

时间连续马尔可夫链的复杂网络上 SIRS 模型分析

陈旭辉^{1,2} 李 尘² 柯 铭² 郝泽龙¹

(厦门理工学院计算机科学与技术系 厦门 361024)¹ (兰州理工大学计算机与通信学院 兰州 730050)²

摘 要 针对传播过程中普遍存在的随机波动特点,以均匀网络上的基本 SIRS 模型为研究对象,建立基于时间连续马尔可夫链的随机网络模型,以平稳分布为研究方法分析了模型的稳态阈值和临界条件,发现所得结果和采用平均场方法所得结果相同;而基于时间连续马尔可夫链建立的传播模型,在对传播过程中存在的随机波动现象的描述方面,给出了较充分的理论解释,这也是概率统计方法在解决此类问题上较平均场方法最明显的优势所在,同时也为分析复杂网络上的传播动力学行为提供了一种基于概率统计方法的思路。

关键词 随机波动,时间连续马尔可夫链,复杂网络,SIRS 模型,稳态分布

中图分类号 TP391 文献标识码 A DOI 10.11896/j.issn.1002-137X.2014.10.027

SIRS Model on Complex Network of Continuous Time Markov Chain Based on Analysis

CHEN Xu-hui^{1,2} LI Chen² KE Ming² HAO Ze-long¹

(Department of Computer Science & Technology, Xiamen Institute of Technology, Xiamen 361024, China)¹

(School of Computer & Communication, Lanzhou University of Technology, Lanzhou 730050, China)²

Abstract Addressing at the general characteristics of random fluctuations in the propagation process, by uniforming network SIRS model for research object, the paper established a random network model based on continuous time Markov chain, and it analyzed the steady-state threshold and critical conditions of random network model. The conclusion of random network model is the same with the result of the mean-field approach. In addition, Propagation model was established based on continuous time Markov chain, in the description of the phenomenon of random fluctuations in the propagation process, the theory explained were given better than the mean-field approach, which is the most obvious advantage compared with the mean-field method in resolving such problems. As well as the paper provided an analysis of the behavior of the transmission dynamics of complex networks of ideas based on probability and statistics methods.

Keywords Random fluctuation, Continuous time markov chain, Complex-network, SIRS propagation model, Steady-state distribution

1 引言

随着复杂网络理论研究的进一步发展,网络上的疾病传播问题已逐渐成为复杂网络传播动力学研究领域中的一个重要的组成部分^[1,2]。现有对传播动力学研究的方法大都是基于平均场理论及其优化以后的方法^[3,4],其主要理论思想是用微分方程组来刻画整个传播过程;然而从微观角度来看,现实中绝大多数的传播现象都具有随机波动特性,即便传播系统达到平衡状态,其传播曲线仍以某一稳态值作随机的起伏,并非是绝对意义上的定值。通过平均场理论建立的传播微分方程组虽然能够较好地解释传播过程,但对于传播过程中存在的随机扰动现象却难以做进一步的描述。对于实体的疾病的爆发和流行问题,概率论与数理统计方法在这方面有着较好

的应用,这给学者们提供了一种思路和借鉴,他们开始尝试将概率论与数理统计方法同复杂系统的传播动力学问题相结合,并取得了不错的研究结果,但这还只是初步研究,有待更进一步深入和广泛的分析和探讨。因此,本文以复杂网络上的 S-I-R-S 模型为对象,研究概率论与数理统计方法应用于这类模型上的疾病传播行为。

2 复杂网络传播行为中的随机波动现象

通常根据“平均场方法”来分析基本的 S-I-R-S 模型,其具体建模步骤和讨论过程如下^[3-5]:

为了便于后续进行数学建模,在讨论均匀网络时,进行如下的假定和近似处理:一方面,对于给定的网络中任意一个节点 i ,忽略其自身节点的度值和网络本身度值的差异性,即近

到稿日期:2013-05-20 返修日期:2013-08-15 本文受国家自然科学基金项目:癫痫患者认知行为的复杂网络及动力学特性研究(61065007),厦门科技计划项目:可穿戴式人体生物信息感知系统及设备研发(3502Z20133033)资助。

陈旭辉(1972—),男,博士生,教授,主要研究方向为复杂网络、智能信息处理,E-mail: xhchen84@hotmail.com;李 尘 硕士生,主要研究方向为智能信息处理;柯 铭 女,博士生,副教授,主要研究方向为复杂网络、计算机图形图像处理;郝泽龙 男,硕士生,主要研究方向为复杂网络、智能电器。

似认为该节点的度值 k 和网络本身的平均度 $\langle k \rangle$ 相等;另一方面,根据网络中的节点在传播行为过程中所处的不同“地位”,将全部节点归为 3 个类型,即处于“易感的状态”、处于“被感染的状态”以及处于“免疫的状态”,为了便于叙述,以 S 、 I 和 R 分别来表示处于这几种状态的节点集合。 S - I - R - S 模型的 3 类节点状态转化过程如图 1 所示。

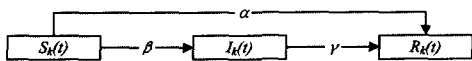


图 1 3 类节点状态之间的转换

其对应的传播流程如图 2 所示。

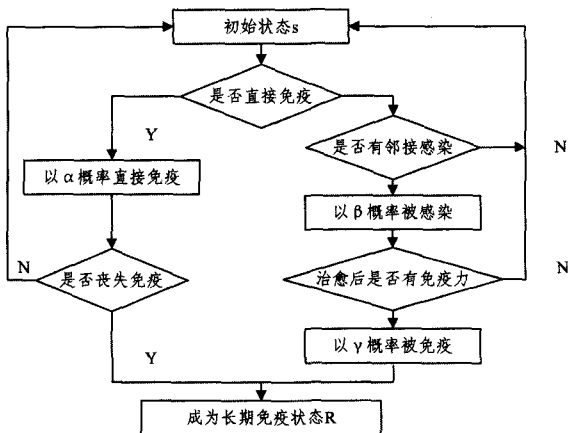


图 2 SIRS 模型的传播流程

由上面的传播流程,依据平均场方法容易得到 S - I - R - S 模型的传播演化方程组为:

$$\begin{cases} \frac{d}{dt}S(t) = \alpha R(t) - \beta \langle k \rangle S(t) I(t) \\ \frac{d}{dt}I(t) = \beta \langle k \rangle S(t) I(t) - \gamma I(t) \\ \frac{d}{dt}R(t) = \gamma I(t) - \alpha R(t) \end{cases} \quad (1)$$

且满足归一化条件 $S(t) + I(t) + R(t) = 1$ 。

求解方程组(1)的定态解,得其两个平衡点 M_1 和 M_2 ,分别为:

$$\begin{cases} M_1 = (1, 0) \\ M_2 = \left(\frac{\gamma}{\beta \langle k \rangle}, \frac{\alpha(\beta \langle k \rangle - \gamma)}{\beta \langle k \rangle (\gamma + \alpha)} \right) \end{cases} \quad (2)$$

设传播阈值 $\beta_0 = \gamma / \langle k \rangle$,当 $\beta < \beta_0$ 时,微分方程组(1)表征的系统稳态解处于无病平衡点;当 $\beta > \beta_0$ 时,方程组(1)表征的系统稳态解处于地方病平衡点。一旦网络中的免疫概率 γ 给定,则传播阈值仅与网络的平均度 $\langle k \rangle$ 有关。

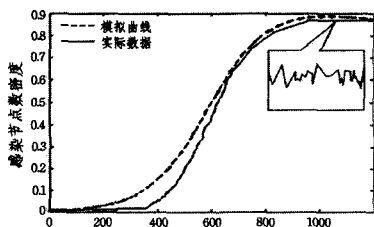


图 3 SIRS 模型的稳态传播曲线

虽然通过平均场方法建立起了 S - I - R - S 模型的传播微分

方程组,并且通过求解系统的平衡点和稳态解得出了稳态传播阈值以及稳态参数指标;但是,通过平均场方法得到的稳态解是对各个状态下对应的节点集合的宏观表征,不含关于时间 t 的参数;而复杂网络上的任何一个传播过程都存在着节点和边之间的随机化连结,这种随机化的概率特性决定了稳态时的各类节点状态密度不可能是定值,应该是一个关于时间 t 的函数^[6-8];而依据平均场方法,达到稳态后各个状态变量随时间 t 的动态变化却无法进一步描述,从图 3 可以看出,稳态后,从宏观上讲系统实际上是处于一种动态平衡^[9],虽然在任意时刻,各个状态节点的数值并非定值,具有随机波动特征^[10-12]。因此要进一步了解稳态后各个状态变量之间的随机扰动关系,仅通过平均场方法还难以做到,本文考虑从随机过程和概率统计的方法着手进行分析。

3 基于时间连续马尔可夫链的复杂网络传播模型分析

3.1 马尔可夫过程的一般定义

为了更好地说明如何利用概率统计方法研究稳态时的随机波动现象,这里先给出马尔可夫过程的一般定义^[9]。

定义 1 已知一个随机过程 $\{X(t), t \in T\}$, 设 n 为随机任意选取的一个正整数,时间参数集合符合 $t_1 < t_2, \dots, < t_n$, $P\{X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}\} > 0$, 且其条件分布满足如下关系:

$$\begin{aligned} P\{X(t_n) \leq x_n | X(t_1) = x_1, \dots, X(t_{n-1}) = x_{n-1}\} = \\ P\{X(t_n) \leq x_n | X(t_{n-1}) = x_{n-1}\} \end{aligned} \quad (3)$$

则称 $\{X(t), t \in T\}$ 为马尔可夫过程。

满足式(3)的随机过程就具有马尔可夫性(也称作无后效性),其实际物理意义表示为:设“系统将来处于某一状态”为一个事件,该事件的发生与否仅仅与目前的状态有关,系统对过去发生的事件不具有“记忆性”。换言之,给定一个系统在 t 时刻所处的状态,设该系统“在 t 时刻以后到达某状态”为事件 A ,又设其“在该时刻之前到达某状态”为事件 B ,则事件 A 和事件 B 相互独立发生。

上面介绍的是马尔可夫过程的一般定义,在实际应用中,通常根据其参数集 (T) 元素和相空间 (I) 元素是连续取值还是离散取值,可将其具体地分为 3 种类型:

- (1)若满足参数集元素和相空间元素的取值都离散,那么就称作“马尔可夫链”。
- (2)若满足参数集元素取值连续而相空间元素取值离散,那么就称作“参数集连续的马尔可夫链”。
- (3)满足参数集元素和相空间元素的取值均连续的“马尔可夫过程”。

3.2 复杂网络上的马尔可夫模型

根据图论的数学思想,可以用若干节点和边来构造和表示任意的一个网络且不论这个网络有多复杂,即任意一个网络可以看作是由节点和边相互连接而成的图。网络中的初始

条件设置为:假定最初网络中节点的状态转换是从 S 状态(即易感状态)节点开始,从网络被计算机病毒入侵开始,节点所处的状态关系按照图 2 所示流程进行转换。

根据定义 1,显然系统达到稳态后,处于各类状态的节点集合之间的随机波动现象,符合马尔可夫过程的一般定义。因为对于一个给定的传播系统,其网络节点规模为 N ,存在感染概率 β 、免疫概率 γ ,在任意时刻 t ,易感状态节点 $S(t)$,感染状态节点 $I(t)$ 以及免疫状态节点 $R(t)$ 均为随机的正整数;另一方面,将来某一时刻 t' 处的感染节点密度仅与当前时刻 t 处的感染节点密度有关,而与之前的感染节点密度无关,显然这样的特点满足马尔可夫性(即无后效性),又由于时间参数集 $t \in T$ 取值连续,状态空间集(即相空间集) $X(t) \in I$ 取值离散,因此其属于典型的时间连续的马尔可夫链。

定义 2 考虑状态集取非负整数的时间连续随机过程 $\{X(t), t \geq 0\}$ 。

设 $\{X(t), t \geq 0\}$ 为随机过程,状态空间集记为 $I = \{i_n, n \geq 0\}$,若对于任意的 $0 \leq t_1 < t_2 \dots < t_{n+1}$ 及 $i_1, i_2, \dots, i_{n+1} \in I$,有

$$P\{X(t_{n+1})=i_{n+1} | X(t_1)=i_1, X(t_2)=i_2, \dots, X(t_n)=i_n\} = P\{X(t_{n+1})=i_{n+1} | X(t_n)=i_n\} \quad (4)$$

则称 $\{X(t), t \geq 0\}$ 为时间连续的马尔可夫链。

因此根据定义 2,任意时刻 t 的感染状态节点密度 $I(t)$ 可以表征为一个时间连续的马尔可夫链。

设概率 $P\{I(t)=i\}=p_i(t)$ 表示在标记时刻 t ,有 i 个节点被感染的概率,经过一个足够小的时间间隔 Δt 之后, $t+\Delta t$ 时

$$P_I = \begin{bmatrix} p_{00} & p_{01} & \dots & \dots & p_{0N} \\ p_{10} & p_{11} & \dots & \dots & p_{1N} \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ p_{(N-1)0} & p_{(N-1)1} & \dots & \dots & p_{(N-1)(N-1)} \\ p_{N0} & p_{N1} & \dots & \dots & p_{NN} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & \dots & \dots & 0 \\ \mu_1 \Delta t & 1 - (\mu_1 + \lambda_1) \Delta t & \lambda_1 \Delta t & \ddots & 0 & \vdots \\ 0 & \mu_2 \Delta t & 1 - (\mu_2 + \lambda_2) \Delta t & \lambda_2 \Delta t & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & \ddots & \mu_{N-1} \Delta t & 1 - (\mu_{N-1} + \lambda_{N-1}) \Delta t & \lambda_{N-1} \Delta t \\ 0 & \dots & \dots & 0 & \mu_N \Delta t & 1 - \mu_N \Delta t \end{bmatrix} \quad (8)$$

设

$$p(t) = \{p_0(t), p_1(t), \dots, p_N(t)\}$$

则可得

$$p_i(t) P_I = p_i(t + \Delta t)$$

所以转移概率 $p_i(t)$ 和 $p_i(t + \Delta t)$ 之间满足如下的差分关系:

$$p_i(t + \Delta t) = p_{i-1}(t) \lambda_{i-1} \Delta t + p_i(t) (1 - \mu_i \Delta t - \lambda_i \Delta t) + p_{i+1}(t) \mu_{i+1} \Delta t$$

且初始概率分布 $p_i(0) = 0$; 讨论两个极限状态的情况,即 $i = 0, i = N$, 可得:

$$\begin{aligned} p_0(t + \Delta t) &= p_0(t) + p_1(t) \mu_1 \Delta t \\ p_N(t + \Delta t) &= p_{N-1}(t) \lambda_{N-1} \Delta t + p_N(t) (1 - \mu_N \Delta t) \end{aligned} \quad (9)$$

根据时间参数连续的马尔可夫链的特点可知, $i = 0$ 是矩阵 P_I 的吸收状态,在网络系统随着时间参量进行演化的过程中,系统中处于其它状态的节点演化到吸收状态后,就达到稳定。因此当 $t \rightarrow \infty$ 时绝对概率分布 $p_0(t) = 1$ 。由此可得时间连续的马尔可夫链的稳态分布应为 $p = (1, 0, 0, \dots, 0)_N$ 。这表明经过多步状态演化后,网络以 1 概率收敛,此时与网络中的参数设置及所处的传播状态均无关。

刻网络中易感节点 S 被感染的概率为

$$\frac{\beta_i}{\gamma_i} \Delta t = \lambda_i \Delta t = \frac{N-i}{N} i \beta \langle k \rangle \Delta t \quad (5)$$

而该时刻,感染节点被免疫的概率为 $\gamma_i \Delta t$, 丧失免疫的概率为 $\alpha \Delta t$ 。根据定义的足够小 Δt 内各个状态的转移概率,可以表示时间参数连续的马尔可夫链在此条件下的转移概率 P 为:

$$P\{I(t + \Delta t) = i + 1 | I(t) = i\} = P_{i, i+1}(\Delta t) = \frac{N-i}{N} i \beta \langle k \rangle \Delta t = \lambda_i \Delta t$$

$$P\{I(t + \Delta t) = i - 1 | I(t) = i\} = P_{i, i-1}(\Delta t) = \gamma_i \Delta t = \mu_i \Delta t \quad (6)$$

$$P\{I(t + \Delta t) = i | I(t) = i\} = P_{i, i}(\Delta t) = 1 - [P_{i, i+1}(\Delta t) + P_{i, i-1}(\Delta t)] = 1 - (\lambda_i \Delta t + \mu_i \Delta t)$$

当系统开始处于传播过程时,一般取 $0 < i < N$, 在这期间任意时刻的状态转移概率均可由式(6)表征;而对于初始状态和极限状态,分别令 $i = 0$ 和 $i = N$, 这两种状态需要单作说明,根据时间连续的马尔可夫链表述定义,可得:

$$P\{I(t + \Delta t) = 0 | I(t) = 0\} = P_{0,0}(\Delta t) = 1$$

$$P\{I(t + \Delta t) = N - 1 | I(t) = N\} = P_{N, N-1}(\Delta t) = \gamma_N \Delta t = \mu_N \Delta t \quad (7)$$

$$P\{I(t + \Delta t) = 1 | I(t) = 0\} = P_{0,1}(\Delta t) = 1 - \mu_N \Delta t$$

对于式(6)和式(7), λ_i 和 μ_i 均小于 1 且时间间隔 Δt 取值足够小,则可满足 $(\lambda_i + \mu_i) \Delta t \ll 1$ 。根据以上表达关系,不难得出式(8)所表示的感染状态 $I(t)$ 的状态转移概率矩阵,记其为 P_I 。

3.3 时间连续马尔可夫链的稳态分布

根据上面分析可知,当系统中节点处于 0 状态时,网络以 1 概率收敛,即达到稳态。所以对于时间连续的马尔可夫链,令 $q(t) = p_i(t) / (1 - p_0(t))$, 代入式(9)中可得

$$q_i(t + \Delta t) (1 - q_i(t) \gamma \Delta t) = q_{i-1}(t) \lambda_{i-1} \Delta t + q_i(t) (1 - \mu_i \Delta t - \lambda_i \Delta t) + q_{i+1}(t) \mu_{i+1} \Delta t \quad (10)$$

式(10)中,状态集合 $I \in \{1, 2, 3, \dots, N\}$, 且 $\sum_i q(t) = 1$ 。 $\{q_1, q_2, q_3, \dots, q_N\}$ 是方程(10)的平稳态解。这里,虽然可以通过多次迭代和归一化条件解得平稳态解的解析形式,但由于迭代步数过多以及实际分析中网络节点规模数 N 值较大,导致运算量太大,因此这里有必要对约束条件进行简化,以减少运算量。

下面进行简化处理,其理论依据是随机过程中的正则性原理,即任意一个随机过程刚进入某一状态时不可能立即又跳跃到另一状态,这正好说明了一个实际存在的物理系统,要在极其有限的时间内发生接近无限多次的状态跳转,从而消耗无穷多能量的情况是不可能发生的。所以,这里假定有:

$$P_{0,0}(\Delta t) = P\{I(t + \Delta t) = 0 | I(t) = 0\} = 0 \quad (11)$$

再将式(11)代入式(8)中的状态转移概率矩阵 P_I , 记新

的状态转移矩阵为 P_I' , 即

$$P_I' = \begin{bmatrix} p_{00} & p_{01} & \cdots & \cdots & p_{0N} \\ p_{10} & p_{11} & & & p_{1N} \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ p_{(N-1)0} & p_{(N-1)1} & \cdots & \cdots & p_{(N-1)(N-1)} \\ p_{N0} & p_{N1} & \cdots & \cdots & p_{NN} \end{bmatrix} = \begin{bmatrix} 0 & 0 & \cdots & \cdots & \cdots & 0 \\ \mu_1 \Delta t & 1 - (\mu_1 \Delta t + \lambda_1 \Delta t) & \lambda_1 \Delta t & \ddots & & \vdots \\ 0 & \mu_2 \Delta t & 1 - (\mu_2 \Delta t + \lambda_2 \Delta t) & \lambda_2 \Delta t & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & 0 & & \ddots & \mu_{N-1} \Delta t & 1 - (\mu_{N-1} \Delta t + \lambda_{N-1} \Delta t) & \lambda_{N-1} \Delta t \\ 0 & \cdots & \cdots & \cdots & 0 & \mu_N \Delta t & 1 - \mu_N \Delta t \end{bmatrix} \quad (12)$$

显然, 通过前面的假定, 矩阵 P_I' 和式(8)中的矩阵 P_I 相比, 其第一行元素全为零, 第一列元素中仅有一个非零元素。为了能对矩阵 P_I' 进行降阶, 这里略去第一行和第一列, 并假定原来矩阵 P_I 中的元素 $p_{10} = 0$, 这样得到的转移概率矩阵

P_I' 比 P_I 阶数要低, 同时也得到了一个和原来近似的平稳分布关系。

通过简化和降阶, 所得的新转移概率矩阵 P_I' 可表征为:

$$P_I' = \begin{bmatrix} 1 - \lambda_1 \Delta t & \lambda_1 \Delta t & 0 & \cdots & \cdots & 0 \\ \mu_2 \Delta t & 1 - (\mu_2 \Delta t + \lambda_2 \Delta t) & \lambda_2 \Delta t & 0 & \cdots & 0 \\ 0 & \mu_3 \Delta t & 1 - (\mu_3 \Delta t + \lambda_3 \Delta t) & \lambda_3 \Delta t & \ddots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mu_{N-1} \Delta t & 1 - (\mu_{N-1} \Delta t + \lambda_{N-1} \Delta t) & \lambda_{N-1} \Delta t \\ 0 & \cdots & \cdots & 0 & \mu_N \Delta t & 1 - \mu_N \Delta t \end{bmatrix} \quad (13)$$

设式(13)转移概率矩阵 P_I' 的平稳态解为 $q' = (q_1', q_2', q_3', \dots, q_N')$, 按照对 P_I 的平稳态解的分析方法和步骤, 可得其平稳分布表达式为:

之间的关系, 不难得出随机变量 τ_i 的分布函数 $F_i(\tau_i)$ 服从 $[0, 1]$ 区间上的均匀分布; 当随机过程达到 τ_i 以后, 模型中处于感染状态的节点开始从原来状态向另一状态转移, 这里对模型中的节点转移个数做了这样的限定: 一步转移后只能从 i 跳到 $i+1$ 或 $i-1$, 不发生多个状态节点的转移, 且各个节点之间的状态转移相互独立发生; 对于初始情况 $i=1$, 规定模型中感染节点密度至少不减少; 对于极限情况 $i=N$, 规定模型中感染节点密度至少不增加。那么, 任意时刻模型中每增加一个感染状态节点的概率应为:

$$q_n' = \frac{1}{n} q_1' \prod_{k=2}^{n-1} \frac{\lambda_k}{\mu_k} \quad (14)$$

式中, 满足 n 取 $2, 3, \dots, N$, 归一化条件为 $q_1' + q_2' + q_3' + \dots + q_N' = 1$ 。

3.4 时间连续马尔可夫链在复杂网络上的模拟

前面分析了时间连续马尔可夫链的稳态分布规律, 这里考虑对传播过程的仿真模拟。根据随机过程理论可知, 时间连续马尔可夫链的状态转移矩阵 P 所对应的 Q 矩阵是由各个状态之间的转移速率构成的。又由马尔可夫性质可知, 在满足 s 时刻处于 i 状态条件下的随机过程, 其在时间序列 $[s, s+t]$ 中保持处于 i 状态的概率与其处于 i 状态至少保持 t 个时间单位的概率恰好相等, 将该随机过程在演化到其它状态之前保持在 i 状态的时间记作 τ_i , 那么对于所有的 $s, t \geq 0$, 其满足如下关系^[9]:

$$P\{\tau_i > s+t | \tau_i > s\} = P\{\tau_i > t\}$$

这说明, 随机变量 τ_i 满足“无记忆”特性, 所以随机变量 τ_i 的概率分布函数属于指数分布。

所以设时间连续的马尔可夫链模型对应的 Q 矩阵为:

$$Q =$$

$$\begin{bmatrix} -\lambda_1 & \lambda_1 & 0 & \cdots & \cdots & 0 \\ \mu_2 & -\mu_2 - \lambda_2 & \lambda_2 & 0 & \cdots & 0 \\ 0 & \mu_3 & -\mu_3 - \lambda_3 & \lambda_3 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \mu_{N-1} & -\mu_{N-1} - \lambda_{N-1} & \lambda_{N-1} \\ 0 & \cdots & \cdots & 0 & \mu_N & -\mu_N \end{bmatrix} \quad (15)$$

假设在 t 时刻网络中的感染节点数为 $i \in [1, N]$, 那么网络中存在有 i 个感染节点的停留时间 τ_i 是一个随机变量, 其概率分布满足指数分布 $F_i(t)$ 。

又根据随机变量的概率分布和随机变量函数的概率分布

$$p_{\uparrow} = \begin{cases} 1, & i=1 \\ \frac{\lambda_i}{\mu_i + \lambda_i}, & \text{else} \\ 0, & i=N \end{cases}$$

同理, 任意时刻每减少一个感染节点的概率应为:

$$p_{\downarrow} = \begin{cases} 0, & i=1 \\ \frac{\mu_i}{\mu_i + \lambda_i}, & \text{else} \\ 1, & i=N \end{cases}$$

且概率之间满足归一化条件: $p_{\uparrow} + p_{\downarrow} = 1$ 。这样就得出任意时刻下, 各个状态节点其状态发生跳转的概率。

为了便于对比平均场方法和马尔可夫过程方法下模型的传播曲线特点, 对所有的网络参数设置做归一化标准。归一化后网络节点密度总数为 1, 网络平均度 $\langle k \rangle = 8$, 初始时刻 $t=0$ 时, $S(0) = 0.90, I(0) = 0.05, R(0) = 0.05, \beta = 0.1, \gamma = 0.2, \alpha = 0.1$ 。这里给出的仿真实例背景来自于 CAIDA (the Co-operative Association for Internet Data Analysis) 提供的红虫病毒采集信息。仿真结果如图 4 所示。

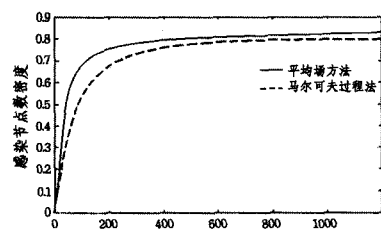


图 4 传播曲线

从图4中可以看出,时间连续的马尔可夫链方法和平均场方法在从开始传播一直到网络达到平稳态的整个过程中传播曲线拟合得比较好,但又略有差异性。根据前面一节随机化模型的建立过程可知,两者的曲线差异性主要源自以下两方面因素:

一方面,在对时间连续马尔可夫链的平稳态分析中,为了减少运算量,对其状态转移概率矩阵 P_t 做了近似简化和降阶,这样得到的平稳态解实际上是近似解;

另一方面,虽然模型是依据时间连续的马尔可夫链建立的,但是无论随机模型中引入的充分小状态跳转时间 Δt 的取值有多短,它仍然是一个离散参数,不可能遍历整个时间数轴,这和平均场法有所差别。

3.5 传播过程的稳态阈值和临界条件

目前已有的结论指出了基于平均场方法下的复杂网络传播动力行为中稳态条件的存在,并通过理论计算得出了具体的表达式;与这类似,下面考察时间连续的马尔可夫链下模型的传播阈值。因为根据概率统计方法可知,随机过程的数学期望是对若干个具体样本观测的平均值,因此可以利用随机变量的均值来得出随机变量函数的均值,从而分析随机过程的稳态条件。

设随机变量 $p(t)$ 和 $q(t)$ 的数学期望分别为:

$$Ep(t) = \sum_{i=0}^N i_i p(t)$$

$$Eq(t) = \sum_{i=0}^N i_i q(t)$$

为了研究在微小状态跳转时间 Δt 内,随机变量的数学期望随时间变化的趋势,令 $Ep(t)$ 和 $Eq(t)$ 分别关于时间间隔 Δt 前向作差,可得:

$$\begin{aligned} Ep(t+\Delta t) - Ep(t) &= \sum_{i=0}^N (\lambda_i \Delta t p_i(t) - \gamma_i \Delta t p_i(t)) \\ &= \gamma \Delta t [\lambda \langle k \rangle \xi(t) - 1] Ep(t) \\ Eq(t+\Delta t) [1 - \gamma q_1(t) \Delta t] - Eq(t) &= \gamma \Delta t [\lambda \langle k \rangle \eta(t) - 1] Eq(t) \end{aligned} \quad (16)$$

式(16)中, $\lambda_i = \frac{\beta_i}{\gamma_i}$, $\xi(t) = \frac{\sum_{i=0}^N \lambda_i p_i(t)}{\beta \langle k \rangle Ep(t)}$, $\eta(t) = \frac{\sum_{i=1}^N \lambda_i q_i(t)}{\beta \langle k \rangle Eq(t)}$, 且满足对于所有的 t 都有 $\xi(t), \eta(t) \in (0, 1)$ 。

所以,当且仅当式(16)中,因子 $\lambda \langle k \rangle \xi(t) - 1 > 0$ 时,代入可得随机变量 $p(t)$ 数学期望的前向差分 $Ep(t+\Delta t) - Ep(t) > 0$, 网络中传播行为发生,此时对应 $\lambda \langle k \rangle > 1$, 即 $\lambda_c > 1/\langle k \rangle$, 说明满足此条件下的网络传播过程持续进行; 而当 $\lambda \langle k \rangle \xi(t) - 1 < 0$ 时,同理可得随机变量 $p(t)$ 数学期望的前向差分 $Ep(t+\Delta t) - Ep(t) < 0$, 对应 $\lambda_c < 1/\langle k \rangle$, 网络中感染状态节点的增长率趋于零,增长曲线趋于平缓并随时间保持相对稳定的数值。显然,分析时间连续马尔可夫链所得到的传播过程稳态临界条件为: $\lambda_c = 1/\langle k \rangle$, 这与通过平均场方法得到的结论相同。

综上所述,通过考察基于时间连续马尔可夫链下随机模型的稳态传播条件,发现其对传播阈值和临界条件的描述与平均场方法所得结果类似;另一方面,依据随机过程建立的网络模型从概率统计学的角度出发,对传播过程中存在的微观随机扰动现象给出了一定程度的分析和解释,从模型的微观

角度方面讲,这一点显然要优于平均场方法所建立的传播模型。

结束语 本文通过阐述平均场方法在描述传播过程微观粒度方面上的不足,并针对一般传播过程中普遍存在的微观随机扰动现象,首先分析了随机扰动现象状态跳跃的条件和特点,即未来某一时刻的感染节点密度仅与当前时刻的感染节点密度有关,而与之前的感染密度无关,这样的特点符合随机过程中对于马尔可夫过程的一般定义;然后再具体地利用基于时间连续的马尔可夫链来对复杂网络上的传播过程进行建模和模拟,并与平均场方法进行比较,发现由时间连续马尔可夫链得到的传播特性曲线和平均场方法所得的传播特性曲线相似;同时,考察显示了两种方法对于稳态阈值和临界条件的描述是一致的;而基于时间连续马尔可夫链建立的传播模型在对于传播过程中存在的随机波动现象的描述方面,给出了较好的理论说明,这也是概率统计方法较平均场方法最明显的优势所在。

需要指出的是,文章具体针对复杂网络上一种典型传播过程中存在的随机波动现象做了分析和研究,而其余的几种传播过程中存在的这一现象是否也有和本文类似的结果,还需要做进一步的研究。

参考文献

- [1] Masuda N, Konno N. Multi-state epidemic process on complex networks [J]. Journal of Theoretical Biology, 2006, 243(1): 64-75
- [2] Kitsak M, Gallos L K, Havlin S, et al. Identification of influential spreaders in complex networks [J]. Nature Physics, 2010, 6: 888-893
- [3] Zhou Tao, Fu Zhong-qian, Wang Bing-hong. Epidemic Dynamics on Complex Networks [J]. Progress in Natural Science, 2006, 16(5): 452-456
- [4] Perdisci R, Lanzi A, Lee W. Classification of packed executables for accurate computer virus detection [J]. Pattern Recognition Letters, 2008, 29(14): 1941-1946
- [5] Gómez S, Gómez-Gardeñes J, Moreno Y, et al. Nonperturbative heterogeneous mean-field approach to epidemic spreading in complex networks [J]. Physical Review E, 2011, 84(3): 253-260
- [6] 赵敬, 夏承遗, 孙世温, 等. 复杂网络上同时考虑感染延迟和非均匀传播的 SIR 模型 [J]. 智能系统学报, 2013, 2(8): 128-134
- [7] 肖丽, 包骏, 冯丽萍. 一种新的计算机病毒模型的稳定性分析 [J]. 湘潭大学自然科学学报, 2012, 34(3): 94-96
- [8] 苑卫国, 刘云, 程军军, 等. 微博双向“关注”网络节点中心性及传播影响力的分析 [J]. 物理学报, 2013, 62(3): 038901-1-10
- [9] 杨慧, 唐明, 许伯铭. 自适应网络中的流行病传播动力学研究综述 [J]. 复杂系统与复杂性科学, 2011, 9(4): 63-83
- [10] 冯丽萍, 王鸿斌, 冯素琴. 改进的 SIR 计算机病毒传播模型 [J]. 计算机应用, 2011, 31(7): 1891-1893
- [11] 葛新, 赵海, 张君. 网络度相关及其传播特征研究 [J]. 计算机研究与发展, 2013, 4(50): 741-749
- [12] 范永青, 王银河, 王青云, 等. 具有相似节点的耦合时滞复杂网络的稳定性与同步控制分析 [J]. 控制与决策, 2013, 2(28): 247-252