

DOI:10.19431/j.cnki.1673-0062.2022.04.014

基于节点重要性的复合免疫策略研究

戴剑勇^{1,2,3}, 范梦鸽^{1,3}

(1. 南华大学 经济管理与法学学院, 湖南 衡阳 421001; 2. 南华大学 资源环境与安全工程学院, 湖南 衡阳 421001; 3. 湖南省核能经济与管理研究基地, 湖南 衡阳 421001)

摘要:为有效抑制病毒在网络中扩散,提出一种基于节点度、介数和聚类系数的复合免疫策略,该算法综合计算三类特征量得出节点重要性并免疫重要性高的节点,在人工生成网络和真实农产品供应链网络上分别进行仿真实验,结果表明:实施该复合免疫后网络中感染节点密度值低于同等免疫比例下的度免疫、节点介数免疫和聚类系数免疫策略,具有更高的免疫效率。此外,该复合免疫算法中节点度占比大小会影响免疫效果,存在一个最优的度权值使免疫效果最优。

关键词:免疫策略;病毒传播;复杂网络

中图分类号:TP393.08 **文献标志码:**A

文章编号:1673-0062(2022)04-0088-06

Compounded Immune Strategy Based on Node Importance

DAI Jianyong^{1,2,3}, FAN Mengge^{1,3}

(1. School of Economics, Management and Law, University of South China, Hengyang, Hunan 421001, China; 2. School of Resource Environment and Safety Engineering, University of South China, Hengyang, Hunan 421001, China; 3. Hunan Economics and Management Research Base of Nuclear Energy, Hengyang, Hunan 421001, China)

Abstract: In order to effectively inhibit the spread of virus in the network, a compound immune strategy based on node degree, intermediate number and clustering coefficient is proposed. The algorithm comprehensively calculates three kinds of characteristic quantities to obtain the nodes with high node importance and immune importance. Simulation experiments are carried out on artificially generated network and real agricultural product supply chain network respectively, the results show that after compound immunization, the density of infected nodes in the network is lower than the degree immunization, node intermediate immunization and cluster coefficient immunization strategies under the same immu-

收稿日期:2022-03-03

基金项目:湖南省教育厅重点资助科研项目(18A235)

作者简介:戴剑勇(1969—),男,教授,博士,主要从事安全系统工程与风险管理方面的研究。E-mail: daijy13@163.com

nization proportion, which has higher immunization efficiency. In addition, the proportion of node degree in compound immune algorithm will affect the immune effect. There is an optimal degree weight to optimize the immune effect.

key words: immunization strategy; virus propagate; complex network

0 引言

病毒这一词汇在生物学领域往往与疾病相关联,但在现实社会中,计算机病毒和供应链风险等都与生物病毒一样具有强感染性和强破坏性,2020年中国互联网网络安全报告显示,全年检测出的恶意程序数量逾4200万个,日均传播次数高达482万次^[1],互联网技术的发展使人与人之间联系更加紧密,与此同时计算机病毒对网络安全的威胁也更加无法忽视,同样地,各类风险在供应链网络中传播也为人类社会带来难以估量的损失。因而,如何对复杂网络中病毒或风险传播进行高效的控制成为学术界极具价值的议题,网络免疫策略通过将“抗体”赋予网络中一定比例的节点,使其对病毒具有免疫力,在现实情况中,受限于成本与资源,对网络中所有的节点企业都实施免疫显然不可能,识别网络中关键节点并使其免于病毒感染是减小病毒传播规模的重要措施。

节点重要性与网络结构密不可分是现有研究中被诸多学者认同的观点^[2-5]。基于不同网络特征量指标的免疫策略不断涌现,其中最直观的为度中心性免疫^[6],将节点度值作为衡量一个节点影响力的指标,计算节点所有邻居节点个数,认为度值大的节点拥有更高的重要度排名,度中心性免疫被证明在无标度网络中具有较好的免疫效果。改进的熟人免疫^[7]就是在该思想的基础上对经典熟人免疫策略进行优化,对随机选取的一组节点,选取其度最大的邻居节点进行免疫,使经典熟人免疫策略的免疫效果得到大幅提升。

介数中心性^[8]也是中心性免疫中重要指标,该方法思想是计算连接节点 i 、 j 间最短路径中通过节点 m 的数目与所有最短路径条数的比值,通过免疫该值较高的节点来抑制病毒在网络中的传播,将该策略应用对象变换为网络中连边即可得到基于边的介数免疫^[9],考虑时间因素可以得到动态介数^[10],介数中心性免疫可以有效控制病毒在网络中传播范围。

复杂网络聚类系数^[11]可以描述节点的各邻居节点间存在连边的概率,即一个企业的两个合

作者间也存在合作关系的可能性,基于节点聚类系数的免疫策略更多考虑邻居节点间是否存在紧密关系。K-核免疫^[12]和基于路径信息的图覆盖免疫^[13]、路径免疫^[14]等策略都是学者在复杂网络免疫领域不断探索的结果,这些免疫策略有各自的优点但也存在不足的地方。K-核免疫在面对单一传染源时表现出高效率,但当网络中存在多传染源时,该方法免疫效果差。图覆盖免疫策略将免疫过程划分为多个步骤,根据“种子节点”位置和覆盖范围综合确定免疫节点,免疫效果较好,但该方法更多基于静态网络实现,对于路径信息随时间发生变化的动态网络而言,其免疫效果还有待考证。此外,将加权网络纳入考虑的改进重要性评估算法^[15]也受到学者青睐。

现有研究中基于节点重要性的免疫策略将关注点更多放在度和介数等单一节点信息上,但现实中网络中个体由于内外环境的不稳定往往呈现为一个复杂的综合体,其在复杂网络中的影响力程度是由多指标共同决定的,很难通过单一标准来准确衡量。在现实网络中,一个节点被感染的概率不仅受到其邻居节点数目多少的影响,同时也和其邻居节点往来密切程度有关。若邻居节点间关系紧密,即使该节点拥有的邻居节点很少,其被感染的概率也可能相对较高,此外,若一个节点是网络中重要的“中转站”,尽管该节点并非度值最高的节点,其依旧具有很高的被感染可能性。因此,本文综合考虑复杂网络特征量指标,提出一种改进的基于节点度、介数和聚类系数的复合免疫策略(immunization based on degree, betweenness and clustering coefficient of node, I-DBCC),并分别在人工生成网络和真实的农产品供应链网络中进行仿真分析。

1 理论和方法

1.1 经典免疫理论

为更加直观地刻画病毒在网络中传播过程,SI模型,SIS模型和SIR模型等理论模型逐渐发展成熟,在这些模型中个体被划分为三个群体,分别为易感态个体(Susceptible),感染态个体(In-

ected)和移除态个体(Recovered),随着感染进程地推进,个体会在以一定概率由易感态转变为感染态,而被感染的个体同样具有一定可能性变为移除状态。基于病毒传播动力学相关理论,学者们提出了复杂网络随机免疫、熟人免疫和目标免疫三类经典免疫策略。

随机免疫或称均匀免疫,是操作最为简单的一种免疫算法,其完全忽略节点间差异,免疫节点的选取遵循随机原则,对于无标度网络而言,若想完全抑制病毒的扩散,需要免疫几乎所有节点。熟人免疫分两步实施免疫,第一步随机选择一定比例节点,第二步在被选择节点的所有邻居节点中随机选择一个作为免疫节点,熟人免疫摆脱了对全局信息的依赖,其仅需部分节点的邻居节点信息即可实现免疫。目标免疫根据一定的标准来选取网络中重要节点进行免疫,经典的目标免疫将度作为评价节点重要性的指标,据前文分析可知,节点介数和聚类系数等指标都是影响节点重要度的因素。

在此定义节点*i*介数如下:

$$B(i) = \sum_{n,j \in v} \frac{\delta_{nj}(i)}{\delta_{nj}} \quad (1)$$

其中 δ_{nj} 和 $\delta_{nj}(i)$ 分别表示节点*n,j*间最短路径条数和这些最短路径中经过节点*i*的条数。

度为 k_i 的节点*i*的聚类系数 C_i 可以描述为:

$$C_i = \frac{2E_i}{k_i(k_i - 1)} \quad (2)$$

式中 E_i 表示节点*i*所有邻居节点间实际存在的连边数, $\frac{k_i(k_i-1)}{2}$ 则表示这 k_i 个节点间可能存在的最大边数,简单整理即可得到式(2)。

1.2 基于节点度、介数和聚类系数的复合免疫策略

相当多的研究都证实度是评价节点影响力大小的重要指标,若网络中被感染的某个节点具有较多邻居节点,则病毒经由该节点向外扩散后影响的可能范围较大。但节点度值的大小并非评判节点是否重要的唯一因素,若被感染节点的邻居节点间接触频繁,体现在供应链网络中即企业间资金流、信息流等交互频次高,企业合作关系紧密,此时被感染节点仅将病毒传染给一个邻居节点就可能造成大范围的传播,节点间这种关系紧密程度可以用聚类系数来量化表示。同样地,网络中存在介数值高的节点,这类节点相当于交通图中的“关键路口”,很多路线都要途径该节点,

若其成为感染节点,波及范围大。为证明上述观点,分别以图1和图2为例探讨聚类系数和介数在节点传播影响力方面发挥的作用。

比较图1中节点*i*和*j*可以发现,二者度均为4,节点*j*有别于节点*i*的地方在于其邻居节点间存在连边,也就是说节点*j*聚类系数高于节点*i*。现假设节点*i*被病毒感染,则其邻居节点仅可能由节点*i*感染;若假设节点*j*为感染节点,其邻居节点除了可能直接被节点*j*感染外,其他已经被节点*j*感染的邻居节点也有可能成为该节点的感染源,由此可见,节点*j*的病毒传播力要大于具有相同度的节点*i*。

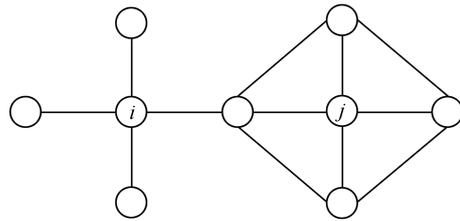


图1 聚类系数对节点传播力影响示例

Fig.1 Illustration of the influence of clustering coefficient on node propagation force

观察图2可知,节点*m*度值为4,节点*n*度值为3,图中A、B、C三个部分内任意节点到达另一个部分内任意节点间最短路径必然要经过节点*n*,即与节点*m*相比,节点*n*度值小但介数值高。在进行免疫节点选择时,若被病毒感染的节点位于A或C部分,即使节点*m*具有最多的邻居节点,选择其作为免疫节点来控制病毒传播收效甚微。相较而言,节点*n*作为免疫节点可以更高效地抑制病毒在网络中的传播,当节点*n*可以免疫病毒后,A、B、C部分间连接断裂,病毒传播被控制在尽可能小的范围内。

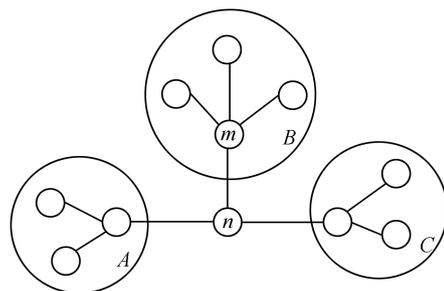


图2 介数对节点传播力影响示例

Fig.2 Illustration of the influence of betweenness on node propagation force

基于上述分析,本文综合考虑节点度值 D_i 、介数值 B_i 与聚类系数值 C_i ,提出了改进的复合免疫策略,将节点 i 的 P_i 定义为:

$$P_i = D(i) \times B(i) + C(i) \quad (3)$$

运用 I-DBCC 免疫策略评估节点重要性时不再依赖单一指标,而是依式(3)计算各节点 P_i 值,并按设定的免疫比例选取该值最高的节点进行免疫。图 1 情形表明对于拥有相同数量邻居节点的感染节点而言,免疫聚类系数更高的那个效果更优;图 2 情形表明即使一个节点度值较小,但若它具有较高介数值,免疫该节点仍可对病毒在网络中传播起到很好的抑制效果。因此,I-DBCC 免疫将节点度、介数和聚类系数同时纳入考虑,取三者综合值来评价节点重要程度,力求更准确识别网络中免疫节点,以更小的免疫比例得到更优的免疫效果。

2 实验仿真与结果分析

多位学者的研究验证了大多数实际网络都具有无标度特性的结论^[16-18],因此,本文利用 R-studio 软件生成一个服从幂律分布的人工无标度网络,同时选取一个真实农产品供应链网络数据进行仿真实验,病毒传播过程仿真基于 SIR 模型,假设网络初始态仅包含易感态节点,网络中移除态节点均由感染态节点转化而来,由此本文以移除态节点变化情况衡量免疫策略效果。

2.1 人工和真实网络上的仿真结果

本节分别在人工生成网络和真实网络上进行仿真,人工生成网络参数设定为 $N=200$,真实农产品供应链网络共包含 1 653 个节点,两个网络主要参数见表 1,表中 N 代表网络规模, $\langle k \rangle$ 表示网络平均度值, k_{\max} 指网络中节点度最大值, C 对应网络平均聚类系数。对农产品供应链网络度分布整理分析结果如图 3 所示,在双对数坐标系中度分布近似呈现幂律分布。

表 1 网络特征参数表

Table 1 Table of network characteristic parameters

网络	N	$\langle k \rangle$	k_{\max}	C
人工生成网络	200	3.940	32	0.040
农产品供应链网络	1 653	2.051	99	0.004

为探讨 I-DBCC 免疫算法效果,分别在上述两个网络中实施度中心性免疫、介数中心性免疫、

聚类系数中心性免疫和 I-DBCC 免疫,假定网络中初始仅存在一个感染节点,感染概率为 0.36,恢复概率为 0.1,设定时间步长为 80,取 10 次仿真结果平均值,改变免疫比例重复进行多次实验得到图 4 和图 5,分别展示四类免疫策略在人工生成网络和真实农产品供应链网络上免疫效果随免疫比例变化情况。从图 4 中可以看出,当免疫比例为 0.05 时,I-DBCC 免疫与度免疫和介数免疫效果相差不多,且三者效果优于聚类系数免疫;当免疫比例为 0.05 ~ 0.15 时,I-DBCC 免疫效果远优于度免疫、介数免疫和聚类系数免疫三类免疫策略,与此同时,观察四条曲线斜率变化情况不难发现,当免疫比例处于区间(0.05,0.15)内时,I-DBCC 免疫效果变化曲线更陡峭,即免疫比例在原有基础上增加相同百分比时,I-DBCC 免疫策略可以使感染节点密度更快减小,换言之,当免疫比例为 0.05 ~ 0.15 时,提高 I-DBCC 免疫策略免疫比例可以更快达到降低感染规模的效果;当免疫比例大于 0.15 时,I-DBCC 免疫效果同样优于度免疫、介数免疫和聚类系数免疫。

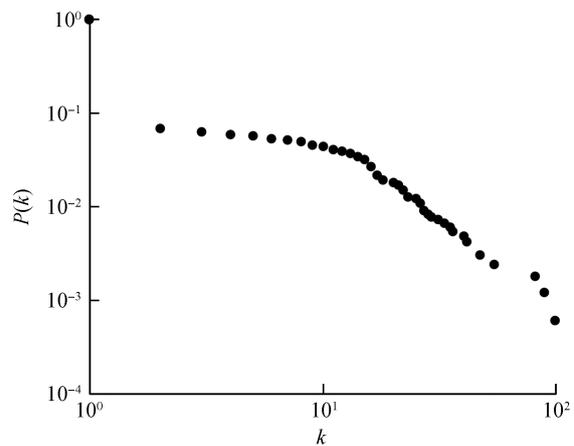


图 3 农产品供应链网络度分布

Fig. 3 Network degree distribution of agricultural products supply chain

图 5 展示了免疫比例小于 0.05 时各免疫策略效果情形,对比图 5 中四条曲线可以看出,免疫比例小于 0.01 时,即免疫节点较少时,实施 I-DBCC 免疫可以比同比例的度免疫、介数免疫和聚类系数免疫取得更好的免疫效果;当免疫比例为 0.01 ~ 0.05 时,I-DBCC 免疫优越性有所下降,但效果仍优于其他三类。综上所述,本文所提 I-DBCC 免疫策略效果能更好地抑制网络中病毒传播,免疫相同节点时,该策略效果优于度免疫、介

数免疫和聚类系数免疫策略。

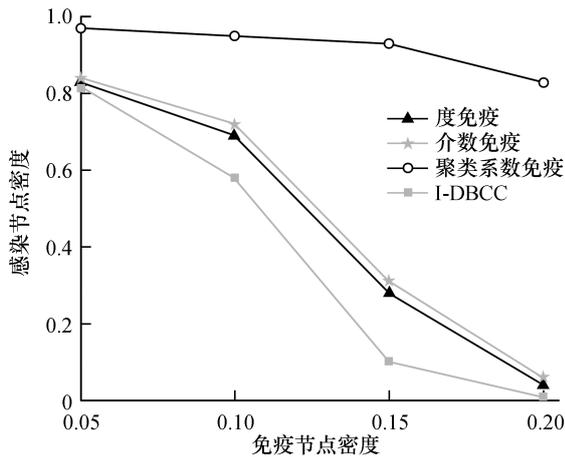


图4 人工生成网络免疫效果

Fig. 4 Immune effect in artificial generation network

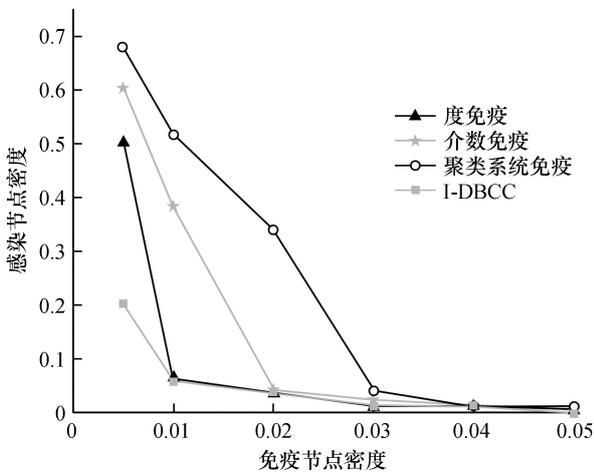


图5 农产品供应链网络免疫效果

Fig. 5 Immune effect in agricultural products supply chain network

2.2 度占比对免疫效果的影响

复杂网络中节点度与介数间关系也引起很多学者的讨论,不可否认网络中存在具有高度值同时具有低介数值的节点,同样也存在度值低但拥有高介数值的节点,究竟免疫哪类节点会得到更高的免疫效率,或者,在对节点重要性进行衡量时,应赋予度和介数哪个指标量更高的权重,为进一步探究该问题,本文在公式(3)的基础上定义如下公式:

$$P_i = D(i)^\alpha \times B(i)^{1-\alpha} + C(i) \quad (4)$$

式中参数 $\alpha \in [0, 1]$, 该值代表度占比, α 增大时, $D(i)^\alpha$ 减小, 而 $B(i)^{1-\alpha}$ 增大, 即节点度所占比重变小, 节点介数比重增加, $\alpha = 0$ 表示节点度权重

为0, 当 $\alpha = 1$ 时, 节点介数权重为0, 若 $\alpha = 0.5$, 则表示节点度和介数比重相同, 此时等同于 I-DBCC 免疫。

本节分别取免疫比例 5%、10%、15% 和 20% 在人工生成网络中进行仿真实验, 调整 α 值大小, 根据公式(4)计算并选取免疫节点, 取 10 次仿真数据平均值, 结果见图 6。如图所示, 当免疫比例为 5% 和 10% 时, 随着 α 增加, 感染节点密度呈现先降后升趋势, 而当免疫比例增加至 15% 和 20%, α 增加使感染节点密度先增后减, 到达最低点后又逐步上升, 曲线最低点处对应的 α 值即效果最优的选项。观察图 6 发现, 当免疫比例为 5%, 10% 和 15% 时, α 取值在 0.8 附近可使感染节点密度到达最小值, 此时免疫效果最好; 而当免疫比例为 20% 时, α 取值在 0.6 ~ 0.8 之间可取得最好的免疫效果。因此, 比重参数 α 取值会对免疫效率产生影响, 且存在一个最佳的 α 值使感染节点密度最小, 但该值在免疫比例不同时会在数值大小上存在差异。

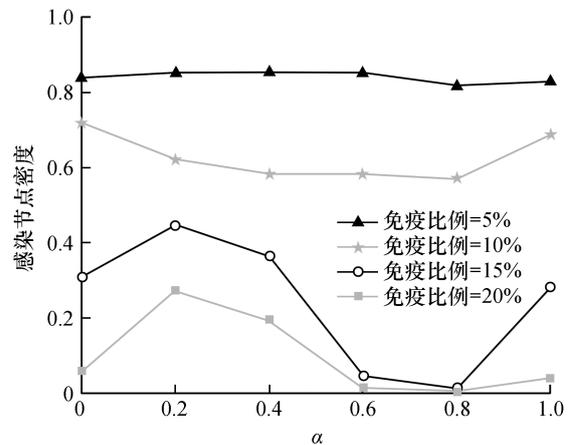


图6 感染节点密度与关系

Fig. 6 Relationship between infection node density and alpha

3 结论

识别出复杂网络中具有较强病毒传播力的节点对于抑制病毒扩散具有重要现实意义, 本文利用病毒传播 SIR 模型, 综合计算节点度值、介数值和聚类系数值, 提出一种基于这三类特征量的复合免疫策略(I-DBCC), 取 P_i 值大的节点进行免疫, 并在人工生成网络和真实农产品供应链网络进行仿真分析。实验结果表明, I-DBCC 免疫算法相较于于仅考虑单一指标的节点度免疫、介数免疫

和聚类系数免疫对病毒传播的抑制效果更优。通过改变 I-DBCC 免疫算法中度与介数权重占比发现,存在一个最优的度占比使得感染节点密度最低,免疫效果最好,但该值并非固定值,其值大小与免疫节点密度相关。

参考文献:

- [1] 国家互联网应急中心. 2020 年中国互联网网络安全报告 [R/OL]. (2021-07-21) [2022-02-25]. <https://www.cert.org.cn/publish/main/upload/File/2020%20Annual%20Report.pdf>.
- [2] 朱大锐,王睿,程文姬,等. 基于改进 PageRank 算法的输电网关键节点辨识方法研究[J]. 电力系统保护与控制,2022,50(5):86-93.
- [3] NIAN F Z, HU C S. The most common friend first immunization[J]. Chinese physics B,2016,25(12):534-542.
- [4] 马媛媛,韩华. 基于有效距离的复杂网络节点影响力度量方法[J]. 复杂系统与复杂性科学,2022,19(1):12-19.
- [5] 闫涵,康海燕. 有向重叠社区下考虑兴趣度的谣言免疫策略研究[J]. 郑州大学学报(理学版),2020,52(4):82-88;95.
- [6] YAN X L, CUI Y P, NI S J. Identifying influential spreaders in complex networks based on entropy weight method and gravity law[J]. Chinese physics B,2020,29(4):048902.1-048902.21.
- [7] HOLME P. Efficient local strategies for vaccination and networks attack[J]. Europhysics letters,2004,68(6):908-914.
- [8] LIU Y Y, SLOTINE J J, BARABÁSI A L. Controllability of complex networks[J]. Nature,2011,473(7346):167-173.
- [9] 刘建友,李代平. 基于三元组结构的在线社交网络边免疫策略研究[J]. 计算机工程,2018,44(6):130-135;140.
- [10] 叶焯星,杨飞. 基于介数中心性的交通拥堵指数计算[J]. 测绘通报,2021(5):86-90.
- [11] 林冰轩,徐明兴,陈志刚,等. 基于复杂网络蜂群无人机网络拓扑结构分析[J]. 火力与指挥控制,2022,47(1):38-42;49.
- [12] QIU L Q, ZHANG J Y, TIAN X B. Ranking influential nodes in complex networks based on local and global structures[J]. Applied intelligence,2021,51(7):4394-4407.
- [13] HUANG X L, ZOU F T, MA F Y. Targeted local immunization in scale-free networks[J]. Journal of computer science and technology,2007,22(3):457-468.
- [14] 林兵,郭文忠,陈国龙,等. 无标度网络中基于最短路径免疫策略的病毒传播研究[J]. 计算机科学,2012(增刊):136-138.
- [15] 孙兴龙,李亚雄,邱艳粉. 基于改进融合算法的交通网络节点重要性评估[J]. 火力与指挥控制,2021,46(4):65-71.
- [16] 宋旭妍,彭甜甜,张高军. 粤港澳大湾区水路客运网络特征[J]. 热带地理,2022,42(2):247-255.
- [17] 邢佳亮,郭强,刘建国. 融合局部聚类特征的银行间网络重构研究[J]. 电子科技大学学报,2021,50(5):782-787.
- [18] 董苏雅拉图. 证券市场中具有流动性人口特征的恐慌情绪传播模型[J]. 系统科学与数学,2021,41(10):2919-2931.
- (上接第74页)
- [7] HEALTH D O. Supplement 1 Isolation facilities for infectious patients in acute settings:Health Building Note 04-01[S]. UK: The UK Department of Health,2013:9-10;15.
- [8] 凌继红,于会洋,李猛,等. 气流组织对负压隔离病房排污效率的影响[J]. 天津大学学报(自然科学与工程技术版),2014,47(2):174-179.
- [9] 赵越. 气流组织对负压隔离病房污染物扩散的影响研究[D]. 天津:天津大学,2011:36-50.
- [10] 嵇赞喆,王晓杰. 呼吸道传染性负压隔离病房带菌飞沫和 CO₂ 分布特性的模拟比较研究[J]. 护理学杂志,2011,26(21):1-5.
- [11] CHO J Y. Investigation on the contaminant distribution with improved ventilation system in hospital isolation rooms:Effect of supply and exhaust air diffuser configurations[J]. Applied thermal engineering,2019,148:208-218.
- [12] HANG J, LI Y G, JIN R Q. The influence of human walking on the flow and airborne transmission in a six-bed isolation room:Tracer gas simulation[J]. Building and environment,2014,77:119-134.
- [13] 段栋丹. 呼吸强度对颗粒物在下呼吸道暴露量影响的研究[D]. 西安:西安建筑科技大学,2018:29-36.