

基于熟人免疫的复杂网络免疫策略

葛新^{1,2} 赵海¹ 张君¹

(东北大学信息科学与工程学院 沈阳 110819)¹ (大连海事大学信息科学技术学院 大连 116026)²

摘要 针对目标免疫和熟人免疫的不足,提出一种综合的免疫算法,即随机选择一些节点,根据节点的不同特征采取合适的免疫措施。本算法保留了熟人免疫的优点,即完全基于本地信息、不需要获知网络的全局结构和识别高连接度节点,在需要免疫的节点数量与目标免疫相同的情况下,具有理想的免疫效果。通过对几种算法在无尺度网络、随机网络以及真实世界网络上进行仿真比较,验证了本算法具有更好的免疫效果和更广泛的应用条件。

关键词 复杂网络,免疫策略,目标免疫,熟人免疫,综合免疫

中图分类号 TP391.9 文献标识码 A

Complex Networks Immune Strategy Based on Acquaintance Immunization

GE Xin^{1,2} ZHAO Hai¹ ZHANG Jun¹

(Information Science and Engineering College, Northeastern University, Shenyang 110819, China)¹

(School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China)²

Abstract We proposed one synthesis immunization algorithm aiming at the drawbacks of acquaintance and target immunization and improve them separately. This immunization strategy randomly chooses nodes in network and adopts different action according to different characteristics. This strategy remains the advantage of acquaintance that is based on local information without global structure and high degree nodes. It is more effective in the situation that the number of node need to be immunized is the same as target immunization. The better effectivity and wider situation of application were verified throughout simulation on scale free, random and some real-world networks.

Keywords Complex networks, Immunization strategy, Target immunization, Acquaintance immunization, Synthesis immunization

随着复杂网络理论研究的兴起,复杂网络中的病毒传播已成为当前复杂网络研究的一个热点问题^[1,2]。现实复杂网络中,往往开始只有很微少的病毒感染源,但是如果不加控制,则可能爆发大规模的病毒流行。在以往的研究中,人们针对免疫策略作了较多的研究,常常根据节点在网络中的地位(即以节点的度来衡量)来关注节点的选择。最典型的免疫控制策略包括随机免疫^[3](完全随机地选取网络中的部分节点进行免疫)、目标免疫^[4](选取少量度最大的节点免疫)、熟人免疫^[5](对随机选出的节点的邻居进行免疫)。

本文在熟人免疫策略和目标免疫策略的基础上,提出一种对两种免疫策略分别改进后联合应用的免疫策略。通过仿真比较和理论推导,验证了本文算法具有更好的免疫效果和更广泛的适应性。

1 目标免疫与熟人免疫

采取随机免疫,需要对网络中几乎所有节点都实施免疫,才能保证最终消灭病毒传染。目标免疫虽然比较有效,但这种方法需要了解网络的整体结构信息,至少需要比较清楚地认识网络中各个节点的度,才能找出度大的关键节点进行免

疫。在许多情况下,例如对于庞大复杂且不断发展变化的人类社会和 Internet 来说,这是难以做到的。因此, Cohen 等人提出一种称为熟人免疫的策略。

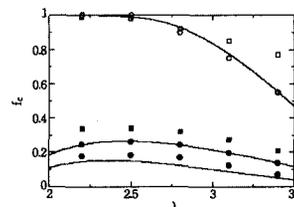


图 1 熟人免疫效果

熟人免疫算法的基本思想是:从 N 个节点中随机选取比例为 p 的节点,再从每一个被选出的节点中随机选择一个邻居节点进行免疫。假设被免疫节点占节点总数的比例为 f ,这种策略只需要知道被随机选择出来的节点以及它们直接相连的邻居节点,从而巧妙地回避了目标免疫中需要知道全局信息(每个节点的度)的问题。由于在无尺度网络中,度大的节点意味着有许多节点与之相连,若随机选取一点,再选择其邻居节点,则度大的节点比度小的节点被选中的概率要大得多。因此,如图 1 所示,熟人免疫策略比随机免疫策略的效

果好得多,其中中线为熟人免疫。

2 综合免疫策略

熟人免疫是一种在免疫开销和免疫效果之间取得平衡的一种策略,其应用条件是全局信息不易获知或开销过大。实际上,在很多情况下,和整个网络的全局信息相比较,局部信息更容易得到或者通过较小代价就可以计算得到。在这种情况下,熟人免疫策略就有了较大的改进空间。

对于一个有 N 个节点的网络,按比例 p 随机选取节点 v_i ,根据不同条件采取不同措施进行免疫。

条件(1):已知节点 V_i ,连接度 K_i ,当 K_i 达到某一阈值 K_{cut} 时,对 V_i 进行免疫,并随机选择一个相邻节点进行免疫。

条件(2):若节点 V_i 的度信息未知或已知其连接度 $k_i \leq k_{cut}$,则对其全部相邻节点中的 n 个节点进行随机免疫,并从这 n 个节点,选择一个节点对其相邻节点进行免疫。 n 的具体确定方法及其值将在后面的分析和仿真中给出。

实际上,当所有节点均满足条件(1)时,是一种根据阈值而定的不完全的目标免疫和熟人免疫策略;对于条件(2),存在这样一种情形:节点度未知,但其值已达到了设定的阈值。由于我们对该节点邻居的邻居进行免疫,类似于熟人免疫,存在概率正比于该节点度的事件,因此该节点会得到免疫处理。我们认为在未知节点连接度信息时进行多重强化的熟人免疫会更有效。综合上述分析,本文分别改进了目标免疫和熟人免疫,提出了一种混合免疫策略。

3 仿真模型及参数

由于无尺度网络的传播率 r 连续而平滑地过渡到零,因此无尺度网络存在正临界点。这表明在 BA 无尺度网络中,只要传播率大于零,病毒都能传播并最终维持在一个平衡态^[6]。由于 Internet 具有持续增长的无尺度网络特征,计算机病毒可以在 Internet 上通过感染很小一部分节点而传播开来。

在现实世界中,许多网络都具有无尺度特征,而且文献[7]中已经证明熟人免疫策略对无尺度网络具有更好的免疫效果。因此,选择 $p(k) \sim k^{-\gamma}$ ($2 < \gamma \leq 3.5$) 的 BA 无尺度网络模型对本文的算法进行仿真和分析。为了不失一般性,考虑一类具有 N 个节点的可变幂率指数 γ 的无尺度网络,由文献[8]可知其度分布函数用如下标准函数来表征:

$$p(k) = (1 + \mu) j^{1+\mu} k^{-2-\mu} = (\gamma - 1) j^{\gamma-1} k^{-\gamma} \quad (1)$$

式中, j 为网络中节点的最小的度,在我们的模型中, j 为 1。对于具有 N 个节点的无尺度网络, $\phi(k)$ 为网络度分布,设度为 k 的节点具有可感染性事件为 s_k 。建立具有 $N=10^6$ 个节点的 BA 无尺度网络,幂率参数 $\gamma=2.5 \sim 3.5$ 。

为了衡量免疫效果,需要计算使得传染停止的关键变量 p_c 和 f_c 。随机选择免疫节点,其比率大于 p_c 时,传染不会扩散。整体感染比率小于 f_c 时,传染也不会扩散。实际上,在后面的解析中可以发现, f_c 最终由 p_c 推导得出,故在本文中,主要对 f_c 进行推导及其仿真结果分析。

4 临界值

因为本文算法主要根据已知条件的不同采取不同的免疫

措施,而条件的不同可以看作是符合某一种概率分布的随机事件,所以不妨先单独分析两种免疫措施的临界值。首先,对于条件(1),一个节点被随机选择的概率为 $1/N$,设该节点连接度为 k ,选中某一个邻居节点的概率为 $1/k$ 。由概率公式有:

$$p(k > k_{cut}) = 1 - p(k < k_{cut}) \approx 1 - \frac{k_{cut}}{\langle k \rangle} \quad (2)$$

式中, $\langle k \rangle$ 为网络中节点的平均度,且有:

$$\langle k \rangle = \sum_k k P(k) \quad (3)$$

则:

$$p(k > k_{cut}) = 1 - \frac{k_{cut}}{\sum_k k P(k)} \quad (4)$$

由文献[6]可知:

$$p(s_k | k) = (e^{-p/k})^k \quad (5)$$

同样地,

$$p(s_{k_{cut}} | k_{cut}) = (e^{-p/k})^{k_{cut}} \quad (6)$$

$$p(s_{k'} | k') = (e^{-p/k'})^{k'} \quad (7)$$

一个熟人节点未被选中的概率为 $1 - \frac{1}{Nk}$,因为条件(1)只选择一个熟人节点进行免疫处理,经过 Np 次免疫后:

$$v_p(k) \equiv \left(1 - \frac{1}{Nk}\right)^{Np} \approx e^{-p/k} \quad (8)$$

又由文献[5]可知, $p(k)$ 满足:

$$\sum_k \frac{p(k) k(k-1)}{\langle k \rangle} v_{p_c}^k e^{-2p_c/k} = 1 \quad (9)$$

根据被免疫节点的分布可以得到未免疫节点解析表达式:

$$f_c^1 = 1 - \sum_k P(k > k_{cut}) p(s_k | k > k_{cut}) v_p^k, \quad (10)$$

式中, $p(s_k | k > k_{cut}) = \frac{p(s_k | k) / p(s_{k_{cut}} | k_{cut})}{p(k > k_{cut})}$ 。

由式(5)、式(6)、式(8)代入式(10)可得:

$$\begin{aligned} f_c^1 &= 1 - \sum_k P(k > k_{cut}) p(s_k | k > k_{cut}) v_p^k \\ &= 1 - \sum_k P(k > k_{cut}) \frac{p(s_k | k) / p(s_{k_{cut}} | k_{cut})}{p(k > k_{cut})} (e^{-p/k})^k \\ &= 1 - \sum_k p(s_k | k) / p(s_{k_{cut}} | k_{cut}) (e^{-p/k})^k \end{aligned}$$

化简后得:

$$f_c^1 = 1 - \sum_k (e^{-p/k})^{2k - k_{cut}} \quad (11)$$

对于满足条件(2)的情况,与条件(1)的主要不同在于邻居节点的选择不再局限于一个,而且将对邻居节点的邻居进行免疫。实际上,若 $n=1$,则退化为传统的熟人免疫策略;若 $n>1$,则为一种强化了熟人免疫策略。

类似于上述分析,一个节点被随机选择的概率仍为 $1/N$,设该节点连接度为 k ,其相邻节点的度为 k_1', k_2', \dots, k_k' ,并且有:

$$\phi(k') \equiv p(k' | k) = k' P(k') / \langle k \rangle \quad (12)$$

实际上条件(2)中又分为两种情况,被选择节点的度小于阈值或未知,因为度小于阈值的节点不再进行免疫,又由式(12)可以看到 $\phi(k')$ 独立于 k ,所以节点度已知或未知不影响 k' 的分布。

若对直接选择的节点不再进行免疫,而是选择 k 个相邻节点中的 n 个进行免疫,则一个邻居节点被选中的概率为

n/k ,有:

$$P(s_k | k') = \frac{n}{k} \quad (13)$$

在满足条件(2)时,式(7)变为 $v_p(k) \equiv \left(1 - \frac{n}{Nk}\right)^{Np}$ 。但是当 N 足够大时,其极限不变,仍为式(7)右端值 $(e^{-p/k})^{k'}$ 。

同样地,根据被免疫节点的分布可以得到未免疫节点解析表达式。同时,考虑到被选择节点邻居节点的邻居节点也进行免疫,则初始选择节点被免疫的概率为 n/k' ,那么:

$$p(k' | k) = \frac{\phi(k') e^{-p/k'}}{e^{-p/k}} = \frac{k' P(k') e^{-p/k'}}{\langle k \rangle e^{-p/k}}$$

且 $p(k') = (\gamma - 1) k'^{-\gamma}$ 。

又由式(8)、(12)、(13),代入后化简,得:

$$f_c^2 = 1 - \sum_k n^2 (\gamma - 1) k^{-\gamma} e^{-p/k} \quad (14)$$

综合分析两种情况:一个度为 k 的节点被选中且符合第一种免疫条件的概率为 q :

$$f_c^2 = 1 - \sum_k P(k' | k) P(s_k | k') v_p \frac{n}{k}$$

$$q = \frac{k P(k)}{N \langle k \rangle} P(K > k_{cut})$$

$$q = \frac{k P(k) - N k_{cut}}{N \langle k \rangle}$$

由以上分析和推导,得出本文免疫策略的临界值:

$$f_c = q f_c^1 + (1 - q) f_c^2 \quad (15)$$

5 免疫效果分析

因为随机免疫已经被文献[9]证明是一种开销过大、基本不具备实际应用价值的免疫策略,故在仿真比较中,只对目标免疫、熟人免疫算法以及本文的算法进行仿真比较。

首先对每种条件的免疫进行单独比较分析。因为条件(1)中的免疫策略是一种有选择的目标免疫和熟人免疫相结合的策略,为了更好地进行横向比较,在目标免疫时,节点度达到阈值,即进行免疫处理。

在对几种算法进行横向比较分析之前,首先需要确定本文算法中的一些关键参数对免疫效果有什么样的影响。影响免疫效果的参数主要有:

(1) 阈值 k_{cut} 。当能够获得节点度信息, k_{cut} 的值决定了应用哪种免疫算法。显然, k_{cut} 越大,免疫越有目标性,免疫也将越有效。为了排除其他参数的干扰,不妨将 γ 取为 2.5, 随机选择出来的节点全部满足条件(1)。图 2 显示了 k_{cut} 分别为 1~10 时对应的 f_c , 并且给出了随机选择节点比例不同的 4 种情况。可以看到, k_{cut} 在一定范围内增加,免疫也将逐渐更有效。选择节点的比例也对免疫效果有较大影响。当随机选择的节点为总数 10%, k_{cut} 达到 4 以后,再提高阈值,对免疫效果基本没有影响。也就是说,随着选择节点比例的提高,取得“最好”免疫效果的阈值也随之增加。但 k_{cut} 达到 7 以后,其值的增加将无益于改善免疫效果。需要说明的是, $p=0.7$ 已经是一个比较高的选择比例,进一步提高随机选择节点的数量对免疫已经没有太大意义,因为节点比例的增加对免疫效果的改善并非线性的,而且过多地选择节点将导致免疫开销过大。综合以上分析,在我们的算法中, k_{cut} 确定为 6 或 7 是一个比较合理的值。在后面的仿真中,也将其确定为此值。

(2) 需要免疫的邻居节点数量 n 。在节点信息未知的情况下, k_{cut} 也就无法确定。这时候通过对其相邻节点免疫,实际上是一种强化的熟人免疫策略。 n 为进行免疫的相邻节点的数目,当相邻的点的数目小于 n 时,相当于对所有相邻节点进行免疫处理了。图 3 显示了 n 对 f_c 的影响。首先,和条件(1)的免疫效果比较起来,条件(2)的免疫效果不如前者,尤其是 n 取值比较小时。而且 n 越小,对免疫效果的影响越大。同样地, n 存在一个有效上限,即 n 为 4。当 n 达到 4 后,随机选择节点的比例对免疫效果影响较小。也就是说,选择较少数量的节点也能够取得较好的免疫效果,这是此方法更优的地方。

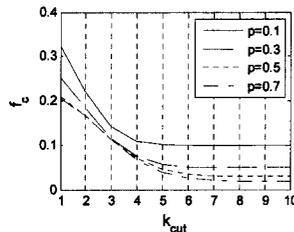


图 2 k_{cut} 对 f_c 的影响

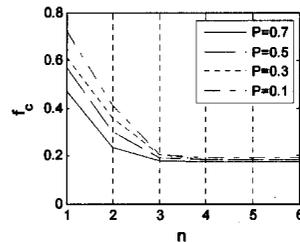


图 3 n 对 f_c 的影响

(3) 随机选择节点中度已知并且大于阈值的比例 q 。现实应用中,随机选择的节点全部满足条件(1)或条件(2)的概率是极低的。从前面对 k_{cut} 和 n 的分析可以看出,在节点度信息已知并且全部满足阈值时,免疫效果是最好的。那么从整体考虑的话,是不是满足条件(2)的节点越多,免疫效果越好呢?事实上并非如此。在这里,随机选择 50% 的节点进行免疫处理,如图 4 所示, q 为满足条件(1)的节点百分比。很明显, $q=1$ 比 $q=0$ 具有更小的 f_c 值,即免疫效果更好。但中间的变化过程不是线性的,也不是单调递减的。随着 q 的增加, f_c 达到一个峰值,之后才单调减小。也就是说,当两种免疫措施都应用时,免疫效果出现了小的反弹。

我们分析,产生这种现象的主要原因可能是当条件(1)、(2)两种免疫措施混合时,有一部分节点虽然度信息未知,但其度值实际上已经满足了 k_{cut} ,可以应用第一种免疫措施,但失去了应用更好的免疫措施的机会。

另外,虽然 q 在大于 0.5 时,免疫效果更好,但其稳定性较差。对前后两部分的区间计算了标准差,发现后半部分的标准差更大一些。

在分析了主要参数之后,对本文免疫策略和其他两种策略进行综合比较。在这里, $k_{cut}=7, n=4, p=0.3, q=0.5$ 。因为无尺度网络的幂率参数 λ 决定了其主要结构和性能特征,选择 $\lambda=2\sim 3.5$ 作为网络结构参数。图 5 为综合比较的仿真结果。可以看到,本文算法比熟人免疫算法具有更好的免疫效果,已经比较接近目标免疫算法。

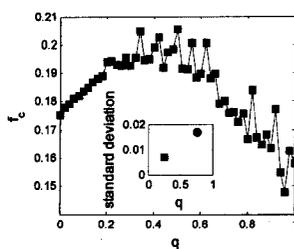


图 4 q 对免疫效果的影响

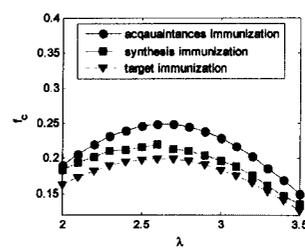


图 5 3 种免疫算法的比较

除临界值以外,稳态感染率和病毒传播速度也是衡量免疫效果的重要指标。为此,我们又在模型网络和真实网络中进行了仿真,比较3种免疫策略。从图6(b)中可以看出,综合免疫效果已经很接近目标免疫,在社会关系网络 Dolphin 上有效地降低了传播速度和稳态感染率。而在 ER 模型和 Internet AS 级拓扑上免疫效果稍差,我们猜测这和网络本身的拓扑结构有关。

最后,对解析解与仿真值进行了比较。因为 k_{cut} , n 等参数是为了观察对仿真效果的影响,故在这里不再给出解析解与仿真值的比较,只给出最后的综合免疫时免疫效果的比较。结果如图7所示,图中为多次仿真的平均统计结果。可以发现,仿真结果能够较好地符合理论推导值。

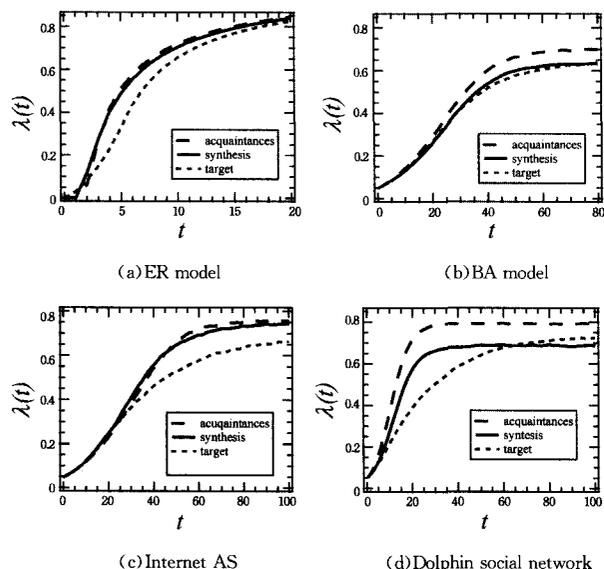


图6 不同网络上的免疫效果
采用 SIR 模型,感染率 $\nu=0.03$,恢复率 $\delta=0.01$ 。其中,(a) ER 随机网络模型, $N=1000, L=4000$; (b) BA 无尺度网络模型, $N=1000, L=4000$; (c) Internet AS 级拓扑¹⁾, $N=21034, L=42392$; (d) Dolphin, 社会关系网络²⁾, $N=62, L=159$ 。综合免疫参数 $k_{cut}=7, n=4, p=0.3, q=0.5$ 。仿真结果为 50 次的平均值。

图6 不同网络上的免疫效果

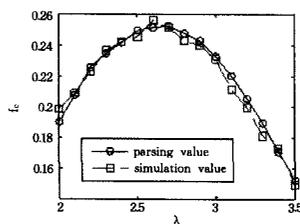


图7 解析值与仿真值比较

与熟人免疫和目标免疫相似, λ 同样对本文算法有影响,这是由无尺度的网络特性决定的。另外, λ 较大时,能使本文

算法的免疫效果更加接近目标免疫算法。至于出现该现象的具体原因,有待进一步分析。

结束语 本文通过对熟人免疫和目标免疫两种策略的分析,找出其局限性和各自的应用条件。在此基础上,提出一种在两种策略中进行平衡的混合免疫策略。本策略主要基于网络局部信息做出决定,在无法得知本地信息的情况下,给出相应的免疫方法,且被选择的节点不需要知道相邻节点信息。为了和其他免疫策略比较,选择具有代表性的 BA 无尺度网络、ER 随机网络以及两个真实网络进行仿真,通过稳态感染率以及传播时间的比较,验证了本免疫策略比熟人免疫效果更好,而且具有更好的适应性,其应用条件进一步符合实际网络要求。

参考文献

- [1] Pastor-satorras R, Vespignani A. Immunization of complex networks[J]. Physical Review E(Statistical Physics, Plasmas, Fluids, and Related Interdisciplinary Topics), 2002, 65 (32A): 36104-1
- [2] Yu L, Xue H, Gao X, et al. Epidemic spread model based on cellular automata [J]. Computer Engineering and Applications, 2007, 43(2): 196-237
- [3] Cohen R, Havlin S, Ben-avraham D. Efficient immunization strategies for computer networks and populations[J]. Physical Review Letters, 2003, 91(24): 247901-247901
- [4] Gomez-Gardenes J, Echenique P, Moreno Y. Immunization of real complex communication networks[J]. European Physical Journal B, 2006, 49(2): 259-264
- [5] Gallos L K, Liljeros F, Argyrakis P, et al. Improving immunization strategies[J]. Physical Review E(Statistical, Nonlinear, and Soft Matter Physics), 2007, 75(4): 45104-1
- [6] De L R. Models of complex networks and how diseases spread on them[C]//Perpignan. France: Springer, 2006
- [7] Zio E, Rocco C M, Salazar D E, et al. Complex networks vulnerability: a multiple-objective optimization approach[C]//Orlando. FL, USA: IEEE, 2007
- [8] Li Y, Liu Y, Shan X M, et al. Dynamic properties of epidemic spreading on finite size complex networks[J]. Chinese Physics, 2005, 14(11): 2153-2157
- [9] Xu D, Li X, Wang X. On virus spreading in local-world complex networks and its immunization control[J]. Control and Decision, 2006, 21(7): 817-820

1) http://www.caida.org/data/request_user_info_forms/ipv4_allpref24-aslinks.xml

2) <http://www-personal.umich.edu/~mejn/netdata/>