

复杂网络中初始节点对传染病传播的影响分析

苏长明¹ 陈端兵² 傅彦²

(成都大学信息科学与技术学院 成都 610106)¹

(电子科技大学互联网科学中心 计算机科学与工程学院 成都 611731)²

摘 要 长期以来,传染病对人类造成持续不断的威胁。研究工作者在传染病传播上投入了大量的精力进行研究,比如:传播模型、传播阈值以及影响传播的因素等方面。主要研究复杂网络中初始染病节点如何影响传染病的传播,对 5 种初始染病节点选择方法进行了深入分析并在两个大规模无向网络上进行了仿真实验。结果表明,传播速度以及传染病影响规模不仅与初始染病节点的选择有关,而且与网络结构有关。本研究成果对传染病控制可提供有价值的借鉴与参考作用。

关键词 传染病传播,复杂网络,初始染病节点,传染病控制

Effect of Initial Infected Individuals on the Spread of Epidemics in Complex Networks

SU Chang-ming¹ CHEN Duan-bing² FU Yan²

(College of Information Science and Technology, Chengdu University, Chengdu 610106, China)¹

(Web Sciences Center, School of Computer Science, University of Electronic Science and Technology, Chengdu 611731, China)²

Abstract Infectious diseases incessantly threaten human beings for a long time. Many researchers pay much attention on the spread of epidemics, such as spread model, threshold spread rate, and affecting factors of spread, etc. Different initial infected individuals in the network how to affect the disease spreading was studied in this paper. Five initial infected individuals selecting methods were analyzed and two large-scale undirected networks were used to simulate the disease spreading under different initial selected individuals. Simulation results demonstrate that the spread rate and the affected population size are related to initial infected individuals as well as network structure. The results can offer some valuable suggestions on disease controlling.

Keywords Epidemic spreading, Complex network, Initial infected individuals, Disease controlling

1 引言

长期以来,传染性疾病对人类造成巨大威胁,如早期的天花、麻疹,近期的艾滋病(AIDS)、非典型肺炎(SARS)、甲流(H1N1)等。最近几年,传染病的传播与控制引起了足够重视^[1-7],提出了各种各样的传染病传播模型^[8-13]。其中最为经典的是基于平均场理论的微分方程模型,它在传染病传播的早期研究中起到了重要的作用。但是,个体之间的接触不可能完全混合,由此,各种其它传播模型相继提出,如基于元胞自动机的二维晶格模型^[14]、具有层次结构的复合种群模型^[15]等。近年来,为更深入地研究传染病的传播,提出了更符合实际情况的传播模型,其中最为经典的是基于小世界网络的传播模型^[10]和基于无标度网络的传播模型^[5,11,12]。最近有关传染病传播速度的影响因素的研究也越来越深入,Moore 和 Newman^[10]指出,网络中即便只加入少量的长程边也会极大地加快传染病的传播。Volchenkov 等人^[16]研究指出,传染病的传播阈值和影响规模与如下 4 个因素密切相关:无标度网络中的幂指数,传染率,伙伴选择策略以及治疗方

案。研究还发现,网络中一般会存在重叠的或不重叠的社区结构^[17-19],而网络中的社区结构对传染病的传播也有较大的影响,社区结构的存在降低了传播阈值但影响规模却减小了,进一步的研究还表明,社区结构的存在会导致感染人数出现周期振荡。周涛等人^[20]对复杂网络中的传播动力学进行了详细的综述,提出了一些传染病传播研究方面极具价值的研究问题。

本文在以上研究的基础上,分析了在无向无权网络中初始染病节点对传染病传播的影响。分析了按度中心、点介数、接近中心度、page rank 和 HITs 等 5 种节点排序策略选择的初始染病节点如何影响传染病的传播速度及规模。仿真实验表明,按不同策略选择初始染病节点,其传播速度和传染病影响规模具有一定的差异,与此同时,传染病影响规模与传播速度不仅与初始染病节点选择有关,与自身的网络结构也有关系。

2 节点排序及初始染病节点选择

2.1 节点排序指标

本文主要考虑 5 种排序指标:度中心,节点点介数,接近

到稿日期:2010-12-19 返修日期:2011-03-22 本文受国家自然科学基金(60973069,90924011,60903073,60973120),四川省国际合作与交流项目(2010HH0002)资助。

苏长明 硕士,讲师,主要研究方向为数据挖掘、图论;陈端兵 博士,副教授,硕士生导师,主要研究方向为数据挖掘、社会计算、NP 难问题高效求解等;傅彦 教授,博士生导师,主要研究方向为数据挖掘、信息安全、模式识别等。

中心度, page rank 和 HITs。度中心实际上就是节点的度, 这里不详细讨论, 下面主要介绍其余 4 种排序指标。

对于一个节点集为 V , 边集为 E 的网络, 其节点 v 的点介数 $C_B(v)$ 为:

$$C_B(v) = \sum_{s \neq v, t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (1)$$

式中, σ_{st} 是从节点 s 到节点 t 的最短路径数量, $\sigma_{st}(v)$ 是从节点 s 到节点 t 并且经过节点 v 的最短路径数量。Brandes^[21] 提出了一种计算节点点介数的快速算法, 其计算复杂度为 $O(|V| \cdot |E|)$ 。

对于接近中心度, 有许多定义方法, 但本质上都是反映一个节点到网络中其它节点的平均最短路径, Sabidussi^[22] 提出了一种计算接近中心度的方法:

$$C_G(v) = \frac{1}{\sum_{t \in V, t \neq v} d_G(v, t)} \quad (2)$$

式中, $d_G(v, t)$ 是节点 v 到节点 t 的最短路径长度。

对于大规模的复杂网络, 计算点介数和接近中心度非常耗时, 为此, 本文在计算点介数和接近中心度时, 采用 Brandes 和 Pich^[23] 提出的近似计算方法, 此方法的复杂度为 $O(|E|)$, 相比精确计算方法, 其复杂度要低得多。

Page rank 和 HITs 通常用于计算有向网络中节点的重要性, 为了应用于无向网络, 文献[24]对 page rank 的计算方法进行了修正:

$$S(v_i) = (1-d) + d \sum_{v_j \in N_{v_i}} \frac{S(v_j)}{D(v_j)} \quad (3)$$

式中, $S(v_i)$ 是节点 v_i 的 page rank 值, d 是介于 0 到 1 之间的衰减因子, $D(v_j)$ 是节点 v_j 的度。

节点的 HITs 值也主要是衡量有向网络中节点的权威性 & 中心性, 对于无向网络, 采用文献[25]的方法对其进行修正:

$$\begin{aligned} a &= (1, 1, \dots, 1)^T; \\ \text{while}(eps > \epsilon) \text{ do} \\ & b_i = \sum_{(v_i, v_j) \in E} a_j, i=1, 2, \dots, |V|; \\ & eps = \max_i |a_i - b_i / \sum_{j=1}^{|V|} b_j|; \\ & a_i = b_i / \sum_{j=1}^{|V|} b_j, i=1, 2, \dots, |V| \\ \text{end while} \end{aligned}$$

式中, a_i 是节点 v_i 的 HITs 值, ϵ 是迭代精度。

2.2 初始染病节点选择

首先按度中心、节点点介数、接近中心度、page rank 和 HITs 的值从大到小对网络中的节点进行排序, 然后选择排序结果的前 m (本文取 $m=20$) 个节点作为初始染病节点。本文将从这 m 个节点出发, 分析传染病在网络中的传播情况, 从而分析不同初始染病节点对传染病传播速度及传染病影响规模有何影响。

2.3 传播影响的评价方法

本论文采用无差异接触的 SIR 传播模型对 5 种排序指标进行分析评价。实际上, 一个个体不可能在短时间内接触它的所有邻居, 因此, 研究者总是假设个体接触能力正比于它的所有邻居。2004 年, Joo 和 Lebowitz^[26] 以及 Olinky 和 Stone^[27] 指出这种假设是不合理的, 他们证明了如果个体的接触能力存在细微差异, 将降低传染病传播的危险性。周涛等人^[28] 也研究了个体无差异接触的传染病传播规律, 发现在疾病爆发及持续期, 疾病的危害降低了。

为了简化模型, 本文假设一个个体在一个时间步内只接触一个个体。易感者和已感者接触后将以概率 α 变为已感者, 与此同时, 已感者在每一时间步将以概率 β 康复变为移除个体, 而传染病的传播速率 $\lambda = \alpha/\beta$ 。

3 仿真实验

本文选择两个大规模的复杂网络进行传染病的传播仿真实验, 一个是预印本文献的科学家合作网络 Cond-mat-2003^[29], 另一个是天体物理的科学家合作网络^[30], 这两个网络的一些基本网络特征参数如表 1 所列。

表 1 网络 cond-mat-2003 和 CA-astrophysics 的一些基本网络特征

网络	节点数	边数	平均最短路径长度	平均度	平均聚集系数
Cond-mat-2003	27519	116181	5.7667	8.4437	0.6546
CA-astrophysics	17903	196972	4.194	22.0044	0.6328

对每种排序指标, 我们取指标值最大的前 20 个节点作为初始染病节点进行分析, 选出的节点中有部分节点是重叠的, 表 2 及表 3 列出了各种排序指标选出的节点中相互之间重叠的数目。

表 2 网络 cond-mat-2003 中各种排序指标选出的 20 个节点中重叠点数目

	点介数	接近中心度	度	Page rank	HITs
点介数	—	9	12	14	3
接近中心度	9	—	12	12	7
度	12	12	—	11	6
Page rank	14	12	11	—	4
HITs	3	7	6	4	—

表 3 网络 CA-astrophysics 中各种排序指标选出的 20 个节点中重叠点数目

	点介数	接近中心度	度	Page rank	HITs
点介数	—	9	8	10	5
接近中心度	9	—	14	19	13
度	8	14	—	14	14
Page rank	10	19	14	—	13
HITs	5	13	14	13	—

从不同的初始染病节点出发, 用计算机模拟了疾病在两个网络中的传播情况, 部分仿真结果如图 1—图 4 所示, 图 1—图 4 中, 所有结果均为 100 次运行结果的平均值。图 1 是网络 cond-mat-2003^[29] 中染病节点数和移除节点数随时间的变化情况, 图中传播速率 $\lambda = 1.3$ 。从图 1 可看出, 如果初始染病节点选择度大的节点, 传播速度较快, 而如果选择 HITs 值较大的节点作为初始染病节点, 传播速度则较慢。实际上, 按 HITs 选出的初始染病节点和其它 4 种方法选出的初始节点只有很少的公共节点, 如表 2 所列。因此, 对于 cond-mat-2003 而言, 按 HITs 选择并不是一个好的选择策略。

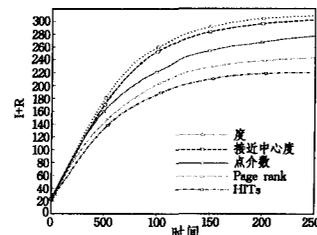


图 1 网络 cond-mat-2003 中, 传染病的扩散速度, $\lambda = 1.3$

图 2 是网络 cond-mat-2003 中不同传播速度 λ 下, 传播达

到稳定状态时移除节点的数量。从图 2 可以看出,传播速度 λ 的阈值在 1 附近,当 $\lambda < 1$ 或 $\lambda > 1.8$ 时,从 5 种方法选择的初始染病节点出发,得到的传染病传播规模差异不大,而当 λ 介于 1 和 1.8 之间时,初始节点选择方法不同,对传染病传播规模的影响也不同,其中 HITs 方法影响最小。

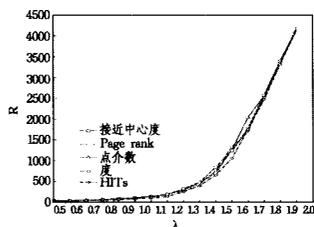


图 2 网络 cond-mat-2003 中,不同传播速率 λ 下传播达到稳定状态时的影响规模

对天体物理研究领域的科学家合作网络 CA-astrophysics^[30],图 3 展示了网络中染病节点数和移除节点数随时间的变化情况,图中传播速率 $\lambda = 1.3$ 。从图 3 可看出,每种方法对传播只有一些细微差异,其原因在于 5 种方法选出的初始节点有较多的重叠,特别是 page rank 和接近中心度之间重叠的节点较多,如表 3 所列。

图 4 是网络 CA-astrophysics 中不同传播速度 λ 下,传播达到稳定状态时移除节点的数量。从图 4 可以看出,传播速度 λ 的阈值也大致等于 1,当 $\lambda < 1$ 或 $\lambda > 1.6$ 时,从 5 种方法选择的初始染病节点出发,得到的传染病传播规模差异很小,当 λ 介于 1 和 1.6 之间时,传播规模存在一定的差异,但差异很小。

两个网络的仿真实验结果表明,传染病在网络中的扩散不仅与初始节点的选择有关,还与网络自身的结构有关系。例如,在网络 cond-mat-2003 中,如果选择大度节点作为初始染病节点,扩散速度比其他方法要快些,而在网络 CA-astrophysics 中却不是这样,如图 1 和图 3 所示。

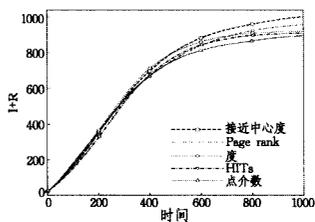


图 3 网络 CA-astrophysics 中,传染病的扩散速度, $\lambda = 1.3$

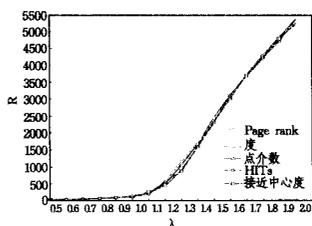


图 4 网络 CA-astrophysics 中,不同传播速率 λ 下传播达到稳定状态时的影响规模

结束语 本文分析了不同的初始染病节点选择策略对传染病在复杂网络中的传播有何影响。用两个大规模的复杂网络对传染病的传播进行了仿真模拟。仿真实验结果表明,传染病在网络中的扩散不仅与初始染病节点的选择有关,而且

与网络本身的结构有关。通过仿真模拟,我们可以发现哪些节点会导致传染病的快速传播,从而在传染病控制时,将重点关注这些节点的状态。本文仅针对无向无权网络进行了模拟,在今后的研究工作中,将在有向有权网络中,研究初始节点对传染病传播的影响规律。

参考文献

- [1] Sanz J, Floría L M, Moreno Y. Spreading of persistent infections in heterogeneous populations[J]. Physical Review E, 2010, 81: 056108
- [2] Zhang H F, Li K Z, Fu X C, et al. An efficient control strategy of epidemic spreading on scale-free networks[J]. Chinese Physics Letters, 2009, 26: 068901
- [3] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks[J]. Physical Review Letters, 2001, 86: 3200-3203
- [4] Zhao H, Gao Z Y. Modular epidemic spreading in small-world networks[J]. Chinese Physics Letters, 2007, 24: 1114-1117
- [5] Xu X J, Wang W X, Zhou T, et al. Geographical effects on epidemic spreading in scale-free networks[J]. International Journal of Modern Physics C, 2006, 17: 1815
- [6] Duan W Q, Chen Z, Liu Z R. Epidemic spreading in contact networks based on exposure level[J]. Chinese Physics Letters, 2006, 23: 1347-1350
- [7] 陈端兵, 黄晟, 高明生. 复杂网络模型及其在疫情传播和控制中的应用研究[J]. 计算机科学, 2011, 38(6): 118-121
- [8] Juher D, Ripoll J, Saldaña J. Analysis and monte carlo simulations of a model for the spread of infectious diseases in heterogeneous metapopulations[J]. Physical Review E, 2009, 80: 041920
- [9] Balcan D, et al. Multiscale mobility networks and the spatial spreading of infectious diseases[C]// Proceedings of the National Academy of Sciences of USA. 2009, 106: 21484-21489
- [10] Moore C, Newman M E J. Epidemics and percolation in small-world networks[J]. Physical Review E, 2000, 61: 5678-5682
- [11] Huang W, Li C. Epidemic spreading in scale-free networks with community structure[J]. Journal of Statistical Mechanics: Theory and Experiment, 2007
- [12] Moreno Y, Vázquez A. Disease spreading in structured scale-free networks[J]. The European Physical Journal B, 2003, 31: 265-271
- [13] Zhou J, Liu Z. Epidemic spreading in communities with mobile agents[J]. Physica A, 2009, 388: 1228-1236
- [14] Sirakoulis G C, Karafyllidis I, Thanailakis A. A cellular automaton model for the effects of population movement and vaccination on epidemic propagation[J]. Ecological Modelling, 2000, 133: 209-223
- [15] Watts D J, Muhamad R, Medina D C, et al. Multiscale, resurgent epidemics in a hierarchical metapopulation model[C]// Proceedings of the National Academy of Sciences of USA. 2005, 102: 11157-11162
- [16] Volchenkov D, Volchenkova L, Blanchard P. Epidemic spreading in a variety of scale free networks[J]. Physical Review E, 2002, 66: 046137
- [17] Shang M, Chen D, Zhou T. Detecting overlapping communities based on community cores in complex networks[J]. Chinese Physics Letters, 2010, 27: 058901

- [18] Chen D, Shang M, Fu Y. Detecting overlapping communities of weighted networks via a local algorithm[J]. *Physica A*, 2010, 389:4177-4187
- [19] Chen D, Fu Y, Shang M. A fast and efficient heuristic algorithm for detecting community structures in complex networks[J]. *Physica A*, 2009, 388:2741-2749
- [20] Zhou T, Fu Z, Wang B H. Epidemic dynamics on complex networks[J]. *Progress in Natural Science*, 2006, 16:452-457
- [21] Brandes U. A faster algorithm for betweenness centrality[J]. *Journal of Mathematical Sociology*, 2001, 25:163-177
- [22] Sabidussi G. The centrality index of a graph[J]. *Psychometrika*, 1966, 31:581-603
- [23] Brandes U, Pich C. Centrality estimation in large networks[J]. *International Journal of Bifurcation and Chaos*, 2007, 17:2303-2318
- [24] Wang J H, Liu J Y, Wang C. Keyword extraction based on pagerank[C]//11th Pacific-Asia Conference Advances in Knowledge Discovery and Data Mining. Nanjing, China, 2007:857-864
- [25] 赵明昌, 崔霞, 李耀东. 基于共引分析的中草药相互关系可视化算法[J]. *计算机工程与应用*, 2008, 44(22):10-12
- [26] Joo J, Lebowitz J L. Behavior of susceptible-infected-susceptible epidemics on heterogeneous networks with saturation[J]. *Physical Review E*, 2004, 69:066105
- [27] Olinky R, Stone L. Unexpected epidemic thresholds in heterogeneous networks; The role of disease transmission[J]. *Physical Review E*, 2004, 70:030902
- [28] Zhou T, et al. Behaviors of susceptible-infected epidemics on scale-free networks with identical infectivity[J]. *Physical Review E*, 2006, 74:056109
- [29] Newman M E J. The structure of scientific collaboration networks[J]. *Proceedings of the National Academy of Sciences of USA*, 2001, 98:404-409
- [30] Leskovec J, Kleinberg J, Faloutsos C. Graph Evolution; Densification and shrinking diameters [J]. *ACM Transactions on Knowledge Discovery from Data*, 2007, 1:1-41

(上接第 143 页)

如图 4 所示。

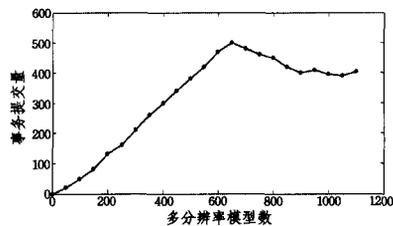


图 4 性能测试结果

我们发现每秒内的事务提交量在一开始是随着模型数量的增加而增加的。当增加到 500 时,随着模型数量的增加,事务的提交量会随之下降,最后平稳在每秒 400 次左右。而当模型的数量增加到 700 时,为维护多分辨率模型的数据一致性,某些事务被迫取消,此时事务的提交量就稳定在 400 到 500 这个范围之内。从中可以看出,为维护分布式仿真系统多分辨率模型数据的一致性,随着多分辨率模型的大幅度增加,交互事务的提交量会受到一定的限制。

结束语 随着计算机网络技术的快速发展,分布式仿真中的多分辨率建模技术是目前研究的热点。通过分析现有分布式仿真的建模方法,本文提出了适应分布式仿真环境的多分辨率模型树状结构建模方法。这个建模方法由平台级实体模型和聚合级实体模型组成。平台级实体模型是实体的高分辨率模型,而聚合级实体模型是实体的低分辨率模型,而且聚合级实体模型是平台级实体模型的父节点,两者之间采用属性映射的方法进行交互。为维护分布式仿真系统中多分辨率模型之间的数据一致性,我们将分布式数据库事务的概念运用到分布式仿真交互中,提出了一个嵌套两段式提交协议来保证分布式交互仿真中交互事务的原子性,确保一个交互要么不发生,要么发生。并且该交互所发生的影响必须反映到同一实体的所有分辨率模型上,从一定程度上避免了多分辨率模型对实体的不一致描述。

今后的研究可以继续借鉴分布式数据库的知识,将其应

用到分布式仿真领域中。例如,由于嵌套两段式提交协议网络的实时性有一定的依赖,并且可能要夭折掉一些交互事务,可研究运用数据库的日志和事务重启等技术来改进所提出的嵌套两段式提交协议,使其更好地应用于分布式仿真的多分辨率建模中。

参考文献

- [1] Davis P K, Bigelow J. Introduction to Multi-resolution Model (MRM) with an Example Involving Precision Fires[C]//Proceeding of SPIE AeroSense on Enabling Technology for Simulation Science(II). 1998
- [2] Komman B D, Marion P B. Cross-model Consistency in JSIMS [R]. Lockheed Martin Information Systems Advanced Simulation Laboratory
- [3] Reynolds Jr P F, Natrajan A, Srinivasan S. Consistency Maintenance in Multi-resolution Simulations [J]. *ACM Transactions on Modeling and Computer Simulation*, 1997, 7(3)
- [4] Biddle M. A Proposed Scheme for Implementing Aggregation and Disaggregation in HLA[C]//Processing of 2000 Fall SIW. 2000
- [5] Natrajan A. Consistency Maintenance in Concurrent Representations[D]. School of Engineering and Applied Science at the University of Virginia, 2000
- [6] David P K, Bigelow J. Experiments on Multi-resolution Modeling(MRM)[R]. RAND Report-MR1004. 1998
- [7] Natrajan A, Reynolds Jr P F, et al. A Flexible Approach to Multi-resolution Modeling [J]. *Parallel and Distributed Simulation*, June 1997
- [8] 刘宝宏, 黄柯棣. 多分辨率建模的研究现状与发展[J]. *系统仿真学报*, 2004, 16(6):1150-1154
- [9] 郭齐胜, 杨立功. 基于 HLA 的聚合级 GGF 初探[J]. *计算机工程与应用*, 2001(19):41-43
- [10] 刘宝宏, 黄柯棣. 分布交互式仿真中的多分辨率模型并发运行问题研究[J]. *系统仿真学报*, 2007, 19(5):1038-1077