿时间: 2012-08-15

Key words: epidemic model; epidemic network; autonomy oriented computing; multi-agent system; network inference; Monte Carlo simulation; spatial-temporal data mining

摘要: 现有的传播网络结构推断方法大都面向信息传播过程,所能处理的数据与可获得的流行病学监控数据形式和特性均不相同,不适合处理具有粗粒度、时空多尺度和数据缺失等特性的流行病学监控数据.针对该问题,提出了基于自治计算的流行病学传播网络建模方法和网络结构推断方法.该方法采用多自治体建模传播网络结构和流行病学传播过程,采用蒙特卡罗模拟结合群智能优化的反馈过程调节系统参数,以缩小模拟系统涌现行为与真实监控数据间差异为目标,改变自治体的行为,促使模拟系统向真实系统逐步演化,以此方式推断出传播网络结构及与流行病学相关的主要生物学参数.采用 2009 年 H1N1 猪流感在香港爆发的真实监控数据分析验证了所提出的模型与方法的有效性和适用情况,并以香港地区流行病学风险评估为例介绍了流行病学传播网络推断的一种应用模式.

关键词: 流行病学传播模型;流行病学传播网络;自治计算;多 Agent 系统;网络推断;蒙特卡罗模拟;时空数据挖掘

中图法分类号: TP18 **文献标识码:** A

流行性传染病(简称流行病)的每次爆发都会给人类社会带来巨大的损失,如 2003 年的严重急性呼吸综合征 SARS 和 2009 年的 H1N1 流感^[1,2].建立理论模型理解和模拟流行病的传播与爆发过程,建立监控系统收集流行病的传播数据,进而基于模型与监控数据进行准确的风险评估,帮助决策者制定相应的预防和应急方案,是有效控制流行病传播与爆发、最大限度减少生命财产损失的新途径,相关研究正日益受到各国政府、世界卫生组织(World Health Organization,简称 WHO)和学术界的重视^[2,3].

流行病的传播过程由流行病模型和传播网络共同决定.医学领域对流行病的研究有较长的历史,提出了多种数学模型,其中最具代表性的是基于动力系统的 SIR 模型和 SIS 模型^[4,5]以及基于它们的多种改进(如考虑潜伏期的 SEIR 模型和混合模型 SIRS 等^[4]).对这类模型的早期研究大都假设流行病的传播媒介均匀分布,没有考虑人口在空间上分布不均对流行病传播过程的影响.人口等流行病传播媒介在空间上的分布可采用网络结构进行抽象与建模.1983 年,Grassberger 将流行病传播问题建模为键逾渗(bond percolation)问题,最先在网络框架下研究了 SIR 模型的基本性质^[6].1998 年,小世界网络模型提出后,Newman 最早研究了小世界网络结构对 SIR 模型的影响^[7,8].在此基础上,后续工作深入研究了小世界网络模型的各种参数(如随机长连概率)对 SIR 模型的影响^[9]、小世界网络传播动力方程的分形、不动点、混沌和周期分叉等多种数学性质^[10].一些工作还将 SIS 和 SEIR 等模型推广到网络上,讨论这些流行病模型在小世界网络上的传播特性^[11].1999 年,无标度网络模型提出后,人们发现包括社会网络在内的很多实际网络都是非均匀网络,其度分布满足幂率而非泊松分布.此后,各种传播模型在无标度网络上的临界值分析成为新的研究热点.2001 年,Pastor-Satorras 和 Vespignani 利用平均场方法给出了 SIS 模型在无标度网络上的传播临界值^[11].在此基础上,后续工作相继给出了 SIS 模型在有限规模无标度网络的临界值^[12]、SIS 模型在关联网络中临界值^[13,14]、级联传播模型在无标度网络上的临界值^[15].

以上述工作为代表的相关研究在分析网络结构对流行病传播产生的影响时,大都事先假定传播网络结构已知(如某种规则网络)或者服从某种特定的网络生成模型(如小世界网络和无标度网络模型),进而分析流行病模型在不同网络结构上的临界值及各种宏观动力学性质.然而实际问题所涉及的传播网络具有更高的复杂性,往往不能满足这些假设.实际传播网络与假设网络结构上的差异会显著影响分析的准确性与可靠性.基于真实的传播网络结构,能够更加精确地刻画流行病传播过程,更加可靠地评估传播趋势,从而更加有针对性地制定控制策略.但真实的传播网络结构通常是隐含的,我们能观测到的大多是流行病在传播过程中所产生的最终影响,而不是个体间直接的相互影响关系.例如,在 SARS 爆发过程中,很难确定一名患者在出行途中直接传染了哪些个体,而仅能观测和统计在不同时间和空间内新增的 SARS 病例数目.可见,真实的传播网络难以直接构建.随着信息技术在医疗领域的广泛应用,记录各种流行病传播过程的监控数据(surveillance data)与日俱增,这些数据记录了与流行病传播和爆发相关的时空信息,包括入院、确诊、出院、死亡等医疗过程的统计数据、地理位置信息以及当地人口的统计数据.如果能够从这些监控数据中推断出导致传播现象发生的传播网络结构,就有望解

决真实传播网络难以建造的问题。

基于观测到的传播数据推断传播网络结构,是研究流行病传播规律的新思路,目前流行病等医学领域还未见到相关研究,类似问题最先受到计算机领域研究者的关注.这方面的早期工作主要针对信息传播现象,提出能够根据传播现象推断出隐含信息传播网络的学习算法.2004年,HP公司的Adar等人较早地开展了这方面的尝试,基于博客(blog)被话题(topic)感染时间的先后次序、博文的文本信息以及博客主页间的超链接结构推断出博客间的隐含影响关系^[16].该工作将传播网络结构学习问题建模为分类问题,采用支持向量机和逻辑回归方法推测每对博客间的影响关系,进而学习出隐含的主题传播网络.有监督的学习方法需要大量正确标注的样本,而在实际问题中,这些样本往往难以获取.IBM公司Gruhl等人提出了模拟博客阅读和撰写博文行为的主题传播模型^[17],进而提出了类似EM算法的非监督参数估计算法,根据博文发表时间的先后次序估计出博客间的影响概率,建立隐含的主题传播网络.2010年,斯坦福大学计算机系的研究小组发表了从多组时间序列中推断出隐含传播网络结构的论文^[18,19].该工作推广了信息传播网络推断问题,使其能够描述流行病传播,并针对独立级联模型给出一种基于子模块(submodular)贪心优化的近似学习算法,可推断出近似最优的隐含传播网络.在此基础上,该小组又相继提出了缺失观测数据的处理方法^[20]和基于凸规划的传播网络的推断算法^[21,22].

以上工作从算法和实验的角度验证了从传播数据中学习传播网络的可行性,可为研究流行病传播网络推断问题提供参考.然而现有的传播网络推断方法不能有效应用于流行病监控数据.其主要原因如下:

- ① 现有方法所能处理的传播数据与可获得的流行病监控数据形式不同.现有方法处理的传播数据都是基于个体(individual based)的时间序列,以“某个体在某时间被感染”的方式依次记录信息在网络上的传播过程.然而,如此细粒度的流行病监控数据通常很难获得,各级疾病预防控制中心(Centers for Disease Control,简称CDC)和各地区医院提供的流行病监控数据通常是基于空间位置(location based)的时间序列,以“某区域在某个时间段内感染、出院或死亡多少人”的方式记录流行病的传播和爆发过程.
- ② 现有传播网络推断方法假设被观察到的传播数据足够多,需要多次传播过程对应的时间序列数据才能较为准确地推断传播网络结构.然而在实际应用中,可获得的流行病监控数据通常十分有限.
- ③ 流行病监控数据具有空间和时间上的多尺度(multi-scale)特性(具体分析详见后文),这些特性都为现有的传播网络推断方法带来困难.
- ④ 现有传播网络推断方法主要针对信息传播过程,大多采用级联和阈值模型建模信息的传播机制,难以扩展到SIR、SIS或SEIR等模型,而这些模型非常适合于刻画流行病的传播和爆发过程.

从以上角度出发,本文提出一种基于自治计算的流行病传播网络建模和推断方法.该方法不仅能够从流行病监控数据中推断出描述流行病传播途径和方式的网络结构,而且能够推断出与流行病相关的主要生物学参数.

本文第1节提出基于自治计算的流行病传播网络建模和推断算法.第2节通过实验对所提出的模型和算法进行验证和评价,并以流行病风险评估为例说明所提出的模型和方法的应用.第3节给出全文工作的总结,并讨论拟进一步开展的研究工作.

1 模型与算法描述

流行病的传播过程是复杂系统上的动力学(dynamics)过程.自治计算(autonomy oriented computing,简称AOC)^[23,24]为建模和分析复杂系统及发生在其上的动力学行为提供一种新的理论和方法,采用自底向上的白盒方法分析和发现隐藏在复杂现象背后的基本机制和规律.多Agent计算与自治计算都是自底向上的系统分析方法.多Agent计算强调Agent根据自己的信念、愿望和意图改变自身状态和行为,与之不同的是,自治计算主要用于刻画自然界与人工系统中的自组织、正负反馈和涌现等自治行为的基本特征.为此,我们采用自治计算方法建模流行病传播网络.具体而言,采用自治计算的机制复现(AOC-by-fabrication)和原型构建(AOC-by-prototyping)机制建模流行病的传播网络,模拟流行病的传播动力学行为,采用自治计算的自发现机制(AOC-by-self-discovery)从流行病监控数据推断出传播网络的主要参数.

一个典型的 AOC 系统包括 4 个主要部分:自治体(agent 或 entity)、自治体的运行环境(environment)、系统的涌现行为(emergent behavior)和系统目标(objectives).针对流行病传播问题,本文建立的 AOC 系统称为 D-AOC 系统(diffusion-AOC).在 D-AOC 系统中,自治体表示人;自治体的运行环境表示为疾病传播网络,网络节点表示包含若干同质(homogeneous)自治体的空间区域或者复合群体(meta-population),网络链接表示疾病在区域或群体间的传播途径和方式;系统的涌现行为表示为传播动力学行为.D-AOC 系统的主要目标是:

- (1) 建模流行病的传播网络,模拟流行病的传播动力学行为;
- (2) 根据流行病监控数据推断出传播网络结构、感染率、潜伏期和恢复期等模型参数.

D-AOC 系统的第 1 个子目标通过自治计算的机制复现和原型构建机制实现.在此过程中,首先对真实系统的工作机制进行一定程度的认知,然后在模型设计的过程中进行简化和抽象,利用真实世界中的自组织集群行为设计和构建仿真模型,模拟自治体的状态转换和交互行为,并通过一个多次尝试的过程修正模型,最终使其变得足够细致和逼真.

D-AOC 系统的第 2 个子目标通过自治计算的自发现机制实现.在此过程中,通过正或负反馈过程来调节系统参数,改变个体的自治行为,从而逐步缩小观察到的真实系统涌现行为与当前模拟系统涌现行为之间的差异,促使模拟系统向真实系统逐步演化.实现反馈机制的一种有效方法是群智能优化算法.在群智能优化算法控制下完成的反馈过程,能够使模型中的参数根据观察数据的动态变化自适应地调整.下面具体说明如何基于自治计算构建传播网络模型,以及如何基于观察到的监控数据,自适应地估计出模型参数.

1.1 D-AOC 系统的环境

D-AOC 系统的环境建模为一个刻画流行病传播途径和方式的传播网络.假定所讨论的传播网络包含 N 个节点,传播网络表示为一个 N 维矩阵 $A=[a_{ij}]_{N \times N}$,其中,网络链接权重 a_{ij} 抽象刻画了网络节点间的局部交互作用对全局传播动力学行为的影响程度,满足 $0 \leq a_{ij} \leq 1$ 和 $\sum_{j=1}^N a_{ij} = 1$.对于不同类型的流行病,抽象的网络链接能够刻画不同的传染途径和方式,如 SARS、A/H1N1 和流行性感冒等通过空气(air-bound)传播疾病的方式,疟疾通过蚊虫(vector-bound)传播疾病的方式.

在传播网络中,每个节点可以表示一个空间区域(如一个村庄或一所学校),也可以表示一个复合群体(如一户人家或者一个年龄段的人口).通过对网络节点的细致划分,可使其满足同质或近似同质的假设.即每个网络节点包含一组同质个体,这些个体具有相似的生物学特征(如对疾病的抵抗力、感染疾病的潜伏期、感染疾病后的恢复能力)和相似的社会学特征(如日常移动规律、与其他个体的接触和交互频率).为了满足节点同质的假设,传播网络可以组织成一个层次结构,低层节点嵌套于高层节点中,使得低层节点满足同质假设,高层节点近似满足同质假设.传播网络的层次结构不仅可以刻画空间区域的多尺度(multi-scale)特性(如,从底层至高层的节点分别表示户、村、乡、镇、县等具有不同大小和行政级别的空间区域),而且可以刻画某些流行病的再次爆发(resurgent)行为(在一段时间内反复出现多次感染高峰).由于从监控数据推断出多层传播网络结构比推断出单层传播网络结构更为复杂,限于篇幅,本文将重点讨论单层传播网络结构的推断方法.

传播网络中的每个节点具有多种状态,采用数字信号表示.令 $S_i^{(t)}$, $E_i^{(t)}$, $I_i^{(t)}$ 和 $R_i^{(t)}$ 分别表示节点 i 在 t 时刻处于易感(susceptive)、潜伏(exposed)、感染(infected)和恢复(recovered)等状态的人口.节点的某些状态可被观测和记录,如节点的感染和康复状态可通过医院提供的就诊、出院和死亡记录获得;而某些状态(如,易感和潜伏)则属于隐含状态,难以被观测和记录.根据传播机制和传播网络结构,传播网络中的每个节点状态会发生动态改变,可被观测状态随时间变化的数字信号共同构成了监控数据集(surveillance data).

1.2 D-AOC 系统的自治体

D-AOC 系统的自治体定义为

$$agent=(id,c,s,\beta,\gamma,\tau,actions),$$

其中,分量分别表示自治体所在网络节点的标识符、自治体维护的时钟、自治体当前所处的状态、自治体的 3 个生物学特征以及自治体可执行的动作集合.D-AOC 系统中的时钟有两种:系统时钟和个体时钟.系统时钟 t 模

拟真实时间,其单位可以表示不同的时间粒度(如,天、周、月等),用于同步系统中的所有自治体.个体时钟 c 表示自治体当前状态的持续时间,用于控制各个自治体的状态转换.在系统的运行过程中, c 根据 t 的时间单位累加,但在自治体进入一个新状态时自动归零.

$\beta, 1/\gamma, 1/\tau$ 分别表示自治体对疾病的抵抗力、潜伏期和恢复期; $R_0 = \beta/\tau$ 表示基本再生数(basic reproduction number),刻画自治体在患病期间平均传染的人数.在节点同质的假设下,节点中的自治体具有相似的生物学特征,因此,节点 i 中所有自治体的生物学特征都可表示为 $\beta_i, \gamma_i, \tau_i$.

根据 SEIR 仓室模型(SEIR compartmental model)^[4],自治体在每个时刻处于以下 4 种状态之一: S (易感)、 E (潜伏)、 I (感染)和 R (恢复).自治体的状态转换过程由环境和自治体的生物学特征共同控制,如图 1 所示.其中,虚线和实线表示的状态分别对应于不可观测和可观测状态.自治体的状态转换规则描述如下:

规则 1. 当自治体处于 S 状态时,以 $\lambda_i^{(t)}$ 的概率进入 E 状态. $\lambda_i^{(t)}$ 称为节点 i 在 t 时刻的感染强度(infection force),由自治体自身对疾病的抵抗力 β_i 和环境(传播网络中其他自治体的状态)共同决定,具体分析详见后文.

规则 2. 当自治体处于 E 状态已经持续 $1/\gamma_i$ 时间后进入 I 状态.

规则 3. 当自治体处于 I 状态已经持续 $1/\tau_i$ 时间后进入 R 状态.

规则 4. 当自治体进入 R 状态后则永久停留在该状态,意味着它获得了对流行病的终身免疫.

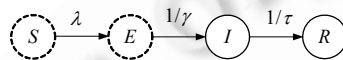


Fig.1 SEIR state transition model of the agent in D-AOC system

图 1 D-AOC 系统中自治体的 SEIR 状态转换模型

为定义自治体的动作,我们引入反映-扩散模型(reaction-diffusion model,简称 RD 模型)^[25,26].如图 2 所示,在 RD 模型中,自治体的动作包括两种:diffusion(扩散)和 reaction(反应).根据传播网络节点表示的具体含义,diffusion 动作可对应两种不同的解释.当网络节点表示空间区域时,diffusion 表示 move(移动),即自治体可从一个空间移动到其他空间;当网络节点表示复合群体时,diffusion 表示 select(选择),即自治体可选择一个群体进行交互.在两种情况下,reaction 都表示同质混合(homogeneous mixing),即自治体和当前节点中的自治体进行等概率交互,交互过程中可能导致疾病传播.

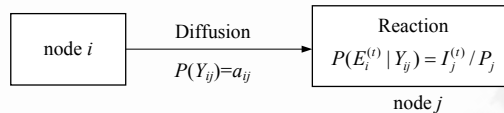


Fig.2 Reaction-Diffusion action model of the agent in D-AOC system

图 2 D-AOC 系统中自治体的 Reaction-Diffusion 动作模型

D-AOC 系统中,节点 i 在 t 时刻的感染强度(infection force)定义为如下的条件概率:

$$\begin{aligned}
 \lambda_i^{(t)} &= P(\text{infected} | \text{susceptible}) \\
 &= P(\text{infected} | \text{susceptible, exposed})P(\text{exposed}) \\
 &= \beta_i P(E_i^{(t)})
 \end{aligned}
 \tag{1}$$

其中, $P(E_i^{(t)})$ 表示节点 i 中自治体在 t 时刻与一个感染者进行交互的概率, β_i 表示节点 i 中的易感自治体与感染者交互后被感染的概率. β_i 刻画了节点 i 中个体对疾病的平均抵抗力.

若不考虑网络效应,传播网络中的每个节点被视为一个封闭世界,则 $P(E_i^{(t)})$ 仅与 t 时刻节点 i 中处于感染状态的总人数 $I_i^{(t)}$ 有关;若考虑网络效应,计算 $P(E_i^{(t)})$ 时则应同时考虑来自节点 i 内和网络中所有其他节点的感染威胁.由以上定义的 RD 模型可知,自治体首先扩散到某个节点,然后与该节点中的自治体进行同质混合.由全概率公式可得:

$$P(E_i^{(t)}) = P(E_i^{(t)} \wedge (\bigvee_{j=1}^N Y_{ij})) = \sum_{j=1}^N P(E_i^{(t)} | Y_{ij}) P(Y_{ij}) \quad (2)$$

其中, Y_{ij} 表示节点 i 中自治体扩散到节点 j 中的随机事件, $P(E_i^{(t)} | Y_{ij})$ 表示 i 中自治体扩散到节点 j 后与 j 中感染自治体发生交互的条件概率. 由于假设自治体与节点 j 中自治体进行同质混合, 节点 i 中自治体扩散到节点 j 后与其中感染自治体进行交互的概率正比于 j 中感染的总人数, 即

$$P(E_i^{(t)} | Y_{ij}) = I_j^{(t)} / P_j^{(t)} \quad (3)$$

其中, $P_j^{(t)}$ 表示 t 时刻节点 j 中的自治体总数.

$P(Y_{ij})$ 表示节点 i 中个体扩散到节点 j 中的概率, 或者说, i 中个体选择节点 j 作为感染源的概率. 实际上, $P(Y_{ij})$ 刻画了疾病的传播方式以及传播网络中每对节点间局部传播效应对全局传播动力学的影响程度, 我们将其定义为传播网络中节点 i 与节点 j 之间链接权重, 即

$$P(Y_{ij}) = a_{ij} \quad (4)$$

当节点表示空间区域时, a_{ij} 表示自治体在不同空间的移动概率; 当节点表示复合群体时, a_{ij} 可理解为不同群体间的接触概率. 基于以上分析, 我们有

$$\lambda_i^{(t)} = \beta_i P(E_i^{(t)}) = \beta_i \sum_{j=1}^N P(E_i^{(t)} | Y_{ij}) P(Y_{ij}) = \beta_i \sum_{j=1}^N \frac{I_j^{(t)}}{P_j^{(t)}} a_{ij} \quad (5)$$

根据以上对自治体状态和动作的讨论, 图 3 给出了节点 i 中自治体的生命周期.

```

/*初始化阶段*/
1. agent.id ← i;
2. agent.c ← 0;
3. agent.s ← 'S' or 'I';
4. agent.β ← βi; agent.γ ← γi; agent.τ ← τi;
/*活动阶段*/
5. 执行 diffusion 动作, 根据概率 aij 扩散到节点 j;
6. 执行 reaction 动作, 与节点 j 中的自治体进行同质混合;
7. 执行状态转移
7.1 IF agent.s == 'S' THEN
    计算当前时刻 t 下的感染强度 λi(t);
    以概率 λi(t) 从 S 状态进入 E 状态;
7.2 IF agent.s == 'E' THEN
    IF agent.c == 1/γi THEN 进入 I 状态;
7.3 IF agent.s == 'I' THEN
    IF agent.c == 1/τi THEN 进入 R 状态;
7.4 IF agent.s == 'R' THEN 继续保持 R 状态;
8. 修改时钟
IF 自治体进入了新状态 THEN
    agent.c ← 0;
ELSE
    agent.c ← agent.c + 1;

```

Fig.3 Life-Cycle of the agent within node i

图 3 网络节点 i 中自治体的生命周期

1.3 D-AOC系统的涌现行为

D-AOC 系统的涌现行为表示为疾病在传播网络上的传播动力学行为, 由流行病监控数据集刻画. 如图 4 所示(其中, “√”和“?”分别表示可获得数据和缺失数据), 时间窗 $[1, T]$ 中的流行病监控数据可表示为一个 $T \times N$ 的二维矩阵 $D = (D^{(1)}, \dots, D^{(T)})$, 其中, 第 t 个列向量表示为 $D^{(t)} = (d_1^{(t)}, \dots, d_N^{(t)})^T$, 列向量中的每个分量 $d_i^{(t)}$ 对应一个二元组 $(c_i^{(t)}, r_i^{(t)})$, 分别表示 t 时刻节点 i 中所观察到的新增病例数目(可由入院和确诊数据估计)和恢复个体数目(可由出院和死亡数据估计). 在实际应用中, 监控数据中 D 通常具有如下几个特性:

① 时空特性

监控数据 D 由 N 个时间序列构成, D 矩阵的第 i 行对应在节点 i 上观察到的时间序列, 记录了节点 i 所表示的空间区域或复合群体新增病例和恢复人数随时间的变化情况. 值得说明的是, 与信息传播网络学习方法分析的数据不同, D 表示的时间序列是基于群体而非个体的. 基于个人的时间序列由多个二元组 (r, t) 构成, 记录不同个体 r 的具体感染时间 t . 然而对于疾病控制而言, 各级疾病预防控制中心(CDC)和各地的医院很难提供如此详尽的监控数据, 只能提供某段时间内感染、恢复和死亡人数的粗略统计数据.

② 时空多尺度特性

传播网络中的节点可具有不同的空间大小, 如图 4 所示, 一些节点对应于村和乡等较小的空间区域, 另一些节点对应于县和市等较大的空间区域. 不同空间尺度的区域对应的时间序列可能具有不同的时间粒度, 如图 4 所示, 村和乡等小尺度区域以天或者周为时间单位普查和统计流行病数据, 而县或市等大尺度区域以月或者季度为时间单位普查和统计流行病数据.

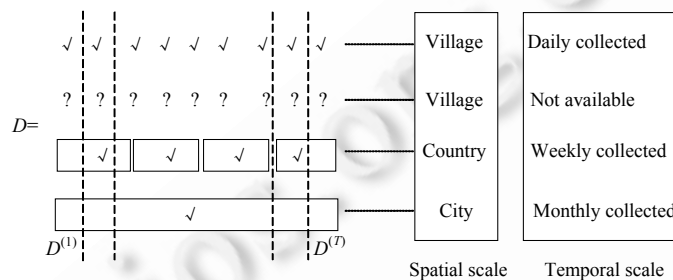


Fig.4 Surveillance data depicting the emergent behavior of D-AOC system

图 4 刻画 D-AOC 系统涌现行为的流行病监控数据

③ 数据缺失

实际的监控数据常常存在数据缺失的情况, 大体上可分为 3 种情况:

第一, 为了降低获取监控数据的成本, 通常不对所有节点而仅对部分节点进行监控和普查, 因而会造成某些节点对应的时间序列数据整组缺失. 如图 4 中第 2 行所示, 该节点对应的时间序列全部缺失.

第二, 如特性②中所讨论的, 空间范围大的节点仅提供时间粒度较大的数据, 造成细粒度数据缺失. 如某些节点仅能提供以周或月为单位的统计数据, 而非每天的统计数据.

第三, 某些情况下, CDC 和医院仅提供新增病例的统计数据, 而缺失有关恢复人数的统计数据.

监控数据虽然可能缺失, 但它们对流行病传播具有实实在在的影响. 因此, 一个好的传播网络推断方法应该能够考虑缺失数据的各种情况, 不仅能够基于现有监控数据推断出传播网络结构, 而且能够同时猜测出缺失的传播数据.

1.4 D-AOC系统的目标

针对 D-AOC 系统的第 1 个目标(建模流行病的传播网络和模拟流行病的传播动力学行为), 第 1.1 节~第 1.3 节给出了相应的自治体模型、传播网络模型和系统涌现行为的描述. 根据以上模型, D-AOC 系统的第 2 个目标(根据流行病监控数据推断出传播网络结构和重要模型参数)可形式化地描述如下:

给定自治体模型和刻画流行病传播动力学行为的监控数据 D , 推断出传播网络 $A=[a_{ij}]_{N \times N}$ 和网络中各节点的生物学参数 (β, γ, τ) .

特别地, 当传播网络中的每个节点仅包含 1 个自治体时, 所估计出的传播网络即为描述博客/个体交互关系的信息传播网络/社会网络. 因此, 本文提出和研究传播网络结构推断问题是对现有信息传播网络结构学习问题的一个推广. 用于估计 D-AOC 系统参数 (A, β, γ, τ) 的自发现过程基于蒙特卡罗模拟的负反馈机制实现, 基本过程描述如图 5 所示.

1. 随机初始化系统参数(A, β, γ, τ);
2. 基于当前参数多次运行 D-AOC 系统,获得该系统涌现行为的期望值 \hat{D} ;
3. 如果 \hat{D} 足够逼近真实监控数据 D ,则停止算法并返回当前的系统参数;
4. 采用某种负反馈机制修正当前系统参数;
5. 返回步骤 2.

Fig.5 Self-Discovery process of D-AOC system

图 5 D-AOC 系统的自发现过程

由于存在随机性,即使在同一组参数下,每次运行 D-AOC 系统得到的模拟结果也会有所不同,因此需要多次运行得到涌现行为的期望值用于分析.当自治体数目很多时,多次运行 D-AOC 系统的蒙特卡罗模拟很耗时,为此,我们采用如下状态转换方程近似求解 D-AOC 系统涌现行为的期望值 \hat{D} .对于节点 i 有:

$$\begin{cases} \Delta S_i = -\lambda_i^{(t)} S_i^{(t)} \\ \Delta E_i = \lambda_i^{(t)} S_i^{(t)} - \gamma_i E_i^{(t)} \\ \Delta I_i = \gamma_i E_i^{(t)} - \tau_i I_i^{(t)} \\ \Delta R_i = \tau_i I_i^{(t)} \end{cases} \quad (6)$$

其中, $\lambda_i^{(t)}$ 是由公式(5)计算的感染强度.

图 6 给出了一个实例,给出了 3 次运行 D-AOC 系统所得到的爆发曲线(分别对应图中的 first simulation, second simulation 和 third simulation)和采用公式(6)得到的爆发曲线(对应图中 approximate expectation).其中, x 轴表示时间单位, y 轴表示单位时间内新增感染人数的比例.从中可看出,各次的模拟结果围绕在近似期望值周围,但都具有不同程度的偏差.

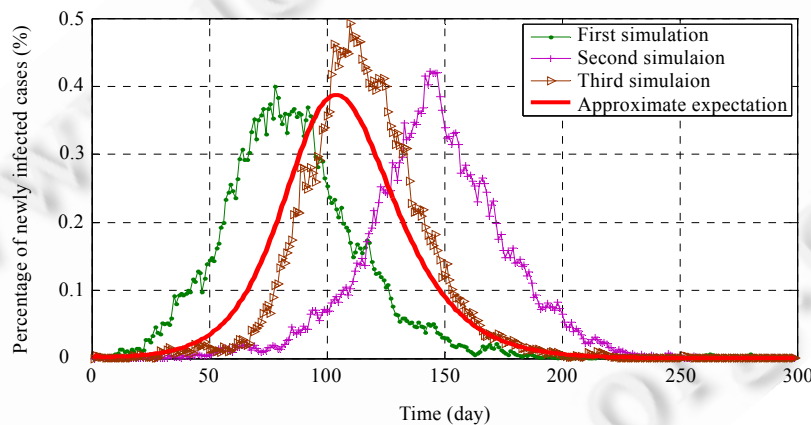


Fig.6 Emergent behavior of the D-AOC system and its approximate expectation

图 6 D-AOC 系统的涌现行为及其近似期望值

本文采用粒子群优化算法作为调节参数的负反馈机制,基本过程描述如图 7 所示.

2 实验

2.1 基于真实数据的模型验证

首先检验所提出的传播网络模型的合理性.实验数据采用 2009 年 H1N1 猪流感在香港爆发的真实监控数据^[27],该数据记录了 2009 年 5 月~2010 年 3 月累计病例的变化情况.如图 8 所示,5 月 1 日~6 月 12 日,已出现 H1N1 患者,但并未开始大规模传播,6 月 12 日之后感染人数呈快速增加.在 H1N1 爆发的中后期,香港政府进行了患者隔离和疫苗接种等措施,对 H1N1 传播进行了有效控制,新增病例逐渐减少.由于当前的 D-AOC 模型没有考虑人为控制措施对流行病传播的影响,因此我们选用 H1N1 爆发过程中人为干预较少的前 3 个月数据来验证模型的

合理性.

- ```

/*初始化过程*/
1. 设置 m 个粒子,每个粒子 P_i 的位置表示一组系统参数 (A, β, γ, τ) ;
2. 初始化每个粒子 P_i 的位置和速度 v_i ;
3. 初始化参数 w, c_1, c_2 ;
/*参数的调整过程*/
4. 对每个粒子 i 执行蒙特卡罗模拟:
 4.1 根据公式(6)计算 D-AOC 系统涌现行为的近似期望值 \hat{D}_i ;
 4.2 计算 \hat{D}_i 偏离 D 的程度 $E_i = \|D - \hat{D}_i\|_2$;
5. 根据各个 E_i 为每个粒子 i 挑选局部最优粒子 LP_i 和全局最优粒子 GP ;
6. 对每个粒子 i 调整速度和位置:
 6.1 $v_i \leftarrow wv_i + c_1 \cdot \text{rand} \cdot (LP_i - P_i) + c_2 \cdot \text{rand} \cdot (GP - P_i)$ //rand 是随机数发生器
 6.2 $P_i \leftarrow P_i + v_i$
7. 返回步骤 4,直至达到规定的迭代次数或者 E_i 小于规定的阈值.

```

Fig.7 PSO-Based self-discovery process of D-AOC system

图 7 基于粒子群优化的 D-AOC 系统自发现过程

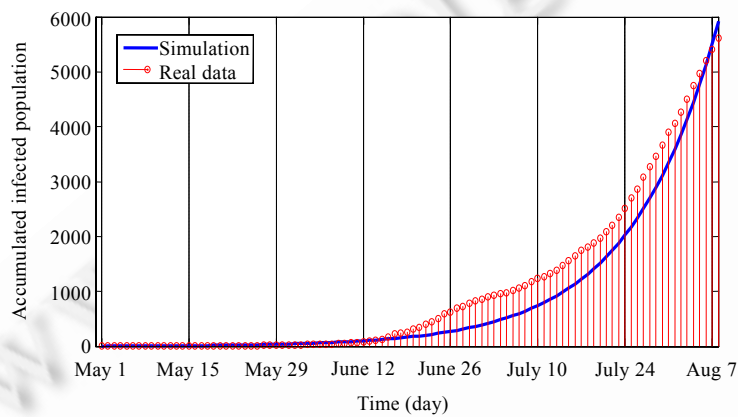


Fig.8 Real data of H1N1 swine-flu epidemic in Hong Kong during 2009 and corresponding simulation of the D-AOC system

图 8 2009 年香港地区 H1N1 猪流感传播的真实数据与 D-AOC 系统的模拟数据

根据年龄,香港最受 H1N1 影响的人口可划分为 5 个年龄组:5~14 岁( $A_1$ ),15~24 岁( $A_2$ ),25~44 岁( $A_3$ ),45~64 岁( $A_4$ ),65 岁及以上( $A_5$ ).相应地,传播网络节点表示为由对应年龄组构成的复合群体,网络链接表示复合群体间的交互概率.基于文献[27-31]等工作讨论的 H1N1 生物学参数(如不同年龄段的感染率、潜伏期和恢复期等)以及不同年龄组的交互频率,我们对 D-AOC 模型参数进行了设置,见表 1.基于这些参数进行模拟得到累计新增病例的近似期望值,如图 8 中 Simulation 曲线所示.从中可以看出,该 D-AOC 系统能够较好地模拟 H1N1 前 100 天感染人数的增长趋势,但由于数据噪声和人为干预的影响,模拟结果距离真实数据还存在一定误差.以下实验采用表 1 中的参数验证所提出的模型和算法的其他方面.

Table 1 Parameters of D-AOC system for simulating H1N1 swine-flu epidemic in Hong Kong during 2009

表 1 模拟 2009 年 H1N1 猪流感在香港爆发过程的 D-AOC 系统参数

| 传播网络          | $A_1$   | $A_2$   | $A_3$   | $A_4$   | $A_5$   | $\beta$ | $\tau$ | $\gamma$ | 人口(万) |
|---------------|---------|---------|---------|---------|---------|---------|--------|----------|-------|
| $A_1$ (5~14)  | 0.640 5 | 0.293 4 | 0.019 0 | 0.028 9 | 0.018 2 | 0.434   | 0.16   | 0.25     | 94    |
| $A_2$ (15~24) | 0.271 8 | 0.413 8 | 0.119 7 | 0.084 2 | 0.110 5 | 0.158   | 0.16   | 0.25     | 91    |
| $A_3$ (25~44) | 0.189 7 | 0.195 0 | 0.341 7 | 0.202 6 | 0.071 1 | 0.118   | 0.16   | 0.25     | 230   |
| $A_4$ (45~64) | 0.084 0 | 0.176 5 | 0.218 8 | 0.325 6 | 0.195 1 | 0.046   | 0.16   | 0.25     | 186   |
| $A_5$ (65+)   | 0.111 5 | 0.159 6 | 0.160 5 | 0.245 2 | 0.323 3 | 0.046   | 0.16   | 0.25     | 85    |

## 2.2 爆发趋势和模型参数估计

本节采用如下实验方案验证 D-AOC 系统的参数估计方法:

- ① 首先,给定一组模型参数  $\theta = \langle A, \beta, \gamma, \tau \rangle$ ;
- ② 基于  $\theta$  运行 D-AOC 系统生成流行病的实际爆发曲线(感染和恢复病例的时间序列),并从中采样作为观测到的流行病监控数据  $D$ ;
- ③ 基于  $D$ ,采用图 7 中方法估计参数  $\hat{\theta}$  和爆发曲线  $\hat{D}$ ;
- ④ 对比实际值与估计值,分析爆发趋势和参数的估计情况.

为了在一定程度上消除随机性对参数估计的影响,步骤②以多次运行(如 20 次)D-AOC 系统的期望值作为实际爆发曲线.如图 9(a)所示,其中的条形图表示通过步骤②得到实际爆发曲线(新增病例期望值的时间序列),曲线表示通过步骤③得到估计爆发曲线(新增病例估计期望值的时间序列).为了度量估计爆发曲线与实际爆发曲线的趋势是否吻合,我们采用公式(7)计算  $\hat{D}$  与  $D$  的趋势偏差  $E(D, \hat{D})$ :

$$E(D, \hat{D}) = 1 - \frac{D \cdot \hat{D}^T}{|D| \cdot |\hat{D}|} \quad (7)$$

其中,  $0 \leq E(D, \hat{D}) \leq 1$ .

本实验采用表 1 中参数作为待估计的参数  $\theta$ .图 9(b)给出了 D-AOC 系统的自发现过程,随着迭代步数的增加,当前参数下估计的爆发曲线  $\hat{D}$  逐步逼近观察值  $D$ ,直至两者的偏差小于规定阈值.从图中可看出,通过自发现过程演化而成的 D-AOC 系统能够较好地模拟监控数据所描述的流行病爆发趋势(偏差为  $10^{-5}$  级).表 2 给出了自发现过程结束时得到的参数估计值  $\hat{\theta}$ .比较表 1 和表 2 可知,图 7 中的估计方法能够较为准确地估计出各网络节点的生物学参数,但由于 D-AOC 系统的随机性对监控数据生成的影响较大,导致部分链接权重的估计值距离实际值还存在一定的偏差.

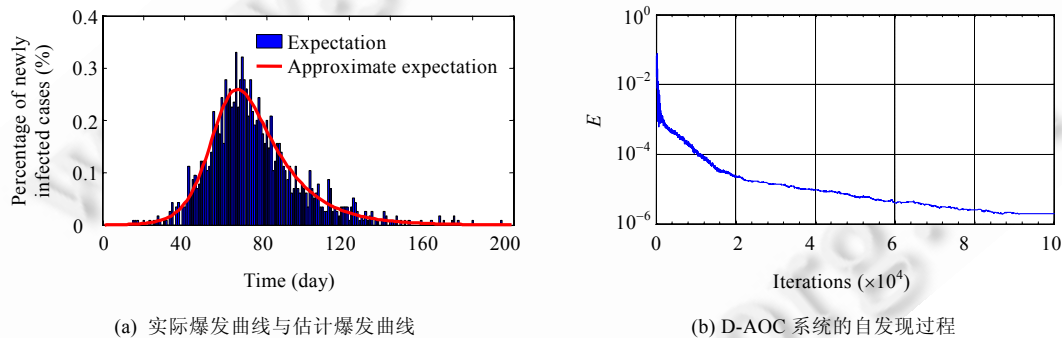


Fig.9 Estimating epidemic trend and the parameters of the D-AOC

图 9 流行病爆发趋势和 D-AOC 参数估计

Table 2 Estimated parameters

表 2 估计出的参数值

| 传播网络          | $A_1$   | $A_2$   | $A_3$   | $A_4$   | $A_5$   | $\beta$ | $\tau$  | $\gamma$ |
|---------------|---------|---------|---------|---------|---------|---------|---------|----------|
| $A_1$ (5~14)  | 0.486 0 | 0.204 3 | 0.019 5 | 0.197 1 | 0.093 2 | 0.566 5 | 0.161 3 | 0.266 2  |
| $A_2$ (15~24) | 0.191 9 | 0.254 6 | 0.027 4 | 0.197 3 | 0.328 8 | 0.227 1 | 0.161 3 | 0.266 2  |
| $A_3$ (25~44) | 0.126 9 | 0.200 5 | 0.000 0 | 0.405 7 | 0.266 9 | 0.174 2 | 0.161 3 | 0.266 2  |
| $A_4$ (45~64) | 0.085 4 | 0.164 6 | 0.280 5 | 0.349 4 | 0.120 1 | 0.044 8 | 0.161 3 | 0.266 2  |
| $A_5$ (65+)   | 0.169 6 | 0.292 8 | 0.292 6 | 0.171 7 | 0.073 3 | 0.031 2 | 0.161 3 | 0.266 2  |

## 2.3 缺失数据对爆发趋势估计的影响

由第 1.3 节的讨论可知,实际可得的监控数据可能会有不同程度的缺失,缺失数据虽不可见,对流行病传播

的影响却实际存在.基于缺失监控数据估计出流行病的爆发趋势后,可从估计数据中猜测出缺失的传播数据.本节研究不同情况下的数据缺失对爆发趋势估计产生的影响,具体实验方案如下:

- 同上一节实验,首先采用表 1 中的参数运行 D-AOC 系统,生成实际爆发曲线,进而采用不同时间尺度从中采样,得到监控数据;
- 基于采样所得的监控数据估计参数和爆发曲线;
- 根据公式(7)计算估计所得的爆发趋势与实际爆发趋势的偏差,进而分析不同情况的数据缺失对爆发趋势估计产生的影响.

实验中讨论的采样策略分别是:

- ① 对每个网络节点对应的爆发曲线每天采样;
- ② 对每个网络节点对应的爆发曲线以 2 天为间隔采样(即获取 2 天数据之和);
- ③ 对每个网络节点对应的爆发曲线以 7 天为间隔采样(即获取 7 天数据之和);
- ④ 对 5 个网络节点对应的爆发曲线分别以 1 天、2 天、7 天、15 天和 30 天为间隔采样.

第 1 种采样策略对应无数据缺失的情况,作为基准对比数据缺失情况.图 10(a)给出了不同采样策略对应的参数估计过程,从中可看出:① 相对于无数据缺失的情况(对应图中 Sampling I 曲线,图中曲线自下而上依次为 Sampling I, Sampling II, Sampling III, Sampling IV),基于缺失数据估计出的爆发趋势与实际爆发趋势具有一定的偏差;② 采样的时间尺度越大,估计出的爆发趋势与实际爆发趋势的偏差越大,但都控制在  $10^{-2}$  以下,这意味着即使监控数据具有不同程度的缺失,D-AOC 的自发现过程也可以较为准确地估计出流行病的爆发趋势.

在某些情况下,CDC 和医院仅能提供新增病例的统计数据,而缺失恢复病例的统计数据.为此,我们还进一步讨论了恢复病例数据缺失对爆发趋势估计的影响.如图 10(b)所示,相对于无数据缺失的情况,在缺失恢复病例数据的情况下估计出的爆发趋势与实际爆发趋势具有较大的偏差,但仍控制在  $10^{-3}$  以下.

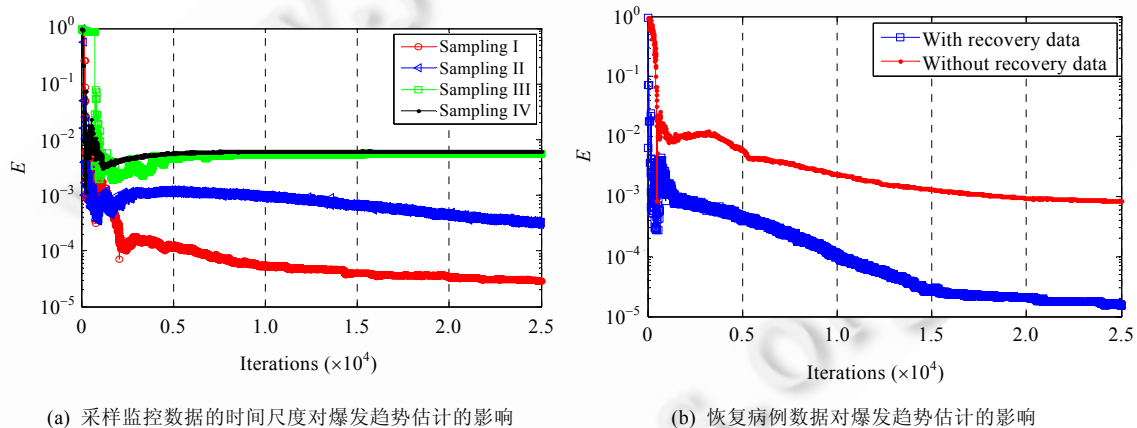


Fig.10 Influence of missing data on the estimation of epidemic trend

图 10 缺失数据对爆发趋势估计的影响

### 2.4 时间窗对爆发趋势估计的影响

流行病传播具有实时性,因而希望快速且准确地估计出流行病的爆发趋势,从而尽早制定相应的防范措施.如第 1.3 节讨论, $[1, T]$ 表示获取流行病监控数据的时间窗,本节研究  $T$  的取值对爆发趋势估计的影响.如果能够以较小的时间窗估计出较为准确的爆发趋势,则意味着能够在爆发高峰之前准确地估计出流行病的爆发趋势.

具体实验方案如下:

- 首先采用表 1 中的参数运行 D-AOC 系统,生成实际爆发曲线,进而采用不同时间窗 $[1, T]$ 从中采样,得到监控数据;

- 基于采样所得的监控数据估计 $[T, T+\Delta t]$ 时期内的爆发趋势;
- 根据公式(7)计算估计所得的爆发趋势与实际爆发趋势的偏差,进而分析不同 $T$ 和 $\Delta t$ 的取值对爆发趋势估计产生的影响.

图 11 中,粗实线表示流行病的爆发趋势,图中其他 3 条曲线分别表示 $\Delta t$  取值 60 天、90 天和 120 天对应的估计偏差.从中可看出:

- ① 可获得的监控数据越多( $T$  越大),估计出的传播趋势越准确;
- ② 预测的时间越短( $\Delta t$  取值越小),估计出的传播趋势越准确;
- ③ 在传播初期( $T < 114$ ),感染人口比例小,爆发趋势不明显,基于当前监控数据估计出的传播趋势具有较大偏差(大于 0.3);
- ④ 而当传播过程刚刚进入开始明显增长的阶段( $T > 134$ )时,基于当前监控数据能够较好地估计出今后 2 个月~3 个月的爆发趋势(偏差小于 0.2),这意味着本文算法能够在爆发高峰到来之前较为准确地估计出今后一段时间内的流行病爆发趋势;
- ⑤ 当传播过程离开高峰期后( $T > 214$ ),监控数据累计得足够充分,无论 $\Delta t$  取值大小,都能够较好地估计出后期的爆发趋势.

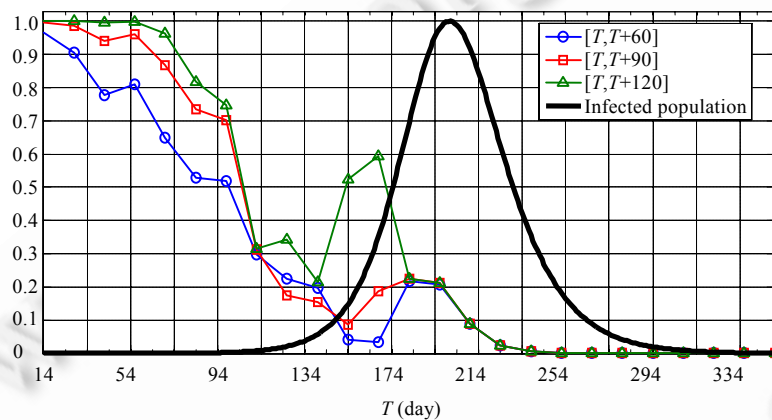
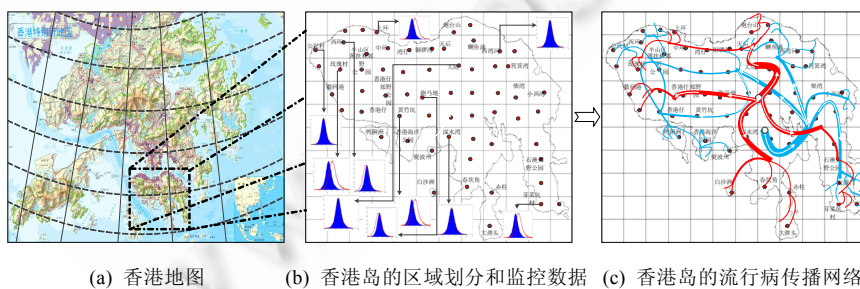


Fig.11 Influence of time window on the estimation of outbreak trend

图 11 时间窗对爆发趋势估计的影响

## 2.5 流行病感染风险评估

流行病传播网络推断的一个重要应用是流行病感染的风险评估,即基于当前的监控数据估计出不同地区的人口在后续时间内被感染的概率.图 12 和图 13 以香港地区为例说明了基于传播网络推断评估感染风险的基本原理.



(a) 香港地图 (b) 香港岛的区域划分和监控数据 (c) 香港岛的流行病传播网络

Fig.12 Inferring epidemic network from surveillance data

图 12 基于监控数据的流行病传播网络推断

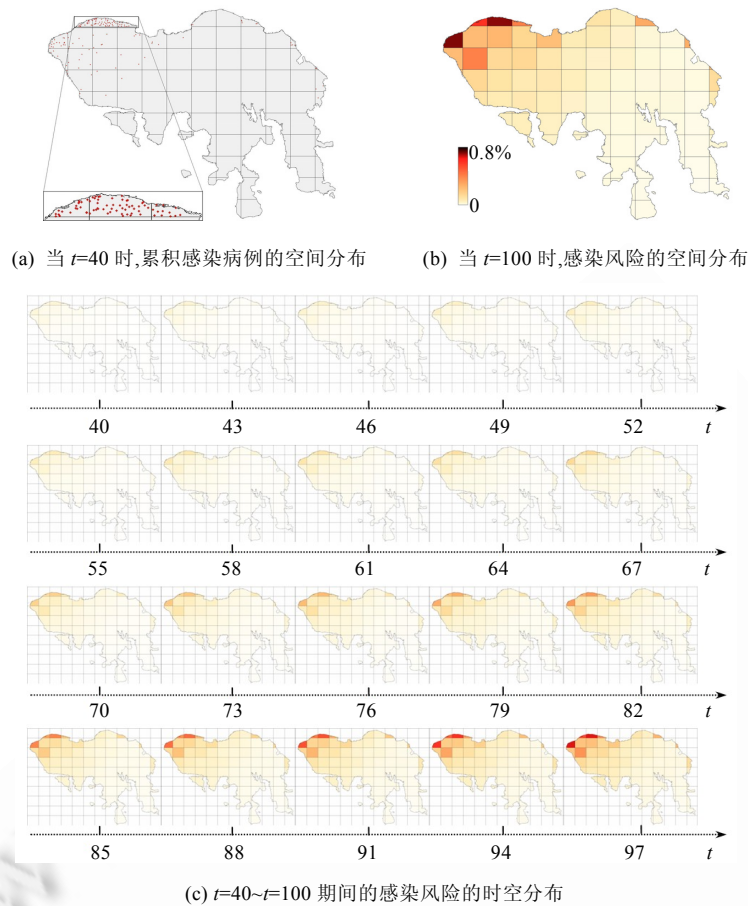


Fig.13 Evaluation of infection risk based on epidemic network inference

图 13 基于传播网络推断的流行病感染风险评估

- ① 首先,利用经纬线对全香港地区进行分区,形成层次网络的第 1 层,如图 12(a)所示.
- ② 由于该分区粒度大,不能满足同质假设,为此,需要对第 1 层中的每个分区再次划分得到层次网络的第 2 层.本例选择香港岛,将其进一步细分为  $12\times 12$  个小区域,使得每个小区域近似满足同质假设,如图 12(b)所示.图中每个区域中心的实心点表示布设的监控点(医院或诊所),提供实时监控数据(图中表示为相应的时间序列).
- ③ 融合各监控点的监控数据推断出传播网络结构(如图 12(c)所示)和模型参数,进而执行 D-AOC 系统模拟一定时间内(如 2 个月内)流行病在香港岛的传播过程.
- ④ 利用模拟数据和公式(8)计算出各地区人口在不同时刻被感染的风险  $r_i^{(t)}$ ,并在地图的相应区域用不同深浅的颜色表示感染风险的大小,如图 13 所示.

### 3 结束语

从流行病监控数据中推断出的流行病传播网络可用于评估流行病在不同时间和空间的传播风险以及拟采取的人工干预策略是否有效.基于现有的条件,可获得的流行病监控数据大都是基于位置的时间序列,“以某空间在某段时间内新增多少病例”的方式粗略地记录流行病的传播过程,具有空间多尺度、时间多尺度和数据缺失等特点,使得现有的信息传播网络推断方法难以胜任.针对该问题,本文提出一种基于自治计算的传播网络建

模和推断方法,采用自治计算的机制复现和原型构建方法建模流行病传播网络和传播过程,采用自发现机制估计模型参数.具体而言,该方法基于流行病模型 SEIR 建模自治体的状态及状态转换规则,基于扩散-反应模型建模自治体的行为,以缩小模拟系统涌现行为和监控数据差异为目标,通过蒙特卡洛模拟结合粒子群优化的负反馈机制调节模型参数,改变自治体行为,使模拟系统不断向真实系统演化,从而推断出导致监控数据出现的传播网络结构和其他模型参数.

采用实际监控数据对所提出的传播网络模型和网络推断方法进行了验证,并从不同方面分析了其有效性和适用性,主要结论如下:

- ① 在监控数据充分的情况下,所提出的方法能够以较高的精度拟合观测到的监控数据,并能较准确地估计各网络节点的主要生物学参数和传播网络的多数权重信息.
- ② 在监控数据部分缺失的情况下,如以不同时间尺度记录爆发过程的监控数据,或者仅能获得爆发过程的新增病例信息而无法获得恢复病例信息,所提出的方法仍然能够较好地估计出流行病爆发趋势,但精度会有不同程度的下降.
- ③ 所提出的方法能够在爆发高峰到来之前较为准确地估计出今后一段时间内(如后 2~3 个月内)的流行病爆发趋势.

本文提出的流行病传播网络建模和推断方法还存在一些局限性,有待在今后的工作中加以深入研究和改进.具体讨论如下:

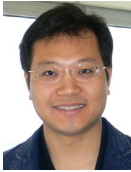
- ① 现有的传播网络模型没有考虑人工干预措施(如隔离和接种疫苗)对传播过程产生的影响,因此仅能较好地模拟出 2009 年 H1N1 在香港爆发的前期过程,在今后的工作中,我们拟考虑这些因素拓展模型,增强自治体的自治能力,一方面使模型能够更好地模拟真实情况,另一方面也使其可用于流行病控制策略的评估.自治体的自治体现在自治体可通过与环境交互改变自身属性或行为,如,接种疫苗可提高自治体对某种疾病的抵抗力,隔离策略可降低自治体移入疫情严重地区的概率.
- ② 从实验中可看出,文中方法能够较为准确地模拟流行病的爆发趋势,但估计出的模型参数(特别是网络权重)还存在一定的误差.我们拟进一步改进参数估计过程,以提高参数估计的准确性.
- ③ 文中仅讨论了单层传播网络的建模和推断方法,但由于监控数据的空间多尺度特性,传播网络组织成多级层次结构更好,不仅可使网络节点近似满足同质假设,而且能够刻画流行病的再爆发现象.相对于单层次传播网络,层次传播网络结构更复杂,待估计的参数更多.如何从监控数据中学习出传播网络的层次结构及权重信息具有很大难度,是我们今后的重点研究内容.

## References:

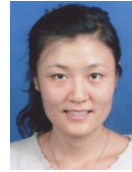
- [1] Lipsitch M, Cohen T, Cooper B, Robins JM, Ma S, James L, Gopalakrishna G, Chew SK, Tan CC, Samore MH, Fisman D, Murray M. Transmission dynamics and control of severe acute respiratory syndrome. *Science*, 2003,300(5627):1966–1970. [doi: 10.1126/science.1086616]
- [2] Birrell PJ, Ketssetzis G, Gay NJ, Cooper BS, Presanis AM, Harris RJ, Charlett A, Zhang XS, White PJ, Pebody RG, Angelis DD. Bayesian modeling to unmask and predict influenza A/H1N1 PDM dynamics in London. *Proc. of the National Academy of Sciences of the United States of America*, 2011,108(45):18238–18243. [doi: 10.1073/pnas.1103002108]
- [3] Germann TC, Kadau K, Longini IM, Macken CA. Mitigation strategies for pandemic influenza in the US. *Proc. of the National Academy of Sciences of the United States of America*, 2006,103(15):5935–5940. [doi: 10.1073/pnas.0601266103]
- [4] Keeling MJ. *Modeling Infectious Diseases in Humans and Animals*. Princeton: Princeton University, 2008. 15–52.
- [5] Vynnycky E, White R. *An Introduction to Infectious Disease Modelling*. Oxford: Oxford University, 2010.
- [6] Grassberger P. On the critical behavior of the general epidemic process and dynamical percolation. *Mathematical Biosciences*, 1983,63(2):157–172. [doi: 10.1016/0025-5564(82)90036-0]
- [7] Newman M, Watts DJ. Scaling and percolation in the small-world network model. *Physical Review E*, 1999,60(6):7332–7342. [doi: 10.1103/PhysRevE.60.7332]

- [8] Moore C, Newman M. Epidemics and percolation in small-world networks. *Physical Review E*, 2000,61(5):5678–5682. [doi: 10.1103/PhysRevE.61.5678]
- [9] Kuperman M, Abramson G. Small world effect in an epidemiological model. *Physical Review Letters*, 2001,86(13):2909–2912. [doi: 10.1103/PhysRevLett.86.2909]
- [10] Yang XS. Fractals in small-world networks with time-delay. *Chaos Solitons & Fractals*, 2002,13(2):215–219. [doi: 10.1016/S0960-0779(00)00265-4]
- [11] Pastor-Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. *Physical Review E*, 2001,63(6):1–9. [doi: 10.1103/PhysRevE.63.066117]
- [12] Pastor-Satorras R, Vespignani A. Epidemic dynamics in finite size scale-free networks. *Physical Review E*, 2002,65(3):1–4. [doi: 10.1103/PhysRevE.65.035108]
- [13] Boguna M, Pastor-Satorras R. Epidemic spreading in correlated complex networks. *Physical Review E*, 2002,66(4):1–5. [doi: 10.1103/PhysRevE.66.047104]
- [14] Boguna M, Pastor-Satorras R, Vespignani A. Absence of epidemic threshold in scale-free networks with connectivity correlations. *Physical Review Letter*, 2003,90(2):028701. [doi: 10.1103/PhysRevLett.90.028701]
- [15] Chakrabarti D, Wang Y, Wang CX, Leskovec J, Faloutsos C. Epidemic thresholds in real networks. *ACM Trans. on Information and System Security*, 2008,10(4):3–26. [doi: 10.1145/1284680.1284681]
- [16] Adar E, Zhang L, Adamic LA, Lukose RM. Implicit structure and the dynamics of blog space. In: *Proc. of the 13th Int'l World Wide Web Conf. (WWW 2004)*. New York, 2004. 17–22.
- [17] Gruhl D, Guha R, Liben-Nowell D, Tomkins A. Information diffusion through blog space. In: *Proc. of the 13th Int'l Conf. on World Wide Web (WWW 2004)*. New York, 2004. 491–501. [doi: 10.1145/988672.988739]
- [18] Gomez-Rodriguez M, Leskovec J, Krause A. Inferring networks of diffusion and influence. In: *Proc. of the 16th Int'l Conf. on Knowledge Discovery and Data Mining (KDD 2010)*. Washington, 2010. 1–10. [doi: 10.1145/1835804.1835933]
- [19] Gomez-Rodriguez M, Leskovec J, Krause A. Inferring networks of diffusion and influence. *ACM Trans. on Knowledge Discovery From Data*, 2012,5(4):3–10.
- [20] Sadikov E, Medina M, Leskovec J, Garcia-Molina H. Correcting for missing data in information cascades. In: *Proc. of the 4th ACM Int'l Conf. on Web Search and Data Mining (WSDM 2011)*. Hong Kong, 2011. 55–64. [doi: 10.1145/1935826.1935844]
- [21] Myers S A, Leskovec J. On the convexity of latent social network inference. In: *Proc. of the 24th Int'l Conf. on Neural Information Processing Systems (NIPS 2010)*. Vancouver, 2010. 1–9.
- [22] Gomez-Rodriguez M, Balduzzi D, Schoelkopf B. Uncovering the temporal dynamics of diffusion networks. In: *Proc. of the 28th Int'l Conf. on Machine Learning (ICML 2011)*. Bellevue, 2011. 1–8.
- [23] Liu JM, Tsui KC. Toward nature-inspired computing. *Communications of the ACM*, 2006,49(10):59–64. [doi: 10.1145/1164394]
- [24] Liu JM, Jin XL, Tsui KC. *Autonomy Oriented Computing: From Problem Solving to Complex Systems Modeling*. New York: Springer-Verlag, 2005.
- [25] Kondo S, Miura T. Reaction-Diffusion model as a framework for understanding biological pattern formation. *Science*, 2010, 329(5999):1616–1620. [doi: 10.1126/science.1179047]
- [26] Vespignani A. Modelling dynamical processes in complex socio-technical systems. *Nature Physics*, 2012,8(1):32–39. [doi: 10.1038/nphys2160]
- [27] Liu JM, Xia S. Toward effective vaccine deployment: A systematic study. *Journal of Medical Systems*, 2011,35(5):1153–1164. [doi: 10.1007/s10916-011-9734-x]
- [28] Knipf DH, Röst G. Modelling the strategies for age specific vaccination scheduling during influenza pandemic outbreaks. *Mathematical Biosciences and Engineering*, 2011,8(1):123–139. [doi: 10.3934/mbe.2011.8.123]
- [29] Wu JT, Ma ESK, Lee CK, Chu DKW, Ho PL, Shen AL, Ho A, Hung IFN, Riley S, Ho LM. The infection attack rate and severity of 2009 pandemic influenza (H1N1) in Hong Kong. *Clinical Infectious Diseases*, 2010,51(10):1184–1191. [doi: 10.1086/652984]
- [30] Miller E, Hoschler K, Hardelid P, Stanford E, Andrews N, Zambon M. Incidence of 2009 pandemic influenza A H1N1 infection in England: A cross-sectional serological study. *The Lancet*, 2010,375(9720):1100–1108. [doi: 10.1016/S0140-6736(09)62126-7]

- [31] Mossong J, Hens N, Jit M, Beutels P, Auranen K, Mikolajczyk R, Massari M, Salmaso S, Tomba GS, Wallinga J, Heijne J, Sadkowska-Todys M, Rosinska M, Edmunds WJ. Social contacts and mixing patterns relevant to the spread of infectious diseases. PLoS Medicine, 2008,5(3):e74. [doi: 10.1371/journal.pmed.0050074]



杨博(1974—),男,河南新乡人,博士,教授,博士生导师,CCF 会员,主要研究领域为多 Agent 系统,数据挖掘,复杂网络分析.



白媛(1986—),女,博士生,主要研究领域为数据挖掘,复杂网络分析.



刘际明(1962—),男,博士,教授,博士生导师,CCF 会员,主要研究领域为 Agent 系统,自治计算,Web 智能.



刘大有(1942—),男,教授,博士生导师,CCF 高级会员,主要研究领域为知识工程与专家系统,Agent 系统,数据挖掘,时空推理.



杨建宁(1987—),男,硕士生,CCF 学生会员,主要研究领域为数据挖掘,复杂网络分析.

www.jos.org.cn

www.jos.org.cn