

用复杂网络理论研究疾病的传播*

孙 胜 秋

(北京交通大学 理学院,北京 100044)

摘 要 概述了用复杂网络的理论来研究两个传染病模型,即 SIR 模型和 SIS 模型,以及在这两个模型中如何计算疾病的传播阈值。由复杂网络的局部渐近性质——在图中节点的总个数趋于 ∞ 时,图具有局部树形结构,可知 SIR 模型可以用分支过程理论来研究,从而根据母函数的方法可得到一些很好的性质。而 SIS 模型可以通过平均场的理论得以理解,且知不论在度相关 Scale-Free 网络中还是度不相关 Scale-Free 网络中都不存在非零传播阈值。

关键词 复杂网络;度分布;渗流;传播阈值

中图分类号: TP311.12;R181.2⁺5

文献标识码: A

文章编号: 1672-6693(2005)04-0001-05

Study of Disease Spreading Using the Theory of Complex Networks

SUN Sheng-qiu

(School of Science, Beijing Jiao Tong University, Beijing 100044, China)

Abstract In this paper we review briefly the recent discussion in literature about the two infectious diseases models: SIR model and SIS model and the problem of how to calculate the transmission threshold in value of the two models. According to the local asymptotic properties of complex networks, when the pitch point number in the graph goes to infinity, the graph is locally shaped into a tree-like structure. SIR model can be studied using the theory of branching process. It leads to get some good properties by using the method of generating functions. SIS model can be understood by the mean field theory, and whether in degree correlated Scale-Free networks or in degree uncorrelated Scale-Free networks, there does not exist nonzero transmission threshold.

Key words complex networks; degree distribution; percolation; transmission threshold

疾病的传播可看作是一个随机复杂网络,在这里用点代表可能通过传染或感染影响疾病传播行为的个体,如果两个个体之间通过某种方式可能直接发生传染与被传染的关系,就在他们之间连一条边。

1 两个传统的传染病模型

先介绍两个传统的传染病模型,此模型被称为充分混合模型。在此模型中,假定在网络中任意两个节点之间都可能存在传播疾病的边,相当于在完全图上讨论问题^[1~8]。

1.1 SIR 模型

19 世纪 20 年代 Lowell Reed 和 Wade Hampton Frost^[1]最初构建了 SIR 模型(充分混合模型)。该

模型把人划分为 3 类:易感类(susceptible,此类人并未染病但会被传染类个体传染),传染类(infective,此类人已患有疾病并会传染给其他人),康复类(recovered,此类人已从疾病中康复过来,并具有免疫力)。SIR 模型用于研究这 3 类人的数量随着时间的变化规律。

用 $S(t)$ 、 $I(t)$ 、 $R(t)$ 表示这 3 类人在某时刻 t 的数量, $s(t) = S(t)/N$, $i(t) = I(t)/N$, $r(t) = R(t)/N$, 用 β 表示单位时间内任意易感类个体被一传染类个体传染的一致概率, γ 表示单位时间内传染类个体康复并具有免疫力的比率,则此模型可用下列微分方程来描述:

* 收稿日期: 2005-09-05

资助项目: 国家自然科学基金重点项目(No. 10531070)

作者简介: 孙胜秋(1982-),女,山东临沂人,硕士研究生,研究方向为复杂网络。

$$\frac{ds}{dt} = -\beta is, \frac{di}{dt} = \beta is - \gamma i, \frac{dr}{dt} = \gamma i$$

1.2 SIS 模型

并不是任何疾病其康复者都会对这种病产生免疫力,一些不能自我控制但能通过药物治疗的疾病,康复的人通常会被病人再次传染上。结核病和淋病是两个大量研究的例子。计算机病毒也属于这种类型(计算机可以通过杀毒软件清除其中的计算机病毒,却抵御不了病毒的再次袭击)。

所谓的 SIS 模型就是携带这种病的人治愈后不是从传染类转到康复类,而是转回到易感类。此模型可用微分方程描述:

$$\frac{ds}{dt} = -\beta is + \gamma i, \frac{di}{dt} = \beta is - \gamma i$$

在上述 2 个模型中,虽然给出了疾病传播的基本的动力系统,但在实际生活中是不现实的,因为疾病只能在有实际身体接触的人之间传播,而在上述模型中任意人都有可能被所有病人传染,亦即疾病传播网络中所有结点具有相同的度(结点的度是指它连接边的个数)。在现实世界的疾病传播中,有必要考虑度分布不均匀的情况^[2](度分布是指具有某个度的结点出现的概率分布)。近年来文献中有不少关于疾病在任意度分布网络中传播的研究^[8-12]。

笔者已经严格证明了在度序列给定的随机图中,有一个渐近局部性质:在图中节点的总个数趋于 ∞ 时,随机选择一个点 V ,距离它小于等于 m 的所有点形成的导出子图 $T_m(V)$ 无圈。在巨大分支还没出现的情况下,分支的大小都是有限的,结果表明有限分支包含圈的概率是趋于 0 的,是一个树形结构,所以可用 G-W 分支过程的理论来讨论疾病在任意度分布网络中传播的性质。本文简要介绍如何把传统的疾病传播模型推广为任意度分布网络上的传染过程。

2 现实世界中的传染病模型

2.1 任意度分布的 SIR 模型

把传统的 SIR 模型推广到任意度分布网络之上的传染过程,在数学上可以把它看作在同一个网络上的边渗流过程^[3]。在疾病传播网络中,边以一定概率 T 传播疾病相当于边渗流理论中边以一定概率 T 被占据,从而可以根据渗流模型得到有关传染病的预测情况。比如疾病爆发规模的分布,疾病的传播阈值,以及相变发生后传染病规模的大小。

下面所介绍的这个 SIR 模型,是讨论随机出现一个病人,在给定度分布的网络上传播疾病的过程,而源于很多实际情况。它描述某疾病有一个传染源,再在人群中传播开来的过程,例如 SARS 疾病的传播。作为初步近似,假定个体对之间的传播概率 r_{ij} 与传染类个体的患病周期 τ_i 均是独立同分布的。

2.1.1 度不相关网络 在本小节简要介绍度不相关网络中几个重要的量。以 $p(k'|k)$ 记以从网络中 k 度点出发到达 k' 度点的概率。所谓度不相关网络,是指 $p(k'|k)$ 与 k 无关。假设网络中具有某个度 k 的节点出现的概率(度分布)是 p_k 。对 $x \in [0, 1]$, 记 $G_0(x) = \sum_k p_k x^k$, 称 $G_0(x)$ 为 $\{p_k\}$ 的母函数(母函数是研究非负整数值随机变量的有力数学工具)。在度不相关网络中,从任意点出发,到达一个度为 k 的邻居点的概率为 $\frac{kp_k}{\sum_j jp_j}$ 。记 $q_{k-1} = \frac{kp_k}{\sum_j jp_j}$ 或 $q_k =$

$$\frac{(k+1)p_{k+1}}{\sum_j jp_j} q_k$$

表示以任意点出发到达一个邻居点,而这一邻居点另外还有 k 个邻居(称 k 为出度)的概率(在这里邻居是指有边相连的结点)。记 $\{q_k\}$ 的母函数为 $G_1(x) = \sum_k q_k x^k$ 。以 $H_1(x)$ 记在网络中随机选择一条边,该边所到达的有限分支的节点个数分布所对应的母函数。以 $H_0(x)$ 表示在网络中随机选择一点,该点所在分支的节点个数所对应的母函数。可以得到

$$\begin{aligned} H_1(x) &= xq_0 + xq_1 H_1(x) + xq_2 [H_1(x)]^2 + \dots = xG_1(H_1(x)) \\ H_0(x) &= xp_0 + xp_1 H_1(x) + xp_2 [H_1(x)]^2 + \dots = xG_0(H_1(x)) \end{aligned}$$

笔者已证明网络中渐近地无圈,因此上述公式渐近地成立。由上述二公式可得随机选择点所在有限分支的平均大小

$$s = H'_0(1) = 1 + \frac{G'_0(1)}{1 - G'_1(1)}$$

当 $G'_1(1) = 1$ 时, $s = \infty$, 相变发生,出现巨分支。巨分支所占比例大小 $S = 1 -$ 有限分支所占比例大小,因此

$$H_1(1) = G_1(H_1(1)),$$

$$S = 1 - H_0(1) = 1 - G_0(H_1(1))$$

设 u 是 $G_1(u) = u$ 的最小非负实数解,由上式知 $H_1(1) = u$, 故 $S = 1 - G_0(u)$ 。而 $u = H_1(1)$ 是随机边到达有限分支的概率。

2.1.2 任意度分布网络中疾病的传播 具体描述这个传播过程:随机出现的一个病人,通过与其邻居

接触,把病传给这些邻居,然后每个邻居再各自进行接触传播,如此进行下去,就可知道在整个网络中疾病的传播情况。

假设 i 是传染类个体 j 是易感类个体 i 把病传播给 j 的概率是 r_{ij} , 个体 i 的患病周期为 τ_i , 个体康复后不会再传染他人, 记 T_{ij} 为 i 传播疾病给 j 的概率。则 i 与 j 之间不传播疾病的概率为

$$1 - T_{ij} = \lim_{\delta_i \rightarrow 0} (1 - r_{ij}\delta_i)^{\tau_i/\delta_i} = e^{-r_{ij}\tau_i}$$

因此 $T_{ij} = 1 - e^{-r_{ij}\tau_i}$ 。

设 r_{ij} τ_i 均是独立同分布的随机变量, 设其分布分别为 $p(r)$ $p(\tau)$, 则 T_{ij} 也是独立同分布的随机变量, 由此可得任意两个人之间传播疾病的概率为

$$T = T_{ij} = 1 - \int_0^\infty p(r)p(\tau)e^{-r\tau} dr d\tau \quad (0 \leq T \leq 1)$$

从而在此网络中, 每条边传播疾病的概率为 T , 对应到渗流理论中, 每条边被占据的概率为 T 。从而若随机选择一个点的度为 k , 与它相连的 m 条边被占据的概率为 $C_k^m T^m (1 - T)^{k-m}$, 故随机出现的一个病人, 传染疾病给其邻居的个数分布的母函数为

$$G_0(x, T) = \sum_{m=0}^\infty \sum_{k=m}^\infty p_k C_k^m T^m (1 - T)^{k-m} x^k = \sum_{k=0}^\infty p_k (1 - T + xT)^k = G_0(1 + (x - 1)T)$$

因为在任意度分布的网络中, 随机选择点的邻居的出度分布的母函数为 $G_1(x) = \sum_k q_k x^k$, 所以随机出现的一个病人, 其每个邻居传播疾病人数分布的母函数为

$$G_1(x, T) = \sum_{m=0}^\infty \sum_{k=m}^\infty q_k C_k^m T^m (1 - T)^{k-m} x^k = \sum_k q_k (1 - T + xT)^k = G_1(1 + (x - 1)T)$$

同理

$$H_1(x, T) = xG_1(H_1(x, T), T)$$

$$H_0(x, T) = xG_0(H_1(x, T), T)$$

$H_0(x, T)$ 表示由随机出现的一个病人所在有限分支中被传染上疾病人数的分布的母函数, 所以疾病爆发规模的平均大小为

$$s = H'_0(1, T) = 1 + \frac{G'_0(1, T)}{1 - G'_0(1, T)} = 1 + \frac{TG'_0(1)}{1 - TG'_0(1)}$$

用 s 表示随机变量 s 的均值, 当 $1 - TG'_0(1) = 0$ 时,

$$\text{即 } T_c = \frac{1}{G'_0(1)} = \frac{G'_0(1)}{G''_0(1)} = \frac{\sum_k k p_k}{\sum_k k(k-1)p_k} \text{ 时发生相变。即该疾病在此网络的传播阈值为}$$

$$T_c = \frac{\sum_k k p_k}{\sum_k k(k-1)p_k}$$

当 $T > T_c$ 时, 巨分支产生, 在这里称之为有瘟疫发生, 即疾病爆发。瘟疫影响人所占比例大小

$$\mathcal{S}(T) = 1 - H_0(1, T) = 1 - G_0(u, T)$$

这里 $u \equiv H_1(1, T)$ 是 $u \equiv G_1(u, T)$ 的最小非负实数解。 u 表示随机边到达一有限分支的概率。

一个一度点不受瘟疫影响的概率 $v = 1 - T + Tu$, 这是因为与这点相连的边没有被占据的概率为 $1 - T$, 与这点相连的边被占据但又指向一有限分支的概率为 Tu , 故一度为 k 的点不受瘟疫影响的概率为 v^k , 所以随机选择一度为 k 的点且没受瘟疫影响的

的概率为 $\frac{p_k v^k}{\sum_j p_j v^j}$, 其母函数为 $\frac{G_0(vx)}{G_0(v)}$, 故不受瘟疫影响的点(巨分支以外的点)的平均度为

$$z_{\text{out}} = \frac{vG'_0(v)}{G_0(v)} =$$

$$\frac{vG_1(v)}{G_0(v)} z = \frac{u(1 - T + Tu)}{1 - S} z \leq z(1 - S \leq u)$$

随机选择一度为 k 的点且受瘟疫影响的概率为 $\frac{p_k(1 - v^k)}{\sum_j p_j(1 - v^j)}$, 母函数为 $\frac{G_0(x) - G_0(vx)}{1 - G_0(v)}$, 所以受瘟疫影响的点(巨分支以内的点)的平均度为

$$z_{\text{in}} = \frac{1 - vG_1(v)}{1 - G_0(v)} z = \frac{1 - u(1 - T + Tu)}{S} z \geq z。$$

上式表明受瘟疫影响的点的平均度比不受瘟疫影响的点的平均度要大或至少相等。这跟实际情况吻合, 在瘟疫发生时受瘟疫影响的人平均接触的人数要比不受瘟疫影响的人平均接触的人数要多。

2.1.3 度相关网络中的讨论 在度相关网络中两邻点之间有某种相关性, 即 $p(k' | k)$ 与 k 和 k' 都有关。以联合概率分布 $e_{kk'} = p(k' | k)q_k$ 表示随机选择的边的两端顶点的出度分别为 k 和 k' 的概率 $\sum_{k'} p(k' | k) = 1$ $e_{kk'} = e_{k'k}$ 。

从 k 度点出发到达有限分支大小的母函数为

$$H_k(x) = x \frac{\sum_{k'} e_{kk'} [H_{k'}(x)]^{k'}}{\sum_{k'} e_{kk'}}$$

小的母函数为 $H(x) = p_0 x + x \sum_{k=1}^\infty p_k [H_{k-1}(x)]^k$ 。当有巨分支出现时, 巨分支所占比例大小 $S = 1 - p_0 -$

$$G_0(u_{k-1}) \mu_k = H_k(1) = \frac{\sum_{k'} e_{kk'} u_{k'}^{k'}}{\sum_{k'} e_{kk'}}$$

这里 u_k 表示从 k 度点出发的随机边到达有限分支的概率。类似 1.2 节其它讨论, 在此不再详细表述。

2.2 现实世界的 SIS 模型

因为康复的病人不具有免疫力,有可能再次被传染上,疾病的传播过程对应到网络中就相当于存在圈,因此不能用分支过程的思想来刻画疾病的传播过程。文献中用另外一种统计物理的平均场方法来研究在任意度分布网络中 SIS 模型疾病的传播。

Pastor-Satorras 和 Vespignani^[6]针对 SIS 传染病给出了一个详细的平均场解法。用 i_k, s_k 分别表示度为 k 的点中传染类个体和易感类个体各占比例。

对于一个 k 度的点,用 $k\lambda\theta(\lambda)$ 来代替传统模型中的 β_k (见第二节)。它表示的是单位时间内间易感类个体被传染所占的比例,这里 λ 表示经接触一个传染类个体而被传染上疾病的概率。 $\theta(\lambda)$ 表示随机边到达的点属于传染类个体的概率。一般而言 $\theta(\lambda)$ 与顶点度数有关,但在后面的讨论 $\theta(\lambda)$ 中对所有结点都相同(这也是统计物理中平均场近似的思想)。作为初步近似,在下面的讨论中还假设传染类个体的康复概率 $\gamma = 1$ 。

2.2.1 度不相关情形 在度不相关网络中,因为随机边到达一 k 度点的概率为 $\frac{k p_k}{\sum_j j p_j}$,而此点为传染类个体的概率为 i_k ,从而随机边到达的点是传染类个体的平均概率为 $\theta(\lambda) = \frac{\sum_k k p_k i_k}{\sum_k k p_k}$ 。从而在 $\gamma = 1$ 时有以下微分方程成立。

$$\frac{d i_k(t)}{d t} = k \lambda \theta(\lambda) (1 - i_k(t)) - i_k(t)$$

考虑系统在一段时间以后的稳定状态,此时

$$\frac{d i_k(t)}{d t} = 0, i_k = \frac{\lambda k \theta(\lambda)}{1 + \lambda k \theta(\lambda)}, \text{代入 } \theta(\lambda) = \frac{\sum_k k p_k i_k}{\sum_k k p_k} \text{得}$$

$$\theta(\lambda) = \frac{1}{k} \sum_k k p_k \frac{\lambda k \theta}{1 + \lambda k \theta} \quad (1)$$

k 表示随机变量的均值。

下面求 i_k 的非零稳定解。对(1)式两端求导,有

$$\frac{d}{d \theta} \left(\frac{1}{k} \sum_k k p_k \frac{\lambda k \theta}{1 + \lambda k \theta} \right) \Big|_{\theta=0} \geq 1$$

即 $\frac{\sum_k k p_k \lambda k}{k} = \frac{k^2}{k} \lambda \geq 1$ 。记 $\lambda_c = \frac{k}{k^2}$, 当 $\lambda \leq \lambda_c$ 时,

$i_k(t)$ 在有限时间内减小到 0, 表明系统处于稳定状态时, 疾病不会传播开来, 而当 $\lambda \geq \lambda_c$ 时, $i_k(t)$ 有非零解, 表明系统处于稳定状态时, 网络中疾病以大于 0 的概率持续传播, 因此有疾病爆发。称 λ_c 为传播阈值。

在 scale-free 网络中不存在非零传播阈值。因为在 SF 网络中 $k^2 \rightarrow \infty$, 所以 $\lambda_c = 0$, 表明无论传播概率有多小, 只要大于 0, 总会用瘟疫发生, Internet 网上疾病的传播就是这种情况。

2.2.2 度相关情形 本小节在度相关的网络中运用度的平衡性来讨论是否存在传播阈值。所谓网络度的平衡性是指, 从 k 度点出发到达 k' 度点的总边数等于从 k' 度点出发到达 k 度点的总边数, 用公式表述为

$$k p_k P(k'|k) = k' p_{k'} P(k|k')$$

在度相关网络中, 从一 k 度点出发的边到达一 k' 度点的概率为 $P(k'|k)$, 而该 k' 度点为传染类个体的概率为 $i_{k'}$, 故从 k 度点出发的边到达度为 k' 的传染类个体的概率为 $P(k'|k) i_{k'}$, 所以从 k 度点出发的边到达的点为传染类个体的平均概率为 $\theta(\lambda) = \sum_{k'} P(k'|k) i_{k'}$ 。从而在康复概率 $\gamma = 1$ 时, 有微分方程

$$\frac{d i_k}{d t} = k \lambda \theta(\lambda) (1 - i_k) - i_k$$

即

$$\frac{d i_k(t)}{d t} = -i_k(t) + \lambda k [1 - i_k(t)] \sum_{k'} P(k'|k) i_{k'}(t) \quad (2)$$

成立。

若 $i_k(t)$ 的值很小, 则(2)式的一阶近似为

$$\frac{d i_k(t)}{d t} \approx \sum_{k'} L_{kk'} i_{k'}(t),$$

其中 $L_{kk'} = -\delta_{kk'} + \lambda k P(k'|k)$ 为 Jacobian 矩阵($\delta_{kk'}$ 是 Kronecker delta 函数)。当 Jacobian 矩阵 $L_{kk'}$ 至少存在一个非负特征值时, $i_k = 0$ 不是上述方程的稳定解。而如果 $L_{kk'}$ 的特征值都严格小于 0, 则易知 $i_k = 0$ 是上述方程的稳定解。

设 Λ_m 是矩阵 $C_{kk'} = k P(k'|k)$ 的最大特征值, 则 $L_{kk'}$ 的特征值为 $-1 + \lambda \Lambda_m$, 所以当 $-1 + \lambda \Lambda_m < 0$ 时, 方程的解 $i_k = 0$ 是稳定解。记 $\lambda_c = \Lambda_m^{-1}$, 由上述讨论知 λ_c 是疾病的传播阈值。根据 Frobenius 定理, 对任意的非负不可约矩阵($A_{kk'}$), 令 Λ_m 为其最大特征值, 有

$$\Lambda_m \geq \min_k \frac{1}{\varphi(k)} \sum_{k'} A_{kk'} \varphi(k'), \quad (3)$$

其中 φ 为任意的 m 维向量。在上述讨论中特别令 $\varphi(k) = k$, $A = C^2$ 得

$$\Lambda_m^2 \geq \min_k \sum_{k'} \sum_{l} k' l P(l|k) P(k'|l)$$

根据网络度的平衡性, 有

$$k p_k P(k'|k) = k' p_{k'} P(k|k'), \quad (4)$$

对(4)式两边同乘以 k , 再对 k, k' 求和, 得到

$$k^2 = \sum_{k'} k' p_{k'} \sum_k k P(k|k')$$

令 $\bar{k}_{nn}(k') = \sum_k k P(k'|k)$ 表示 k' 度点的最近邻的平均度, 故 $k^2 = \sum_{k'} k' p_{k'} \bar{k}_{nn}(k')$.

在度相关 scale-free 网络中, 可以证明如果幂率指数 $\gamma < 3$ 时, k^2 随着节点增加而趋于 ∞ , 此时 $\bar{k}_{nn}(k') \rightarrow \infty$, 故有

$$\Lambda_m^2 \geq \min_k \sum_l l P(l|k) \bar{k}_{nn}(l) \rightarrow \infty,$$

因此由上述的讨论知 $\lambda_c = 0$, 即此时疾病传播的阈值为 0.

3 结束语

通过复杂网络的理论来研究疾病的传播, 能深入理解到网络的拓扑结构对疾病传播有重要影响, 从而找到控制疾病传播的更有效的方法, 这对于社会上的很多流行病提供了一些理论研究的基础。本文介绍的两个模型只是真实情况的简单抽象, 能够说明一些问题, 当然也存在一些与实际不符的地方, 比如缺少与实际传染病学和计算机病毒学所对应的参数等。因此建立更加真实的模型, 并在此基础上研究疾病的传播是目前一个很重要的课题。

参考文献:

[1] ANDERSON R M ,MAY R M. Infectious Diseases of Humans[M]. Oxford :Oxford University Press ,1991.
 [2] BARAB A' S. ALBERT R , JEONG H. Meanfield Theory for Scale-free Random Networks[J]. Physical A ,1999 , 272 :173-187 .

[3] CALLWAY D S ,NEWMAN M E J ,STROGATZ S H ,et al. Network Robustness and Fragility :Percolation on Random Graphs [J]. Phys Rev lett 2000 85 5468-5471.
 [4] COHEN R ,EREZ K ,BEN-AVRAHAM D , et al. Resilience of the Internet to Random Break Downs[J]. Phys Rev lett 2000 86 3682-3685.
 [5] NEWMAN M E J ,STROGATZ S H ,WATTS D J. Random Graphs With Arbitrary Degree Distributions and Their Applications[J]. Phys Rev E 2001 64 :026118.
 [6] PASTOR-SATORRAS R , VESPIGNANI A. Epidemic Spreading in Scale-free Networks[J],Phys Rev lett 2001 , 86 3200-3213.
 [7] BOGU ÑA' M ,PASTOR-SATORRAS R ,VESPIGNANI A. Absence of Epidemic Threshold in Scale-free Networks With Connectivity Correlations. [EB/OL]. http ://www. arxiv. org/abs/cond-mat/0208163 2002.
 [8] PASTOR-SATORRAS R , VESPIGNANI A. Immunization of Complex Networks [J]. Phys Rev E 2002 65 036104.
 [9] COHEN R ,BEN-AVRAHAM D ,HAVLIN S. Efficient Immunization of The Population and Computers[EB/OL]. http ://www. arxiv. org/abs/cond-mat/0207387 2002.
 [10] NEWMAN M E J. Spread of Epidemic Disease on Networks[J]. Phys Rev lett 2002 89 208701.
 [11] VAZQUEZ A ,MORENO Y. Resilience to Damdae of Graphs With Degree Correlations[J]. Phys Rev E 2003 , 67 015101.
 [12] NEWMAN M E J. The Structure and Function of Complex Networks[EB/OL]. http ://www. arxiv. org/abs/cond-mat/0303516 2003.

(责任编辑 游中胜)