

复杂网络病毒传播的局域控制研究^{*}

许 丹 李 翔[†] 汪小帆

(上海交通大学自动化系, 上海 200240)

(2006 年 6 月 27 日收到, 2006 年 7 月 17 日收到修改稿)

从复杂网络的节点路径长度范围的角度来研究病毒传播的局域控制, 分析了在不同拓扑结构的复杂网络中进行局域控制的有效性. 研究表明, 局域控制对 WS 小世界网络、BA 无标度网络和 ER 随机网络三类复杂网络均有效, 但只有 WS 小世界网络存在零感染的控制范围最优值 $d=3$; 对于长程连边的分布存在距离偏好的 Kleinberg 小世界网络, 随着依赖度的增大, 病毒传播率临界值增加, 同时局域范围控制的效果得到加强.

关键词: 复杂网络, 病毒传播, 局域控制, 路径长度

PACC: 0565, 0570J

1. 引 言

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^[1-9]. 现实复杂网络中, 往往开始只有很微少的病毒感染源, 但是如果加以控制, 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)而关注节点的选择. 最典型的免疫控制策略包括随机免疫(完全随机地选取网络中的部分节点进行免疫), 目标免疫(选取少量度最大的节点免疫^[10]), 熟人免疫(对随机选出的节点的邻居进行免疫^[11]).

上述方法选取节点时, 都是以节点的度为衡量标准. 如果换一个角度, 不考虑节点的度, 仅从感染源出发, 在感染源周围的一定距离范围内进行局域控制, 这样的控制范围如何选择? 这里的控制意义, 包含着预防、免疫、杀毒、隔离等各种实际应对的措施. 当一个实际网络利用图论来抽象刻画时, 我们考虑用节点间的路径长度来表征控制的距离范围. 对于特定拓扑结构的复杂网络中的病毒传播, 是否存在控制范围的最小值? 或者是否能在一定的小范围内进行控制, 就能很好地控制大面积的感染? 我们发现, 在 BA 无标度网络^[12]、WS 小世界网络^[13]和 ER

随机图网络^[14]中, 为了抑制病毒的大范围传播, 对距离感染源 d 步范围内的节点进行控制, 均存在很小的路径长度范围 d , 但只有 WS 小世界网络在 $d=3$ 时可以通过少量的控制节点达到零感染的最优效果. 而对于根据个体间的距离生成的 Kleinberg 小世界网络^[15, 16], 我们发现, 长程连接在距离相对近的节点间较集中时, 这种网络拓扑在一定程度上抑制了病毒的传播, 局域控制的效果更加明显.

2. 局域范围控制

随机免疫策略往往带有一定的盲目性, 特别对于无标度网络, 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. 目标免疫法^[10]能够克服这一缺陷而提高免疫效果, 但它常常需要知道网络的全局信息, 对于庞杂的并且不断发展变化的网络来说, 并非易事. 文献 [11] 中提到的熟人免疫属于一种局域控制方法, 这种策略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, 其目的在于找出度大的节点进行免疫. 而本文的局域范围控制旨在对网络全局信息未知的情况下, 利用被感染节点周围的局部信息, 控制被感染节点周围的部分节点, 从而达到控制病毒在整个网络大范围传播的目的. 它不需要考虑节点的度, 仅考虑感染源周围一定路径长度内的所有节点, 因此只需要知道部

^{*} 国家自然科学基金(批准号: 70431002, 60504019, 90412004, 60225013), 上海市青年科技启明星计划(批准号: 05QMX1436)资助的课题.

[†] 联系作者, E-mail: xli@sjtu.edu.cn

分节点之间的连接关系.

考虑无权无向网络,每条边代表距离为 1,若节点 i 至少通过 d 条边到达节点 j ,则称从 i 至 j 的距离(路径长度)为 d .从被感染节点出发,用步数 d 表征网络被控制的范围.进行局域控制后,在距离感染源一定范围内的个体都受到良好的保护,这体现被感染节点 d 步路径范围内所有节点被感染的概率大大降低.本节研究的网络规模为 $N = 2000$.我们采用 SIS 模型,网络中的节点分为易染(susceptible)状态和感染(Infected)状态,设定易染节点被感染的概率 $\nu = 0.5$,不失一般性,假设感染节点恢复(治愈)成为易染节点的概率 $\delta = 1$,因此有效传播率 $\lambda = \nu / \delta^{[17]}$.在控制范围内,易染节点的感染概率 $q = 0.01$,表示在这个控制范围内的节点被很好地控制和保护.

2.1. 随机网络、小世界网络、无标度网络比较

ER 随机网络^[12]、WS 小世界网络^[13]和 BA 无标度网络^[14]是复杂网络中三类最典型的例子,它们分别从不同的方面体现了现实复杂网络的拓扑特征,我们选取这三种网络模型来进行局域范围控制的研究.为了具有可比性,三种网络取相同的网络规模和平均度.

在这几种网络中,对被感染节点周围几步之内的节点进行控制,才能够达到很好的控制效果?究竟局部范围控制能否遏制病毒的大范围传播呢?图 1 表示控制范围 d 从 0 (没有任何控制措施)变化到 4,三种拓扑结构的网络中最终被感染节点数的变化.由图可知,三种网络中, $d = 3$ 时,病毒传播得到了很好的控制; $d \geq 3$ 时,网络中的病毒感染被彻底消灭.

但这里需要进一步考虑一个代价问题.我们的目的是想通过局部的控制达到控制整个网络的大范围传播,这里的局域控制一方面体现在距离步数不能太长,另一方面体现在所需控制的个体数不能太多,否则仍然是高成本的策略.只有小范围少数节点的控制,才是真正意义上的优化控制.因此,图 2 描绘了控制范围变化时(d 取不同的值时),被控制的节点数占网络总节点数的比例是如何变化的.随机网络和 BA 网络中, d 超过 2 时,被控制的节点数迅速上升, $d = 3$ 时虽然出现了零感染,但此时需要被控制的节点数分别占整个网络的 90% 和 70% 以上,这样的代价就失去了最初设想的意义.只有 WS 小

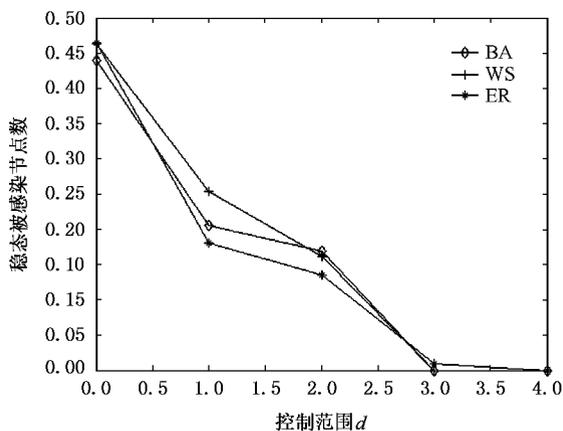


图 1 三种网络模型控制范围比较

世界网络中这一比例只维持在 10% 左右,因此,只有在 WS 小世界网络中才存在 $d = 3$ 这样的最优范围,即控制范围的下界.

对于许多实际的无标度网络或随机网络,虽然达到最优的零感染控制需要付出很大的代价,但注意到控制范围 $d = 1$ 或 2 时,最终被感染数将会有大幅下降,所以通过采取这种 $d = 1$ 或 2 的局域控制来抑制病毒传播,将病毒控制在小范围内流行,达到控制代价和效果之间的一种次优平衡.

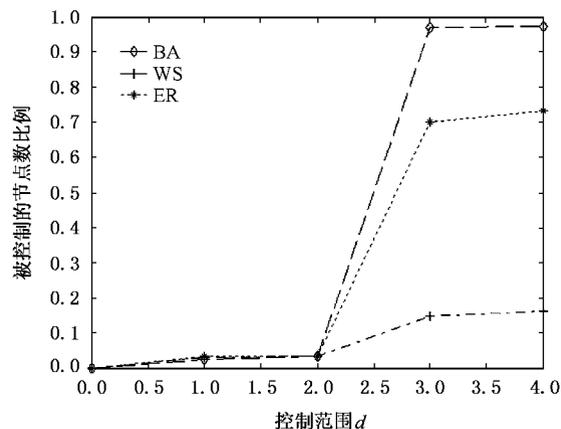


图 2 三种网络被控制的节点数比例

复杂网络研究中一个重要的概念是网络的聚类系数.假设网络中的一个节点 i 有 k_i 条边将它和其他节点相连,这 k_i 个节点就称为节点 i 的邻居.这 k_i 个节点之间实际存在的边数 E_i 和总的可能的边数 $k_i(k_i - 1)/2$ 之比定义为节点 i 的聚类系数 C_i ,整个网络的聚类系数 C 为所有节点 i 的聚类系数 C_i 的平均值.我们计算比较了实验中不同拓扑结构的

网络的聚类系数,见表 1.可见,随机网络和 BA 无标度网络的聚类系数明显比 WS 小世界网络小很多.因此,WS 小世界网络有别于另两种网络模型而存在范围控制最优值,是由聚类系数的差别引起的.聚类系数衡量网络的紧密程度,具有类似于社会关系网络中“物以类聚,人以群分”的特性.因为网络中的节点的集结性,控制感染节点 3 步以内的节点,这些节点彼此是“朋友”的概率很大,所以 3 步内找到的节点的邻居或邻居的邻居很大程度上是同一批节点,因而总的被控制节点数的比例相对较小.因此,对于病毒的传播,对局部集聚的节点进行控制显然要比对分散稀疏的节点进行控制更为容易,代价也更低.

表 1 不同网络的聚类系数比较

网络	WS 小世界网络	BA 网络	随机网络
聚类系数	0.5830	0.0206	0.0030

2.2. 传播率结合局域范围控制

由于只有 WS 小世界网络中才存在范围控制的最优值,我们这里只讨论 WS 小世界网络中的情况. Pastor-Satorras 和 Vespignani 等人的研究表明^[17],均匀网络中,当病毒有效传播率超过一个正的临界值时,大规模传播才有可能. WS 小世界网络也属于一种均质网络,因此我们考察局域范围控制对传播率临界值的影响,以及两者共同影响的结果.

图 3 表示不同的传播率和局域控制范围的共同作用对病毒传播的影响,图中为 50 次平均的结果.当 $d=0$ 时,表示不进行任何控制, $\lambda < 0.5$ 时,仍会出现最终感染数为 0 的状态,这是因为实际传播率小于传播率临界值造成的;当 d 不断增加,控制开始起作用,使得传播率临界值增大,并且最终被感染节点数也有显著下降(见 $d=1, 2$),约为不加任何控制作用时的一半,可见对被感染节点周围很小的局域范围进行控制,能够有效地抑制病毒的传播;当 $d=3$ 时,局域范围控制起了主导作用,无论传播率 λ 如何变化,即使达到最大值 $\lambda=1$,最终都能抑制病毒的扩散,使得病毒无法大范围传播.

3. Kleinberg 小世界网络上的病毒传播及局域控制

本节由 Kleinberg 模型出发,在一类距离偏好的加边概率生成的小世界复杂网络中,探索网络中长

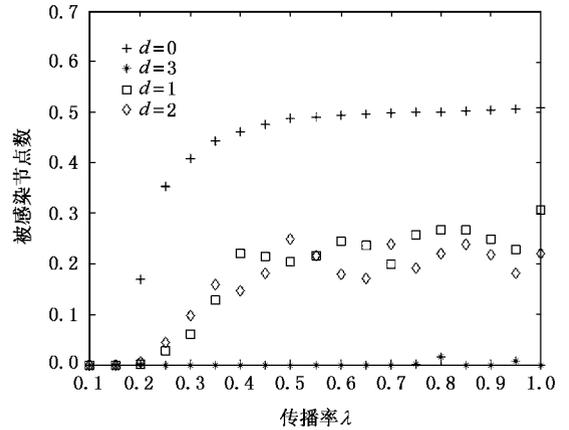


图 3 传播率变化结合范围控制的效果图示

程连接的分布对病毒传播及局域范围控制的影响.

3.1. 一维 Kleinberg 小世界网络模型构造

Kleinberg 小世界网络模型源于社会网络搜索问题.该模型中的 N 个节点分布在一个二维网格上^[15,16].网络中两个节点 u 和 v 之间的网格距离 $d(u, v)$ 定义为两节点之间的网格步数.在 Kleinberg 模型中,节点 u 和 v 之间有长程连接的概率与 $[d(u, v)]^{-\alpha}$ 成正比,这里 α 是一个参数.随着参数 α 值的增加,两个在网格距离上相隔较远的节点之间有长程连接的概率越来越小.

根据 Kleinberg 模型的思想,考虑最简单的一维环状网络,该小世界模型是通过用“按距离概率加边”取代 NW 小世界模型构造^[18]中的“随机化加边”而得到的.具体构造算法如下.

1) 从规则图开始,考虑一个含有 N 个点的最近邻耦合网络,它们围成一个环,其中每个节点都与它左右相邻的各 $k/2$ 个节点相连, k 是偶数.

2) 按距离概率加边:以概率 $[d(u, v)]^{-\alpha} / \sum_v [d(u, v)]^{-\alpha}$ 在随机选取的一对节点之间加上一条边.其中,任意两个不同的节点之间至多只能有一条边并且每一个节点都不能有边与自身相连.其中 $d(u, v)$ 为节点 u 和节点 v 之间的距离.

3.2. 传播率临界值

一维的 Kleinberg 小世界网络同样存在传播率临界值.仿真中我们选择网络规模 $N=2000$,平均次数取为 50.图 4 反映了 α 的改变对传播率临界值的影响.可见,在这一类小世界网络中,都存在传播率临界值 λ_c .随着 α 的增大, λ_c 有所增加,这是因为当

α 增大时,节点之间添加长程连接的概率减小,同时由于长程连接更倾向于在距离小的节点之间,不会有太多的跨度很大的长程连接,这在一定程度上减少了传播途径,从而阻碍了病毒的大面积传播.

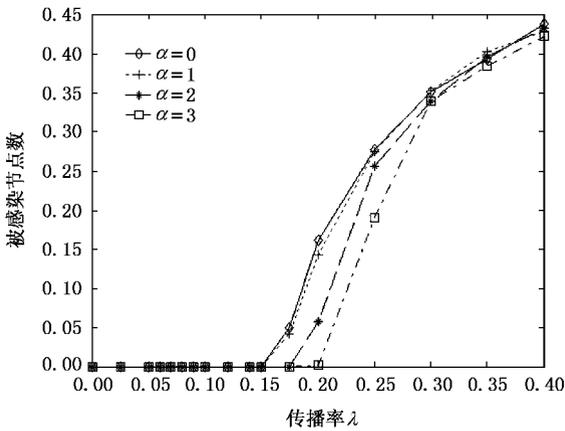


图 4 传播率 λ 与 α 关系比较

3.3. 局域范围控制

仍然采用第 2 节提到的局域范围控制方法,在仿真实验中选取固定传播率 $\lambda = 0.5$,始终大于传播率临界值,因而仿真分析中可以排除传播率临界值因素的影响,而只考虑 d 和 α 的作用.

在不加入任何控制措施时,由图 5(a)可见,无论 α 如何变化,被感染节点比例几乎没有变化,略呈下降趋势.当加入控制措施,即 d 逐渐增大(见 $d = 1, 2$),最终被感染节点数与没有局域范围控制作用时相比有明显下降.同时随着 α 的增加,被感染节点数迅速下降,当 $\alpha = 4$ 时,可以达到近似零感染的效果.原因在于 α 的增大使长程连接的分布更偏向于距离近的节点之间,从而加剧了网络的局部长程连接,使形成更加紧密的局域网络,因而控制小范围内的节点,甚至只有一步或两步之内的节点,即可将传播限制在此范围附近,而通过跨度大的长程连接又使遥远的地方产生新的传染源的概率将有所下降.当 $d \geq 3$ 时,无论 α 如何变化,传播都可以被有效抑制住.

图 5(b)反映了总的被控制节点数比例随 α 和 d 的变化曲线. $d \leq 2$ 时被控制节点比例没有明显的变化,当 $d = 3$ 时出现了明显不同的情形: α 很小时,需要控制相当大的节点数;当 α 增大,被控制节点比例呈大幅下降.而 $d > 3$ 与 $d = 3$ 的情况类似,这说明 $d = 3$ 是一个临界的状态.我们计算发现网络聚

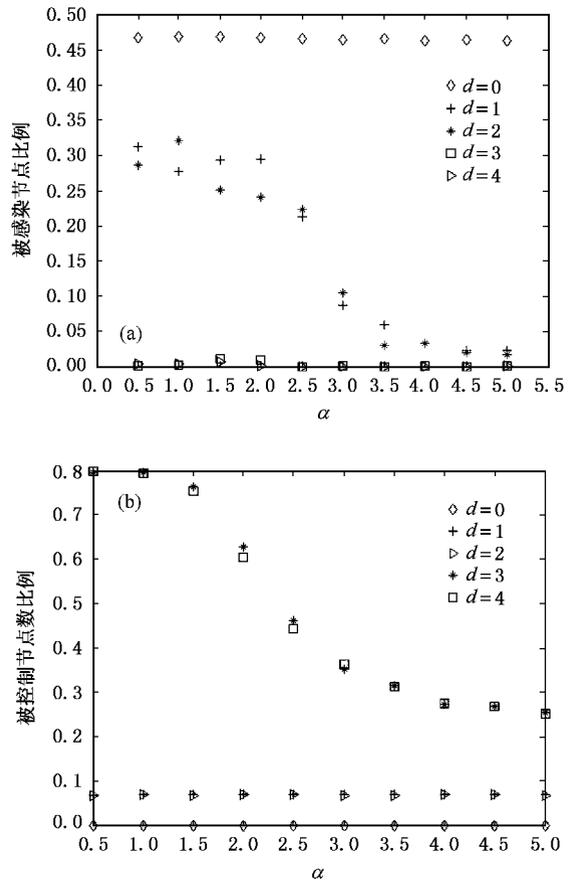


图 5 一维网络中被感染节点数和被控制节点数随 α 和 d 的变化曲线 (a)被感染节点数随 α 和 d 的变化曲线;(b)被控制节点数随 α 和 d 的变化曲线

类系数随 α 增大有很明显的增加,这可以解释当 $d = 3, 4$ 时,随 α 增加被控制节点数显著减少的原因.而当 $d = 1, 2$ 时,因为涉及的范围最多为邻居的邻居,所以网络的聚类系数变化对此并无多大影响.当 d 增大,涉及到更多的节点和连边,节点的选取和网络的紧密程度密切相关.随着聚类系数的增大,某感染源节点的 d 步邻居很可能也是 $d - 1$ 步邻居或 $d - 2$ 步邻居.因此,选择感染源周围的节点加以控制时,由于局部网络的紧密性,就会出现节点被重复选取的情况,例如对距离感染源 $d = 3$ 范围内的节点进行控制时,可能在 $d = 2$ 范围内已经包含了 $d = 3$ 范围内的大部分节点,所以被控制节点数比例只占整个网络总节点数的一小部分.因此,大的聚类系数特性将有效减少控制成本,提高控制效率.

4. 结 论

网络中发生病毒传染事件后,需要采取措施进

行有效控制.对于庞大复杂的网络,如果对网络的大部分节点进行控制,无疑会付出很大的代价.为避免病毒在整个网络的大范围传播,同时又降低成本,本文研究如何利用局部信息来控制感染源周围小范围内的节点,使得最终能抑制整个网络发生大范围病毒传播的局域范围控制策略.我们发现:对于随机图网络和 BA 无标度网络采用局域范围控制可以在 d

$= 1, 2$ 达到次优控制的效果,而只有在 WS 小世界网络中,存在这种控制范围的最优值,即最优控制范围 $d = 3$;而对于存在距离偏好的 Kleinberg 小世界网络,加入局域范围控制后, $d \geq 3$ 使得网络最终零感染.我们发现网络聚类系数影响受控节点比例.因此,加强网络的紧密程度可以增强局域范围控制的效果.

- [1] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. E* **63** 066117
- [2] Wang X F, Chen G R 2003 *IEEE Circuits and Systems Magazine* **3** 6
- [3] Li X, Chen G R, Li C G 2004 *Int. J. of Systems Science* **35** 527
- [4] Li X, Wang X F 2006 *IEEE Trans. Automatic Control* **51** 534
- [5] Volchenkov D, Volchenkova L, Blanchard Ph 2002 *Phys. Rev. E* **66** 046137
- [6] Xu D, Li X, Wang X F 2004 *J. of Complex Systems and Complexity Science* **1** 10 (in Chinese] 许 丹、李 翔、汪小帆 2004 复杂系统与复杂性科学 **1** 10]
- [7] Xu D, Li X, Wang X F 2006 *J. of Control and Decision* **21** 817 (in Chinese] 许 丹、李 翔、汪小帆 2006 控制与决策 **21** 817]
- [8] Li Y, Liu Y, Shan X M, Ren Y, Jiao J, Qiu B 2005 *Chinese Physics* **14** 2153
- [9] Zhen J, Liu Q X 2006 *Chinese Physics* **15** 1248
- [10] Pastor-Satorras R, Vespignani A 2001 *Phys. Rev. E* **65** 036134
- [11] Cohen R, Havlin S, Ben-Avraham D 2003 *Phys. Rev. Lett.* **91** 247901
- [12] Erdős P, Rényi A 1959 *Publ. Math. Inst. Hung. Acad. Sci.* **5** 17
- [13] Watts D J, Strogatz S H 1998 *Nature* **393** 440
- [14] Barabási A L, Albert R 1999 *Science* **286** 509
- [15] Kleinberg J 2000 *Nature* **406** 845
- [16] Kleinberg J 2000 *Proceedings of the 32nd Annual ACM Symposium on Theory of Computing* 163
- [17] Pastor-Satorras R, Vespignani A 2003 *Handbook of Graphs and Networks* Bornholdt S, Schuster H G (eds.) (WILEY-VCH publisher)
- [18] Newman M E J, Watts D J 1999 *Phys. Lett. A* **263** 341

An investigation on local area control of virus spreading in complex networks *

Xu Dan Li Xiang[†] Wang Xiao-Fan

(Department of Automation, Shanghai Jiaotong University, Shanghai 200240, China)

(Received 27 June 2006 ; revised manuscript received 17 July 2006)

Abstract

The local area control of virus spreading is studied in the light of the path length in complex networks, and the efficiency of local area control for complex networks with different topologies is analyzed. The research suggests that local area control method is effective in the WS small world networks, the BA scale-free networks, and the ER random networks; but the optimal radii $d = 3$ of the local control area for zero infection applies only to the WS small world networks. In the Kleinberg small world networks, when the distance bias of long-range links increases, the epidemic threshold increases, and the effect of local area control strengthens.

Keywords: complex networks, epidemics, local area control, path length

PACC: 0565, 0570J

* Project supported by the National Natural Science Foundation of China (Grants Nos. 70431002, 60504019, 90412004 and 60225013), and the Shanghai Rising-Star Program (Grant No. 05QMX1436).

[†] Corresponding author, E-mail: xli@sjtu.edu.cn