

基于拓扑路径的网络演化传播机制研究

张林姿 贾传亮

(中央财经大学管理科学与工程学院 北京 100081)

摘要 现有的社会网络信息传播模型主要分析传播的途径,将传播过程与节点的度相结合,而传播媒介常常被忽略。在现实世界的网络中,传播源作为一个物理传播媒介通常由特定的路径从一个节点传播到另一个节点(基于路径的传播)。本研究不再局限于节点的总体行为分析,而是分别考虑每个节点的状态转换,用连续状态的马尔科夫链分析来模拟传播源和路径对传播行为的影响。该方法通过引入平均场近似,将基于路径的传播机制的计算复杂度从指数级别降低到多项式级别;定义了同时包含路由选择和交通信息的传播特性矩阵,并得出了基于路径传播的关键传播阈值。当有效传播率低于关键传播阈值时,传播就会逐渐消亡,因此可以运用该关键传播阈值来促进或抑制基于路径的传播。最后,除了随机无标度网络,引入了现实世界网络交通作为研究案例来对比基于连接和基于路径的传播行为,结论表明所提模型在社交网络中的传播具有高度持续性和极强的稳定性。

关键词 社会网络,信息传播,马尔科夫理论,路由路径,平均场理论

中图分类号 TP393 文献标识码 A

Study of Propagation Mechanism in Networks Based on Topological Path

ZHANG Lin-zi JIA Chuan-liang

(School of Management Science and Engineering, Central University of Finance and Economics, Beijing 100081, China)

Abstract The existing information dissemination model of social network mainly analyzes the ways of dissemination, and combines the process of communication with the degree of nodes. However, the media is often ignored. In real-world networks, the propagation source, a physical propagation medium, usually propagates from one node to another via a specific path. This paper was no longer limited to analyze the overall behavior of nodes, but considered each node separately, and used the continuous Markov chain to simulate the influence of propagation sources and paths on propagation. By introducing a mean field approximation, the computational complexity of the path-based propagation is reduced from an exponential level to a polynomial level. This paper also defined a propagation characteristic matrix containing both routing and traffic information, and derived a key propagation threshold based on path propagation. When the effective transmission rate is below the threshold, the propagation will gradually die out, so we can use this critical propagation threshold to promote or suppress path-based propagation. Finally, in addition to the stochastic scale-free network, this paper introduced the real-world network traffic as a research case to compare the connection-based and path-based propagation behaviors. The conclusions show that the model's propagation in social networks is highly persistent and extremely stable.

Keywords Social networks, Information dissemination, Markov theory, Routing paths, Mean field theory

1 研究背景及意义

1.1 研究背景

自 1990 年互联网被发明以来,其呈现与日俱进的趋势,现已在人们的日常生活中扮演着必不可少的重要角色。信息传播方式也从传统的报纸、书刊演变为在线网络的传播,人们的社交渠道也在逐渐从线下向线上延展。全球性社交网络代理营销机构 WeAreSocial 的报告^[1]指出,2015 年的全球范围互联网用户为 34.2 亿,在社交网络媒体上活跃的用户规模已经达到 23.1 亿,与 2014 年同期相比增长了 10%,而这个变化在移动端(手机)方面更为显著,其社交媒体用户已达到 20 亿,较 2014 年同比增长 17%。全球的网民规模与日俱增,手

机网民在其中起着不可忽略的重要作用,而中国的社交网络发展也呈现出极为相似的趋势。中国互联网络信息中心(CNNIC)于 2017 年 1 月发布的报告^[2]显示,“到 2016 年 12 月为止,我国互联网的普及率已经达到 53.2%,网民规模为 7.31 亿,其中移动端(手机)网民规模依然庞大,占到了总体网民比例的 95.1%,为 6.95 亿”。

网络惠及了人们日常衣食住行的各方面,从学习到娱乐,从吃饭到打车,无一不充斥着互联网的身影。其中最为显著的还是以微信、微博为首的社交网络的发展,如今人们仅需连通网络,便可以和不同城市甚至不同国家的朋友、家人进行交流,分享自己生活的点滴,十分便捷。CNNIC 的报告显示,2016 年网民最常用的 APP 中,微信、QQ 列居首位,直观地反

映了互联网对人们社交方式的改变。同时,根据 2016 年 11 月发布的《2015 微博用户发展报告》^[3]，“截至 2015 年 9 月 30 日,微博的每月活跃的用户数(MAU)已经达到了 2.22 亿,较 2014 年同期增长 33%,微博已成为网络信息传播中不可忽视的重要媒介”。

1.2 研究意义

1.2.1 增强网络信息传播的可控性

不可否认,高速发展的互联网带来了不可计数的快捷方便之处,但另一方面,其两面性也决定了必然会产生的与其优点相对应的一系列缺陷。较传统的传播方式,互联网传播具有速度快、强度高、范围广的特点,因此,谣言、邪教、黄赌毒等不良信息一旦通过互联网蔓延开来,将造成难以估计的巨大影响^[4]。若要更好地实现对社会网络中传播的管理和掌控,就必须仔细对其传播机制进行研究,了解传播运行机理,才能使网络管理者按照自己的意愿采取措施来促进或抑制特定信息传播的发生。

1.2.2 改进传统网络传播模型

传统的网络传播模型大多假设传播是基于连接发生的,只要两节点相连,它们之间就会存在一个固定的感染率。也就是说,每个被感染节点都会不断感染与它最近的邻接节点,尽管并不是所有节点都会一直保持在活跃状态(传感器节点常常停留在“睡眠”模式来节省电量)。这种无方向的基于连接的传播过程无法解释传播动力学中的两个现象:1)在许多现实世界网络中,信息传播都遵循特定的路径;2)基于连接的传播过程没有考虑到传播源需要物理接触来感染节点的事实,它默认传播在节点之间没有互动的情况下也会发生^[5]。因此,本文研究的基于路径的传播模型与现实网络更为符合,更具代表性。

受以上观察的启发,本文将引用最前沿的疾病传播理论来模拟基于路径的信息传播,并解释由不同路由协议所构建的路径和传播源的“传播携带者”身份所带来的方向效应。首先,本文将网络拓扑结构、路由协议和交通分布作为输入变量来对传播源进行建模。然后,本文将基于最具代表性的疾病传播模型之一(SIS 模型),引用连续状态的马尔科夫链分析来模拟基于路径的传播机制,并进一步采用平均场近似将总体的分析复杂度从指数级别降为多项式级别,从而研究出更贴合现实世界中网络系统的传播模型,并提供给网络管理者以方便其对网络传播根据意愿进行调节。

2 基本概念与分析框架

2.1 状态的定义

疾病传播模型将一个“疾病”分解为不同的状态(阶段),网络中的每一个节点在任意给定时间均处于其中一个状态上。应用最广泛的两个模型为 SIS 模型和 SIR 模型,其对可能状态的定义如下:

- (1)易感态(S):健康的未被感染的但易于被感染的个体。
- (2)感染态(I):被感染的并具有传染性的个体。
- (3)隔离态(R):免疫后的既不会被感染也不会感染其他节点的个体。

研究文献中还存在一些其他的状态,如 SEIS/SEIR 模型中添加了个体被感染却不具有感染性的状态^[6]、SAIS 模型中添加了因个体提高警惕而被感染几率降低的状态等。

2.2 传播机制

传播机制描述了网络信息是如何从一个节点传播到另一个节点的,本质上是个体状态的转换。这样的状态转换常常和有效传播率 $\tau = \beta/\delta$ 联系在一起,其中 β 为感染率(有时也称转换率), δ 为治愈率。在本文的假设下,有效传播率还受系统中交通状况的影响。早期的研究多认为节点是基于质量作用定律均匀混合的,其中个体有相等的概率与被感染个体接触,而近来的研究开始关注不均匀的情况^[7]。

本文主要研究通过路径传递信息数据的社交网络,认为传播必须由传播源执行,并明确按照传播动力学的规范发生。在此条件下,处于易感态的节点会有不同的感染几率,其被感染的潜在几率是由其在系统中的活跃程度决定的。一个节点的被感染几率与以它为终点和经由它的传播源的总量成正比。图 1 简单地对比了基于连接和基于路径的传播过程。在基于连接的传播模型(见图 1(a))中,节点 0 是最初被感染的节点,只有节点 1,2,3,4 会立即处于易感状态,而其他节点在之后出现了被感染的邻接节点的情况下也会处于易感态。与之对比,假定节点 0 和 6 之间存在互动,受系统中信息交流的影响,基于路径的传播模型(见图 1(b))就会有不同的易感节点集合(节点 3,5 和 6)。在这种情况下,节点 1,2 和 4 虽然和被感染节点直接相连,却处于安全状态。

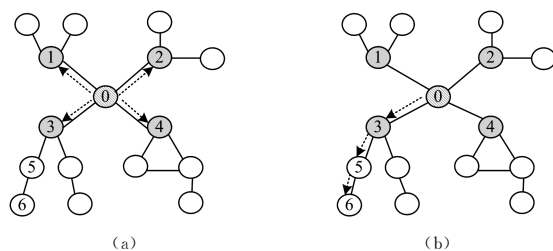


图 1 基于路径传播的模型马尔科夫链状态图解

3 模型与方法

3.1 基于路径的传播模型

3.1.1 一般基于路径的传播过程

以有 3 个节点的线图为例,在图 2 中引用 SI, SIS 和 SIR 模型的马尔科夫链图解来说明基于路径和基于连接的传播过程的区别。

在 $N=3$ 时,SI 模型(见图 2(a))的马尔科夫链有 $2^3=8$ 种状态,SIS 模型(见图 2(b))有 $2^3=8$ 种状态,SIR 模型(见图 2(c))有 $3^3=27$ 种状态。尽管不同节点可选的状态的数目保持一致,但它们彼此的状态的转换过程是不同的。由于任意两节点间一般只有一条有效路径,图 2 对任意路由协议均具有代表性。一次转换中包括多个节点的状态转换是可能的(如状态 SSI 到状态 III),而这是传统的基于连接的传播所不允许的。注意,由于数据包的流动,在线图中从状态 SIS 到状态 III 的直接转换是不可能的,因此传播是具有方向性的。一个处于感染态的位于中间的节点只能在任意时间感染它左边或右边的节点。然而,如果我们考虑同样规模的环状拓扑结构,且路由协议选择最长的路径来传递数据包,从状态 SIS 到 III 的转换就是可能的(如图 2 中的虚线箭头)。因此,可以得出,尽管拓扑结构会影响传播过程,但真正的传播动力学是由路由协议所决定的。

3.1.3 基于路径传播的 SIS 模型

$X_n(t)$ 表示节点 n 在时间点 t 的状态,在 SIS 模型中 $X_n(t)$ 只能是感染态或易感态。本文进一步用 $i_n(t) = \Pr[X_n(t)=1]$ 表示节点 n 在时间点 t 处于感染态的概率,其中 1 表示感染态,0 表示易感态。因此,一个节点处在健康状态的概率为 $s_n(t) = \Pr[X_n(t)=0] = 1 - i_n(t)$ 。应用马尔科夫理论,这个两状态马尔科夫链的无穷小生成元 $Q_n(t)$ 可以被表示成如下形式:

$$Q_n(t) = \begin{bmatrix} -q_{1,n} & q_{1,n} \\ q_{2,n} & -q_{2,n} \end{bmatrix} \quad (9)$$

由于状态转换中的治愈率与其他节点的状态无关,因此有 $q_{2,n} = \delta$ (具体在 3.3 节讨论)。

另一方面, $q_{1,n}$ 是一个依赖于网络中其他节点活动的随机变量,为了继续马尔科夫分析, $q_{1,n}$ 的随机性必须被消除,其中一个可行方法是使 $q_{1,n}$ 遍历所有节点 X_n ($1 \leq n \leq N$) 的所有可能状态的组合,然而这却导致了马尔科夫链结果的指数级复杂度。

本文沿用 Miegheem^[12] 研究的多项式简化法,即用平均场近似为随机感染率 $q_{1,n}$ 寻找一个有效替代率。假定感染率 β 和治愈率 δ 均为常数,则期望感染率可通过对所有状态取平均得出:

$$\begin{aligned} E[q_{1,n}] &= \beta \sum_{m=1}^N \Pr[X_m(t)=1] c_{n,m} = \beta \sum_{m=1}^N i_m(t) c_{n,m} \text{ or} \\ &= \beta \sum_{m=1}^N i_m(t) \lambda_m b_{n,m} \end{aligned} \quad (10)$$

其中, $c_{n,m}$ 是感染特性矩阵 C 的元素。因此,可以得出有效无穷小生成元的形式:

$$\overline{Q_n(t)} = \begin{bmatrix} -E[q_{1,n}] & E[q_{1,n}] \\ \delta & -\delta \end{bmatrix} \quad (11)$$

经过以上步骤,本文可用 $\overline{Q_n(t)}$ 继续马尔科夫分析,将其引入马尔科夫微分方程,当 $X_n(t)=1$ 时可以得出如下非线性微分方程:

$$\frac{di_n(t)}{dt} = \beta \sum_{m=1}^N c_{n,m} i_m(t) - i_n(t) (\beta \sum_{m=1}^N c_{n,m} i_m(t) + \delta) \quad (12)$$

或

$$\frac{di_n(t)}{dt} = \beta \sum_{m=1}^N \lambda_m b_{n,m} i_m(t) - i_n(t) (\beta \sum_{m=1}^N \lambda_m b_{n,m} i_m(t) + \delta) \quad (13)$$

基于网络交通分布信息,这个非线性的微分方程可通过其矩阵形式求解:

$$\frac{dI(t)}{dt} = \beta CI(t) - \text{diag}(i_n(t)) (\beta CI(t) + \delta u) \quad (14)$$

其中, u 是所有项均为 1 的向量, $\text{diag}(i_n(t))$ 是以 $i_1(t), i_2(t), \dots, i_N(t)$ 为主对角线元素的对角矩阵。

将 $I(t) = \text{diag}(i_n(t)u)$ 代入式(14)可得:

$$\frac{dI(t)}{dt} = (\beta \text{diag}(1 - i_n(t))C - \delta \Pi) I(t) \quad (15)$$

$$= (\beta \text{diag}(1 - i_n(t)) \text{diag}(\Lambda)B - \delta \Pi) I(t) \quad (16)$$

其中, Π 是 $N \times N$ 阶单位矩阵。

进而,系统中感染态节点的瞬时比例可表示为:

$$\rho(t) = \frac{1}{N} \sum_{n=1}^N i_n(t) \quad (17)$$

在稳态情况下, $\frac{di_n(t)}{dt} |_{t \rightarrow \infty} = 0$, 令 $i_{n\infty} = \lim_{t \rightarrow \infty} i_n(t)$, 由

式(12)可得:

$$i_{n\infty} = \frac{\beta \sum_{m=1}^N c_{n,m} i_{m\infty}}{\beta \sum_{m=1}^N c_{n,m} i_{m\infty} + \delta} = 1 - \frac{1}{1 + \tau \sum_{m=1}^N c_{n,m} i_{m\infty}} \quad (18)$$

其中, $\tau = \frac{\beta}{\delta}$, 根据式(18), $i_{n\infty} = 0$ 是一个平凡解,这显然可以

由观察有限且拥有所有节点可达的吸收率(即所有节点处在健康态)的马尔科夫链得出。然而,对任何一个规模为 N 的现实网络,这个稳态只能在相当长的一段时间之后达到。同时,系统会呈指数型收敛并长时间停留在一个亚稳态(式(18)除过平凡解的另一个正解),而 Miegheem^[14] 指出这个亚稳态更接近于现实世界中的传播。因此,本文将重点研究这个亚稳态并将其视作稳态。

$i_n(t)$ 和 $i_{n\infty}$ 的可计算性使本文可以接着对系统中个体节点的易感性进行研究,而不是仅仅观察整体网络。

通过式(17)、式(18),被感染节点的稳态比率可表示为:

$$\rho_{\infty} = \frac{1}{N} \sum_{n=1}^N i_{n\infty} \quad (19)$$

文献[9]的研究发现纯随即图,如具有二度分布无标度网络图及呈幂律分布的节点度的 ER 模型,其呈现极为不同的传播行为。本文将对 ER 模型进行蒙特卡罗(Monte-Carlo)仿真,以展示分析框架的预测能力。本文对之前提到的每个模型图生成一个样本图集合,并在假定非加权连接条件下,用 Dijkstra 算法预先算出每个图中的路由路径及每个节点对,当 $t=0$ 时有 10 个节点被随机选择设定为感染态,其他所有节点均假设为健康态。在每个单位时间内,有 λN 个数据包生成,对于每个新生成的数据包,会有特定的源-目的地节点对与之相对,然后数据包按照这对节点间预先计算好的路径传递,因此每个节点的感染和治愈过程将遵循第 3.2 节中的说明。

图 3 和图 4 给出了 $N=100$ 的交通均匀分布的网络中感染比率瞬时演变的两个具有代表性的结果。

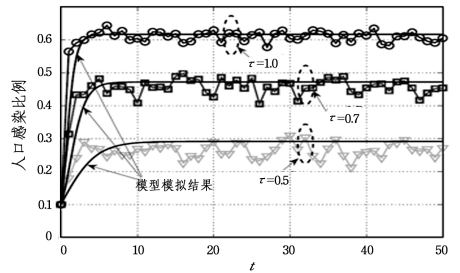


图 3 随机图的感染比率瞬时演变

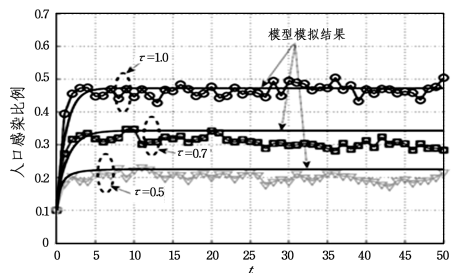


图 4 无标度图的感染比率瞬时演变

两个图分别考虑 τ 为 0.5, 0.7, 1.0 时的 3 种情况, 其中平滑的曲线为由式(15)计算得出的结果, 带有记号的折线为 Monte-Carlo 仿真的结果。可以发现, 两种情况下的演变极为相似, 感染比率会趋于稳定在一个由 τ 决定的水平上。图 5 进一步展示了由模型和仿真得到的不同规模 ($N = \{100, 200, 300, 400, 500\}$) 的样本网络的稳态。

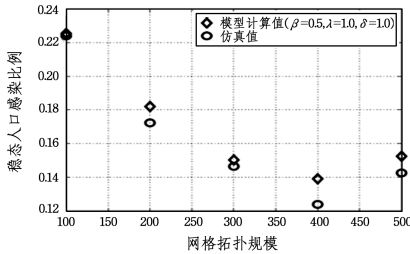


图 5 不同规模样本网络的稳态感染比率

3.2 基于路径传播的阈值

第 3.1 节已经简单提到, 先前的研究^[12]已经发现了理论上的关键阈值 τ_c , 并认为有效传播率低于 τ_c 时传播就会逐渐消亡。然而, 近期的一些研究^[11]对现实世界中阈值的存在性表示了质疑。

为了研究传播是否会消亡, 本文将沿用与 Mieghe^[12] 相似的方法, 考虑所有节点处于感染态的概率 $i_n(t)$ 接近于 0 的时间, 因为只有当 $i_n = 0, \forall n \in N$ (所有节点都处于健康态, 即吸收态) 时, 网络中才会不存在感染。在此条件下, 忽略非线性项, 式(15)可以改写为:

$$\frac{dI(t)}{dt} = (\beta C - \delta II) I(t) \quad (20)$$

解式(20), 可以得到 $I(t)$ 关于时间 t 的函数:

$$I(t) = e^{Ht} I(0) \quad (21)$$

其中, $H = \beta C - \delta II = \beta \text{diag}(\Lambda) B - \delta II$ 。通过特征分解:

$$H = U M^C U^T \quad (22)$$

其中, $M^C = \text{diag}(\mu_n^C)$ 是以矩阵 C 的第 n 个特征值 μ_n^C 为第 n 个对主角线元素 $M_{n,n}^C$ 的对角矩阵, U 是以与特征值 μ_n^C 相对的特征向量为第 n 列得到的标准正交矩阵, 因此有:

$$H = \beta C - \delta II = U(\beta M^C - \delta II) U^T = U \text{diag}(\beta \mu_n^C - \delta) U^T \quad (23)$$

将式(23)代入式(21), 可以得到:

$$I(t) = U \text{diag}(e^{(\beta \mu_n^C - \delta)t}) U^T I(0) \quad (24)$$

由于 $I(t)$ 是一个概率向量, 所有的特征值都必须满足 $\beta \mu_n^C - \delta \leq 0$ 。进而可计算出传播阈值的表达式为:

$$\tau_c = \frac{\beta}{\delta} \leq \frac{1}{\mu_{\max}^C} \quad (25)$$

也就是说, 当 τ 等于或小于矩阵 C 的谱半径 μ_{\max}^C (见 3.3 节) 的倒数时, 传播会以指数型的速度衰退至 0。因此, C 描述了基于路径传播的传播强度。

图 6 给出了 5 个不同规模 ($N = \{100, 200, 300, 400, 500\}$) 的随机图的稳态被感染人口比例随 τ 的变化情况, 而这些随机图的矩阵 C 的谱半径由图 6 右下方的小图给出。当 $\tau = 0.5$ 时, 所有节点向其他所有节点发送相等的交通量。从图中可以看出, 不论何时, 当 $\tau < \tau_c$ 时, 传播即是无法残存的; 而当 $\tau \geq \tau_c$ 时, 有效传播率 τ 的增加会导致稳态被感染人口比例的单调递增。本文强调这个关键阈值只定义了传播的存活和消亡的转折点, 并没有涉及在阈值之上传播到达亚稳态的

规模 (即 $\tau_c(C) > \tau_c(C')$ 并不意味着 $\rho_{\infty}(C) > \rho_{\infty}(C')$, 反之亦然)。

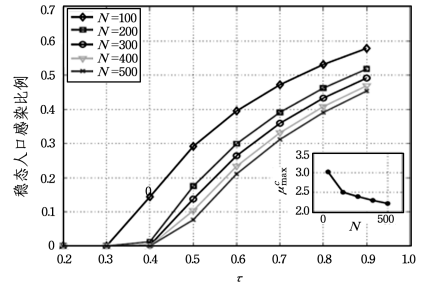


图 6 不同规模随机图的传播亚稳态规模

由式(25)可以看出, 关键阈值和系统的初始状态相互独立。而图 7 给出了 $N=100$ 时具有不同初始被感染节点数量 ($\sum i_n(0) = \{10, 30, 50, 70, 90\}$) 的网络的瞬时演变, 进一步展示了无论初始状态如何, 系统总会收敛到一样的亚稳态 (即稳态被感染比例 ρ_{∞} 也和初始状态相互独立)。

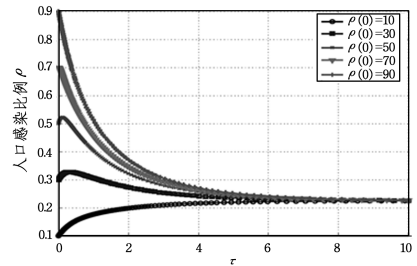


图 7 不同初始状态网络的感染比率瞬时演变

对于相关主特征向量 z_{\max}^C , 其与 μ_{\max}^C 相对应, 定义如下:

$$\mu_{\max}^C (z_{\max}^C)_n = \sum_{m=1}^N c_{n,m} (z_{\max}^C)_m \quad (26)$$

其中, $(z_{\max}^C)_n$ 是主特征向量 z_{\max}^C 的第 n 项。观察式(26)可得出, $(z_{\max}^C)_n$ 直接与节点 n 在交通传输中的参与程度成比例, 因为 $c_{n,m}$ 是由矩阵 R 和 Γ 计算得出的。

3.3 基于路径和基于连接的传播动力学对比

本文通过对比相关网络的 2^N 态马尔科夫链来比较两种传播方式, 并用与 Mieghe^[12] 研究相似的方法采用二元表述描述 2^N 态马尔科夫链, 来对比两种传播方式的无穷小生成元 Q^{path} 和 Q^{contact} 。简单来说, 网络的每个状态都用一个 n 位二元字符串来表示, 第 n 位字符就表示网络中的第 n 个节点, “1”表示节点处于感染态, “0”表示易感态。

在 SIS 模型中, 网络的整体状态是由感染和治愈两个过程所决定的, 由于这两个过程相互独立, 因此可以分别对其进行研究。为了简化起见, 本文将矩阵 Q 分解成 3 个 $N \times N$ 矩阵之和, 即 $Q = Q_{L\Delta} + Q_{U\Delta} + Q_{\text{diag}}$, 其中 $Q_{L\Delta}$ 和 $Q_{U\Delta}$ 分别是 Q 的下三角和上三角部分, 而 Q_{diag} 取了 Q 的主对角线元素。由于状态空间的二元表述, $Q_{L\Delta}$ 和 $Q_{U\Delta}$ 分别表示治愈 (与 δ 相关) 和感染 (与 β 相关) 的状态转换。介于基于连接和基于路径传播之间的治愈过程是一致的并与感染过程相独立, 有 $Q_{L\Delta}^{\text{path}} = Q_{L\Delta}^{\text{contact}}$, 并且它们不受矩阵 A, R 和 Γ 的影响。Mieghe^[12] 指出当 $\beta = 0$ 时, 矩阵 Q 的 C_N^2 重特征根 $\mu^Q = -k\delta$ 同样适用于基于路径的传播。

然而, 矩阵 Q 的上三角部分却不相同 (即 $Q_{U\Delta}^{\text{path}} \neq Q_{U\Delta}^{\text{contact}}$), $Q_{U\Delta}^{\text{contact}}$ 是由矩阵 A 决定的, 而 $Q_{U\Delta}^{\text{path}}$ 却是基于矩阵 R 和 Γ 计算

得出的。正如 3.1 节中提到的,基于路径的传播中会有更多可能的状态转换,因此 Q_{UD}^{path} 与 $Q_{UD}^{contact}$ 相比会有更多的非零项,从而 Q^{path} 比 $Q^{contact}$ 更为密集。然而,因为 $\det(Q)$ 依然是 0,所以两个矩阵 Q 的最大特征值也均为 0。

进一步将基于路径传播中的式(15)和式(16)与相对应的基于连接的模型^[12]对比,可以发现两套公式具有相似的形式。基于路径传播的模型最显著的不同便是邻接矩阵 A 不再是直接影响因素,取而代之的是特性矩阵 C , C 同时考虑了交通强度 Γ 和路由协议 R ,而事实上 R 又受拓扑结构 A 的连通性的影响。

根据定义,矩阵 A 和 C 是非负矩阵。一个连通图的矩阵 A 的最大特征值被限制在顶点的平均度和最大度之间,当连通图是有 N 个顶点的完全图时,其有最大值 $N-1$,因此式(1)可以应用于基于连接的传播,但对于基于路径的传播却不适用。

由于矩阵 C 不再是一定对称的,其谱半径也许是个复数。对于这样的矩阵来说,最大特征值在数值上是个实数,且其实部是在所有特征值中最大的,即 C 的谱半径 $\mu_{\max}^C = R(\mu_{\max}^C)$ 。而形如 C 的矩阵的谱半径已由如下不等式限定:

$$\min_{1 \leq n \leq N} \left(\sum_{m=1}^N |c_{n,m}| \right) \leq \mu_{\max}^C \leq \max_{1 \leq n \leq N} \left(\sum_{m=1}^N |c_{n,m}| \right) \quad (27)$$

定理 1 (τ_c 的一般性边界) 给定一个由矩阵 C 描述的基于路径的传播,它的关键阈值受以下不等式约束:

$$\frac{1}{\max_{1 \leq n \leq N} \left(\sum_{m=1}^N |c_{n,m}| \right)} \leq \tau_c \leq \frac{1}{\min_{1 \leq n \leq N} \left(\sum_{m=1}^N |c_{n,m}| \right)} \quad (28)$$

证明:该式可直接由式(27)和式(29)得出。

定理 1 的物理意义是 τ_c 的上界(下界)由网络中传输交通量最小(最大)的节点决定。这个定理的一个应用实例是,通过传播相关的约束,将不同交通工程最优化问题公式化现已成为可能,因此传播是否会发生就取决于矩阵 C 的无穷范数 $\|C\|_{\infty}$ 。

3.4 路由协议、交通及网络拓扑结构的效应

本节将研究网络拓扑结构(A)、路由协议(R)及系统中的交通运载量(Γ)对基于路径的传播阈值(τ_c)的影响。首先确定 τ_c 的单调性。

定理 2 (τ_c 的单调性) 给定任意两个感染特性矩阵 C 和 C' ,若 $0 \leq C \leq C'$ (定义 $C \leq C'$, 若有 $c_{n,m} \leq c'_{n,m} : \forall n, m$), 则 $\tau_c(C) \geq \tau_c(C')$ 。

证明:因为矩阵 C 和 C' 是非负方阵,其遵循 Perron-Frobenius 定理,即 $\mu_{\max}^C \leq \mu_{\max}^{C'}$,若 $0 \leq C \leq C'$ 。因为传播阈值与谱半径成反比,所以有:

$$\tau_c(C) = \frac{1}{\mu_{\max}^C} \geq \tau_c(C') = \frac{1}{\mu_{\max}^{C'}} \quad (29)$$

当矩阵 C 为素矩阵时,单调是严格的。

定理 2 指出矩阵 C 的每次正向变动都会对 τ_c 有相反的影响(即任意项 $c_{n,m}$ 的增加都会导致 τ_c 的单调下降,反之亦然)。因此,在网络拓扑结构和路由协议固定不变的情况下,更高的交通强度总会导致更低的传播阈值,进而促进并维持传播的进行,相反,在更低的交通强度下传播更容易消亡。图 8 给出了稳态感染比率与交通量的正向关系,其系统总运量的提高是通过增加 λ 来体现的。注意, ρ_{∞} 的增加与 λ 的变化并不是线性的关系。图 8 中的左上角的小图给出了不同 λ_n

随着 n 的变化所对应的 C 的谱半径。

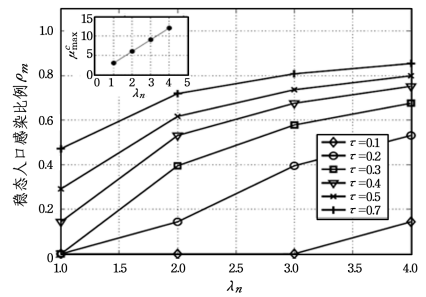


图 8 稳态感染比率与交通量的正向关系

然而,网络中的交通量通常是不可控制的(最好的情况下,也只能是通过交通整顿或政策实现部分可控),因为它依赖于使用者的行为及其他许多方面的因素。相反,路由协议由网络经营者所设定,因此为可控的,本文现转而研究路由协议 R 的效应。首先,本文通过与顶点度无关的非均匀连接加权构建出具有无序的边的 ER 图集(加权网络方案)。然后,固定图集中每个图的度分布(即矩阵 A 保持不变),同时改变其无序制度(通过改变连接权重分布),从而得到同一个图所对应的不同矩阵 R 及节点的交通传输参与度 C 。

特别地,本文附加考虑非负独立同分布(i, i, d)的连接权重,在同一个图的无序状况控制不变的情况下,对其构建出不同的交通分布。为此,本文沿用 Miegheem 等^[14]应用多项式表示连接权重分布的方法。

$$F_w(x) = x^\alpha 1_{x \in [0,1]} + 1_{x \in (1,+\infty)}, \alpha > 0 \quad (30)$$

其中,若 x 为真, $1_x = 1$, 否则 $1_x = 0$ 。当 $\alpha \rightarrow 0$ 时,信息传递路径主要由组成连接中的最高连接权重所决定,这与强无序的限制一致;在这个无序体系中,每两个节点间的路径都由沿路径的最大连接权重所描述,因此最短路径就指两节点间最短的具有最大连接权重的路径。而构建一个弱无序约束,只需将其改为统一的单位连接权重(即非加权网络)。在弱无序系统中,一条路径上的大部分或全部连接共同决定了两节点间的最短路径。改变连接权重分布在本质上会导致同一个图产生不同的最短路径集合,从而得到不同的无界限定,因此本文就可以独立于度分布来研究路径的影响。

图 9 给出了同一个网络的强无序和弱无序体系(通过改变节点间信息传递的路径来实现)中被感染比例随时间的变化情况,其中强无序限制下网络展现出了更高的被感染比率。由于同一个 ER 图可对应不同的被感染比率,可以得出度分布不再直接决定网络中的信息有效传播情况,这进一步证实了传播是由传递路径而非节点的度所直接影响的。

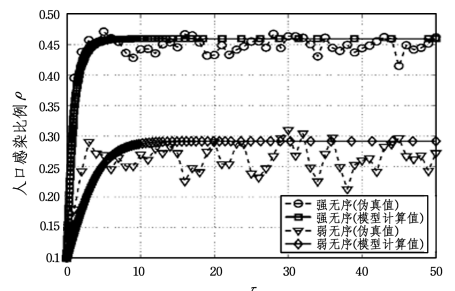


图 9 不同路径体系下的稳态感染比率

定理 3 (给定 A 和 Γ 时的最大可达 τ_c) 网络拓扑结构 A

和交通量 Γ 已知, 当 $R=R^{USPT}$ 时, 有 $\tau_c = \tau_c^{\max}$, 其中 R^{USPT} 是 A 中所有可能节点间通过跳数计数法所得的最短路径组成的路由矩阵。

证明: 令 R^* 表示任意路由协议所对应的二元路由矩阵, 已知当且仅当 $R^* = R^{USPT}$ 时 $\sum_{n=1}^N \sum_{k=1}^{N(n-1)} r_{n,k}^*$ 取得最小值, 在 Γ 固定的条件下, 应用式(8)可得出:

$$\sum_{m=1}^N C_{n,m}^{USPT} \leq \sum_{m=1}^N C_{n,m}^* : \forall n \quad (31)$$

这反映了若不使用最短路径, 节点整体传递传播源的参与便会膨胀, 传递同量的传播源便会需要更多的节点参与(即 $\rho^{USPT} \leq \rho^*$), 因此有 $\tau_c^{\max} = \tau_c^{USPT} \geq \tau_c^*$, 其中 τ_c^{USPT} 表示使用通过跳数计数法得出最短路径时系统的关键阈值。

定理3定义了一个既定运载量的网络的 τ_c 的上界, 当路由由协议仅使用跳数计数法作为度量来计算所有节点间的最短路径时达到上界。然而, 现实世界中的路由选择并不总是最小化跳数。例如, 尽管IP网络中使用的开放式最短路径优先(OSPF)域内的路由协议基于Dijkstra最短路径算法, 但它常常受使用连接权重改变结果路由的交通工程操作干预。这种干预对使用基于政策的路由协议的域内路由来说更为强烈, 这种路由协议并不试图最小化任意特定长度标准。因此, 基于定理3, 若改变路由协议使得 $R^* \rightarrow R^{USPT}$, 则降低网络中的传播诱因通常是可行的。

定理4(不同网络的 τ_c) 对规模为 N 的两个不同网络, 用 A 和 A' 表示。

$$\tau_c(A) \leq \tau_c(A') \quad (32)$$

若在同样运量下:

$$\frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{N(n-1)} r_{n,k} \geq \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{N(n-1)} r'_{n,k} \quad (33)$$

其中, $\tau_c(A)$ 表示网络 A 的关键传播阈值。

证明: 在两个网络具有相同的 Γ 的条件下, 根据式(10), τ_c 的决定因素是 R , 然而 R 的构建(即节点间的路径计算)受拓扑结构 A 的限制。由于 A 决定了可能的矩阵 R , 根据定理3, 有:

$$\mu_{\max}^C \propto \frac{1}{N} \sum_{n=1}^N \sum_{k=1}^{N(n-1)} r_{n,k} \quad (34)$$

即所有节点间的路径长度均值越大, 矩阵 C 的谱半径就越大。因此, 有更大平均路径长度的 A 将会导致更大的 C 的谱半径 μ_{\max}^C 。又因为阈值与 μ_{\max}^C 成反比, 且 τ_c 具有单调性(见定理2), 定理得证。

该定理在网络拓扑结构能够根据特定要求(如数据中心网络)或特定规则(如自组织无线传感网络)灵活构造的情形中特别有用。根据实际中的应用方案, 拓扑结构的构建可以用来促进或阻碍传播的进行。

例如, 在数据中心(DC)网络中, DC 拓扑结构的设计是一个复杂的问题, 其由多种因素限制(如资源碎片、超额利用率等)。定理4允许不同DC拓扑结构设计间根据传播容量限度进行直接比较。另一方面, 自组织传感网络常在具有高度不确定(如温度的变化、间歇连通性、故障、能量储存需求等)的不稳定环境中运行, 导致了随时间而变化的拓扑结构。在这种情况下, 一个能促进信息传播的网络通常需要保证信息的高度可获得性和持续性。

4 实验分析

本节将基于路径传播的分析框架应用于3个现实世界一级网络(即Level-3(AS1), Sprint(AS1239)和AT&T(AS7018)3家公司的人网点(POP)层面网络)中, 来研究在基于路径的传播中, 它们是如何有效运行的。表1列出了这些网络的相关特性, 并针对非加权和加权路径分别计算出了两组相关参数的值。从表1中可以看出, 基于连接的传播比基于路径的传播有着更小的关键阈值, 这意味着基于路径的传播更容易消亡。这也许具有一定的误导性: 对于一个基于连接的传播来说, 相邻节点间的传播总是可能的(即每个被感染节点都在一致感染它的相邻节点); 而基于路径的传播是依赖于传播源进行的, 因此增加系统中的交通量 Λ 将会导致 τ_c 的单调下降(见定理2), 从而使得在某一时刻基于路径的传播将有更小的 τ_c 。图10给出了这3个现实网络(加权路径条件下)的阈值和传播规模(ρ_{∞})随交通量(λ_n)和感染率(β)的变化关系。由图10可以看出, 增加系统中的交通量或感染率将会增加传播的流行度并降低关键传播阈值。

表1 现实网络样本的拓扑性质

性质	Level-3 (AS1)	Sprint (AS1239)	AT&T (AS7018)
N	42	52	113
d	2.6190	3.2380	2.6018
d_{\max}	8	14	25
平均路径长度	3.7271(非加权)	3.4721(非加权)	3.3475(非加权)
	3.9826(加权)	3.5765(加权)	3.4603(加权)
μ_{\max}^A	3.9669	5.7166	6.9066
$\tau_c^{contact}$	0.2521	0.1749	0.1448
μ_{\max}^C	3.1716(非加权)	3.0252(非加权)	2.7743(非加权)
	3.4090(加权)	3.1123(加权)	2.8676(加权)
τ_c^{path}	0.3153(非加权)	0.3306(非加权)	0.3605(非加权)
	0.2933(加权)	0.3213(加权)	0.3486(加权)

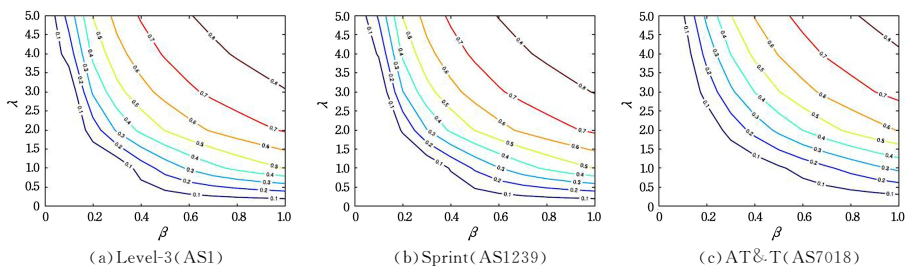


图10 被感染节点比率和关键阈值的变化

- [9] HU Y, LI X. An improvement of DV-Hop localization algorithm for wireless sensor networks[J]. *Telecommunication Systems*, 2013, 53(1):13-18.
- [10] 陶志勇, 魏强, 刘影. 多功率锚节点辅助的 DV-Hop 定位算法[J]. *计算机工程与应用*, 2014(21):121-124, 156.
- [11] 夏少波, 邹建梅, 朱晓雨, 等. 无线传感器网络 DV-Hop 定位算法的改进[J]. *计算机应用*, 2015(2):340-344.
- [12] MASS-SANCHEZ J, RUIZ-IBARRA E, CORTEZ-GONZÁLEZ J, et al. Weighted Hyperbolic DV-Hop Positioning Node Localization Algorithm in WSNs[J]. *Wireless Personal Communications*, 2017, 96(1-23):5011-5033.
- [13] NICULESCU D, NATH B. DV Based Positioning in Ad Hoc Networks[J]. *Telecommunication Systems*, 2003, 22(1-4):267-280.
- [14] NICULESCU D, NATH B. Ad hoc positioning system (APS) using AOA[C]// *Joint Conference of the IEEE Computer and Communications*. IEEE, 2003:1734-1743.
- [15] NAGPAL R, SHROBE H, BACHRACH J. Organizing a Global Coordinate System from Local Information on an Ad Hoc Sensor Network[C]// *Information Processing in Sensor Networks, Second International Workshop (IPSN 2003)*. Palo Alto, CA, USA, 2003:333-348.
- [16] 邹佳顺, 张永胜. 无线传感器网络中关于 DV-Hop 定位算法的改进[J]. *计算机应用与软件*, 2016, 33(3):147-150.
- [17] SHI W, JIA C, LIANG H. An improved DV-Hop localization algorithm for wireless sensor networks[J]. *International Journal of Online Engineering*, 2010, 9(6):2232-2236.
- [18] WU H, DENG M, XIAO L, et al. Cosine Theorem-based DV-Hop Localization Algorithm in Wireless Sensor Networks[J]. *Information Technology Journal*, 2011, 10(2):239-245.
- [19] 唐静静. 无线传感器网络非测距定位算法研究[D]. 南宁: 广西师范大学, 2016.

(上接第 314 页)

进一步观察可以发现, 在每个案例的基于连接的传播模型中, 通过加权图计算出的路线会导致网络更易受感染。然而, 在所有案例中, 尽管并不是所有方案都采用最短路径, 总体的平均路径长度仍相对接近, 且关键阈值并没有明显偏离 τ_c 的下界(见定理 3)。

结束语 许多传播过程都是沿着特定路径发生的, 且依赖于特定传播源作为物理传播媒介从一个节点传到另一个节点。本文通过考虑传播源的量和分布建立了基于路径的传播模型, 尽管建模方法具有一般性, 但本文重点研究社交网络环境, 并用数据包代替传播源。在以 Miegheem 研究^[12]为基础的建模上延伸, 本文不再考虑节点的整体行为, 而是分别对每个节点进行分析, 并使分析框架的复杂度降为多项式级别。本文引入了传播特性矩阵 C , 来表示网络交通强度及分布和路由协议计算所得路径的共同影响, 从而描述传播行为, 进而发现基于路径传播的关键阈值等于矩阵 C 的谱半径的倒数(即 $\tau_c = 1/\mu_{\max}^C$)。基于连接的传播中, 传播在很大程度上是由拓扑结构 A 决定的, 而基于路径的传播不同, 其影响因素主要与交通量和通往终点的路由路径有关, 这就为在同一个网络结构中控制传播行为提供了两个“调节钮”: 1) 通过交通工程、规范政策和进入率控制等技巧来调整交通强度; 2) 通过改变系统中的路由协议来构建不同的路径选择。基于此, 本文进一步推导出了 τ_c 的条件边界, 其受系统交通量和路由协议信息的可获得性所限制, 因此网络管理者可利用这些手段修改边界来根据需要控制传播过程。此外, 由于对每个节点分别进行考虑, 本文还可以很容易地列出网络中最容易受传播感染的节点, 该节点层面的信息可以被用作设计免疫或保护机制时的新型中心度量。本文通过推断连接权重集合和跟踪路由数据, 将该体制应用于 3 个现实网络中, 来体现其适用性。基于本文模型, 关键传播阈值随着 λ 的增加而逐渐减小, 这再次证实了 Paster 的研究^[11]中关于社交网络中的传播衰减具有极高的鲁棒性的观点。由于本文建模方法具有一般性, 因此可以很容易地延伸并用来模拟不同的传播模型。

参考文献

- [1] We Are Social. Digital in 2016 Report[R/OL]. <http://wearesocial.com/special-reports/digital-in-2016>, 2016-01-27.
- [2] 中国互联网信息中心(CNNIC). 第 39 次中国互联网络发展状况统计报告[R/OL]. <http://www.cnnic.cn/>, 2017-01-22.
- [3] 新浪微博数据中心. 2015 微博用户发展报告[R/OL]. <http://data.weibo.com/report/reportDetail?id=333>, 2016-11-10.
- [4] KEPHART J O, WHITE S R. Measuring and modeling computer virus prevalence[C]// *1993 IEEE Computer Society Symposium on Research in Security and Privacy*, 1993. Proceedings. IEEE, 2002:2.
- [5] NIKOLOSKI Z, DEO N, KUCERA L. Correlation Model of Worm Propagation on Scale-Free Networks[J]. *Complexus*, 2006, 3(1-3):169-182.
- [6] COLIZZA V, VESPIGNANI A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: theory and simulations[J]. *Journal of Theoretical Biology*, 2008, 251(3):450-467.
- [7] ONNELA J P, CHRISTAKIS N A. Spreading paths in partially observed social networks[J]. *Phys. Rev. E*, 2012, 85(3 Pt 2):036106.
- [8] MELONI S, ARENAS A, MORENO Y. Traffic-driven epidemic spreading in finite-size scale-free networks[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, 106(40):16897-16902.
- [9] YANG H X, WU Z X, WANG B H. Suppressing traffic-driven epidemic spreading by edge-removal strategies[J]. *Physical Review E Statistical Nonlinear & Soft Matter Physics*, 2013, 87(6):064801.
- [10] XIONG F, LIU Y, ZHANG Z J, et al. An information diffusion model based on retweeting mechanism for online social media[J]. *Physics Letters A*, 2012, 376(30/31):2103-2108.
- [11] ZHAO L, WANG Q, CHENG J, et al. Rumor spreading model with consideration of forgetting mechanism: A case of online blogging LiveJournal[J]. *Physica A Statistical Mechanics & Its Applications*, 2011, 390(13):2619-2625.
- [12] PANTAZOPOULOS P, KARALIOPOULOS M, STAVRAKAKIS I. Centrality-driven scalable service migration[C]// *International Teletraffic Congress*. International Teletraffic Congress, 2011:127-134.
- [13] MIEGHEM P V. Performance Analysis of Communications Networks and Systems[M]. Cambridge University Press, 2006.
- [14] Rocketfuel. An ISP Topology Mapping Engine[R/OL]. <http://research.cs.washington.edu/networking/rocketfuel>.