

doi:10.11835/j.issn.1000-582X.2018.07.009

基于活跃度的网络免疫策略

张乐延,高树鹏,张自力,高 超

(西南大学 计算机与信息科学学院,重庆 400715)

摘要:免疫策略是抑制复杂网络传播过程的有效方法。虽然基于网络拓扑的免疫策略取得了良好的免疫效果,但这些策略都基于同构性假设,即:网络中的节点具有相同的属性。然而,越来越多的研究揭示了网络节点的异构特征,如传播过程中节点具有不同的活跃度,活跃的节点能够促进传播。笔者结合网络结构和节点活跃度,提出了一种新的免疫策略,在真实网络和人工网络中的仿真结果表明,所提免疫策略能够有效抑制传播扩散。

关键词:复杂网络;网络免疫;病毒传播;活跃度;异构

中图分类号:TP393.01

文献标志码:A

文章编号:1000-582X(2018)07-067-06

Network immunization based on the node activities

ZHANG Leyan, GAO Shupeng, ZHANG Zili, GAO Chao

(School of Computer and Information Science, Southwest University, Chongqing 400715, P.R.China)

Abstract: Network immunization is an effective strategy for restraining the propagation process in complex networks. Although lots of strategies based on the topological structure of a network have achieved good immunization effects, they have the same homogeneous assumption, i. e., nodes have same characteristic. However, more and more existing studies have found the heterogeneous characteristics of nodes in a network. For example, nodes have different activity during the propagation process and a node with higher activity can promote the propagation process. Combining the network topology and the activity of node, this paper proposes a novel strategy for restraining virus propagation. Some experiments have been conducted in benchmark and synthetic networks. The simulation results show that the proposed strategy can restrain propagation effectively.

Keywords: complex network; network immunization; virus spreading; activity; heterogeneous

免疫策略是抑制病毒在复杂网络传播的有效手段^[1]。如果不能及时控制病毒传播,会造成严重的破坏性后果。由于资源限制,无法免疫网络中的每一个节点。如何免疫一组在病毒传播中具有影响力的重要节点已经成为科学家们关注的挑战性课题^[2]。

已有研究认为节点的影响力是由网络结构决定的。学者们提出了许多基于网络拓扑结构的节点中心性

收稿日期:2017-10-23

基金项目:国家自然科学基金资助项目(61403315,61402379),中央高校基本业务费专项资金资助项目(XDJK2016A008,XDJK2016B029)。

Supported by National Natural Science Foundation of China(61403315,61402379) and the Fundamental Research Funds for the Central Universities(XDJK2016A008,XDJK2016B029).

作者简介:张乐延(1992—),男,西南大学硕士研究生,主要从事复杂网络研究。

高超(联系人),副教授,主要从事复杂网络传播动力学研究,(E-mail)cgao@swu.edu.cn。

指标。比如度中心性^[3]是对一个节点的所有邻居进行计数,并认为拥有邻居越多的节点会有更强的影响力;介数中心性^[4]是评估通过节点的最短路径条数占全图最短路径总数的比例;核免疫^[5]认为,节点的位置比其邻居更重要,通过从外部到内部除去边缘节点选出核心节点;特征向量中心性^[6]假设一个节点的影响力也是由其邻居的影响力决定,以上4个策略都需要网络的全局信息。图覆盖免疫(D-step)^[7]使用局部信息去寻找离源节点 d 步内拥有最大度中心性的节点进行免疫。

然而,最近的研究表明,病毒的传播也受节点活动的影响。例如,邮件蠕虫传播过程中的用户行为会影响传播的范围和速度^[1]。具体地说,当用户用网络中的节点表示时,用户行为即是一种节点活动,其频率定义为节点活跃度。因此,在网络拓扑结构基础上,有必要对节点活跃度能否运用免疫策略进行研究。

一般而言,设计免疫策略的步骤,首先,从网络中选择有影响力的节点进行保护,然后,选择一个传播模型作为病毒传播的仿真平台,最后,通过模拟结果来判断免疫策略的优劣。然而现有验证免疫策略而使用的病毒传播模型,很少考虑节点活跃度。这使得以往研究仅注重于如何基于网络拓扑结构来识别有影响力的节点,忽略了节点活跃度对免疫策略效率的影响。因此,首先基于交互式传播模型刻画节点的异构特征,分析不同类型的节点活跃度对传播影响,揭示节点活跃度在传播过程中的重要性,以此为基础设计一种基于节点活跃度和网络拓扑结构的改进免疫策略,提高免疫效率。

1 节点活跃度与传播模型的选择

1.1 节点活跃度

复杂网络中的传播不会自发进行,需要节点活动激发传播行为。在不同的传播过程中,节点活动的含义是不同的。例如,在社交媒体谣言传播过程中^[8],节点活动被表示为用户对谣言的阅读和转发行为。在流行病疫情传播过程中^[9],节点活动包括人们之间的接触和出行模式变化。在邮件蠕虫传播过程中^[10],节点活动包括用户查看邮箱和对电子邮件的点击行为。节点活跃度是节点活动发生的频率,在传播过程中起着重要的作用。而在人类参与的系统,节点活跃度可以通过人类行为动力学的研究来表达和解释^[11]。

人类行为动力学领域的现有研究揭示了人类行为的规律^[11]。2005年,Barabási揭示了人类行为对泊松过程的偏离^[12],这进一步激发了学者们对人类行为动力学的研究兴趣。通过对实际人类行为数据的统计分析,许多人类行为可以通过数学分布来表示其规律。例如,幂律分布用来刻画用户连续发送2封电子邮件之间的时间间隔^[13],指数分布用来刻画北京出租车的乘车距离^[14],负二项分布用来刻画伦敦地铁乘客乘车距离^[15]等。据此,节点活跃度可以通过统计节点活动的历史数据,使用拟合的方法得到其满足的分布函数。

1.2 传播模型

传播模型在免疫策略的设计中不可缺少。为了评估免疫策略的效率,需要预先选择一种传播模型作为测评平台分析与评估各类免疫策略的效率。为研究传播过程中节点活跃性与免疫效率的关系,选用的传播模型需能刻画节点活跃度。目前,有2种典型的病毒传播模型:1)基于平均场的流行病传播模型。2)基于个体行为交互的传播模型。基于平均场^[16]的经典流行病传播模型包括SI, SIR, SIS^[17-19]等。虽然,流行病传播模型从宏观角度可对病毒的传播规模和速率进行准确刻画,但这些模型无法揭示个体行为对传播的影响。这是因为基于平均场理论的流行病传播模型中,病毒传播的均匀性和传播行为的被动性假设过于理想,在现实环境中很难实现。

为了刻画节点行为的异构特征,Zou^[20]构建了一个基于个体交互的病毒邮件传播模型。该模型中病毒的激活和传播需要由个体行为决定。模型刻画了2种个体行为:用户查看邮箱行为和用户点击病毒邮件行为。由于网络用户众多且各自行为相互独立,所以模型假设不同用户间查看邮箱和点击病毒邮件行为的概率服从正态分布。泊松过程是人类活动经典的模拟刻画工具之一,通常用于描述特定事件在某段时间内出现的次数统计规律,在泊松过程中2个相邻的随机事件时间间隔服从指数分布^[21]。在交互式邮件传播模型

中,Zou 使用泊松过程刻画用户查看邮箱行为,即假设对于某一个固定用户来说,其查看邮箱的时间间隔分布服从指数分布。然而,随着大数据时代到来,通过人类行为动力学研究发现,人类很多行为偏离了泊松过程^[22]。以交互式传播模型为基础,利用现有人类行为动力学统计结果替代原有分布假设,克服原模型的局限性。

已有研究结果表明,当节点行为服从不同分布特征时,将会有不同的传播过程^[1]。以电子邮件网络为例,在交互式传播模型中随机选择 2 个节点作为被感染节点,分析不同节点活跃度下的传播特征,如图 1 所示。结果表明,节点活跃度分布对传播有着重要影响,在设计免疫策略时,应考虑此影响。

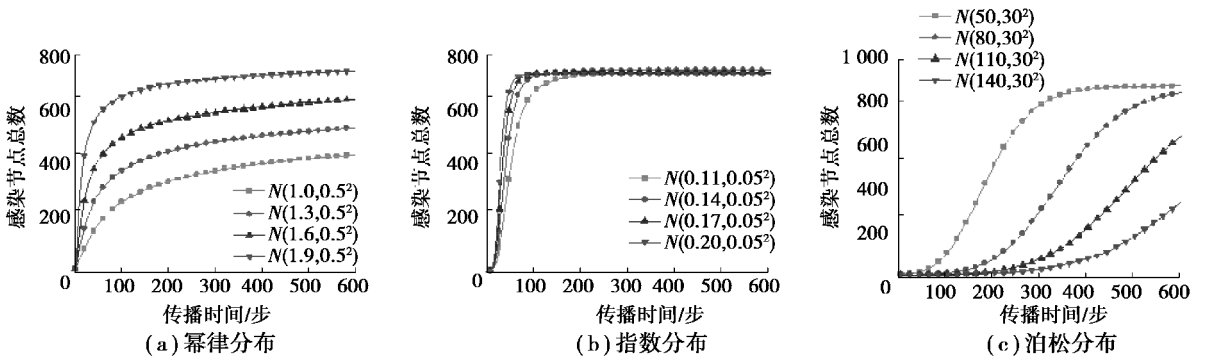


图 1 不同分布下病毒传播结果

Fig.1 Propagation results with different distributions of node activities

每个小图为相同分布,其参数满足不同的正太分布。实验结果表明,不同活跃度对传播过程有着重要影响。

2 基于节点活跃度的免疫策略

2.1 算法设计

网络免疫策略需要选择一组影响力节点进行免疫从而达到抑制病毒传播效果。虽然,学者提出了确定节点影响力的各类中心性算法,但这些算法具有各自的局限性。因为,此类算法是基于同构性假设提出,即:网络的节点具有相同的属性。因此,每个节点的活跃度也相同。然而,越来越多的研究表明,网络的节点拥有异构特征,每个节点的活跃度也是不尽相同,因此,这些算法忽略了节点活跃度对节点影响力的作用。基于图 1 节点活跃度对传播的影响实验发现,当节点活动时间间隔期望越小,病毒传播速度越快。因此,需要对拥有较小活动时间间隔的节点更加关注。研究提出基于节点活跃度和网络拓扑结构的改进免疫策略,算法为网络 $G=(V,E)$,其中 V 为所有节点的集合, E 为所有边集合。其中:网络中第 i 个节点记为 n_i ; L_i 为存放节点 n_i 的活动时间间隔队列,活动时间间隔可由真实数据或其活动行为满足的分布函数取得

$$L_i = \{l_{i1}, l_{i2}, \dots, l_{im}\}, \quad (1)$$

式中: l_{i1} 表示节点 n_i 第一次记录的活动时间间隔; l_{im} 表示节点 n_i 第 m 次记录的活动时间间隔。若 l_{im} 在队末,则说明节点 n_i 总共有 m 条活动的的时间间隔记录。节点 n_i 的活动时间间隔期望为

$$Q(n_i) = \frac{\sum_{z=1}^m l_{iz}}{m}. \quad (2)$$

评价一个节点是否活跃,并没有一个恒定的阈值参考,而是与网络中其他所有节点相比。而由定义可得, $Q(n_i)$ 值越小,其节点越活跃。笔者使用自然对数函数对节点活跃度进行刻画,具体如下

$$A(n_i) = -\ln \frac{Q(n_i)}{\sum_{n_j \in V} Q(n_j)}, \quad (3)$$

通过公式(3),可以将网络中相对活跃的节点赋予较大的值。

同时,需要选择一种节点的拓扑属性值作为节点影响力的评判指标。不同拓扑属性值代表节点不同方面的拓扑性质。比如度中心性、介数中心性、核值或者特征向量中心性等。假设选择度中心性作为节点的拓扑属性值,那么节点 n_i 的拓扑属性值等于节点 n_i 的度中心性值,即

$$\text{Key}(n_i) = n_i \cdot \text{degree}, \quad (4)$$

其中, $n_i \cdot \text{degree}$ 代表节点 n_i 的度。最后,将节点的活跃度与拓扑属性值结合,得出节点 n_i 的影响力,表示为

$$\text{ActiveRank}(n_i) = \text{Key}(n_i)^2 A(n_i), \quad (5)$$

节点拥有越大的 ActiveRank 值,其在传播过程中越活跃,节点的拓扑位置越重要,对传播的影响力也越大。为了能够得到更好的免疫效果,需要优先免疫拥有更大的 ActiveRank 值的节点。改进策略,将节点活跃度列为评价节点影响力的一个重要因素,使得网络拓扑结构不再是唯一判断节点影响力的指标。

2.2 实验网络

表 1 展示使用 8 个真实网络和 2 个人工社团网络的结构特征。真实网络数据集来自斯坦福网络数据库、科布伦茨网络数据库和 Newman 教授个人网站。人工社团结构网络 G_9 和 G_{10} 是由 Andreas 算法^[23]生成,其包含 1 000 个节点,3 个社团,网络幂指数分别为 2 和 2.5。

表 1 实验使用网络

Table 1 Networks used in the experiment

网络	编号	节点数	边数	平均度	聚类系数
电子邮件网络	G_1	1 133	5 451	9.622	0.254
高能物理-理论合作网络	G_2	9 877	25 998	10.521	0.600
高能物理-现象学合作网络	G_3	12 008	118 521	39.477	0.698
天文物理合作网络	G_4	18 771	198 050	21.102	0.677
作家合作网络	G_5	56 276	631 632	22.448	0.687
脸书网络	G_6	4 039	88 234	43.691	0.606
Hamsterste 网络	G_7	2 426	16 631	13.711	0.231
Hamsterster 朋友网络	G_8	1 858	12 534	13.492	0.090
人工网络 1	G_9	1 000	9 944	19.888	0.029
人工网络 2	G_{10}	1 000	7 938	15.876	0.022

2.3 实验结果

首先,根据不同免疫策略对影响力节点的定义,选取一组重要节点进行免疫。鉴于所提策略是结合网络拓扑结构和节点活跃度,故每个传统免疫策略都会对应一个基于节点活跃度的改进策略。研究选取了一些全局免疫策略(如目标免疫、介数免疫、核免疫和特征向量免疫)以及局部免疫策略(如图覆盖免疫)进行对比。为刻画节点活跃度异质特征,实验采用了交互式传播模型,并基于现有人类行为动力学研究成果^[13-15],随机选取 3 种典型的幂律分布、指数分布和负二项分布之一用来刻画单个节点活跃度特征,所选分布参数都基于真实数据统计得到,具体请参考文献[13-15]。实验选取不同免疫比例(10%,20%,30%),在 10 种不同网络中进行对比,初始随机选择 2 个感染节点,模拟传播 600 步。为消除波动,所有结果为 500 次仿真的平均值。

表 2 列出了仿真结果,其中,I%表示免疫节点占总节点的百分比,DS 表示图覆盖免疫策略($d=3$)。Deg 表示目标免疫,Bet 表示介数免疫,KC 表示核免疫,EC 表示特征向量免疫。研究提出的对应免疫策略用带有后缀-AR 表示。

表 2 病毒传播结果
Table 2 Results of virus propagation

Net	$I/\%$	DS/DS-AR	Deg/Deg-AR	Bet/Bet-AR	KC/KC-AR	EC/EC-AR
G_1	10	470/448	478/454	465/444	537/509	502/486
	20	195/162	195/161	185/157	319/282	295/271
	30	25/18	21/20	30/29	128/80	117/94
G_2	10	983/557	326/146	735/625	2 757/2 407	2 093/1 964
	20	11/10	8/8	15/13	322/94	83/49
	30	6/5	5/4	5/5	11/9	10/9
G_3	10	5 411/5 253	5 583/5 436	3 403/3 279	5 954/5 945	5 720/5 685
	20	2 611/2 141	2 697/2 457	353/327	4 164/4 057	3 824/3 680
	30	32/18	17/15	28/20	1 241/779	1 560/1 331
G_4	10	8 717/8 492	8 472/8 342	7 692/7 399	9 504/9 449	9 501/9 402
	20	5 195/4 628	4 838/4 280	3 157/3 027	6 360/6 091	6 192/5 973
	30	773/342	596/270	62/54	3 403/2 923	2 806/2 751
G_5	10	24 183/22 162	25 594/24 139	16 631/16 484	30 878/30 449	29 090/29 086
	20	3 008/1 117	4 046/2 455	1 233/1 072	19 671/18 903	15 135/13 876
	30	12/14	16/15	27/22	8 203/7 574	2 925/2 235
G_6	10	1 690/1 498	2 247/2 014	766/742	2 427/2 363	2 477/2 424
	20	1 215/1 155	1 586/1 350	567/562	1 819/1 772	1 650/1 599
	30	866/835	770/758	387/378	1 299/1 262	1 273/1 093
G_7	10	671/634	550/508	552/547	720/689	648/649
	20	120/97	38/33	119/118	248/222	192/181
	30	17/14	10/8	14/13	33/25	37/26
G_8	10	564/544	556/512	568/544	700/690	647/642
	20	84/62	57/49	164/154	232/224	311/279
	30	11/10	9/9	24/19	35/32	100/77
G_9	10	625/613	623/612	628/618	689/644	628/617
	20	462/435	460/427	461/445	586/487	484/464
	30	208/148	199/153	216/180	338/311	302/257
G_{10}	10	596/586	594/587	595/591	677/635	608/596
	20	409/373	410/374	426/399	590/492	474/447
	30	136/87	128/80	133/99	486/286	286/258

相比传统策略,实验结果表明基于节点活跃度的免疫策略可有效抑制传播,即最终感染的节点数最少,说明节点活跃度可提升免疫效率。当免疫 10% 节点时,基于节点活跃度免疫策略对免疫效果的提升并不十分明显,这是因为当免疫节点较少时,免疫算法选择出的免疫节点拥有较大的拓扑特征值,即:拓扑结构对传播的影响要大于节点活跃度。当免疫强度达到 20% 时,基于节点活跃度免疫策略的免疫效果有较大提升,这是因为当免疫节点较多时,很多节点具有相似的拓扑特征值。这时节点活跃度可以用来区分活跃节点和不活跃节点,从而可以利用宝贵的资源免疫活跃节点,对免疫效果进行提升。

总之,结合了节点拓扑结构和节点活跃度的策略免疫效果优于传统仅基于节点拓扑结构的免疫策略。仿真结果表明,活跃度较高的节点具有较强传播影响力,这在免疫策略的设计中不容忽视。

3 结 论

研究分析了病毒传播过程,在交互式传播模型中比较了当节点活跃度服从不同分布对传播的影响。实验结果表明,节点活动的时间间隔期望值越小,节点越活跃,病毒传播越快。据此提出了一种基于节点活跃度和网络拓扑结构的改进免疫策略。仿真实验验证表明,结合节点活跃的免疫策略有效提高了免疫效率。这项工作说明节点活跃度在传播中的重要性,为免疫策略设计提供了新的视角。后续研究将会进一步对节点活跃度在免疫策略中的潜力进行挖掘。

参考文献:

- [1] Gao C, Liu J, Zhong N. Network immunization and virus propagation in email networks: experimental evaluation and analysis[J]. Knowledge and Information Systems, 2011, 27(2): 253-279.
- [2] Newman M E J. The structure and function of complex networks[J]. Siam Review, 2003, 45(2): 167-256.
- [3] Pastor-Satorras R, Vespignani A. Immunization of complex networks[J]. Physical Review E, 2002, 65(3): 1-8.
- [4] Liu Y Y, Slotine J J, Barabási A L. Controllability of complex networks[J]. Nature, 2011, 473(7346): 167-173.
- [5] Dorogovtsev S, Goltsev A V, Mendes J F. K-Core organization of complex networks[J]. Physical Review Letters, 2006, 96(4): 1-5.
- [6] Bonacich P. Factoring and weighting approaches to status scores and clique identification[J]. Journal of Mathematical Sociology, 1972, 2(1): 113-120.
- [7] Echenique P, Gómez-Gardeñes J, Moreno Y, et al. Distanced covering problems in scale-free networks with degree correlations[J]. Physical Review E, 2005, 71(3): 1-5.
- [8] Moreno Y, Nekovee M, Pacheco A F. Dynamics of rumor spreading in complex networks[J]. Physical Review E, 2004, 69(6): 1-8.
- [9] Watts D J, Muhamad R, Medina D C, et al. Multiscale, resurgent epidemics in a hierarchical metapopulation model[J]. Proceedings of the National Academy of Sciences of the United States of America, 2005, 102(32): 11157-11162.
- [10] Li X, Gao C. A human-oriented virus propagation model in email networks [J]. Journal of Theoretical & Applied Information Technology, 2013, 48(3): 1391-1397.
- [11] Barabási A L. Bursts: the hidden patterns behind everything we do, from your e-mail to bloody crusades[M]. USA: Penguin, 2011.
- [12] Barabasi A L. The origin of bursts and heavy tails in human dynamics[J]. Nature, 2005, 435(7039): 207-211.
- [13] Shetty J, Adibi J. The Enron email dataset database schema and brief statistical report[J]. Information Sciences Institute Technical Report, University of Southern California, 2004, 4(1): 120-128.
- [14] Liang X, Zheng X, Lv W, et al. The scaling of human mobility by taxis is exponential[J]. Physica A: Statistical Mechanics and its Applications, 2012, 391(5): 2135-2144.
- [15] Roth C, Kang S M, Batty M, et al. Structure of urban movements: polycentric activity and entangled hierarchical flows[J]. PLoS one, 2011, 6(1): 1-8.
- [16] Barabási A L, Albert R, Jeong H. Mean-field theory for scale-free random networks[J]. Physica A: Statistical Mechanics and its Applications, 1999, 272(1-2): 173-187.
- [17] Pastor-Satorras R, Vespignani A. Epidemic spreading in scale-free networks[J]. Physical Review Letters, 2001, 86(14): 3200-3203.
- [18] Moore C, Newman M E J. Epidemics and percolation in small-world networks[J]. Physical Review E, 2000, 61(5): 5678-5682.
- [19] Eguiluz V M, Klemm K. Epidemic threshold in structured scale-free networks[J]. Physical Review Letters, 2002, 89(10): 1-4.
- [20] Zou C C, Towsley D, Gong W. Modeling and simulation study of the propagation and defense of internet e-mail worms[J]. IEEE Transactions on Dependable and Secure Computing, 2007, 4(2): 105-118.
- [21] 《运筹学》教材编写组. 运筹学[M]. 北京:清华大学出版社, 2015.
Operational Research Teaching Materials Writing Group. Operational Research[M]. Beijing: Tsinghua university press, 2015. (in Chinese)
- [22] 周涛, 韩筱璞, 闫小勇, 等. 人类行为时空特性的统计力学[J]. 电子科技大学学报, 2013, 42(4): 481-540.
ZHOU Tao, HAN Xunpu, YAN Xiaoyong, et al. Statistical mechanics on temporal and spatial activities of human[J]. Journal of University of Electronic Science and Technology of China, 2013, 42(4): 481-540. (in Chinese)
- [23] Lancichinetti A, Fortunato S, Radicchi F. Benchmark graphs for testing community detection algorithms[J]. Physical Review E, 2008, 78(4): 1-6.