论 文

基于度相关性的病毒传播模型及其分析

徐久强*, 王进法, 张君, 赵海

东北大学信息科学与工程学院, 沈阳 110819

* 通信作者. E-mail: xujiuqiang@ise.neu.edu.cn

收稿日期: 2013-05-03; 接受日期: 2013-09-06; 网络出版日期: 2014-05-12

国家自然科学基金 (批准号: 61101121) 和国家科技支撑计划 (批准号: 2012BAH82F04) 资助项目

摘要 近年来网络病毒传播已对网络安全构成严重威胁. 研究表明, 互联网宏观拓扑结构与病毒传播有很大关系. 度相关性是互联网宏观拓扑的一个重要特征, 度相关性的改变意味着互联网拓扑结构的变化. 通过度相关特征分析, 发现互联网的异配性呈现减弱趋势; 本文使用 DPR 算法构造连续匹配系数的网络拓扑, 以便在具有不同匹配系数上的网络上进行病毒传播实验, 以此研究病毒的传播速度、稳态感染率、传播临界值; 然后根据传统病毒传播模型 SIS, 建立了适合因特网上的病毒传播模型 SIS-DVDI, 并进行病毒传播实验, 分析了病毒传播的稳态特性和瞬态特性; 最后根据先前仿真实验得出的结论探讨了网络病毒的防护措施.

关键词 互联网拓扑 度相关性 病毒传播 DPR 算法 SIS-DVDI

1 引言

随着互联网的发展, 互联网上的病毒传播已经成为全球性问题. 目前互联网作为一种典型的复杂网络^[1], 人们对它的宏观拓扑结构与病毒传播特点的研究并不充分. 近年来的研究表明: 互联网的宏观拓扑结构对网络病毒传播有着重要影响. 度相关性是互联网宏观拓扑的一个重要特征, 度相关性的改变往往意味着互联网整体拓扑结构的变化. 在复杂网络的研究中, 网络被抽象成简单的“节点”与“边”的集合, “节点”指复杂网络中参与交互的个体, 对 Internet 拓扑来说, “节点”指的是一台计算机 (IP 级拓扑)、一个路由器 (路由级拓扑) 或一个自治域 (autonomous system, 简称 AS^[2,3] 级拓扑), 而“边”代表节点间的关系.

目前对于复杂网络的研究主要包括网络结构发现、网络拓扑建模、网络特性分析及网络控制与同步等几个方面. 用来刻画实际网络宏观拓扑结构的网络特征量主要有度分布、平均路径长度、聚类系数、介数、度相关性等等. 传统的病毒传播研究主要为病毒传播模型的构建^[4]以及随机免疫^[5]、目标免疫^[6]、熟人免疫^[7]等免疫策略下病毒传播的控制, 而最近关于病毒传播的研究主要关注在人类的干预下, 采取积极的控制策略, 提高病毒传播的时间延迟, 从而使病毒传播达到无病点或者有病点的局部或全局稳定状态^[8~14], 但是病毒传播与互联网拓扑特性接合的研究并不多^[15], 在度相关性这个网络特征量与病毒传播关系的研究上国际上几乎一片空白. 本文从节点度相关性来研究网络拓扑对病毒传播的影响, 通过 (degree-preserving rewiring, DPR) 算法可以构造具有连续匹配系数的网络拓扑

引用格式: 徐久强, 王进法, 张君, 等. 基于度相关性的病毒传播模型及其分析. 中国科学: 信息科学, 2014, 44: 793–810, doi: 10.1360/N112013-00064

结构. 在不同匹配系数的网络上做了大量的仿真实验发现了病毒传播在不同匹配系数网络上表现出来的特性, 这些挖掘出的特性将对互联网病毒防护提供支持. 本文组织如下: 第 2 节介绍相关知识; 第 3 节介绍互联网宏观拓扑的度相关特征及其演化; 第 4 节基于度相关性研究病毒传播规律; 第 5 节对互联网病毒传播建模; 第 6 节总结.

2 相关知识

定义 1 度与分布: 一个节点 i 的度表征了该节点与网络中其它节点的连接数目, 记为 k_i . 网络节点度分布指网络中所有节点的度的分布情况, 即网络中各度值的节点数目占网络中总节点数目比例情况, 通常用函数 $P(k)$ 来表示.

定义 2 度相关性: 度相关性描述了网络中不同度值节点的连接偏好性. 对此一方面可以通过邻居节点平均度分布 $K_{nn}(k)$ 来进行分析,

$$K_{nn}(k) = \sum_{j=1}^{k_{\max}} j p(j|k), \quad (1)$$

其中, 条件概率 $P(j|k)$ 表示度为 k 的节点连接到度为 j 的节点的概率. 通过对各度值节点的邻居节点平均度分布情况的分析可以了解网络的度相关性质. 另一方面可以通过 Newman^[16] 提出的计算网络的匹配系数来进行量化, 匹配系数的定义如式 (2),

$$r = \frac{M^{-1} \sum_i j_i k_i - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}{M^{-1} \sum_i \frac{1}{2}(j_i^2 + k_i^2) - [M^{-1} \sum_i \frac{1}{2}(j_i + k_i)]^2}, \quad (2)$$

其中, j_i 和 k_i 分别为第 i 条边的两个端点的度数, M 为网络中边的数量.

定义 3 网络的同配性与异配性: 同配性指的是网络中大度值节点更倾向与大度值节点相连, 小度值节点倾向与小度值节点相连, 同配性越大, 这种倾向越明显; 异配性指的是网络中大度值节点更倾向与小度值节点相连, 异配性越大, 这种倾向越明显.

定义 4 极大、极小匹配系数: 在网络 G 中, 如果从同配方向进行边重连操作 n 次后, 匹配系数 r 有 $\lim_{n \rightarrow \infty} r \leq r_{\max}$, 则称 $r_{\max}$ 为网络 G 的极大匹配系数; 如果从异配方向进行边重连操作 n 次后, 匹配系数 r 有 $\lim_{n \rightarrow \infty} r \geq r_{\min}$, 则称 $r_{\min}$ 为网络 G 的极小匹配系数.

定义 5 (susceptible-infected-susceptible, SIS) 病毒传播模型: 在模型中, 处于感染状态 (infected) 的个体以一定概率 v 将疾病传播给易染个体 (susceptible), 同时感染个体也会以一定的概率 δ 恢复到健康状态. 因此, SIS 模型适用于那些处于感染状态的个体恢复健康后还存在被再次感染的情况.

在简化情况下, SIS 模型中病毒传播可以用以下微分方程组来描述:

$$\begin{cases} \frac{ds(t)}{dt} = -v \cdot i(t) \cdot s(t) + \delta \cdot i(t), \\ \frac{di(t)}{dt} = v \cdot i(t) \cdot s(t) - \delta \cdot i(t), \end{cases} \quad (3)$$

其中, $s(t)$, $i(t)$ 分别代表了网络中处于 S 状态 (易感染状态) 和 I 状态 (感染状态) 个体的比例, v , δ 分别代表了节点的感染率和恢复率.

图 1 度相关性分析

Figure 1 The analysis about correlation of degree. (a) Average degree of neighbor nodes distribution; (b) network average degree evolution; (c) network assortativity coefficient evolution

3 度相关特征及其演化

本节将通过 2011 年 2 月 AS 级 Internet 网络的局部特征来分析互联网的度相关性特征。

图 1(a) 是 2011 年 2 月 AS 级拓扑中邻居节点平均度分布情况, 其中横坐标为网络度值对数, 纵坐标为邻居节点平均度对数。从图中可以看出, 不同度值节点的邻居节点平均度分布情况表现出了较明显的差异性, 其中, 低度值节点具有较高的邻居节点平均度, 而高度值节点邻居节点的平均度较低。这说明了网络中低度值节点更加倾向于同高度值节点相连, 高度值节点则更加倾向于同低度值节点相连。这从直观上反映了网络的异配特性。这同总体上通过匹配系数来度量的网络连接偏好性 (度相关性) 的结论是相一致的^[16]。

图 1(b) 可看出, 随着时间变化, 节点平均度值在 4.2~4.8 间震荡, 并存在个别网络振幅较大情况, 但总体上看, 网络中节点平均度值未出现明显增大或减小趋势。这说明 AS 级拓扑中尽管不断有新节点加入和旧节点消亡^[17], 但网络的整体结构无显著变化。

图 1(c) 中可看出 AS 级匹配系数并非一成不变, 其随着时间在不断变化, 总体上看在增大, 即网

络异配特性有弱化趋势. 一般来说, 网络中新加入的节点大部分为小度值节点, 而通过网络异配性有弱化的趋势可推断网络中有很多新加入节点并非都与大度值节点连接.

4 基于度相关性的病毒传播特性

网络的度相关性可以用匹配系数来衡量, 网络匹配系数的值域为 $[-1, 1]$, 随着匹配系数的增大, 网络的度相关性表现为: 异配性在减弱, 同配性在增强. 现实网络的匹配系数往往为某一特定值, 例如, 技术网络的匹配系数一般大于 -1 小于 0 , 而社会网络的匹配系数一般大于 0 小于 1 ^[16]. 因此为观察不同匹配系数下的病毒传播特征, 本文引用边重连算法^[18]的分析过程, 给出了算法的具体实现过程, 并且使用 DPR 算法改变网络的匹配系数并对不同匹配系数下的网络仿真 SIS 病毒传播, 从传播速度、稳态感染率、传播临界值等 3 个方面分析 SIS 模型特性.

4.1 边重连算法

从上面有关 AS 级互联网度相关特性的分析, 可以发现互联网的度相关特征的异配性有弱化的趋势. 为了进一步分析和预测这种变化趋势对病毒传播的影响, 本文在深入分析度相关性的量化标准—匹配系数的基础上, 使用边重连 (degree-preserving rewiring) 算法改造网络的结构, 使网络向同配或异配方向变化. 由文献 [16] 可知, 给定某一度向量, 所构成网络的匹配系数是介于 $-1 \sim 1$ 之间的变量.

4.1.1 算法理论

可以把式 (2) 改写为

$$r(D_i, D_j) = \frac{E(D_i D_j) - \mu_{D_i}^2}{E(D_i^2) - \mu_{D_i}^2}, \quad (4)$$

其中 D_i, D_j 表示节点 i, j 的度, 又因为

$$E(D_i D_j) = \frac{\sum_{i=1}^N \sum_{j=1}^N d_i d_j a_{ij}}{2M} = \frac{\mathbf{d}^T \mathbf{A} \mathbf{d}}{2M}, \quad (5)$$

其中, d_i, d_j 是网络中所有节点度向量 $\mathbf{d}^T = [d_1, d_2, d_3, \dots, d_n]$ 中的元素, a_{ij} 是网络邻接矩阵 $\mathbf{A}$ 中的元素, 根据期望和邻接矩阵的含义有

$$\mu_{D_i} = \frac{\sum_{i=1}^N \sum_{j=1}^N d_i a_{ij}}{2M} = \mu_{D_j}, \quad (6)$$

其中

$$\mu_{D_i} = \frac{\sum_{i=1}^N \sum_{j=1}^N d_i a_{ij}}{2M} = \frac{\sum_{i=1}^N d_i \sum_{j=1}^N a_{ij}}{2M} = \frac{\sum_{i=1}^N d_i^2}{2M} = \frac{\mathbf{d}^T \mathbf{d}}{2M}. \quad (7)$$

又因为 $\sigma_{D_i}^2 = E[D_i^2] - \mu_{D_i}^2$, 且

$$E[D_i^2] = \frac{\sum_{i=1}^N \sum_{j=1}^N d_i^2 a_{ij}}{2M} = \frac{\sum_{i=1}^N d_i^3}{2M} = E[D_j^2]. \quad (8)$$

把 (5), (7), (8) 带入 (4) 中可以得到

$$r = \frac{N_1 N_3 - N_2^2}{N_1 \sum_{i=1}^N d_i^3 - N_2^2}. \quad (9)$$

其中, $N_1 = \sum_{i=1}^N d_i$, $N_2 = \sum_{i=1}^N d_i^2$, $N_3 = \mathbf{d}^T \mathbf{A} \mathbf{d}$. $N_k = \mathbf{d}^T \mathbf{A}^k \mathbf{d}$ 表示网络中所有跳数为 k 的路径数量. 由式 (9) 可以看出, 若节点的度值不变, 匹配系数 r 只与 $N_3 = \mathbf{d}^T \mathbf{A} \mathbf{d}$ 成正比. 于是在不改变网络度分布的基础之上, 重新进行连边可以改变网络的匹配系数.

首先, 从网络中随机选择两条边 $v_1 \sim v_2$ 和 $v_3 \sim v_4$, 它们对应的节点分别是 v_1, v_2, v_3, v_4 , 那么存在的边重连可能性是 $v_1 \sim v_3, v_2 \sim v_4$ 或者 $v_1 \sim v_4, v_2 \sim v_3$. 然后把随机选择的 4 个节点度值由大到小排列 $d_1 \geq d_2 \geq d_3 \geq d_4$, 并假设对应的节点变为 $n_{d_1}, n_{d_2}, n_{d_3}, n_{d_4}$, 此时这 4 个节点只存在 3 种不同连接关系, 即

$$S_a : n_{d_1} \sim n_{d_2}, n_{d_3} \sim n_{d_4} \quad S_b : n_{d_1} \sim n_{d_3}, n_{d_2} \sim n_{d_4} \quad S_c : n_{d_1} \sim n_{d_4}, n_{d_2} \sim n_{d_3}.$$

由文献 [19] 可知

$$\sum_{i=1}^N d_i^3 - N_3 = \sum_{i \neq j} (d_i - d_j)^2 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N a_{ij} (d_i - d_j)^2. \quad (10)$$

由式 (9), (10) 可得

$$r = 1 - \frac{\sum_{i \neq j} (d_i - d_j)^2}{\sum_{i=1}^N d_i^3 - \frac{1}{2M} \sum_{i=1}^N d_i^2}. \quad (11)$$

因为随机选择边并进行重连只会改变式 (11) 中的 $\omega = -\sum_{i \neq j} (d_i - d_j)^2$, 因此 S_a, S_b, S_c 3 种连接对应的匹配系数大小关系为

$$\omega_a - \omega_b = -(d_1 - d_2)^2 - (d_3 - d_4)^2 + (d_1 - d_3)^2 + (d_2 - d_4)^2 = 2(d_2 - d_3)(d_1 - d_4) \geq 0.$$

同理可以得到

$$\omega_a - \omega_c = 2(d_2 - d_4)(d_1 - d_3) \geq 0,$$

$$\omega_b - \omega_c = 2(d_1 - d_2)(d_3 - d_4) \geq 0.$$

上述 S_a, S_b, S_c 3 种状态中对应的匹配系数的大小关系为 $r_a \geq r_b \geq r_c$. 那么如果重连关系按 $S_a \rightarrow S_b, S_a \rightarrow S_c, S_b \rightarrow S_c$ 进行, 会使得匹配系数单调递减; 相反地, 如果按照 $S_c \rightarrow S_a, S_c \rightarrow S_b, S_b \rightarrow S_a$ 进行, 会使得匹配系数单调递增. 从以上可看出通过局部边的重新交换连接可获得渐变的匹配系数值, 同时保持了节点度值的不变. 在重连的过程中, 随着交换边数的不断增加, 符合重连规则的边数将会越来越少. 理论上来说, 如果给予足够充分的重连次数的话, 匹配系数将稳定在某一极值.

4.1.2 算法实现

基于上面算法的理论研究, 对 DPR 算法进行实现. 该算法可以在两个方向上构造具有连续匹配系数的网络集合. 选择同配方向重连可构建匹配系数渐增的连续匹配系数网络; 选择异配方向重连则可构建匹配系数减小的连续匹配系数网络. 当边重连足够充分时可得到具有极大匹配系数或极小匹配系数的网络. 下面是算法的实现过程描述:

1. 加载网络拓扑数据进行原始网络的构建, 从原始网络中随机选择两条边, 假定为 (v_1, v_2) 和 (v_3, v_4) , 记录这两条边两侧节点的索引号 (由一数组 fourNodes[4] 记录) 以及各个节点对应的度值 (由另一数组 fourDegrees[4] 记录), 然后根据这 4 个节点度值的降序序列用一个数组 (数组 orderIndex[4]) 来记录对应节点的数组下标.

图 2 (网络版彩图) 不同网络中进行 DPR 操作的匹配系数变化

Figure 2 (Color online) The changes of assortative coefficient under different network with DPR. (a) Assortativity coefficient changes under sufficient DPR in the BA network; (b) maximal and minimal assortativity coefficient evolution in AS-level Internet

2. 根据选择重连方向 (同配方向或异配方向) 的不同对上述两条边 (v_1, v_2) 和 (v_3, v_4) 进行重连, 并统计已执行重连数.

(a) 同配或异配方向重连时, 需满足如下共同条件:

- i. 原始两条边的 4 个端节点不存在公共节点, 也就是 $v_1 \neq v_3, v_1 \neq v_4, v_2 \neq v_3, v_2 \neq v_4$.
- ii. 原始两条边 (v_1, v_2) 和 (v_3, v_4) 至少存在一条边的两侧节点不是最大和次大度值节点相连或最小和次小度值节点相连.
- iii. 新网络依然是连通的.

(b) 当选择同配方向重连时, 还需满足条件:

原始网络中不存在边 $(\text{fourNodes}[\text{orderIndex}[0]], \text{fourNodes}[\text{orderIndex}[1]])$ 和 $(\text{fourNodes}[\text{orderIndex}[2]], \text{fourNodes}[\text{orderIndex}[3]])$, 即不存在要重连形成的新边.

(c) 当选择异配方向重连时, 还需满足条件: 原始网络中不存在边 $(\text{fourNodes}[\text{orderIndex}[0]], \text{fourNodes}[\text{orderIndex}[3]])$ 和 $(\text{fourNodes}[\text{orderIndex}[1]], \text{fourNodes}[\text{orderIndex}[2]])$, 即不存在要重连形成的新边.

3. 重复执行上述过程 pM 次, 其中 M 是网络的连接总数, p 是一个代表了进行的重连比值. 当 p 取得足够大时可得到极大匹配系数网络和极小匹配系数网络. 在重复执行边重连的过程中, 不断的保存当前的网络, 就可得到所要的连续匹配系数网络集合.

4.1.3 算法验证

图 2(a) 是对 BA 模型^[20] 构造的网络进行充分的 DPR 过程中匹配系数的变化情况, 其中网络节点数为 230, 边为 675. 分别从同配方向和异配方向对原始的 BA 网络进行了充分的重连, 从图中可看出在重连比值超过 400% 时, 重连得到的网络匹配系数已经基本保持不变, 达到了相应的极值. 图 2(b) 为本文对实际的 AS 级网络进行了边重连操作, 并分别从同配方向 (使匹配系数增大) 和异配方向 (使匹配系数减小) 对网络进行了内部边重连, 得到了对应的极大和极小匹配系数以及这个区间的连续值.

表 1 3 种原始网络的基本属性

Table 1 The basic properties of three original networks

Type	Number of nodes	Number of edges	Average degree	Assortativity coefficient
ER	999	3054	6.11	-0.003
WS	1000	3000	6.00	-0.029
BA	1000	2985	5.97	-0.048

结合图 1(c) 中原始网络匹配系数的演化, 可以发现匹配系数的边界值 (极大、极小值) 随着原始实际网络匹配系数的演化而出现逐渐增大趋势, 这从值域的角度说明了互联网 AS 级拓扑的异配性有弱化的趋势; 同时也说明了对于 AS 级网络很难得到一个全局的最大或最小匹配系数, 只能得到某一个时段内的极大或极小匹配系数, 但这种差异非常小, 因此并不影响后面有关度相关性对病毒传播影响的研究.

4.2 传播速度、稳态感染率、传播临界值

为了观察具有不同匹配系数的网络上的病毒传播行为, 需要以 DPR 算法为基础构建具有不同匹配系数的网络集合; 由于不同拓扑模型构造出的网络结构特征的差异性, 需要多种不同类型的网络进行综合比较分析. 本文选取经典网络模型 (erdős-rényi, ER) 模型, (watts-strogatz, WS) 模型和 (barabási-albert, BA) 模型^[1] 作为仿真对象; 选择经典传播模型 -SIS, 网络中个体被抽象分成易感染状态和感染状态类, 可以通过设定不同感染率和恢复率观察病毒传播情况. 现实网络节点状态在时间上是连续的, 但在实验过程中, 为了记录方便, 节点只取特定的时间点上的状态, 特定的时间点称之为时间步. 首先在初始时刻随机选定一定比例的节点为感染节点, 某一节点下一时间步所处的状态由其当前时间步邻居节点的状态以及病毒传播的感染率和恢复率来决定. 实验在具有不同匹配系数的网络上进行, 每次仿真重复 100 次最后取平均值.

4.2.1 传播速度

设定病毒感染率 $v=0.05$, 恢复率 $\delta=0.04$, 初始感染节点比例 2%, r_{org} 为对应模型构造的原始网络的匹配系数, r_{max} 和 r_{min} 分别为在对应原始网络基础上通过 DPR 算法得到的网络的极大和极小匹配系数. 实验采用的网络的基本参量如表 1 所示.

仿真结果如图 3 所示, 其中横坐标表示时间步, 纵坐标表示节点感染比例. 从图 3 中可看出, 无论 ER 模型, WS 模型, BA 模型, 病毒在异配型网络的传播速度要快于在同配型网络中的传播速度. 为了更清晰的观察病毒在异配型网络 and 同配型网络中传播速度的差异性, 并分析其产生的原因, 本文在小规模的 BA 模型网络上进行了相似的实验并描绘出了对应的网络拓扑结构和节点感染状态图.

图 4 为病毒传播在异配型网络 and 同配型网络中的感染状态图. 网络规模: 节点数为 200, 边数为 780, 感染率 $v=0.04$, 恢复率 $\delta=0.05$, 初始感染节点比例 5%, 时间步 $t=30$. 黑色表示节点处于感染状态, 白色表示节点处于易感染状态, 节点的大小与节点度值成正比. 从图中可看出, 在相同时间步内, 异配型网络病毒节点感染比例明显高于同配型网络, 即异配型网络病毒传播速度快于同配型网络. 通过结构特征对比可看出, 在异配型网络中, 大度值更倾向于同小度值节点相连^[21], 由于大度值节点本身具有较多的边与周围的节点相连, 一旦其被感染就会将病毒迅速向周围扩散, 小度值节点也会较容

图 3 (网络版彩图) 3 种网络模型病毒传播速度仿真结果

Figure 3 (Color online) The simulation results of virus spreading speed of the three networks' model. (a) ER; (b) WS; (c) BA

图 4 (网络版彩图) BA 模型网络在不同匹配系数下的传播状态

Figure 4 (Color online) The spreading state of BA model under different assortative coefficient. (a) $r_{\min} = -0.606$, $\rho_0 = 0.805$; (b) $r_{\max} = 0.492$, $\rho_0 = 0.435$

易的将病毒传播给与之相连的大度值节点, 从而加速病毒传播, 因此图 4(a) 中大部分节点被感染, 且度值越大被感染的概率越高. 在同配型网络中, 往往是度值相近的节点更加紧密联系在一起, 从而就更加容易的形成一个个区块, 而在不同的区块之间往往存在着较少的连接, 区块之间这种弱连接性降

图 5 (网络版彩图) 不同有效传播率下匹配系数与稳态感染率的关系

Figure 5 (Color online) The relationship between assortative coefficient and steady infection under different spreading-rate effective. (a) ER; (b) WS; (c) BA

低了病毒在不同区块之间传播, 因此图 4(b) 中只有个别区块被感染. 同时, 由于在大度值区块内, 节点之间的联系更加的紧密, 病毒传播会更加容易, 所以图 4(b) 可清晰地看出只有最上部的一个区块的大部分节点被感染, 其余区块只有部分节点感染.

4.2.2 稳态感染率

从前面的基础知识知道网络中病毒传播的稳态感染率主要是由有效传播率 $\lambda = v/\delta$ 决定的. 为此, 首先对前面的 ER, WS, BA 模型对应的 3 种匹配系数网络进行了仿真实验以研究不同有效传播率下的稳态感染率情况. 通过实验结果发现在较低的有效传播率下, 病毒的稳态感染率为零, 但是随着有效传播率的增加, 病毒也将开始传播; 当有效传播率在 0.2~0.8 之间, 病毒传播迅速, 且随着有效传播率的增加, 稳态感染率也相应的增加并保持在一一定的比例; 当有效传播率大于 0.8 以后, 网络的稳态感染率趋于稳定, 但是始终小于 1, 说明网络中总是会存在未被感染的节点. 同时还发现网络匹配系数越大, 病毒传播的稳态感染率越低, 且在进行大量实验的基础上, 发现该规律在 BA 模型上最明显, ER 模型模型次之, WS 模型最不明显, 就该特性而言, 说明度相关性对病毒传播的影响在无标度网络上更显著.

为了更进一步的分析度相关性对病毒传播的稳态感染率的影响, 我们在连续匹配系数网络集合中对在不同有效传播率下的传播稳态感染率进行了仿真研究. 连续匹配系数网络集合是根据前面的 DPR 算法以固定的重连比值来选择的. 图 5 显示了在 11 种不同有效传播率下匹配系数对病毒传播稳态感

染率的影响, 其中 (a) 为 ER 模型, (b) 为 WS 模型, (c) 为 BA 模型. 下面以 BA 模型为例来分析稳态感染率与匹配系数之间的关系. 从图 5(c) BA 模型中可以看出随着有效传播率的变化, 匹配系数对病毒传播稳态感染率的影响关系并非一成不变的, 当 $\lambda=0.12, 0.15, 0.18, 0.21$ 时, 病毒传播的稳态感染率随着匹配系数的增大出现了先增大后减小的现象; 当 $\lambda=0.24, 0.27, 0.3, 0.5$ 时, 病毒传播的稳态感染率随着匹配系数的增大出现了加速下降的现象; 当 $\lambda=0.7, 1.1, 1.5$ 时, 出现了较明显的线性下降现象, 并且随着有效传播率的增加, 这种线性减小的变化变得十分微弱. 由此现象可以看出, 随着有效传播率的增加, 病毒传播的稳态感染率同网络的匹配系数之间存在着一个不断变化的过程, 即从总体上来说, 一方面, 病毒传播的稳态感染率随着匹配系数的增加有降低的趋势, 而另一方面, 有效传播的增加又弱化了这种变化的趋势. 上述现象在 ER 模型和 WS 模型中都存在, 但相比较而言, BA 模型的变化要明显于其他两种模型.

4.2.3 传播临界值

传播临界值是复杂网络病毒传播研究中重要的一点, 它表示只有当病毒有效传播率大于某一正临界值时, 大规模传播才会出现, 如果小于该临界值时, 病毒无法传播. 从文献 [22] 可知网络传播临界值 $\lambda_c = 1/\Lambda_m(A)$, $\Lambda_m(A)$ 是网络对应邻接矩阵 A 的最大特征值. 从临界值公式中可知, 网络传播临界值同网络对应邻接矩阵最大特征值成反比. 我们对 ER, WS, BA 等网络模型下的匹配系数与对应网络邻接矩阵最大特征值之间关系进行了分析, 发现网络匹配系数与网络邻接矩阵的最大特征值呈现出较明显的正比关系. 说明在异配型网络中尽管病毒传播的稳态感染率较高, 但同时也具有较高的传播临界值, 即在异配型网络中, 如果病毒有效传播率较低, 病毒无法大规模传播; 而在同配型网络中尽管稳态感染率较低, 但传播临界值也相对较低, 即在同配型网络中, 不需要较高的有效传播率, 病毒就会爆发. 对于这种现象, 分析其原因是当匹配系数减小时, 节点对之间添加长程连接 (捷径) 的概率在减小, 与此同时长程连接更倾向于在小距离节点间进行, 减少了跨度较大的长程连接出现的可能, 从而在一定程度上减少了病毒传播的途径, 进而阻碍了病毒的大规模爆发.

5 病毒传播建模

传统研究病毒传播模型主要采用 SI, SIS, SIR 模型, 这些模型假设网络中每个节点感染概率在传播过程中相同, 这同现实网络不相符. 文献 [23] 采用 SI 病毒传播模型, 对无标度网络中不同度值类型的节点病毒传播做了详细的分析, 但是文献 [23] 中每个节点的感染率均相同, 而在互联网中, 个体间是存在差异性, 往往节点越重要, 节点的保护措施越完善, 节点抗病毒能力就越强. 为了突出网络节点免疫能力的差异, 更好的符合实际, 本文在 SIS 传播模型的基础之上, 建立了 SIS-DVDI 传播模型 (different vertex different immunity), 使不同度值节点的感染率不同, 并对 AS 级 Internet 网络上病毒传播特性进行分析. 在该模型中, 节点的易染状态和感染状态依然符合传统 SIS 模型定义. 为简单起见, 不考虑节点由感染状态恢复到易感染状态的时间差异和恢复差异, 恢复率依然表示节点从感染状态恢复到易感染状态概率.

5.1 模型建立

通过对 SIS 模型进行改进, 引入新的特性来描述网络中不同度值节点拥有不同抗感染性这一现象. 采用节点度值作为衡量节点重要性参量, 假定节点度值越大, 节点在网络中越重要, 抗病毒能力就

越强, 并假定相同度值的节点有相同抗病毒能力. 为简单起见, 假定每一个度值为 k 的易感染节点, 如果与感染节点存在着一一条或多条边相连, 那么该节点感染病毒概率为 v_k , 其定义如下:

$$v_k = (k + L)^{\alpha-1}, \quad (12)$$

k 代表节点度值; L 为大于 1 的常数. 设置这一常数考虑的是实际网络中往往存在度值为 1 的节点, 而这些节点的感染概率不可能为 1. 这一常数的设定可以保证网络中度为 1 的节点感染率小于 1, 使模型更符合实际. $\alpha \in [0, 1)$ 是感染因子. 由式 (12) 可知: 度值 k 一定时, 感染率 v_k 随着 α 的增大而增大; 在 α 一定时, 感染率 v_k 随着 k 的增大而减小. 定义 δ 为网络中节点由感染状态恢复到易染状态概率.

在新建模型中, 网络中的节点状态依然遵循易染状态 (S) $\rightarrow$ 感染状态 (I) $\rightarrow$ 易染状态 (S) 的转换. 病毒在网络中传播的演化过程为: 在每一个时间步内, 对于每个度值为 k 的易染节点, 如果它与已经感染的节点存在着一一条或多条边相连, 那么该节点将以 v_k 的概率变为感染状态, 同时已感染节点会以 δ 概率恢复为易感染状态; 重复上述过程直到稳定状态, 也即网络中处于感染状态节点的数目占网络中总节点数目的比例不再发生大的变化.

下面给出在此模型下病毒在无标度网络中的动态演化过程. 假设网络的规模为 N , 函数 $S(t)$ 和 $I(t)$ 分别代表了时间步 t 时, 网络中易染节点的数量和感染节点的数量, 因此满足 $N = S(t) + I(t)$. 采用时间步网络中感染节点数目占网络中总节点数目的比例 $i(t) = I(t)/N$ 来描述网络中病毒的动态传播过程, 不过考虑到无标度网络中节点度的波动性和对病毒抗感染性的不同, 将 $i_k(t)$ 表示为度值为 k 的感染节点在所有度值为 k 的节点中所占的比例, 其具体定义如下:

$$i_k(t) = \frac{I_k(t)}{N_k}, \quad (13)$$

其中 $I_k(t)$ 表示处于感染状态且度值为 k 的节点数目, N_k 表示度值为 k 的节点数目. 同时考虑网络的度分布, 则可推导 $i(t)$ 为

$$i(t) = \frac{\sum_{k=1}^{k_{\max}} I_k(t)}{N} = \sum_{k=1}^{k_{\max}} \frac{N_k}{N} \frac{I_k(t)}{N_k} = \sum_k P(k) i_k(t), \quad (14)$$

其中 $P(k)$ 代表的是网络中节点度分布函数, $k_{\max}$ 为网络中的最大度值.

在这个模型中, $i_k(t)$ 的变化率方程如下 [23]:

$$\frac{di_k(t)}{dt} = v_k k [1 - i_k(t)] \theta_k(t) - \delta i_k(t), \quad (15)$$

其中 v_k 表示度为 k 的节点的感染病毒的概率, $1 - i_k(t)$ 表示的是度为 k 的节点在时间步 t 时, 连接到没被感染节点的概率, $\theta_k(t)$ 代表了度为 k 的节点的相邻节点已处于感染状态的概率. 在此, 不考虑不同节点之间的度关联性, 因此 $\theta_k(t) = \theta(t)$. 忽略掉式 (15) 中的高阶校正项 $O(i^2)$, 得到式 (16)

$$\frac{di_k(t)}{dt} = (k + L)^{\alpha-1} k \theta(t) - \delta i_k(t). \quad (16)$$

实际上, 处于易染状态的节点连接到感染状态节点的概率与该感染节点的度值成正比, 即随机选择的节点更加容易连接到大度值且处于感染状态的节点上 [24, 25]. 再考虑到每个感染节点至少有一条边与另一个感染节点相连, 因此可得到式 (17)

$$\theta(t) = \frac{\sum_k (k-1) P(k) i_k(t)}{\sum_s s P(s)}. \quad (17)$$

图6 (网络版彩图) 感染因子和恢复率对病毒传播的影响

Figure 6 (Color online) The impact of infectious factor and recovery rate to virus spreading. (a) α vs. $i(t)$; (b) δ vs. $i(t)$

根据式 (16) 和 (17) 得出 $\theta(t)$ 的变化率方程

$$\frac{d\theta(t)}{dt} = \frac{\langle k^2(k+L)^{\alpha-1} \rangle - \langle k(k+L)^{\alpha-1} \rangle - \delta \langle k \rangle}{\langle k \rangle} \theta(t), \quad (18)$$

其中 $\langle k \rangle = \sum_k kP(k)$ 为无标度网络的平均度, $\langle k^2(k+L)^{\alpha-1} \rangle = \sum_k k^2(k+L)^{\alpha-1}P(k)$, $\langle k(k+L)^{\alpha-1} \rangle = \sum_k k(k+L)^{\alpha-1}P(k)$. 其中式 (14), (16), (18) 和初始条件 $i_k(0) = i_0$, 得出 $i(t)$ 的函数表达式为

$$i(t) = i_0 \left[1 + \frac{\langle k \rangle \langle k(k+L)^{\alpha-1} \rangle}{\langle k^2(k+L)^{\alpha-1} \rangle - \langle k(k+L)^{\alpha-1} \rangle - \delta \langle k \rangle} (e^{\frac{t}{\varphi}} - 1) \right], \quad (19)$$

其中,

$$\varphi = \frac{\langle k \rangle}{\langle k^2(k+L)^{\alpha-1} \rangle - \langle k(k+L)^{\alpha-1} \rangle - \delta \langle k \rangle}. \quad (20)$$

分析式 (19) 可以看出: 在网络中, δ 一定, 当节点与感染节点存在连接时, α 越大, 节点感染病毒的可能性越大, 病毒在网络中传播的速度也就越快, 病毒达到稳态时的稳态感染率越高; 而当 α 一定, δ 越大, 病毒从感染状态恢复到易感染状态的可能性就越大, 病毒越不容易扩散, 病毒达到稳态时的稳态感染率越低. 同时根据前面的知识也了解到, 无标度网络的度分布符合幂律分布^[22], 也就是网络的结构是不均匀的, 因此一旦病毒爆发, 就会迅速的在网络中扩散开来.

5.2 模型验证与分析

为了验证模型的合理性, 实验选取了 (cooperative association of internet data analysis, CAIDA) 提供的一个完整探测周期的全球 AS 级互联网拓扑数据 (2010 年 10 月 28 日全球范围内 AS 间互联的数据). 网络规模: 节点 16248, 边 31034. 病毒在网络中的动态演化过程为: 初始时刻随机选定一定比例节点为感染节点, 其余节点处于易感染状态, 在每一个时间步内, 对于每个度值为 k 的易染节点, 如果它与已经感染的节点存在一条或多条边相连, 那么该节点将以 v_k 的概率变为感染状态, 同时已感染节点会以 δ 概率恢复为易感染状态; 重复上述过程直到稳定状态, 同时为了消除随机误差的干扰, 仿真实验重复 100 次, 最后取平均值.

图 7 (网络版彩图) 各度值节点的稳态感染率

Figure 7 (Color online) Steady infection of nodes with same degree

图 8 (网络版彩图) 各度值节点的辐射强度

Figure 8 (Color online) Radiation intensity of nodes with same degree

5.2.1 模型验证

网络中对病毒传播起主要作用的参数: 感染因子 α 和恢复率 δ . 下面分别从这 2 个参数入手, 进行模型的合理性和可行性验证.

首先考虑感染因子 α 对病毒传播的影响. 图 6(a) 给出了在不同的感染因子下, 病毒在网络中传播随时间的演化过程. 其中病毒传播其它参数的设置为: 初始感染率 $\rho_0=0.01$, 恢复率 $\delta=0.2$. 从图中看出, 感染因子 α 越大 (节点抗感染能力整体减弱), 病毒传播的速度越快, 达到稳态时, 节点的感染比例也越高. 进一步讲就是病毒传播能力越强 (感染因子 α 越大), 病毒传播的速度就越快. 这与通过式 (19) 得出的理论分析是相符的.

接下来, 考虑恢复率 δ 对病毒传播的影响. 图 6(b) 给出了在不同的恢复率下, 病毒在网络中传播随时间的演化过程. 其它参数设置: 初始感染率 $\rho_0=0.01$, 感染因子 $\alpha=0.2$. 从图中可看出恢复率 δ 越大, 病毒传播的速度越慢, 达到稳态时, 节点感染比例越低. 在两种极端的情况下, 当 $\delta=0.0$ 时, 病毒传播达到稳态时的稳态感染率变为 1, 即网络中的节点最终会被完全感染. 而当 $\delta=1.0$ 时, 病毒传播达到稳态时的稳态感染率最终为 0, 即病毒最终在网络中消亡. 以上说明病毒在传播过程中越容易被治愈那么它就越不易传播. 这与通过式 (19) 得出的理论分析结果也是相符的.

5.2.2 病毒传播稳态特性分析

度值为 k 的节点的辐射强度 R_k 的定义如下:

$$R_k = NP(k)i_k(t)k, \quad (21)$$

其中 N 为网络节点总数, $P(k)$ 为网络度分布函数, $i_k(t)$ 表示度值为 k 的节点的感染比例.

图 7, 8 中, 总体 α 表示在给定感染因子值的情况下病毒达到稳态时的总体稳态感染率, 初始感染率 $\rho_0=0.01$, 恢复率 $\delta=0.2$.

图 7 表示各度值节点的稳态感染率的情况, 高度值节点的稳态感染率要高于低度值节点的稳态感染率, 并且在低度值节点区域稳态感染率随着度值的升高呈现较为明显的平滑上升趋势, 而在高度值节点区域稳态感染率呈现出波动态势, 不同感染因子下绝大多数度值节点的稳态感染率要高于相应的总体稳态感染率, 只有最低度值区的稳态感染率要低于总体稳态感染率. 同时低度值区的稳态感染率

随着感染因子 α 的增大而增大, 而高度值区不存在此规律. 对于总体上稳态感染率随着度值的增大而变大现象, 体现了在 AS 级 Internet 这种无标度网络中高度值节点更加容易被感染 (度大表示连接边数多, 相当于接触病毒机会大, 并且根据前面的分析发现 AS 级 Internet 存在着明显的异配性), 尽管高度值节点具有较高的抗感染性; 而对于高度值区域稳态感染率出现的波动现象, 分析其产生的原因是高度值节点数目的稀少性.

图 8 是不同感染因子下各度值节点的辐射强度情况. 从图中可看出辐射强度最强区出现在最低度值区和最高度值区, 最低度值区呈现出强大的辐射力度的主要原因是存在着大量的低度值节点, 特别是度值为 1 和 2 的节点; 而最高度值区之所以也存在着较强的辐射力度是因为无标度网络中存在着少量的度值相对很高的节点, 这些节点存在着大量的边与周围节点相连. 我们还对网络达到稳态感染率下, 各核的感染节点比例以及初始感染率对稳态感染率的影响做了研究, 通过分析发现低核区节点的稳态感染率随着核数的增加出现了明显的上升趋势, 高核区稳态感染率也出现了波动现象; 病毒传播的稳态感染率与初始感染率是无关的.

5.2.3 病毒传播瞬态特性分析

研究病毒传播的瞬态特性, 最为关键的是研究病毒传播达到稳态时所需要的时间. 因为病毒传播的时间越长, 就能为人们提供更加充分的时间来研究抑制病毒传播的方案和软件^[22,26]. 下面的实验是针对的 AS 级 Internet 网络上的病毒传播瞬态特性的研究, 主要包含初始感染比例的不同对病毒传播达到稳态所需时间的影响.

图 9 为不同初始感染率下病毒传播的瞬态特性, 传播参数设置为: 感染因子 $\alpha=0.1$, 恢复率 $\delta=0.1$; 图 9(a) 为在不同的初始感染率下节点感染比例随时间的变化情况. 从图中可发现随着初始感染比例的增加, 节点感染比例变化曲线近似于向左进行了平移, 也就是说病毒传播达到稳态所需要的时间在缩短, 并且病毒传播在传播一段时间后都以极迅速地速度达到稳态. 这说明了初始感染率越高, 病毒的感染速度也就越快. 此外, 图中不同初始感染率下病毒达到了相同的稳态感染也进一步说明了稳态感染率与初始感染率的无关性. 为了更细致的分析不同初始感染率下病毒传播速度, 对不同初始感染率下感染速率和感染节点的平均度随时间的变化情况进行了分析, 如图 9(b) 和 (c). 其中感染速率的定义如下:

$$V(t) = \frac{di(t)}{dt} \approx i(t) - i(t-1), \quad (22)$$

假定初始时刻感染速率为 0. $V(t)$ 表征了病毒传播过程中的瞬时感染速度. 图 9(b) 是不同初始感染率下病毒节点感染速率随时间的变化情况. 从图中可以看出, 在不同的初始感染率下病毒传播的感染速率随时间都经历了先增大后减小的过程. 对于这一点, 分析其原因是: 在病毒爆发的初期, 处于感染状态的节点比例较低, 感染节点会以较大的概率在互不干扰的情况下将病毒传播给邻居节点, 再加上互联网本身的异配性特征, 小度值节点倾向于同大度值节点相连, 从而加速了病毒在大度值节点中的传播, 因而感染速率会随时间迅速增大; 而随着病毒传播的进一步加深, 处于已感染状态节点比例的增加, 病毒传播的相关性增强, 病毒开始在小度值节点中传播, 从而导致了病毒传播感染速率的下降, 并最终稳定在 0 值附近 (达到稳态), 注意此时感染速率是波动的, 但波动幅度很小. 同时, 初始感染率越高, 感染速率峰值出现的越早、越高, 这是因为初始感染率越大, 在下一时刻感染大度值节点的概率就越大, 感染节点的数目也越多, 因而下一时刻达到的节点感染率就越大. 图 9(c) 是不同初始感染率下网络中感染节点的平均度随时间的变化情况. 从病毒传播过程中感染节点平均度值的变化情况来看, 病毒在互联网中传播可分两个阶段, 开始时会在较短的时间内感染度值大的节点 (对应于网络中感染

图 9 (网络版彩图) 不同初始感染率下病毒传播的瞬态特征

Figure 9 (Color online) Transient characteristics of virus spreading under different initial infection

节点的平均度值较高, 在图中表现为峰值的出现), 而后会感染度值较小的节点 (在图中表现为感染节点平均度值从峰值处下降), 当网络中病毒传播达到稳态时, 网络中感染节点平均度几乎保持不变, 但要高于网络整体平均度值. 对于达到稳态时为什么网络中感染节点平均度要高于网络整体平均度可以从各度值节点的稳态感染率分布情况得到解释, 因为达到稳态时网络中的感染节点集合包含了大多数的大度值节点和部分小度值节点. 不同初始感染率下感染节点平均度峰值的高低差异则受该时刻感染节点的总度值和节点数目的影响, 峰值出现时间的差异则受传播速度的影响.

6 结束语

本文对互联网 AS 级拓扑的度相关特征进行了深入的分析. 通过度相关特征分析验证了 AS 级拓扑的异配性, 通过度相关特征的时间演化分析发现 AS 级拓扑的异配性呈现出弱化的趋势.

通过 DPR 算法构造具有连续匹配系数的网络, 仿真研究了在不同匹配系数的情况下度相关性对病毒传播的影响. 本文发现相对于同配网络, 异配网络中病毒传播的速度更快, 这是因为在异配网络中大度值节点更加倾向于同小度值节点相连, 这样的结构加速了病毒在大度值节点和小度值节点间的传播速度, 因而从总体上加速了病毒的传播速度. 在较低的有效传播率下, 病毒都无法传播开来, 但随着有效传播率的增加, 病毒将传播开并保持在一定的感染比例, 同时, 当有效传播率增大到一定的值后, 网络匹配系数同病毒传播的稳态感染率之间的关系保持一致, 即网络匹配系数越大, 病毒传播的稳态感染率越低. 在进行大量实验的基础上, 还发现该规律在无标度网络上要比在随机网络和小世界网络

上更为明显, 这表明: 度相关性对病毒传播的影响在无标度网络上更加显著. 异配型网络中尽管病毒传播的稳态感染率较高, 但同时也具有较高的传播临界值; 在异配型网络中, 当病毒传播有效传播率较低时, 病毒无法大规模传播; 而在同配型网络中尽管稳态感染率较低, 但传播临界值也相对较低, 不需要较高的有效传播率, 病毒就会爆发.

在考虑网络中不同个体差异性的情况下, 基于传统 SIS 传播模型建立了依赖于互联网宏观拓扑特征的病毒传播模型, 更好的刻画了实际网络中不同个体具有不同抗感染性这一事实. 在验证了上述模型的正确性和可行性的基础上, 在 CAIDA 提供的 AS 级互联网拓扑上进行病毒传播实验, 重点从病毒传播的稳态特性、瞬态特性进行了分析. 对实验结果分析发现, 初始感染率以及初始感染节点类型尽管不影响病毒传播的稳态感染率, 但对于病毒的传播速度有较大的影响.

如果能控制好这些因素, 那么将会为研究抑制病毒传播的方法提供的更多的指导. 同时本文还发现加强节点的抗感染性, 特别是加强重要节点的抗感染性和免疫保护可以十分有效地抑制病毒的传播.

参考文献

- 1 Chen G R, Wang X F, Li X. Complex Network Theory and Application. Beijing: Tsinghua University Press, 2006. 3-4 [陈关荣, 汪小帆, 李翔. 复杂网络理论及其应用. 北京: 清华大学出版社, 2006. 3-4]
- 2 Zhang G Q, Zhang G Q. Research on local clustering of the Internet AS-level topology. Complex Syst Complex Sci, 2006, 3: 34-41 [张国强, 张国清. 互联网 AS 级拓扑的局部聚团现象研究. 复杂系统与复杂性科学, 2006, 3: 34-41]
- 3 Oliveira R, Beichuan Z, Lixia Z. Observing the evolution of Internet AS topology. Comput Commun Rev, 2007, 37: 313-324
- 4 Grabowski A, Kosinski R A. The SIS model of epidemic spreading in a hierarchical social network. Acta Phys Pol B, 2005, 36: 1579-1593
- 5 Anderson R M, May R M, Anderson B. Infectious Diseases of Humans: Dynamics and Control. Oxford: Oxford University Press, 1992
- 6 Pastor-Satorras R, Vespignani A. Immunization of complex networks. Phys Rev E, 2002, 65: 036104
- 7 Cohen R, Havlin S, Ben-Avraham D. Efficient immunization strategies for computer networks and populations. Phys Rev Lett, 2003, 91: 247901
- 8 Li T. Computer virus dynamic detection model based on immune. Sci Sin Inform, 2009, 39: 422-430 [李涛. 基于免疫的计算机病毒动态检测模型. 中国科学 F 辑: 信息科学, 2009, 39: 422-430]
- 9 Han X, Tan Q. Dynamical behavior of computer virus on Internet. Appl Math Comput, 2010, 217: 2520-2526
- 10 Ren J, Yang X, Yang L X, et al. A delayed computer virus propagation model and its dynamics. Chaos Soliton Fract, 2012, 45: 74-79
- 11 Dong T, Liao X, Li H. Stability and Hopf bifurcation in a computer virus model with multi state antivirus. In: Abstract and Applied Analysis. Cairo: Hindawi Publishing Corporation, 2012, 2012
- 12 Feng L, Liao X, Li H, et al. Hopf bifurcation analysis of a delayed viral infection model in computer networks. Math Comput Model, 2012, 56: 167-179
- 13 Zhu Q, Yang X, Yang L X, et al. Optimal control of computer virus under a delayed model. Appl Math Comput, 2012, 218: 11613-11619
- 14 Yao Y, Xiang W, Qu A, et al. Hopf bifurcation in an SEIDQV worm propagation model with quarantine strategy. Discrete Dyn Nat Soc, 2012, 2012
- 15 Ma J, van den Driessche P, Willeboordse F H. The importance of contact network topology for the success of vaccination strategies. J Theor Biol, 2013, 325: 12-21
- 16 Newman M E. Assortative mixing in networks. Phys Rev Lett, 2002, 89: 208701
- 17 Vázquez A, Pastor-Satorras R, Vespignani A. Large-scale topological and dynamical properties of the Internet. Phys Rev E, 2002, 65: 066130

- 18 Van Mieghem P, Wang H, Ge X, et al. Influence of assortativity and degree-preserving rewiring on the spectra of networks. *Eur Phys J B*, 2010, 76: 643–652
- 19 Fiol M A, Garriga E. Number of walks and degree powers in a graph. *Discrete Math*, 2009, 309: 2613–2614
- 20 Barabasi A L, Albert R C C H. Emergence of scaling in random networks. *Science*, 1999, 286: 509–512
- 21 Yang J, Lin H, Wu C X. Preferential spreading on scale-free networks. *Phys A*, 2010, 389: 3915–3921
- 22 Faloutsos M, Faloutsos P, Faloutsos C. On power-law relationships of the Internet topology. *ACM SIGCOMM Comput Commun Rev*, 1999, 29: 251–262
- 23 Barthélemy M, Barrat A, Pastor-Satorras R, et al. Velocity and hierarchical spread of epidemic outbreaks in scale-free networks. *Phys Rev Lett*, 2004, 92: 178701
- 24 Pastor Satorras R, Vespignani A. Epidemic dynamics and endemic states in complex networks. *Phys Rev E*, 2001, 63: 035108
- 25 Romualdo P S, Alessandro V. Epidemic spreading in scale-free networks. *Phys Rev Lett*, 2001, 86: 3200–3203
- 26 Huang W, Li C. Epidemic spreading in scale-free networks with community structure. *J Stat Mech*, 2007, 2007: P01014

Virus spreading model based on degree correlation and its analysis

XU JiuQiang*, WANG JinFa, ZHANG Jun & ZHAO Hai

College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

*E-mail: xujiuqiang@ise.neu.edu.cn

Abstract In recent years the virus spreading has produced a series threat to network security. The research shows that the Internet topology has a great relationship with the virus spreading. The degree correlation is as an important characteristic of the Internet topology, its changes means changing the Internet topology structure. According to analyzing the degree characteristic correlated, we find that the Internet disassortativity is weaken. The experiments about virus spreading are done on the networks what are produced by DPR algorithm under different associativity coefficient aiming to study the spreading rate, steady infection rate and spreading critical value; Then a virus spreading model suited for the Internet is proposed based on the traditional virus spreading model SIS (susceptible-infected-susceptible). We do the experiments about the virus spreading and analyze the characteristic of virus spreading under steady and transient. Finally the protection policy of the virus spreading is discussed according to the simulation experiments results.

Keywords Internet topology, degree correlation, virus spreading, DPR, SIS-DVDI

XU JiuQiang was born in 1966. He received Ph.D. degree in Computer Architecture from the Northeastern University, Shenyang city, in 2000. Currently, he is a Professor of Computer System Institute, Northeastern University. His research interests include computer architecture, computer network and complex network.

ZHANG Jun was born in 1967. She received Ph.D. degree in Computer Architecture from the Northeastern University, Shenyang city, in 2009. Currently, she is a Lecturer of Computer System Institute, Northeastern University. Her research interests include Internet topology structure, computer network and complex network.

WANG JinFa was born in 1988. He received M.S. Degree in network engineering from LuDong University, Yantai, in 2011. Currently he is pursuing his Ph.D. Degree in Computer Application Technology from Northeastern University. His main research interests include Internet topology structure and complex network.

ZHAO Hai was born in 1959. He received Ph.D. degree in Computer Architecture from the Northeastern University, Shenyang city, in 1995. Currently, he is a director of the Embedded Technology Laboratory and a Professor of Computer System Institute, Northeastern University. His research interests include Internet topology structure, software network, wireless sensor network, information/data integration, ubiquitous computing, image processing, embedded technology.