

基于富人俱乐部特性的搜索免疫

纪鹏,葛洪伟

(江南大学信息工程学院,江苏无锡214122)

(jipeng_114@163.com)

摘要:为了使用更少的免疫数量且能更快地消灭病毒,已经提出了很多种免疫策略。熟人免疫策略是局域免疫中最有效的方法。基于BA无标度网络的富人俱乐部特性和广度搜索策略,提出了基于富人俱乐部特性的搜索免疫(RPBSI)。该免疫算法有两种策略,这两种策略对邻居节点的搜索方法不同,免疫的代价也有所差异。实验采用SIS病毒传播模型分别在BA无标度网络和科研合作网中验证:RPBSI算法与熟人免疫算法相比可以通过免疫较小的一部分节点达到消灭病毒的目的。

关键词:熟人免疫;BA无标度网络;病毒传播模型;SIS模型;富人俱乐部特性;免疫临界值

中图分类号: TP183 **文献标志码:** A

Rich-club phenomenon based search immunization

Ji Peng, Ge Hong-wei

(College of Information Technology, Jiangnan University, Wuxi Jiangsu 214122, China)

Abstract: In order to eliminate the disease with few immunized nodes and high speed, a great many immunization strategies have been proposed. Acquaintance immunization strategy is the most effective of local strategies. Based on the BA scale-free network's rich-club phenomenon and breadth first search, the paper proposed search immunization strategy based on the rich-club phenomenon which requires only local degree information. There are two types of immunization strategies, which depend on searching the neighbors' degree differently, and can also be used with different cost required. RPBSI algorithm can lead to the eradication of the epidemic by immunizing a smaller fraction of the nodes than the acquaintance immunization in BA scale-free network and scientific collaboration networks.

Key words: acquaintance immunization; BA scale-free network; virus propagation model; Susceptible Infected Susceptible (SIS) model; rich-club phenomenon; epidemic threshold

0 引言

当前网络动力学研究的热点是怎样用最少的免疫数量对人群网络、计算机网络等进行免疫,并且已经提出了很多种免疫策略,比如均匀免疫、目标免疫^[1]、熟人免疫^[2]等。免疫采用的网络模型有随机网络模型、小世界网络模型^[3]、BA无标度网络模型^[4]等。

均匀免疫是随机选取网络中的一部分节点进行的免疫,故亦称为随机免疫。该免疫对度数大的节点和度数小的节点平等对待,这在不均匀网络中是不适用的。针对BA无标度网络的不均匀性提出的目标免疫需要知道全局信息,这在复杂且不断发展变化的人类社会和因特网中都是很难做到的。基于局部信息的熟人免疫算法可以有效地减小免疫临界值,但是免疫临界值仍然很大。通过研究BA无标度网络的特性,本文提出了基于富人俱乐部特性的搜索免疫(Rich-club Phenomenon Based Search Immunization, RPBSI),该免疫算法与熟人免疫算法相比可以极大地降低免疫临界值。

BA无标度网络具有富人俱乐部特性,即度数大的节点之间比度数小的节点之间相互连接的比例大。在邻居节点度数已知的前提下,RPBSI算法根据富人俱乐部特性通过广度搜索策略搜索度数大的节点进行免疫,可以有效地降低网络的异质性,从而降低网络的免疫临界值。本文实验采用SIS病

毒传播模型分别用RPBSI算法与熟人免疫算法对BA无标度网络和科研合作网中进行免疫,并从免疫临界值、感染密度、搜索代价3方面进行对比。

1 SIS病毒传播模型

在SIS病毒传播模型^[5]中网络中的节点处于两种状态:易感节点(S)和感染节点(I)。感染节点为传染的源头,它以一定的概率 λ 通过节点之间的边进行感染。在传染过程中如果易感节点邻居中至少有一个是感染节点,则该易感节点以一定的概率 λ 转变成感染节点,同时感染节点以概率 μ 转变为易感节点。在传播时间足够长的情况下,恢复率的大小对最终感染密度无影响^[1],所以 μ 可以为1。SIS模型的感染机制可以用式(1)来表示:

$$\begin{cases} S_i + I_j \xrightarrow{\lambda} I_i + I_j \\ I_i \xrightarrow{\mu} S_i \end{cases} \quad (1)$$

其中: S_i 和 I_i 表示节点 i 的易感状态和感染状态, S_j 和 I_j 表示节点 j 的易感状态和感染状态。

易感人群和感染人群充分混合时,其动力学用式(2)表示:

$$\begin{cases} \frac{d(S_i)}{dt} = -\lambda I_i S_i + \mu I_i \\ \frac{d(I_i)}{dt} = \lambda I_i S_i - \mu I_i \end{cases} \quad (2)$$

收稿日期:2009-12-14;修回日期:2010-03-14。

作者简介:纪鹏(1984-),男,山东泰安人,硕士研究生,主要研究方向:人工智能、网络动力学;葛洪伟(1967-),男,江苏无锡人,副教授,博士,主要研究方向:人工智能、嵌入式系统。

其中: S_t 和 I_t 表示在病毒传播时刻 t 处于 S 和 I 状态的密度, 并且 $S_t + I_t = 1$ 。式(2) 存在一个阈值 λ_c , 当 λ 小于 λ_c 时, 病毒很快消亡 $I_t = 0$; 当 λ 大于 λ_c 时, 病毒可以长久地存在, 即 I_t 永远大于 0。

2 BA 无标度网络的生成算法

在现实世界中万维网、路由器层次网、电影合作网、科研合作网都精确或近似地体现增长和优先连接特性, 而 BA 无标度网络能充分反映实际网络中的这两个重要特性:

1) 增长特性。网络的规模不断扩大。

2) 优先连接特性。新的节点更倾向与具有较高连接度的“大”节点相连接, 这种现象也称为“富者更富”或“马太效应”。

BA 无标度网络模型的构造算法如下:

1) 初始化网络规模为 M_0 , 即 M_0 个顶点之间相互连接。

2) 添加一个度为 M 的节点 u 。新的节点 u 连接到现有节点 v 的概率正比于 $k_v / \sum_j k_j$, 其中节点 v 具有 k_v 条边。

3) 如果现有节点总数为 N , 则停止, 否则循环第 2) 步。

Barabási 和 Albert 的数值模拟^[4] 表明如果 N 足够大, 网络会达到稳定演化状态。平均路径长度 $L \approx \frac{\log(N)}{\log(\log(N))}$, 表明 BA 无标度网络具有小世界特性。节点度分布服从指数为 3 的幂率分布 $p(k) = \frac{2M(M+1)}{k(k+1)(k+2)} \propto k^{-3}$ 。平均连接度 $\langle k \rangle = 2M$ 。

3 基于富人俱乐部特性的搜索免疫

3.1 富人俱乐部特性

BA 无标度网络中度数大的节点称为“富节点”, 它们倾向于相互连接构成“富人俱乐部”^[7]。可以用富人俱乐部连通性 $Knn(k)$ ^[8] 来刻画这种现象, $Knn(k)$ 表示网络中度数大于 k 的节点之间, 实际存在的边数 L_k 与 r_k 个节点之间总的可能存在的边数 $r_k(r_k - 1)/2$ 的比值, 即:

$$Knn(k) = \frac{L_k}{r_k(r_k - 1)/2} = \frac{2L_k}{r_k(r_k - 1)}$$

其中 r_k 表示度数大于 k 的节点数目。

如果 $Knn(k) = 1$, 那么度数大于 k 的节点组成的富人俱乐部为一个完全连通子图。在本文第 4 章实验与分析中采用的 BA 无标度网络和科研合作网所展现的富人俱乐部连通性如图 1 所示。

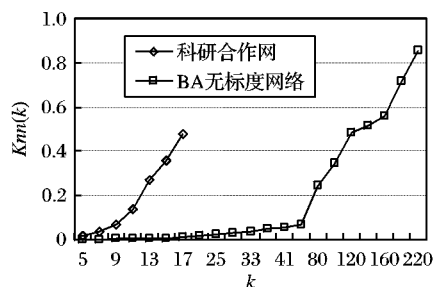


图1 网络的富人俱乐部特性

由图 1 所示节点之间相互连接的比例随着度数的增加而快速增加。在已知邻居节点度数的前提下, 可以由一个随机节点 v 出发, 搜索并免疫其邻居中度数大于节点 v 的节点或者邻居中度数最大的节点。

3.2 基于富人俱乐部特性的搜索免疫

针对 BA 无标度网络的富人俱乐部特性, 即度数大的节点之间倾向于相互连接, 本文提出了基于富人俱乐部特性的搜索免疫 (RPBSI)。

RPBSI 算法在选择邻居节点的过程中可以有两种策略: 策略 1, 选择度数大于中心节点的邻居节点进行免疫; 策略 2, 选择邻居节点度数最大并且度数大于中心节点的邻居节点进行免疫。策略 2 显然在免疫临界值和最终感染密度方面得到了优化, 但是其代价更高, 在实验中可以根据不同的代价要求选择不同的免疫策略。

设 μ 为免疫的节点数目, $adj(u)$ 为节点 u 的邻居节点集合, $degree(u)$ 为节点 u 的度数。

因为假设已知任意一个节点的邻居节点的度数, 所以 RPBSI 算法为局域免疫算法。

RPBSI 策略 1 免疫算法可以用伪代码描述如下:

```

Begin
    免疫的数目 = 0
    队列清空
    While( 免疫数目 <  $\mu$  )
        { 随机选取一个顶点  $v$ 
          选择  $v$  的邻居中度数大于  $v$  的节点放入队列
        Do
            { If ( 队列为空 ) Then
              跳出本循环
            Else
              If ( 免疫数目 =  $\mu$  ) Then
                  结束程序
              Else
                  { 出队列付给节点  $w$ 
                    If ( 节点  $w$  没有被免疫 ) Then
                        { 免疫  $w$ 
                           $v = w$ 
                        }
                    }
                  结束本循环
                End If
            }
        End If
    }
    End If
    选择  $v$  的邻居中度数大于  $v$  的节点放入队列
    While ( 队列不为空 )
        {
        }
    }
End

```

RPBSI 策略 2 免疫算法可以用伪代码描述如下:

```

Begin
    免疫的数目 = 0
    While( 免疫数目 <  $\mu$  )
        { 随机选取一个顶点  $v$ 
          While ( 1 )
              { 选取  $v$  的邻居中度数最大的节点付给  $w$ 
                If ( 节点  $w$  的度数小于节点  $v$  的度数或者节点  $w$  已经被免疫 ) Then
                    跳出本循环
                Else
                    { 免疫  $w$ 
                       $v = w$ 
                    }
                    If ( 免疫数目 =  $\mu$  ) Then
                        结束程序
                    End If
                }
            }
        }
    }
End

```

```

    End If
}
End

```

RPBSI 算法不必获取节点度的准确值,只需要获取邻居节点的度数的相比较值。在现实的人群网络中,节点的度数信息是很复杂并随着时间变化的,很难确定准确值,但是节点度之间是很容易比较出大小的。RPBSI 算法采用广度优先搜索策略,又称宽度优先搜索算法(Breadth First Search, BFS),是从源节点 S 出发,搜索距离节点 S 为 k 的所有节点,然后搜索距离为 $k+1$ 的所有节点,直至满足条件为止,其中 k 从 1 开始。在 RPBSI 算法中源节点 S 是随机选择的,搜索并免疫邻居节点中度数比中心节点大的节点,终止条件为免疫节点的个数达到给定的个数 μ 。

4 实验与分析

本文设定 BA 无标度网络总节点数 $N = 10^4$,初始网络规模 $M_0 = 8, M = 4$,生成 BA 无标度网络模型。采用 SIS 病毒传播模型,设定有一半的节点为感染节点。因为病毒传播的最终感染密度在一定范围内变动,所以感染密度定为一定范围内的平均值^[9]。实验设定传播时间 t 为 15000,选取传播时间从 10000 到 15000 的感染密度的平均值作为最终感染密度。

实验一 主要研究 RPBSI 算法和熟人免疫算法对应的免疫临界值。

BA 无标度网络的均匀免疫临界值^[10] g_c

$\frac{\langle k \rangle}{\langle k^2 \rangle}, \langle k^2 \rangle \rightarrow \infty$ 时免疫临界值为 1,需要免疫几乎所有的节点才能最终消灭病毒。熟人免疫算法的依据是度数小的节点倾向于连接到度数大的节点上,RPBSI 算法在本实验中采用了两种策略与熟人免疫算法进行对比。设定传染概率为 0.25,对比不同免疫下的免疫临界值,如图 2 所示。

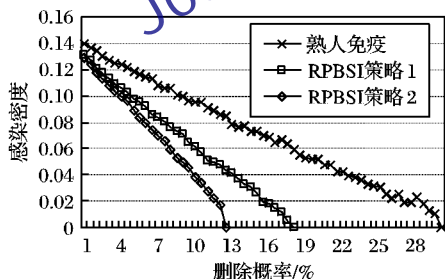


图2 不同免疫策略的临界值比较

熟人免疫算法对应的免疫临界值 $\lambda_c = 0.3$, RPBSI 策略1和策略2对应的免疫临界值 λ_c 分别为 0.18 和 0.125。随着删除概率的增加,各免疫算法的感染密度之间的差别呈线性增加。

实验二 主要研究在删除概率相同和传染率不同的条件下,RPBSI 算法和熟人免疫算法对应的感染密度。

设定删除概率为 0.08,传染率的变化为 0~0.5。研究感染密度的变化如图 3 所示。RPBSI 算法两种策略获得的感染密度都小于熟人免疫算法对应的感染密度。如果删除概率设定为其他值,也可以得到同样的结果。

实验二中当删除概率设定为 0.08 时,不同策略对应的重新随机选择搜索个体的平均次数:熟人免疫算法对应 917 次,

RPBSI 策略1对应 271 次,RPBSI 策略2对应 3229 次。由此可以看出 RPBSI 策略2的代价比 RPBSI 策略1的代价大很多。

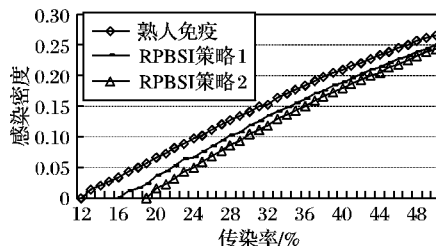


图3 不同传染率下的免疫策略比较(删除概率为 0.08)

实验三 主要研究 RPBSI 算法和熟人免疫在科研合作网中的免疫。

Newman 统计了 1589 位科学家之间的科研合作网^[11-12]。如果两位科学家共同发表过论文,则两位科学家之间存在边进行连接。如图 4 所示,采用 SIS 病毒传播模型,初始感染节点的密度为 0.5,对该实际网络进行免疫,设定传染率为 0.4,比较 RPBSI 算法和熟人免疫的临界值。熟人免疫的临界值为 0.075,RPBSI 算法的两种策略免疫临界值分别为 0.02 和 0.015。

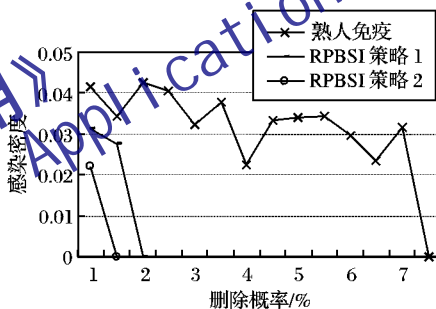


图4 不同免疫策略在科研合作网的临界值比较

设定删除概率为 0.01,研究最终感染率与传染率的关系,如图 5 所示。因为该网络的规模比较小,所以具有一定的不稳定性。

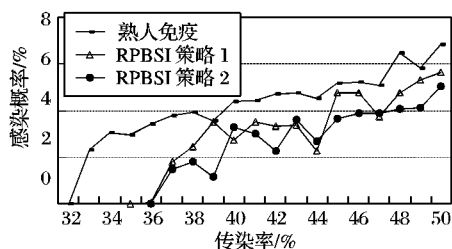


图5 不同免疫在科研合作网中感染概率的比较

熟人免疫和 RPBSI 策略1所对应的搜索平均次数都为 18, RPBSI 策略2对应的搜索平均次数为 31。BA 无标度网络和科研合作网具有小世界特性导致了策略2的平均次数远远多于另外两种策略。

由此实验二和实验三也可以看出 RPBSI 策略1和策略2在搜索代价上的巨大差异,并且当删除概率比较小时,感染密度差别也比较小,所以当免疫代价比较小的情况下选择 RPBSI 策略1比较好。通过实验可以验证 RPBSI 算法适用于具有富人俱乐部特性的所有不均匀网络,与网络的规模无关。

5 结语

基于富人俱乐部特性的搜索免疫是基于局部信息的并且
(下转第 1538 页)

$$\text{SimClassVector}(c_2) = ((u_1), (u_2, u_3, u_4, u_5, u_7, u_8, u_9), (u_2, u_3, u_4, u_5, u_7, u_8, u_9), (u_2, u_3, u_4, u_5, u_7, u_8, u_{11}), (u_2, u_3, u_4, u_5, u_7, u_8, u_9, u_{11}), (u_6, u_{10}, u_{12}), (u_2, u_3, u_4, u_5, u_7, u_8, u_{11}), (u_2, u_3, u_4, u_5, u_7, u_8, u_9, u_{11}), (u_2, u_3, u_5, u_8, u_9), (u_6, u_{10}, u_{12}), (u_4, u_5, u_7, u_8, u_{11}), (u_6, u_{10}, u_{12}))$$

所以 $\text{DCV}(c_2) = (1, 7, 7, 7, 8, 3, 7, 8, 5, 3, 5, 3)$

经计算 c_3 下的各个相似类则可得:

$$\text{SimClassVector}(c_3) = ((u_1, u_6), (u_2, u_5, u_7), (u_3, u_6, u_7, u_9, u_{10}, u_{12}), (u_4, u_5, u_8), (u_2, u_4, u_5), (u_1, u_3, u_6, u_9, u_{10}, u_{12}), (u_2, u_3, u_7, u_9, u_{10}, u_{12}), (u_4, u_8, u_{11}), (u_3, u_6, u_7, u_9, u_{10}, u_{12}), (u_3, u_6, u_7, u_9, u_{10}, u_{12}), (u_8, u_{11}), (u_3, u_6, u_7, u_9, u_{10}, u_{12}))$$

所以 $\text{DCV}(c_3) = (2, 3, 6, 3, 3, 6, 6, 3, 6, 6, 2, 6)$

由各个条件属性的数字特征向量,取 $\delta = 0.02$,根据式(7)可得条件属性间的相似矩阵:

$$[R] = \begin{bmatrix} 1 & 0.56 & 0.56 \\ 0.56 & 1 & 0.32 \\ 0.56 & 0.32 & 1 \end{bmatrix}$$

使用模糊矩阵闭包运算方法^[7,10]可以求得:

$$[t(R)] = \begin{bmatrix} 1 & 0.56 & 0.56 \\ 0.56 & 1 & 0.56 \\ 0.56 & 0.56 & 1 \end{bmatrix}$$

取 $\lambda = 0.8$ 可得:

$$[t(R)]_{\lambda} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

因此,在模糊等价矩阵的截集阈值 $\lambda = 0.8$ 的条件下,各连续条件属性是不相关的,因此表1的主观约简集为 $\{c_1, c_2, c_3\}$,这个结果与文献[7]所得的结果完全一致。

通过这个实例说明,利用本文算法不仅能够解决连续域决策表属性约简问题,而且还可以根据实际需要获得主观的属性约简集和一组模糊规则集,这说明本算法是可行的。

5 结语

针对粗糙集对于连续域属性决策表的处理能力差以及不容易获得模糊集之间关系等问题,提出一种基于连续型属性的 HCM 聚类约简算法。该算法首先引入三角隶属度函数将连续属性值转化为模糊值,并使用 HCM 聚类方法获得数据集之间关系,最后提出了一个新的属性约简算法。实例验证表明,采用该算法,用户可以根据实际决策需要和领域知识更改阈值,从而获得满意的属性结果。

参考文献:

- [1] PAWLAK Z. AI and intelligent industrial applications: the rough set perspective[J]. International Journal of Cybernetics and Systems, 2003, 31(4): 227-252.
- [2] PAWLAK Z, SKOWRON A. Rudiments of rough sets[J]. Information Sciences, 2007, 177(1): 3-27.
- [3] PAWLAK Z. Rough sets: theoretical aspects of reasoning about data [M]. Norwell: Kluwer Academic Publishers, 1992.
- [4] DUBOIS D, PRADE H. Rough fuzzy and fuzzy rough sets [J]. International Journal of General Systems, 1990, 17(2/3): 191-208.
- [5] 赵峰. 一般模糊矩阵传递闭包的计算、简化与应用[D]. 青岛: 中国海洋大学, 2003.
- [6] 何亚群, 胡寿松. 一种基于粗糙-模糊集集成模型的决策分析方法[J]. 控制与决策, 2004, 19(3): 315-318.
- [7] LIU WENJUN, XIAO QI-MEI. Fuzzy decision algorithm based on rough sets [J]. Fuzzy Systems and Mathematics, 2006, 26(2): 127-132.
- [8] CHANG T H. Using the fuzzy multi-criteria decision making approach for measuring the possibility of successful knowledge management[J]. Information Sciences, 2009, 179(4): 355-370.
- [9] 钱进, 叶飞跃, 孟祥萍, 等. 一种基于新的条件信息量的属性约简算法[J]. 系统工程与电子技术, 2007, 12(21): 54-57.
- [10] 胡军, 王国胤. 覆盖粗糙集的模糊度[J]. 重庆邮电大学学报, 2009, 4(21): 112-115.

(上接第1535页)

在现实中很容易实现。该策略可以应用到多种现实的无标度网络中去,比如互联网、人群网络等。利用平均场理论或者渗流^[13]方法可以证明 RPBSI 算法能有效地降低网络的异质性进而降低免疫临界值,是我们下一步的研究工作。

参考文献:

- [1] SATORRAS R P, VESPIGNANI A. Epidemic spreading in scale-free networks [J]. Physical Review Letters, 2001, 86(14): 3200-3203.
- [2] COHEN R, HAVLIN S, ben-AVRAHAM D. Efficient immunization strategies for computer networks and populations [J]. Physical Review Letters, 2003, 90(24): 277901.
- [3] WATTS D J, STROGATZ S H. Collective dynamics of 'small-world' networks [J]. Nature, 1998, 393(6684): 440-442.
- [4] BARABÁSI A L, ALBERT R. Emergence of scaling in random networks [J]. Science, 1999, 286(5439): 509-512.
- [5] EGUILUZ V M, KLEMM K. Epidemic threshold in structured scale-free networks [J]. Physical Review Letters, 2002, 89(10): 108701.
- [6] COHEN R, HAVLIN S. Scale free networks are ultrasmall [J]. Physical Review Letters, 2003, 90(5): 058701.
- [7] ZHOU S, MONDRAGON R J. The rich-club phenomenon in the Internet topology [J]. IEEE Communication Letters, 2004, 8(3): 180-182.
- [8] ZHOU S, MONDRAGON R J. Structural constraints in complex networks [J]. New Journal of Physics, 2007, 9(6): 173-184.
- [9] SATORRAS R P, VESPIGNANI A. Epidemics and immunization in scale free networks [M]// BORNHOLDT S, SCHUSTER H G. Handbook of Graphs and Networks: From the Genome to the Internet. Berlin: Wiley-VCH, 2002.
- [10] ANDERSON R M, MAY R M. Infectious diseases in humans [M]. Oxford: Oxford University Press, 1992.
- [11] NEWMAN M E J. The structure of scientific collaboration networks [J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98(2): 404-409.
- [12] NEWMAN M E J. Finding community structure in networks using the eigenvectors of matrices [J]. Physical Review E, 2006, 74(3): 036104.
- [13] GRIMMETT G. Percolation [M]. 2nd ed. Germany: Springer, 1999.