

# 时序网络上的随机游走免疫策略研究

朱义鑫<sup>1,2</sup>, 张凤荔<sup>2</sup>, 王瑞锦<sup>2</sup>, 秦志光<sup>2</sup>

(1. 新疆财经大学计算机科学与工程学院 乌鲁木齐 830012; 2. 电子科技大学网络与数据安全四川省重点实验室 成都 610054)

**【摘要】**针对传统免疫模型在时序网络中所面临的难以收集、分析网络拓扑信息的困境,提出了基于随机游走机制的免疫策略,一定数量的免疫粒子被随机地分配到网络节点上,当该节点有边激活时,免疫粒子就可以沿着激活边游走到另一节点,获得免疫粒子的节点获得免疫能力,失去免疫粒子的节点转换成非免疫的易感态。根据随机游走者之间在转移时是否相互影响,分别建立了非独立随机游走免疫模型和 $P$ 独立随机游走免疫模型。在这两种免疫模型中,免疫粒子传播所需的网络开销受到事先给定的免疫粒子密度的限制。实验表明,这两种随机游走免疫模型可以获得比熟人免疫模型更好的免疫效果,而与目标免疫模型比较结果取决于网络拓扑结构的异质性程度。

**关键词** 阵发性; 免疫策略; 随机游走; 时序网络; 病毒传播

**中图分类号** TP393.08 **文献标志码** A **doi**:10.3969/j.issn.1001-0548.2017.01.015

## Research on Immunization Strategy Based on Random Walk Mechanism in Temporal Networks

ZHU Yi-xin<sup>1,2</sup>, ZHANG Feng-li<sup>2</sup>, WANG Rui-jin<sup>2</sup>, and QIN Zhi-guang<sup>2</sup>

(1. School of Computer Science and Engineering, Xinjiang University of Finance and Economics Urumqi 830012;

2. Network and Data Security key Laboratory of Sichuan Province, University of Electronic Science and Technology of China Chengdu 610054)

**Abstract** For the problems of traditional immune models arising in collecting and analyzing network topology information in temporal networks, an immune strategy based on random walk mechanism is put forward and it can be implemented without collecting network topology information. A certain number of immune particles were randomly assigned to the network nodes. When a node with immune particles has one or more activated links, the immune particles on the node will walk to another node along an activated link of the node. The nodes with immune particles acquire the immunity, but the nodes losing immune particles will be transformed into the susceptible state. Considering whether the random walkers exert impact upon each other when they move, the dependent random walk immune model and the  $P$ -independent random walk immune model are established, respectively, in which the network transmission overhead of the immune particle is limited by a given immune particle density. Experiments show the two random walk immune models are characterized by their better immune effects with lower immune particle density and the network overhead when compared with acquaintances immune model. In addition, the comparative result with the target immune model depends on heterogeneity of network topology.

**Key words** burstiness; immunization strategy; random walk; temporal network; virus propagation

新型网络病毒不断涌现,不仅威胁到网络和主机的安全,也威胁着网络用户的信息安全。为了阻止或抑制病毒在网络上的传播,各种免疫策略被提出,如环状免疫、目标免疫、熟人免疫、局部免疫及优先删边免疫等<sup>[1-5]</sup>。这些免疫策略往往是选择一些重要节点实施免疫,被免疫的节点既不会感染病毒也不会传播病毒。熟人免疫策略有两个针对时序网络的改编版<sup>[6]</sup>: 1) 权重协议,选择与随机选取的节点交互次数最多的节点加入免疫节点集; 2) 最近

协议,选择与随机选取的节点最后一次交互的节点加入免疫节点集。这些免疫策略实施的一个前提条件是事先收集、分析网络拓扑信息,以确定免疫对象。在时序网络中,节点间的连接是暂时的、反复的且具有阵发性的,即网络结构是不断变化的,故只能根据部分的或是已知的时序网络信息,对节点进行排名,以确定网络节点的重要性。在大型网络中,收集、分析含有时间维度的网络信息,以确定免疫对象十分困难。因此,本文研究无需收集网络

收稿日期: 2015-03-11; 修回日期: 2016-06-10

基金项目: 国家自然科学基金(61440047); 国家自然科学基金青年项目(61502087)

作者简介: 朱义鑫(1974-),男,博士,主要从事计算机网络安全、复杂网络传播方面的研究。

拓扑信息, 就可以快速实施的免疫策略。

### 1 多粒子随机游走过程分析及分类

网络的时间属性会对扩散过程产生怎样的影响, 解决这一问题的第一种方法是在真实或人工合成的带时间戳的事件序列数据上模拟随机游走过程, 并与所定义的空模型比较相应的动态属性<sup>[7-9]</sup>。第二种方法, 研究分析在特定时序网络模型上的随机游走属性<sup>[10-11]</sup>。

文献[10, 12-13]提出将时序网络建模成事件随机序列, 其中每个链接上的事件服从规定的间隔时间分布。考虑  $N$  个节点的一个无向网络, 由节点集  $V = \{1, 2, \dots, N\}$  和边集  $E$  组成, 这种网络往往被称为聚合网络, 它被认为是时序网络的聚合。用更新方式给每个边  $e$  分配一个随机事件间隔时间  $\tau_e$ , 通过在聚合网络上添加时间维度生成随机时序网络<sup>[10, 12]</sup>。用  $\psi_e(t)$  表示  $\tau_e$  的概率密度函数(probability density function, PDF), 并假设  $\tau_e$  的均值有限。随机游走者可以立刻从当前节点跳到一个邻近节点, 只要这两个节点之间的链接出现。

在时序网络中, 根据随机游走者沿着一个链接到达一个节点时, 该节点的其他链接的事件间隔时间<sup>[14]</sup>是否重新初始化, 随机游走过程可以分为主动随机游走和被动随机游走两种模式。1) 主动随机游走过程被定义为一个更新过程, 一个随机游走者到达一个节点后, 这个节点的所有链接的事件间隔时间重新初始化<sup>[10, 12]</sup>。它的非马尔科夫性是由于这样一个事实: 一个事件发生的概率取决于前面的事件。这种随机过程可以用广义主方程表示, 并且可以获得一些属性的解析解, 比如稳态密度<sup>[12]</sup>和弛豫时间。2) 被动随机游走模式, 当一个游走者到达一个节点时, 只初始化刚使用的链接的事件间隔时间。与主动随机游走相比, 被动随机游走被认为是一种更自然的网络扩散模式。因为在现实中, 扩散的实体, 如病毒, 通常不会重新启动一个活跃边两端的节点。被动随机游走比主动随机游走表现出更强的非马尔科夫性, 因为被动随机游走在某种程度上记住了过去的游走轨迹。

本文提出的随机游走模型包含多个随机游走者, 每一个随机游走者都是一个免疫粒子, 它作为一种特殊的蠕虫程序, 可以转移但不能增生, 即它可以从一个节点转移到另一节点, 但不能增加免疫体数量。获得免疫粒子的节点获得免疫力, 失去免疫粒子的节点转换成非免疫的易感态。在多随机游

走者的模型中, 无论是主动随机游走模式还是被动随机游走模式都存在一个同样的问题, 即随机游走者之间在转移时是否相互影响。任意一个免疫粒子在随机游走中不受其他免疫粒子影响的模型被称为独立随机游走免疫模型。免疫粒子在随机游走中相互影响的模型被称为非独立随机游走免疫模型。

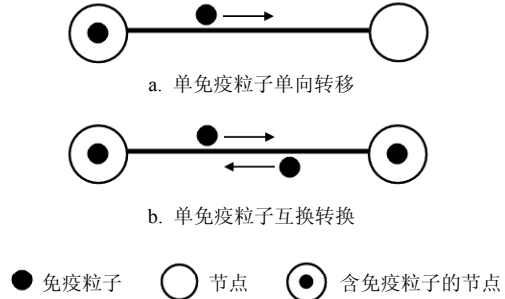


图1 节点含有单个免疫粒子的随机游走

如图1所示, 当一个边被激活后, 这条边的两个端节点中的免疫粒子便可以从一个端节点沿着这条活跃边转移到另一个端节点。图1中的直线表示连接两个节点的边, 直线上方的实心圆数目表示将要转移的粒子数, 直线上方的箭头表示免疫粒子的转移方向, 以上解释同样适用于图2和图3。当端节点中含有的免疫粒子数不超过1个时, 独立与非独立的随机游走方式是相同的。图1b中的双方互换免疫粒子, 看似是一个浪费网络资源的无意义的数据传输, 但它避免了通信双方在传输数据前互换彼此状态信息(含有的免疫粒子数)所带来的网络开销, 同时也简化了免疫粒子转移时的传输协议。

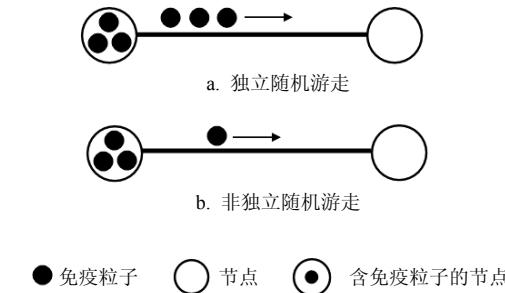
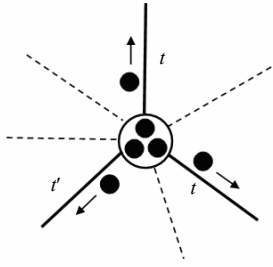


图2 节点含有多个免疫粒子的随机游走

当出现节点含有多个免疫粒子的情形时, 这两种随机游走方式存在区别。由于独立随机游走的免疫粒子互不影响, 当其所在节点有边出现(激活)时, 免疫粒子就会沿着边转移到另一节点, 而无论其所在节点中是否有其他免疫粒子存在或转移。因此, 如图2a所示, 对于独立随机游走方式, 当图中的边被激活后, 该边左侧端节点中的所有免疫粒子都会向右侧端节点转移。而对于非独立的随机游走方式,

如图2b所示,当图中的边被激活时,该边左侧端节点中的多个免疫粒子只有一个会向右侧端节点转移,而其他免疫粒子或者选择该节点同时激活的其他边,或者等待该节点下一个活跃边的到来。如图3所示, $t$ 时刻,节点有两个活跃边,每个活跃边都有一个免疫粒子经此转移。 $t'$ 时刻( $t' > t$ ),节点有一个活跃边,剩下的一个免疫粒子经此活跃边转移至另一端的邻节点。



● 免疫粒子 ○ 含免疫粒子的节点 — 活跃边 ---- 非活跃边

图3 非独立随机游走示例图

将节点上的免疫粒子看作是一种有限资源,用  $(A_i, B_i)$  表示节点  $i$  上的资源信息,其中  $A_i \geq 0$ ,  $B_i \geq 0$ 。 $A_i$  表示节点  $i$  上现有的免疫粒子数,即该节点在本时间步内还可以转移的免疫粒子数;而  $B_i$  表示本时间步内已从邻居节点转移来的免疫粒子数,这些免疫粒子在本时间步内不允许再转移出去,以防止一个免疫粒子在一个时间步内出现多跳,即防止一个免疫粒子在一个时间步内依次完成经历多个节点的转移。比如免疫粒子从一个节点转移到另一节点后,再转移到第3个节点上。

设  $(A_i, B_i)$ 、 $(A_j, B_j)$  分别是边  $e_{ij}$  被激活时节点  $i$ 、 $j$  的资源信息,如果  $A_i > 0$  且  $A_j > 0$ ,则本次数据传输结束时,独立随机游走方式下的节点  $i$ 、 $j$  的资源分别为  $(0, B_i + A_j)$  和  $(0, B_j + A_i)$ ;而非独立随机游走方式下的节点  $i$ 、 $j$  的资源分别为  $(A_i - 1, B_i + 1)$  和  $(A_j - 1, B_j + 1)$ 。如果  $A_i > 0$  而  $A_j = 0$ ,则本次数据传输结束时,独立随机游走方式下的节点  $i$ 、 $j$  的资源分别为  $(0, B_i)$  和  $(0, B_j + A_i)$ ;而非独立随机游走方式下的节点  $i$ 、 $j$  的资源分别为  $(A_i - 1, B_i)$  和  $(0, B_j + 1)$ 。其他情况可以参照图1、图2类推。

## 2 随机游走免疫模型

### 2.1 非独立随机游走免疫模型

在每个时间步  $t$ ,为避免由于激活边的处理顺序使得随机游走实验结果偏好节点序号小或大的节点,每次从第  $t$  个时间步激活的边集  $E_t^a$  中随机取出

一条边,直至集合  $E_t^a$  为空。在每个时间步,所有活跃边都处理完毕后,统一更新所有节点的资源状态,节点  $i$  的资源状态由  $(A_i, B_i)$  更新为  $(A_i + B_i, 0)$ ,  $i = 1, 2, \dots, N$ 。非独立随机游走免疫模型的具体算法如下。

算法1: 非独立随机游走免疫模型算法

输入: 网络中的免疫粒子总数  $m$ , 免疫开始时间  $t_0$ , 时序网络活跃边集序列  $E^a = \{E_0^a, E_1^a, \dots, E_{T-1}^a\}$ , 其中  $E_t^a$  表示  $t$  时刻所有活跃边所构成的集合。

输出: 网络所有节点的免疫粒子分布  $M = (A_0, A_1, \dots, A_{N-1})^T$ 。

从网络的所有节点中随机选出  $m$  个节点;

将每个选出的节点的资源置成  $(1, 0)$ , 其他节点的资源置成  $(0, 0)$ ;

for  $t \leftarrow t_0$  to  $T-1$

while  $E_t^a \neq \emptyset$  do

从活跃边集合  $E_t^a$  中随机选出一条边  $e_{ij}$ ;

$E_t^a \leftarrow E_t^a - \{e_{ij}\}$ ;

if  $A_i > 0$  then

$A_i \leftarrow A_i - 1$ ;

$B_j \leftarrow B_j + 1$ ;

end if

if  $A_j > 0$  then

$A_j \leftarrow A_j - 1$ ;

$B_i \leftarrow B_i + 1$ ;

end if

end while

for  $i \leftarrow 0$  to  $N-1$

$A_i \leftarrow A_i + B_i$ ;

$B_i \leftarrow 0$ ;

end for

end for

### 2.2 P\_独立随机游走免疫模型

独立随机游走模型有一个很明显的问题,如图2a所示,一旦有多个游走者到达同一节点,则它们将永远结伴而行,因为只要有边激活,它们便一同沿边转移,这对免疫一定数量的节点非常不利。一个好的解决办法是:当一个节点有边激活时,该节点上的每个免疫粒子便以概率  $P$  转移,以概率  $1-P$  不转移。每个免疫粒子是否转移与节点上是否有其他免疫粒子存在或转移无关,称这种独立随机游走免疫模型为  $P$ \_独立随机游走免疫模型。具体算法如下:

算法2:  $P$ \_独立随机游走免疫模型算法

输入: 免疫粒子转移概率  $P$ , 免疫粒子总数  $m$ , 免疫开始时间  $t_0$ , 时序网络活跃边集序列  $E^a = \{E_0^a, E_1^a, \dots, E_{T-1}^a\}$ , 其中  $E_t^a$  表示  $t$  时刻所有活跃边所构成的集合。

输出: 网络所有节点的免疫粒子分布  $M = (A_0, A_1, \dots, A_{N-1})^T$ 。

从网络的所有节点中随机选出  $m$  个节点;

将每个选出的节点的资源置成(1,0), 其他节点的资源置成(0,0);

for  $t \leftarrow t_0$  to  $T-1$

while  $E_t^a \neq \emptyset$  do

从活跃边集合  $E_t^a$  中随机选出一条边  $e_{ij}$ ;

$E_t^a \leftarrow E_t^a - \{e_{ij}\}$ ;

Particle\_num  $\leftarrow A_i$ ;

for  $k \leftarrow 1$  to Particle\_num

产生[0,1)之间的随机数RandomNum;

if RandomNum  $< P$  then

$A_i \leftarrow A_i - 1$ ;

$B_j \leftarrow B_j + 1$ ;

end if

end for

Particle\_num  $\leftarrow A_j$ ;

for  $k \leftarrow 1$  to Particle\_num

产生[0,1)之间的随机数RandomNum;

if RandomNum  $< P$  then

$A_j \leftarrow A_j - 1$ ;

$B_i \leftarrow B_i + 1$ ;

end if

end for

end while

for  $i \leftarrow 0$  to  $N-1$

$A_i \leftarrow A_i + B_i$ ;

$B_i \leftarrow 0$ ;

end for

end for

### 3 实验与分析

#### 3.1 以BA网络为底图的实验及分析

如图4所示, 实验所使用的时序网络底图(聚合网络)为BA网络, 其节点总数  $N = 1\ 000$ , 每个新加入节点连接5个已有节点, 边的激活间隔时间服从指

数为2.1的幂律分布。对于随机游走免疫模型而言, 免疫粒子密度是指要设定的免疫粒子数与网络节点总数的比值。对于目标免疫模型与熟人免疫模型而言, 免疫粒子密度是指要选择的免疫节点数与网络节点总数的比值。实验结果中的感染密度是指实验结束时, 网络中处于感染态的节点数与网络节点总数的比值。每次实验的总时间步为5 000, 对于每一个给定的免疫粒子密度值, 图中对应的各模型的感染密度值均为100次重复实验的均值。实验结果的图例中及分析论述中出现的URW、PRW分别表示非独立随机游走免疫模型和P\_独立随机游走免疫模型, 其中  $P = 0.5$ 。

实验以随机游走免疫模型的开始实施时间为0时刻, 目标免疫和熟人免疫开始实施时间为  $t_0$  时刻, 将  $[0, t_0]$  时间内收集到的所有节点间的交互聚合成一个无重边的无向网络, 就获得了时序网络在这个时间段内的拓扑结构信息。这个结构信息是熟人免疫和目标免疫为选择免疫节点而需要收集的网络信息, 可以用一个称之为网络拓扑信息量的定义来量化, 它可以由下式定义:

$$\text{网络拓扑信息量} = \frac{\text{给定时间内所收集的边数量}}{\text{网络底图的总边数}}$$

随机游走免疫模型与熟人免疫模型、目标免疫模型的免疫效果对比如图4所示。图例前缀AI表示熟人免疫, TI表示目标免疫, 图例中的数字表示熟人免疫模型和目标免疫模型实施免疫的开始时刻  $t_0$ 。由于熟人免疫(AI)结果的3条曲线基本重合, 目标免疫(TI)结果的3条曲线非常接近, 需要在图中的曲线旁注明其对应的免疫策略。在本实验中, 在  $t_0 = 1$  时, 网络拓扑信息量为43.8%, 在  $t_0 = 1\ 000$  时, 网络拓扑信息量增加至98.1%, 熟人免疫模型和目标免疫模型分别以这两个不同时刻作为免疫开始时间, 所获得的免疫效果并没有明显的区别, 这个信息表明网络拓扑信息量的持续增加并不能明显地持续改善这两种免疫策略的免疫效果。

如图4所示, 各免疫模型的免疫效果由好至差依次为URW免疫模型、PRW免疫模型、目标免疫模型、熟人免疫模型。与PRW免疫模型和目标免疫相比, URW免疫模型的免疫效果有明显的优势, 而前两者又明显优于熟人免疫模型。PRW免疫模型与目标免疫模型相比较, 随着免疫粒子密度的增大, PRW免疫模型的优势逐渐显现。

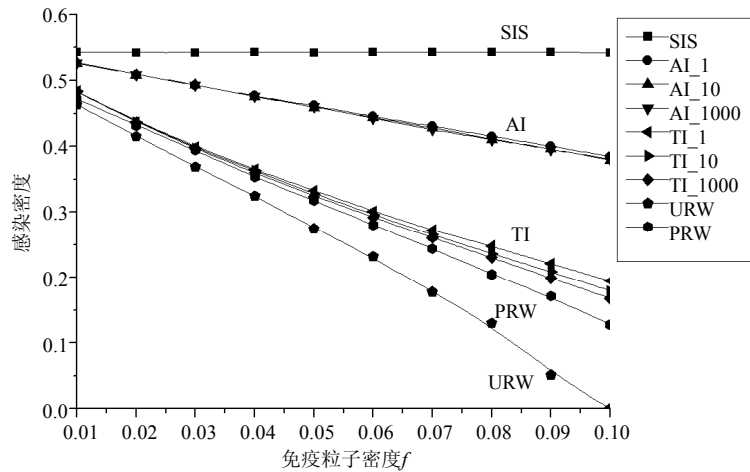


图4 BA网络为底图的免疫效果对比

### 3.2 以AS网络为底图的实验及分析

本实验所使用的真实网络数据是来自俄勒冈州立大学的路由器视角项目的在线数据。这个数据集包含了从1997年11月8日~2000年1月2日共785天中的733天的英特网自治域间的数据交换,所有数据交换关系聚合成一个AS网络。以这个AS网络作为时序

网络的底图,只有底图上存在的边才有可能在某个时刻被激活,并发送数据。本次实验仍设网络边激活间隔时间服从指数为2.1的幂律分布。测试前对网络做预处理,剔除网络中不连通的部分,只保留最大连通子图,处理后的聚合网络节点数  $N=6\ 474$ , 边的总数为26 467,在这里简记为AS20网络。

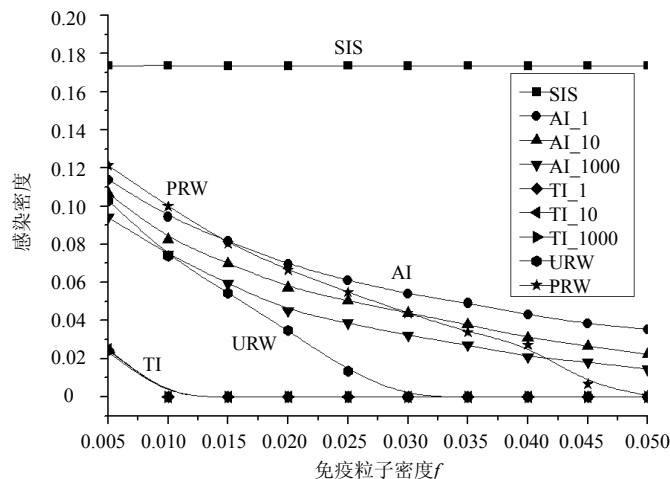


图5 AS20网络为底图的免疫效果对比

如图5所示,以AS20网络为底图的时序网络上,各免疫模型的免疫效果由好至差依次为目标免疫模型(TI)、URW免疫模型、PRW免疫模型、熟人免疫模型(AI)。AS20网络与前面所用的BA网络相比较有一个显著区别:网络度分布的异质性程度。BA网络度分布是服从幂指数为3的幂律分布,而AS20网络度分布的幂指数是小于1.5,即AS20网络的拓扑结构远比BA网络拓扑结构的异质性要强。意味着,与BA网络相比,AS20网络中的节点度集中在更少的节点上。因此,与以BA网络为底图的时序网络上的相应免疫模型的免疫效果(见图4)相比较,目标免疫模型

和熟人免疫模型都有显著提高,尤其是目标免疫的免疫效果有了很大的飞跃,其免疫粒子密度临界值  $f_c < 0.01$ ,成为免疫效果最好的模型。但目标免疫必须要能够搜集到足够的信息,并汇总分析以产生全局信息,只有这样才能对节点排序、免疫。因此,在大的时序网络上比较免疫效果,目标免疫只是作为一个标杆,其难以实施。另一方面,即便在拓扑结构的异质性很强的AS20网络上,URW免疫模型和PRW免疫模型的免疫效果也是优于熟人免疫模型的免疫效果,它们的免疫粒子密度临界值  $f_c$  分别小于或等于0.03和0.05,而熟人免疫模型的免疫粒子密度

临界值  $f_c$  是大于这两个值的(超出了图5的横坐标范围)。

与以BA网络为底图的时序网络上的各模型免疫效果比较, 不仅是目标免疫模型和熟人免疫模型的免疫效果有显著提高, URW免疫模型和PRW免疫模型的免疫效果也有较大的提高。在以AS20网络为底图的时序网络上, URW免疫模型和PRW免疫模型在免疫更少比例的节点基础上, 获得了更好的免疫效果。其原因仍然是与BA网络相比, AS20网络的拓扑结构异质性更强。对于拓扑结构异质性越强的网络, 就会有越少的重要性更强的节点。基于随机游走扩散机制的免疫粒子更倾向于游走到这些更重要的节点上, 所以在以AS20网络为底图的时序网络上, 通过免疫这些少量的重要节点, 就可以获得好的免疫效果。

## 4 结束语

本文提出了基于随机游走机制的免疫模型, 包括非独立随机游走免疫模型和 $P$ -独立随机游走免疫模型, 这两种免疫模型的免疫节点密度都不超过给定的免疫粒子密度。通过实验, 发现随机游走免疫模型的免疫效果优于熟人免疫模型, 与目标免疫模型比较结果取决于网络拓扑结构的异质性程度。随机游走免疫模型具有无需搜集、分析网络拓扑信息, 以及可以随时部署实施的优势。随机游走免疫模型需要考量免疫粒子传播所需要的网络开销, 而这一网络开销可以直接通过事先设定的免疫粒子密度等模型参数得以控制。

本文研究工作还得到了网络与数据安全四川省重点实验室开放课题(NDSMS201603)、新疆维吾尔自治区高校科研计划科学研究重点项目(XJEDU2016I036)以及新疆财经大学博士启动基金(2016BS008)的资助, 在此表示感谢。

## 参 考 文 献

[1] MULLER J, SCHONFISCH B, KIRKILIONIS M. Ring vaccination[J]. *J Math Biol*, 2000, 41(2): 143-171.

- [2] PASTOR-SATORRAS R, VESPIGNANI A. Immunization of complex networks[J]. *Phys Rev E Stat Nonlin Soft Matter Phys*, 2002, 65(3 Pt 2A): 036104.
- [3] COHEN R, HAVLIN S, BEN-AVRAHAM D. Efficient immunization strategies for computer networks and populations[J]. *Physical Review Letters*, 2003, 91(24): 247901.
- [4] HOLME P. Efficient local strategies for vaccination and network attack[J]. *EPL (Europhysics Letters)*, 2004, 68(6): 908.
- [5] CHEN Y, PAUL G, HAVLIN S, et al. Finding a better immunization strategy[J]. *Physical Review Letters*, 2008, 101(5): 058701.
- [6] LEE S, ROCHA L E, LILJEROS F, et al. Exploiting temporal network structures of human interaction to effectively immunize populations[J]. *PLoS One*, 2012, 7(5): e36439.
- [7] STARNINI M, BARONCHELLI A, BARRAT A, et al. Random walks on temporal networks[J]. *Physical Review E*, 2012, 85(5): 056115.
- [8] RIBEIRO B, PERRA N, BARONCHELLI A. Quantifying the effect of temporal resolution on time-varying networks[J]. *Scientific Reports*, 2013, 3(10): 3006.
- [9] ROCHA L E C, MASUDA N. Random walk centrality for temporal networks[J]. *New Journal of Physics*, 2014, 16: 063023.
- [10] HOFFMANN T, PORTER M A, LAMBIOTTE R. Generalized master equations for non-Poisson dynamics on networks[J]. *Physical Review E*, 2012, 86(4): 046102.
- [11] SPEIDEL L, LAMBIOTTE R, AIHARA K, et al. Steady state and mean recurrence time for random walks on stochastic temporal networks[J]. *Physical Review E*, 2015, 91: 012806.
- [12] HOFFMANN T, PORTER M A, LAMBIOTTE R. Random walks on stochastic temporal networks[M]//HOLME P, SARMAKI J. *Temporal Networks*, Berlin: Springer, 2013: 295-313.
- [13] ROCHA L E, BLONDEL V D. Bursts of vertex activation and epidemics in evolving networks[J]. *PLoS Computational Biology*, 2013, 9(3): e1002974.
- [14] LAMBIOTTE R, TABOURIER L, DELVENNE J-C. Burstiness and spreading on temporal networks[J]. *European Physical Journal B*, 2013, 86(7): 1-4.

编辑 蒋 晓