

# 第3章

## 复杂网络上传播模型与动力学行为

近年来,随着网络科学的蓬勃发展,人们开始关注网络结构对传播行为的影响。网络上的传播行为在许多实际网络中都广泛存在,包括人类社会中的疾病传播,通信网络中的计算机病毒传播,社交网络中的信息传播,电力网络中的相继故障,以及经济网络中的危机扩散等。社交网络具有信息和行为协同传播的特点,由于网络信息传播与病毒扩散具有相似之处,因此,本章先介绍较为经典的传染病模型,然后介绍区别于传染病在人群中的传播,网络信息在人群中传播的行为特点。

### 3.1 基本传染病模型

当谈到流行性疾病时,人们会想到由生物病原体引起的传染性疾病,如流感、麻疹等都是通过人际传播的<sup>[1]</sup>。一些流行性疾病前期没有明显症状,潜伏期不定,很可能在人群中暴发,如新冠在全国甚至全世界范围内的传播。因此深入地探究、分析、挖掘传染病的传播规律具有重要的意义。

经典疾病传染病模型把个体分为三种状态:易染状态  $S$ (susceptible),指个体在感染之前处于易染状态,即节点个体有可能被邻居个体感染;感染状态  $I$ (infected),指个体感染上某种病毒,并且会以一定的概率感染其邻居个体;移除状态  $R$ (removed 或 recovery),也称为免疫状态或恢复状态,指个体已经痊愈,并且不会再被感染也不会感染其邻居个体。经典传染病模型的一个基本假设是完全混合(fully mixed),即一个个体在单位时间里与网络中任意其他个体接触的机会都是均等的。本节介绍三种经典的传染病模型:SI(susceptible-infected)模型、SIS(susceptible-infected-susceptible)模型和

SIR(susceptible-infected-removed)模型。

### 3.1.1 SI 模型

目前研究最多且最简单的就是 SI 模型,SI 模型中个体只有两种状态:易染状态 S 和感染状态 I。假设个体接触感染率为  $\beta$ ,即在传染病暴发初期,感染节点会以概率  $\beta$  去感染其他易染节点,同时一个易染节点被感染之后就永远处于感染状态。图 3-1 所示为 SI 模型中处于两种不同状态的节点相互之间的转换关系。

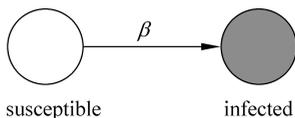


图 3-1 SI 模型中节点状态转换

SI 模型对应的微分方程可以表示为

$$\begin{cases} \frac{dS(t)}{dt} = -\beta I(t)S(t) \\ \frac{dI(t)}{dt} = \beta I(t)S(t) \end{cases} \quad (3-1)$$

式中, $S(t)$ 和 $I(t)$ 分别表示网络中健康节点和感染节点所占比例; $\beta$ 表示感染率,即健康节点被感染的概率。由于在 SI 模型中,节点只有这两种状态,所以 $S(t)+I(t)=1$ 。

在 SI 模型中,随着时间推移,最终所有个体都会被感染,故 SI 模型适用于描述那些受到感染后不可能治愈的疾病,或突然暴发尚没有有效措施对其控制的传染病。然而在现实世界中,感染个体一般不可能永远处于感染状态并永远传染别人。接下来介绍两种更为常见的模型。

### 3.1.2 SIR 模型

在现实中,病毒一般不可能一直存在于个体中。随着时间的推移,有些感染个体可能会被治愈或死亡,这类个体就具有了免疫能力,从而自动转变为不参与传播也不传染的 R 态。SIR 模型中增加了 R 状态,感染节点会以概率  $\beta$  去感染其他易染节点,感染态节点被治愈后也会以定常速率  $\gamma$  变为 R 态,即个体恢复为具有免疫性的个体或者死亡。图 3-2 所示为 SIR 模型中节点状态转换过程。

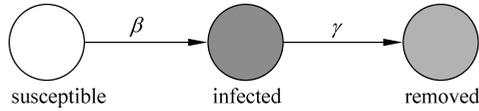


图 3-2 SIR 模型中节点状态转换

SIR 模型对应的微分方程可以表示为

$$\begin{cases} \frac{dS(t)}{dt} = -\beta I(t)S(t) \\ \frac{dI(t)}{dt} = \beta I(t)S(t) - \gamma I(t) \\ \frac{dR(t)}{dt} = \gamma I(t) \end{cases} \quad (3-2)$$

式中,  $S(t)$ 、 $I(t)$ 、 $\beta$  与 SI 模型和 SIS 模型中的定义是一致的;  $R(t)$  表示免疫节点在  $t$  时刻所占的比例;  $\gamma$  表示由感染态变为免疫态的概率。在 SIR 模型中, 节点有 S、I 和 R 这三种不同的状态, 因此  $S(t) + I(t) + R(t) = 1$ 。

记  $\lambda = \beta/\gamma$ ,  $\lambda = 1$  是 SIR 模型的传播临界值: 如果  $\lambda < 1$ , 那么  $\gamma = 0$ , 意味着病毒无法传播; 如果  $\lambda > 1$ , 那么  $\gamma > 0$ , 并且随着  $\lambda$  值的增大,  $\gamma$  值也增大, 意味着病毒在网络中扩散的范围增大。参数  $\lambda$  的一个直观解释是一个感染个体在恢复之前平均能感染的其他易感染个体的数目, 因此也常常称为基本再生数 (basic reproduction number), 文献中常用  $R_0$  表示。

### 3.1.3 SIS 模型

实际生活中还存在一种常见现象。某些个体被感染疾病后, 通过药物或其他方式治愈变成健康个体, 具有了暂时的免疫能力, 但不具有永久性免疫, 即恢复健康的 S 态个体还会被再次感染。SIS 模型就是刻画这种类型疾病传播的模型。它与 SIR 模型的区别在于 I 态个体恢复后的状态。在 SIS 模型中, 每个 I 态个体会以定常速率  $\alpha$  重新恢复为健康 S 态。初始时刻, 在网络中随机选取一些节点作为初始感染源, 被选取的节点会以概率  $\beta$  去感染其邻居, 被感染的节点会以概率  $\alpha$  恢复为健康态。随着时间的推移, 当网络中感染态的节点占一定的比例时, 网络中节点的状态就基本保持不变, 达到一个均衡阶段。图 3-3 所示是 SIS 模型中节点状态转换过程。

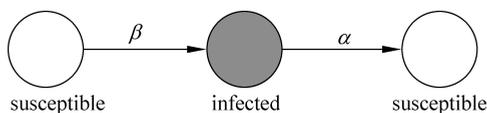


图 3-3 SIS 模型中节点状态转换

SIS 模型对应的微分方程可以表示为

$$\begin{cases} \frac{dS(t)}{dt} = -\beta I(t)S(t) + \alpha I(t) \\ \frac{dI(t)}{dt} = \beta I(t)S(t) - \alpha I(t) \end{cases} \quad (3-3)$$

式中,  $S(t)$ 、 $I(t)$  和  $\beta$  的定义与 SI 模型相同;  $\alpha$  代表恢复率即被感染节点重新恢复为健康态的概率。由于 SIS 模型中也只有 S 状态和 I 状态这两种状态, 因此同样有  $S(t) + I(t) \equiv 1$ 。

模型中  $\alpha$  和  $\beta$  比值的大小通常能够决定传染病传染的规模。假设  $\lambda = \alpha/\beta$ , 那么当  $\lambda > 1$  时, 感染率大于恢复率, 感染个体数目会越来越多, 疾病传染的规模将不断扩大; 当  $\lambda < 1$  时, 恢复率大于感染率, 说明恢复的病人会越来越多, 疾病传染的规模也会越来越小。因此  $\lambda = 1$  是 SIS 模型的传播临界值, 也是 SIS 模型的基本再生数<sup>[2]</sup>。

SIS 模型可用来描述健康个体被感染后, 又能恢复到健康状态的疾病, 如流感等。经典的 SIR 模型和 SIS 模型基于的完全混合假设意味着感染节点把病毒传染给任意一个易感染节点的机会都是均等的。但是在现实世界中, 一个个体通常只能和网络中很少一些节点是直接邻居。也就是说, 一个感染个体通常只可能把病毒直接传染给那些与之直接接触的部分节点。因此, 研究网络结构对于传播行为的影响自然就成为了一个重要的课题。

### 3.2 复杂网络上信息与行为传播模型

现实生活中广泛存在各种传播现象, 除了疾病、信息、舆论等传播主体可以在网络上传播以外, 人的观念、行为、甚至情绪等都可以在复杂网络上传播。网络中信息传播过程十分复杂, 但却可以通过信息传播模型对信息在复杂网络中的传播过程进行刻画。信息传播模型主要对影响最大化问题中的影响力传播过程进行建模, 使其更贴近于真实网络中影响力的传播机制。独立级联模型(independent cascade model, IC 模型)和线性阈值模型(linear threshold model, LT 模型)是

两种经典的影响力传播模型。

### 3.2.1 独立级联模型

独立级联模型最早是在研究市场营销时由 Jacob Goldenberg 提出的一个概率模型<sup>[3]</sup>。在独立级联模型中,节点被分为活跃和不活跃两种状态,且每一条边 $(v, u)$ 上的影响概率为一指定常量。IC 模型的基本假设是,节点  $u$  试图激活其邻居节点  $v$  的行为能否成功是一个概率为  $P_{u,v}$  的事件,且一个处于不活跃状态的节点被刚进入活跃状态的邻居节点激活的概率独立于之前曾尝试过激活该节点的邻居的活动。IC 模型还假设网络中的任意节点  $u$  只有一次机会尝试激活其邻居节点  $v$ ,无论是否成功,在以后的时刻中, $u$  本身虽然仍保持活跃状态,但已经不再具备影响力,这一类节点称为无影响力的活跃节点。

独立级联模型的算法如下:

(1) 给定初始的活跃节点集合为  $A$ 。

(2) 在  $t$  时刻,新近被激活的节点  $u$  对它的邻居节点  $v$  产生影响,成功的概率为  $P_{u,v}$ 。若  $v$  有多个邻居节点都是新近被激活的节点,那么这些节点将以任意顺序尝试激活节点  $v$ 。

(3) 如果节点  $v$  被激活成功,那么在  $t+1$  时刻,节点  $v$  转为活跃状态,将对其邻居不活跃节点产生影响;否则,节点  $v$  在  $t+1$  时刻状态不发生变化。

(4) 该过程不断进行重复,直到网络中不存在有影响力的活跃节点时,传播过程结束。

由于是概率模型,IC 模型中节点的激活过程是不确定的。对于同一个网络,同样的种子节点进行激活得到的结果可能会有较大差异。

独立级联模型抽象概括了社交网络中人与人之间独立交互影响的行为,它通过边上的概率来描述人与人之间发生影响的可能性和强度。很多简单实体,如在线网络上新消息的传播,或人际之间新病毒的传播等,都很符合独立传播的特性。独立级联模型在基于实际数据的影响力学习中也被初步验证是有效的,是目前研究比较广泛和深入的模型。

### 3.2.2 线性阈值模型

线性阈值模型是一种价值积累模型,体现了影响力的累积过程<sup>[4]</sup>。与独立级联模型一样,LT 模型中的节点也被分为活跃和不活

跃两种状态,每个节点都随机地选取 $[0,1]$ 区间中的值作为阈值,用来表示不活跃节点被其他节点激活的难易程度。线性阈值模型适合于描述个体的行为受多个个体的影响,独立级联模型适合于描述个体的行为只受一个个体的影响。

线性阈值模型定义如下:

(1) 每个节点用户  $v_i$  都有一个激活阈值  $\theta_i$ , 该阈值表示节点用户受影响的难易程度。在  $t_i$  时刻其所有邻居节点用户  $v_j$  对其综合影响为  $\varphi_i(t_i)$ ;

(2) 若  $\varphi_i(t_i) \geq \theta_i$ , 则  $v_i$  被激活(决定转发信息),  $t_{i+1}$  时刻  $v_i$  变为激活状态。

在 LT 模型中,当网络中已存在的所有活跃节点中任意活跃节点的影响力之和都不能激活它们的处于不活跃状态的邻居节点时,传播过程结束。LT 模型节点的激活过程是确定的,当对同一个图用同样的种子节点来激活时,最后的传播范围也是完全一样的。

### 3.2.3 复杂网络上的信息传播

信息传播过程演化建模的目的在于把影响信息传播的关键因素及相互作用关系用形式化模型(如概率模型、平均场理论、元胞自动机模型等)刻画出来,通过模型分析、仿真模拟和实际场景数据集验证,有效揭示和深入理解信息传播机理,预测信息传播过程的演化发展趋势。

数学家香农曾在其著名论文《通信的数学理论》中指出,信息是用来消除随机不定性的东西的。即当对信息了解不全面时,会影响个体做出的决策判断,而当个体能时时刻刻把握信息时,会给工作和学习等带来事半功倍的效果。随着网络技术的发展,网络上的信息传递速度越来越快,而人们接受信息的程度却没有改变。人类是独立存在的个体,具有自主的思想及各不相同的接受程度。即虽然接收到的信息相同,但是不同的人会产生不同的行为。行为通常是伴随信息而产生的,即信息与行为是相互作用、相互影响的。

虽然信息传播与疾病传播的传播机制不一样,但是经典 SIR 模型同样适用于舆情传播,只是三种状态表示的个体状态与疾病传播有区别。在舆情传播中,S 态代表网络中的个体不知道舆情,I 态表示个体知道舆情并传播舆情,R 态代表个体已经知道舆情信息,但不再传播舆情。由于对疾病传播的研究比较彻底,因此,用疾病传播模型来考

虑或延伸到信息传播具有重要的现实作用。虽然信息(包括谣言、舆情、行为等)传播与疾病传播有相似的传播机制,但是两者在本质上还存在着不同之处。对疾病传播和信息传播特性的分析如下。

(1) 疾病传播不具有记忆性,而信息传播有记忆性。在疾病传播中,这次没有接触到感染体与下次有没有接触到感染体的结果是互不相关的,而在信息传播中,这次信息传播可能没有改变当前节点的状态,但是会影响到下一次传播的结果。

(2) 疾病传播没有社会加强效果,而信息传播具有社会加强效果。一条信息如果只听到一次可能有所怀疑、没有接受信息,但是听多了就可能会相信,进而把信息传播出去。

(3) 在疾病传播中,传播链路可以多次使用;但在信息传播中,传播链路常常只使用一次。体现在现实世界中即人们可能会多次接触到疾病,但是一般不会多次告诉同一个人相同的消息。

(4) 信息的时效性随着时间衰减迅速,因为人们的兴趣会随着时间衰减。信息的时效性是信息价值的重要部分,人们不会对一个过时的信息予以过多的关注。而疾病可以存在上千年,不会因为时间的推移不在人群中产生、传播。

(5) 不同类型的边的传播能力和方式在信息传播中是不同的,而在疾病传播中没有太大区别。疾病传播中个体的连边值会造成传播概率的差异,而信息传播一般基于关系网络,人与人之间的联系强度是不一样的。

(6) 疾病传播的效果不受疾病本身的影响,而信息传播的传播效果却受到信息内容的影响。人们更倾向于接受和传播那些自己感兴趣的内容;而疾病不管个体的喜恶,都有可能受到感染或者传播出去。

(7) 信息传播中每个节点的影响力是不同的,人们更容易关注那些由名人或者权威专家等发出的信息。而在疾病传播中,经常与这些人接触不会增加被感染的概率。

由此可见,在信息传播中,人们的主观能动性起到很重要的作用,即人们对同一信息的接受程度、感兴趣程度等都可能不一样,这样产生的影响及对信息做出的判断、行为都将会具有浓烈的个人色彩。而在疾病传播中,人们是完全被动的,无论个体喜不喜欢、想不想要都阻止不了个体感染上疾病。但是随着信息技术的不断进步,人们对信息传播的理解也更深入,SIR 传播模型已不再适用于信息传播了。对

此,吕琳媛等建立的信息传播模型综合考虑了以上列举的一些特点,如社会加强效应、信息的记忆性等<sup>[5]</sup>,这些都为进一步了解和研究信息传播科学提供了建议与指导。

#### 3.2.4 复杂网络上的行为传播

行为传播是个体根据接触的信息产生相应或相异行为及行为对周围个体影响的过程,需要通过社会接触才能在社会网络上传播。对于行为在网络中的传播模式存在两种对立的观点<sup>[6]</sup>。一种观点认为,行为在具有高聚集特性且节点间度值差异较大的网络中传播时传播效率较低,但是当把网络中的冗余连接重新调整,即把相距较远的节点连接起来时,行为的传播效率就会得到明显的提高。这种观点将行为传播看作一种简单的疾病传播过程,与感染个体一次简单接触就可以传播该行为,而距离较远节点之间相连接减少了传播过程的冗余,可以使行为快速地从一节点传播到另一节点。

另一种相矛盾的观点认为,行为传播与疾病传播不同。行为传播是一个复杂的蔓延过程,在人们确定是否要采取一个行为之前,往往需要多次接触“传染源”,即社会强化作用。社会强化作用是指个体对接收的信息具有记忆性,接收信息的次数对驱动个体产生相应行为具有累加作用,个体接收信息次数越多,产生相应行为的概率越大。由于聚集网络中存在较多的冗余连接,人们通过冗余的连接会多次接触到信息驱动的行为,进而促进行为在人群中大规模地扩散开来。因此,这种观点认为行为在具有高聚集特性的网络结构中传播效率较高。无论是从哪种观点来研究行为传播,理解行为在网络中的扩散过程都是十分重要的。

2010年,Centola在*Science*上发表了关于行为传播的研究成果<sup>[6]</sup>,引起了人们对信息传播过程中行为作用的关注。对在线社会网络上的健康行为传播实验研究发现,网络拓扑结构对行为传播的过程有显著影响,具有大规模集群和较大直径的网络结构更有利于行为的传播。由于行为传播具有社会强化效应,个体采取行为的可能性是否提高受其所在社交群的多重激励影响,行为在拥有更多冗余连接的集群网络上比小世界网络上传播更快更广。在考虑社会强化效应的基础上,吕琳媛等建立了一个行为传播模型,验证了Centola在*Science*上发表的关于行为传播实验的相关结果<sup>[7]</sup>。阚佳倩等也研究了社会强化效应、边权重和网络结构特性对信息传播的影响<sup>[8]</sup>。虽然这些研

究都关注了行为在传播过程中的作用,但是均未考虑信息和行为的相互作用。在实际的社会网络中,信息与行为是相互作用和影响的。一方面,用户个体的行为会受到传播信息的影响;另一方面,个体行为的示范作用也会影响他人在信息传播的过程中产生跟随行为<sup>[9]</sup>。因此,在线社会网络的信息传播研究需要考虑信息与行为的相互影响。

### 3.3 复杂网络上的动力学行为

复杂网络上的动力学行为主要有网络的同步化和网络上的疾病传播两种。研究复杂网络上动力学行为的目标之一就是理解复杂网络的拓扑结构对其动力学行为的影响。网络的拓扑结构在决定网络动态特性方面起着重要的作用。例如,网络结构对耦合振子的同步化影响,网络结构对疾病传播的影响等。复杂网络小世界特性和无标度特性的发现,使人们开始关注网络的拓扑结构与网络的同步化行为之间的关系<sup>[10-11]</sup>。

#### 3.3.1 网络的同步化

同步是复杂网络的集体行为,是耦合振子之间的同步运动。人们很早就注意到了生活中的耦合振子同步化现象,如1665年荷兰物理学家惠更斯发现两个钟的钟摆同步摆动的现象,1680年荷兰人Kempfer在泰国旅行时发现大量萤火虫同时发光又同时熄灭的同步现象等。大量的同步现象引起了人们的极大关注,而复杂网络作为一个巨大且复杂的相互作用系统,对于研究大量相互作用的物理振子的同步化行为有重要作用。

事实上,同步在很多领域具有广泛的应用,如卫星导航系统(北斗卫星导航系统、全球定位系统GPS等)、电力网络系统设计、网络通信安全、智能控制等。同步现象可能是有益的,如卫星导航系统、激光发射器、心脏起搏器中脉冲耦合振子的同步等;也可能是有害的,如网络上路由器周期性发布的路由信息出现的同步引发拥堵的情况。因此,对于有益的同步,需要研究同步如何发生,满足什么条件或需要设计什么机制才能促使同步发生。对于有害的同步,同样需要研究同步发生的条件或机制,并进一步研究如何消除同步并预防有害同步的发生。

1998年,Pecora和Carroll研究了线性耦合网络同步的稳定性问

题,给出了主稳定函数判据及动力学网络的同步流形的线性稳定性<sup>[12]</sup>。2002年,Barahona和Pecora深入研究了对称耦合的情况,通过动力学分析研究了小世界网络的同步化情形<sup>[13]</sup>。汪小帆和陈关荣研究了耦合振子是连续系统的复杂网络同步稳定性问题,分析了无标度网络的同步现象<sup>[2]</sup>。更多关于网络同步的研究有小世界网络的同步化<sup>[10,13-15]</sup>、无标度网络的同步化<sup>[11,16-17]</sup>,以及社区集团网络的同步化<sup>[18-20]</sup>,这些领域都已经有了许多研究成果。

随着研究人员对网络结构与网络同步能力之间关系理解的深入,学者们提出了很多提高网络同步能力的方法,甚至利用各种方法尝试在一定条件下确定同步能力最强的“同步最优化网络”。这里的提高网络同步能力包含两方面含义:一是使原本在某一网络上不能同步的动力学系统能够同步起来,或者使同步变得容易;二是在保证同步的基础上提高动力学系统的同步稳定性<sup>[21-22]</sup>。关于网络的同步化研究已经获得了丰硕的成果,可以预见在这一领域还会有大量的研究成果出现。

### 3.3.2 网络上的疾病传播

在人类的社会发展进程里,疾病的传播都会对人类的社会生活产生极为重大的影响,疾病在人类社会中的传播极大地危害着人类的生存和生活。随着人类生活的日益网络化,人类的交往活动日益频繁,疾病的传播速度变得越来越快。现代生活中经常使用的计算机同样受到计算机病毒的危害,计算机病毒借助便利的网络可以瞬间侵入到世界各地,对人们的生产和生活产生极大的危害。这样看来,人们生活和工作的网络化很容易使疾病扩散和暴发。

因此,关于网络上疾病传播的研究对我们在实际生活中控制和预防疾病显得极其重要。将生物种群中的生物个体和计算机网络中的计算机都看成节点,将个体之间的相互作用看成节点之间的连边,就可以利用复杂网络理论对其进行研究。复杂网络疾病传播的研究可以使人们认识生物种群疾病和计算机网络病毒的传播机制。一直以来人们都认为,只有当疾病的传播速度超过某一个临界值时,疾病才会大规模地暴发。但是对无标度网络中病毒传播的研究发现,在无限大规模的无标度网络中,病毒传播的临界值趋向于零<sup>[23]</sup>。这是一个很恐怖的结论,这意味着即使是一个非常小的传播速度也会使疾病大规模地暴发,其危害将变得不敢想象。小世界、无标度、高聚类特性和