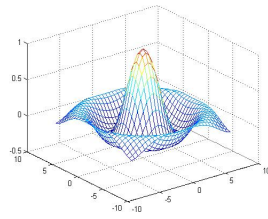


五一数学建模竞赛



题 目 自媒体时代的消息传播

关键词：传播 SEIR LWCS 节点属性 元胞自动机 偏微分方程

摘 要：

新媒体的诞生标志着信息的传递进入了一个全新的时代，各类信息间的相互渗透，媒体平台间的交叉传递，拉近人与人之间的距离，无时无刻地影响着人们生活方式。

本文就信息的传递方式及过程展开研究，主要解决了信息如何在实际生活中传递的问题，并针对与之高度关联信息出现时的对其传播带来的影响，给出了其最优化影响的处理方案，同时考虑信息自身的类型及媒体平台的差异，结合数据分别分析了其各自影响，最后依据所得结论，提出了一份关于新媒体消息传播及管理的报告。

针对问题一，我们基于传统传染病传播模型，就 SNS 网络建立一种新型 SEIR 模型，将消息传播过程中不同职能的媒体平台化作不同节点，分别分析其在实际传播过程中的变化过程，多角度地揭示了实际消息传播的过程。

针对问题二，依据高度相似消息的出现，我们沿用问题一传染病传播的思想，建立基于节点属性和信息内容等多个因素，对信息传播概率和传播延迟两个目标变量建立细粒度的模型，具体分析了新消息对原消息传播带来的影响并求解出最大化影响的条件。

针对问题三，我们建立元胞自动机模型仿真不同信息类型对传播带来的影响，并给出两者间的关系，结果表明受用户欢迎度高的消息具备更广泛的辐射范围及更快的传播速率。对于不同类型的媒体平台，我们建立一种新型 LWCS 模型考量平台类型与消息传播的关系，经数据分析得出平台的类型仅与传播过程中速率相关，同时我们还证明了该指标相对于其他指标的正确性及优越性。

针对问题四，我们梳理前三问中所得结论，就自媒体平台管理的现状，给出切实可行的改进建议。

1. 问题重述

自媒体时代，全民皆媒体。互联网的发展普及，全球成为一个公共信息平台。自媒体时代，人们已不再是被“统一”的告知对或错，更多的是从独立获得的资讯中接受传播和进行传播。自媒体传播颠覆了以往新闻的生产流程和媒体的价值观，逐渐成为新闻信息传播的中坚力量。

问题一：建立消息在自媒体平台上传播的数学模型，并且以“中共中央、国务院决定设立河北雄安新区”为例，对其传播过程进行分析。

问题二：某条消息在自媒体传播的过程中出现了新的与之高度关联的消息，对此建立数学模型，分析并说明新消息出现后传播过程的变化情况。

问题三：建立数学模型分析（1）不同类型的消息在自媒体传播过程中是否存在差异？（2）同一消息在不同自媒体平台上传播是否存在差异？并通过数值算例进行验证。

问题四：结合所建立的模型，写一份分析报告，阐述对自媒体时代消息传播的见解或看法，并提出对自媒体平台管理的建议。

2. 问题分析

2.1 问题一的分析

新媒体时代的信息传播的不是简单的图论网络，还要实际情况，显然鉴于信息的特殊性质，可以将信息类比于流行性传染疾病的传播。但这还是远远不够的，所以在此基础上我们建立一种新型的 SEIR 模型，更加细分传播过程中媒体平台的职能，最后利用 MATLAB 对设立的偏微分方程进行求解。

2.2 问题二的分析

为探究新高度关联消息对原信息的影响，我们必须先分析新信息的特征，为此，沿用问题一的思想我们建立了基于节点和信息传播特征的网络信息的传播模型，鉴于问题二的情况提取出两个主影响因子，得出节点间传播速率与新信息的关系，并利用问题一中的模型得出了节点间传播速率与整体信息传播点密度的关系，进行分析比较可得出影响其重要因素。

2.3 问题三的分析

针对不同类型的数据，必然有着不同的节点传播速率，在此基础上我们利用元胞自动机进行仿真，改变每次仿真时的节点传播速率，并对数据数据分析。对于不同类型的媒体平台，我们没有选用传统的单一指标，而是建立了一符合 LWCS 模型作为评价标准，我们利用问题一中的模型对其进行检验的一方面验证了其正确性，另一方面证明其较与其他指标的优越性，并利用所得的结果进行分析。

2.4 问题四的分析

对于问题四，可根据前三个问题所建立的模型及其结论，针对其存在的问题对自媒体平台管理提出切实可行的建议。

3. 模型假设

- 3.1. 假设在某条消息传播过程不受自然因素影响；
- 3.2. 假设消息传播过程不受政策的影响；

- 3.3. 假设各个自媒体平台之间相互独立；
- 3.4. 假设不同类型信息间不存在竞争；
- 3.5. 假设网络中每个节点本身具有一定的权重。

4. 模型建立与求解

4.1 问题一的求解

4.1.1 模型的建立

信息具有不灭性和无限复制性，显然人类社会的信息传播具有明显的动态性、序列性和结构性。为探究信息在不同自媒体平台上的传播过程，我们根据信息在 SNS 网络（社会性网络服务 social network service, SNS）中的传播规律，在已有的 SIR 模型的基础上，把媒体平台中的传播节点分为四类：易感染节点 (S)、潜伏节点(E)、传播节点(I)、免疫节点(R)四类，建立了更适用于新媒体传播的 SEIR 模型。其中传播节点表示该节点接受了来自其邻居节点的信息，并具有传播该信息的能力。易感染节点表示该节点没有接受过来自其邻居节点的信息，并有机会接受信息，即有概率被感染。潜伏节点表示该节点已接收信息但进行传播。免疫节点表示该节点已经接受了其邻居节点的信息，但是不具有传播能力。

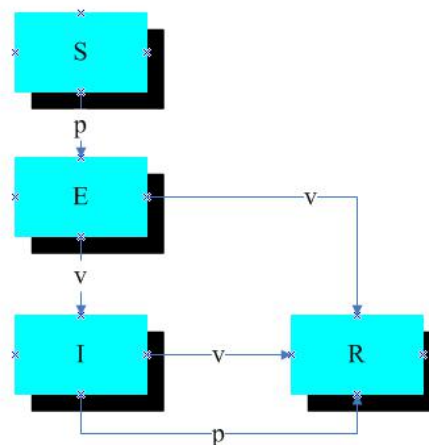


图 1：节点状态转移图

对于 SNS 网络中的某个节点 j ，其状态可能在 S、E、I、R 之间变化，在 $[t, t + \Delta t]$ 时间段内，节点 j 发生状态转移的部分概率定义下表所示：

符号	符号说明
p_{ss}^j	节点 j 保持易感染状态的概率
p_{se}^j	节点 j 从易感染状态转移到潜伏状态的概率
p_{ee}^j	节点 j 保持潜伏状态的概率
p_{ei}^j	节点 j 从潜伏状态转移到传播状态的概率
p_{er}^j	节点 j 从潜伏状态转移到免疫状态的概率
p_{ii}^j	节点 j 保持传播状态的概率
p_{ir}^j	节点 j 从传播状态转移到免疫状态的概率

假设节点j在t时刻处于易感染状态. 显然有:

$$p_{ss} + p_{se} = 1$$

用 $m_1 = m_1(t)$ 表示 t 时刻节点 j 的邻居节点中传播节点的数目, 则有:

$$p_{ss}^j = (1 - \Delta t p_1)^{m_1}$$

其中, $p_1(k, t)$ 为 t 时刻从具有 k 条边的易感染节点连接到一个传播节点的概率:

$$p_1(k, t) = \sum p(k_1 | k) p(Ik_1 | Sk)$$

$p(k_1 | k)$ 为度相关函数, 表示度为 k 的节点与度为 k_1 的节点相邻的条件概率;

基于上述理论我们建立网络信息传播过程中流体动力演化方程。假设在 t 时刻, SNS 中度为 k 的节点总数为 $N(k, t)$, 分别用 $S(k, t)$ 、 $E(k, t)$ 、 $R(k, t)$ 、 $I(k, t)$ 表示度为 k 的四类节点的数量, 则有:

$$S(k, t) + E(k, t) + I(k, t) + R(k, t) = N(k, t)$$

下面我们具体讨论在 $[t, t + \Delta t]$ 时间段内, 各类节点的数量变化情况。

(1) 易感染节点

$$S(k, t + \Delta t) = S(k, t) p_{ss}(k, t);$$

(2) 潜伏节点

$$E(k, t + \Delta t) = E(k, t) + S(k, t)(1 - p_{ss}(k, t)) - E(k, t)(p_{er} + p_{ei});$$

(3) 传播节点

$$I(k, t + \Delta t) = I(k, t) + E(k, t) p_{ei}(k, t) - I(k, t) p_{ir}(k, t);$$

(4) 免疫节点

$$R(k, t + \Delta t) = R(k, t) + E(k, t) p_{er}(k, t) + I(k, t) p_{ir}(k, t);$$

4.1.2 实例分析

针对问题一中“中共中央、国务院决定设立河北雄安新区”的消息, 为直观准确表现其传播过程, 我们参照相关文献, 选取了一组特定的数值带入 SEIR 模型中求解。我们指定 $v_1 = 0.25$ (次/小时), $v_2 = 0.3$ (次/小时), $v_3 = 0.05$ (次/小时), 鉴于该新闻为时政性新闻, 我们选取每个节点的度 $k=15$ 。通过分别对各节点传播数量进行对时间求偏微分方程, 以显示信息在新媒体传播过程中流动的演化的形式, 并用 MATLAB 绘制相关图像。

为检验该模型的正确性, 我们利用已有文献中的数据与该模型仿真结果相比

较，选取一组训练数据。图中节点即为基于实际情况给出的训练数据，图中曲线为该信息传递的仿真结果。

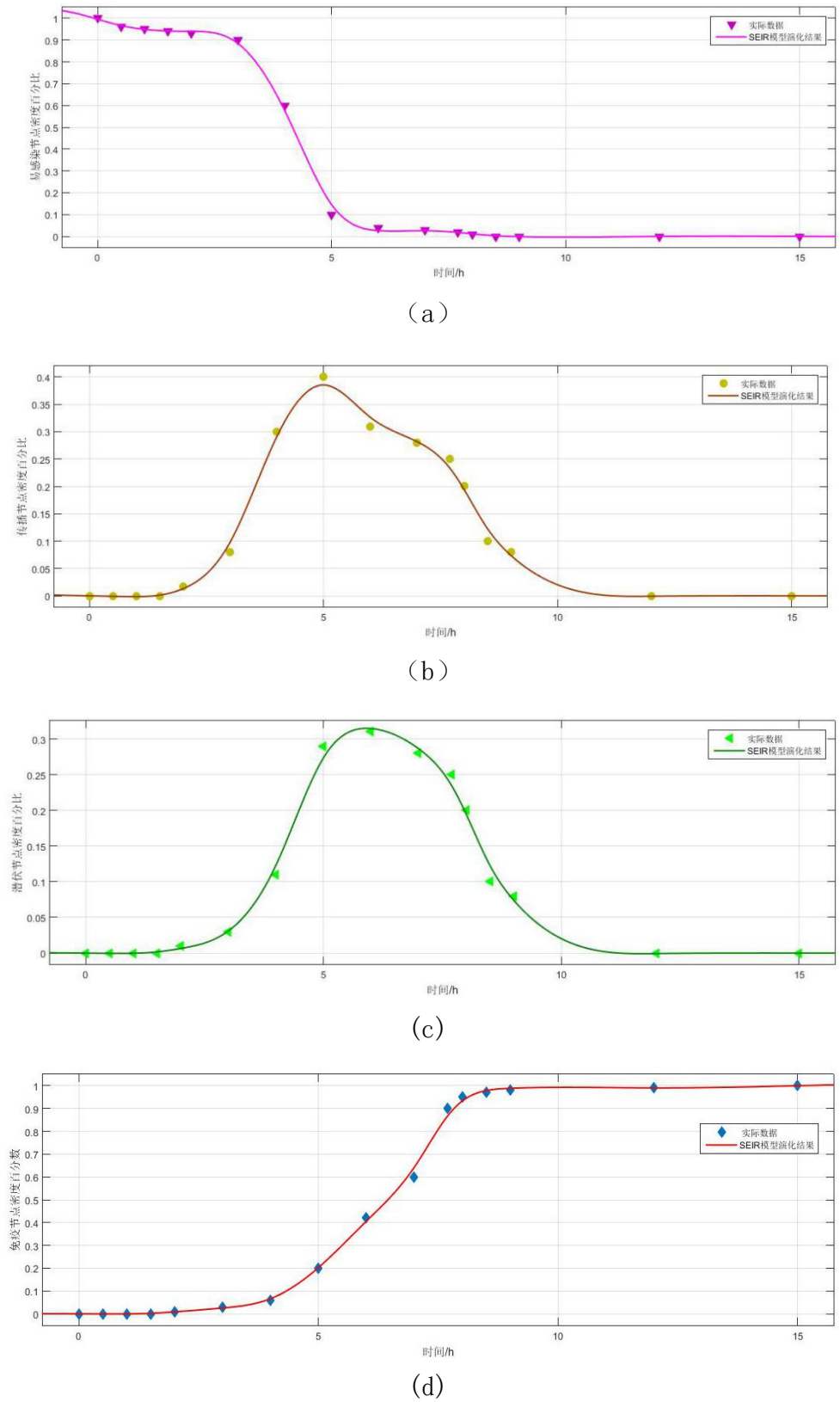


图 2：已有数据与该模型仿真结果对比图

从图中可以看出:易感染节点 $S(t)$ 的密度在初始阶段迅速减少到 0, 表示该信息在社交网络中扩散极快。潜伏节点 $E(t)$ 的密度在初始阶段迅速增大, 达到最高点后迅速减少直到 0。初始阶段, 大量的用户收到信息, 但处于离线状态的用户不会立即传播该信息, 所以潜伏节点 $E(t)$ 的密度快速增长。潜伏节点 $E(t)$ 曲线到达顶点后, 网络中不再有未接收信息的用户, 并且收到信息的用户逐渐迁移到传播状态开始传播信息或者进入免疫状态拒绝传播该信息, 潜伏节点 $E(t)$ 曲线迅速下降。

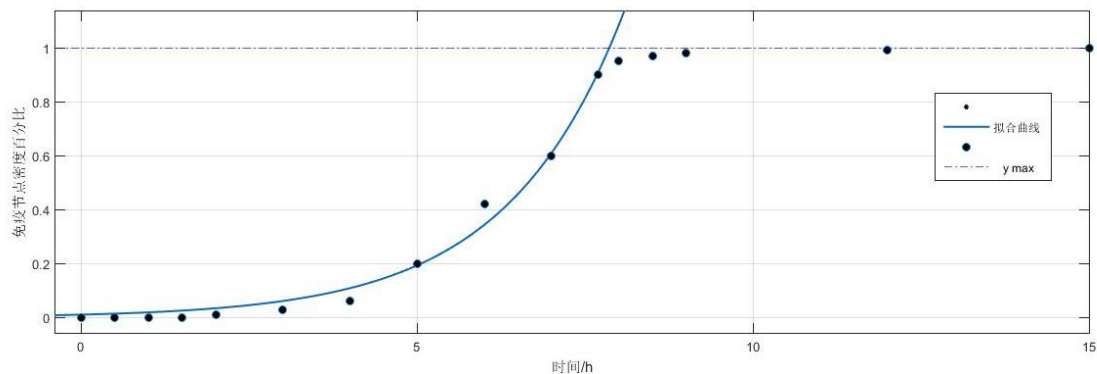


图 3: 免疫节点百分比与时间关系图

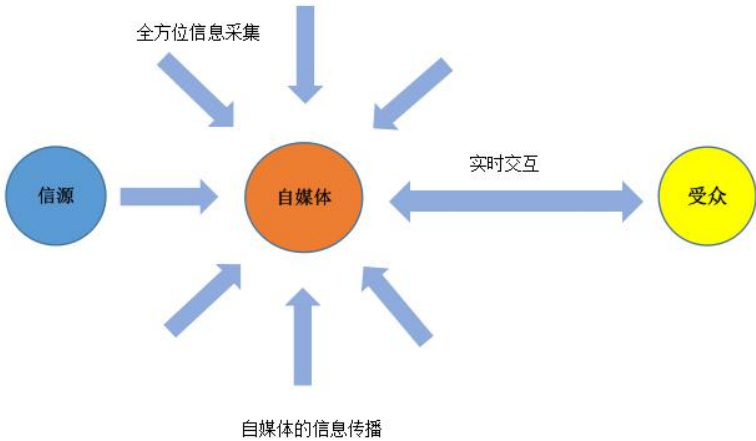
我们通过对免疫节点密度与时间曲线的快速增长部分进行拟合后发现两者关系呈指数型增长, 如图三所示。在信息传播的开始的 6 小时内近乎饱和。即表示该信息自中心节点发出后开始近似流行性传染病快速传播进行扩散, 在一定时间内迅速达到可能覆盖的范围。

4.2 问题二的求解

通过融合节点属性和信息内容等多个因素, 对信息传播概率和传播延迟两个目标变量建立细粒度的基于节点和信息特征的网络信息传播模型, 深入分析了影响自媒体平台上信息传播的主要因素, 包括信息内容、节点传播能力和网络结构等。

在此基础上, 分析了某条信息在自媒体传播过程中, 如果出现了新的与之高度关联的消息, 说明了新消息出现后传播过程的变化。

4.2.1 模型建立



4.2.1.1 特征的定义

首先定义了节点的特征集，其中 S:传播主体，R:传播客体，C:信息

(1) 传播主体特征 Φ_s

节点影响力 **Influ**: 节点影响力用该节点在一段时间内发布的所有消息的影响力之和表示，假设时段内一个节点所发布消息总的被转发次数表示为 x ，则

$$\text{Influ}(u)=\lg(x+1)$$

节点权威度 **Authu**: 用节点入度和出度的比值表示，即一个用户的被关注度 y_1 和关注度 y_2 的比值

$$\text{Authu}=\lg(y_1/y_2)$$

节点活跃度 **Act**: 用节点平均每天发布消息的数量 z 表示。活跃度高的用户其发布的消息更容易被其用户看到并传播

$$\text{Act}(u)=\lg(z+1)$$

(2) 传播客体特征 Φ_r

节点传播意愿: 该特征用于度量用户接收到某条消息后是否愿意传播该消息。用户历史中转发消息次数 n_1 与原创消息次数 n_2 的比值表示其传播意愿，该值越大表示用户传播信息的主动性越强

$$\text{Will}(r)=\log(n_1/n_2+1)$$

(3) 信息特征 Φ_c

采用最基本的二分类方法，根据情感将信息分为两类，即正向和负向两个情感极性。该特征用一个二维向量表示，每一个分量代表信息在这方面的情感倾向度。对特征向量进行归一化，使得向量所有分量之和为 1。例如，一条信息的正面情感分值为 m_1 ，负面情感分值为 m_2 ，则该特征向量为

$$\left(\frac{m_1}{m_1+m_2}, \frac{m_2}{m_1+m_2} \right)$$

(4) 传播主体和传播客体关系特征 $\Phi_{c,r}$

结构相似度: 用两个节点的邻居集合的 Jaccard 距离表示其结构相似度

$$\text{Sim-s}(u,v)=\frac{|N(u)\cap N(v)|}{|N(u)\cup N(v)|}$$

其中， $N(v_i)$ 表示 v_i 的邻居节点，自媒体中用该用户所关注或收听的节点集表示。

(5) 传播客体和传播内容关系特征 $\Phi_{s,r}$

采用传播客体对应文档与信息内容的余弦相似度表示用户对信息内容的传播兴趣

$$\text{Sim}(v,c)=\cos(\theta)=\frac{V\cdot C}{|V|\times|C|}$$

其中， V 、 C 分别代表用户 v 的传播内容向量和传播内容 c 的向量。

4.2.1.2 传播函数的建立

根据提取的特征集，建立特征与传播概率函数和传播延迟函数之间的关系。首先将节点特征和关系特征分别组合为一个特征向量，向量维数分别是 d_n 和 d_c ，

也就是节点特征和边特征的数量。

$$\phi_n = \begin{pmatrix} \phi_s \\ \phi_r \\ \phi_c \end{pmatrix} \quad \phi_e = \begin{pmatrix} \phi_{s,r} \\ \phi_{c,r} \end{pmatrix}$$

基础函数 $f_\alpha = (u, v, c)$ 可以表示为特征向量 Φ_n 和 Φ_e 的线性组合：

$$f_\alpha = (u, v, c) = \alpha_0 + \alpha_{T1}\phi_n + \alpha_{T2}\phi_e$$

传播概率 $p = (u, v, c)$ 用贝叶斯逻辑斯谛函数表示为：

$$p = (u, v, c) = \frac{1}{1 + \exp\{-f_\alpha(u, v, c)\}}$$

其中， α_0 是一个常量， α_1 表示为节点特征的权重， α_2 表示为关系特征的权重，权重越大代表它对应的特征对传播概率 $p = (u, v, c)$ 的影响越大。

传播时间延迟 $\tau(u, v, c)$ 可以表示为特征向量 Φ_n 和 Φ_e 的线性组合：

$$\tau(u, v, c) = \beta_0 + \beta_1^T \phi_n + \beta_2^T \phi_e$$

其中， β_0 是一个常量， β_1 表示为节点特征的权重， β_2 示为关系特征的权重，权重越大代表它所对应的特征对传播时间延迟 $\tau(u, v, c)$ 的影响越大。为表述方便，

设 $\alpha = (\alpha_0 + \alpha_1^T + \alpha_2^T)^T$ ， $\beta = (\beta_0 + \beta_1^T + \beta_2^T)^T$

$p = (u, v, c)$ 充分利用丰富的节点属性和信息内容特征，给出了一个细粒度的信息传播概率基础模型，同时需考虑现实网络中信息传播概率会随着传播时间间隔而演变。由于节点之间的影响力或传播信息的能力随着时间间隔增大而衰减，符合指数衰减规律，设想点 u 在时刻 t_u 感染信息 c^k ，节点 u 的邻居节点集用 $N(u)$ 表示，节点 $v \in N(u)$ ，节点 v 时刻 t_v 被节点 u 感染信息 c^k 的概率为

$f((v, t_v)|(u, t_u); \alpha, \beta)$ ，这个概率不会一成变而是随着时间间隔 $\Delta t = t_v - t_u$ 的增大而降低。对于问题二，考虑到在研究节点间信息或影响力的传播概率的同时融合时间衰减因素的传播概率模型，我们选择指数模型。指数模型中，传播概率随着 Δt 的增大单调递减，用指数模型来刻画传播概率随 Δt 的衰减规律。

其中，指数传播概率模型的公式为

$$\begin{cases} p(u, v, c) \tau(u, v, c) \exp\{-\tau(u, v, c)(t_v - t_u)\} & , \text{ if } t_v > t_u \\ 0 & , \text{ else} \end{cases}$$

在 t_u 到 t_v 时间段内，节点 v 被节点 u 感染的概率用 $f((v, t_v)|(u, t_u); \alpha, \beta)$ 表示，即传播概率密度函数的累积函数为

$$F((v, t_v) | (u, t_u); \alpha, \beta) = \int_{t_u}^{t_v} f((v, t) | (u, t_u); \alpha, \beta) dt$$

节点 v 不被节点 u 感染的概率用 $S((v, t_v) | (u, t_u); \alpha, \beta)$ 表示，即生存概率。

定义生存概率：已知节点 u 在时刻 t_u 感染信息 c^k ，节点 $v \in N(u)$ ，直到时刻 t_v 节点 v 未被节点 u 感染信息 c^k 的概率为生存概率

$$S((v, t_v) | (u, t_u); \alpha, \beta) = 1 - F((v, t_v) | (u, t_u); \alpha, \beta)$$

在时间窗口 $[0, T]$ 内，信息 c^k 在 $G(v, e)$ 内传播，所有的感染节点和感染时间用 $D^k = \{(v_1, t_1), \dots, (v_n(k), t_n(k))\}$ 表示， v 的父亲节点用 $\text{par}(v)$ 表示，时刻 t 之前感染信息 c^k 的节点为 $B^k(t) = \{v_i | (v_i, t_i) \in D^k, t_i \leq t\}$ ，则有可能使得节点 v 在时刻 t 之前感染信息 c^k 的节点集合为 $Q^k(v) = N(v) \cap B^k(t)$ 。

在时刻 t_v 节点 v 被其父节点 $\text{par}(v)$ 感染，概率为 $f((v, t_v) | (u, t_u); \alpha, \beta)$ ，在时刻 t_v 之前，节点 v 未被除 $\text{par}(v)$ 之外的其他节点感染的概率为

$$\prod_{w \in Q^k(v) \setminus \text{par}(v)} S((v, t_v) | (w, t_w); \alpha, \beta)$$

对于目标信息级联 M^k ，目标节点 $(v, t_v) \in D^k, m_{u, v, t_u, t_v}^k \in M^k$ ，节点 v 在时刻 t_v 被节点 u 激活，同时又不被节点 u 之外的其他可能激活它的邻居节点激活的概率为

$$f((v, t_v) | (u, t_u); \alpha, \beta) \cdot \prod_{w \in Q^k(v) \setminus \text{par}(v)} S((v, t_v) | (w, t_w); \alpha, \beta)$$

最后，满足全局似然概率 $f(M | \alpha, \beta)$ 最大化的参数估计 $\hat{\alpha}, \hat{\beta}$ 即为模型的解：

$$\min_{\alpha, \beta} -\lg f(M | \alpha, \beta)$$

其中， $\alpha = (\alpha_0 + \alpha_1^T + \alpha_2^T)^T$ ， $\beta = (\beta_0 + \beta_1^T + \beta_2^T)^T$ 。

4.2.1.3 参数求解

设 ω 代表参数集 $\omega = \{\alpha, \beta\}$ ，目标函数 $L(\alpha, \beta) = -\lg f(M | \alpha, \beta)$ ， $L(\alpha, \beta)$ 关于参数 α 在 \mathbb{R}^n 上具有连续一阶偏导数。同样， $L(\alpha, \beta)$ 关于参数 β 在 \mathbb{R}^n 上也具有连续一阶偏导数。

公式的求解可以采用随机梯度下降算法进行。随机梯度下降算法是一个迭代算法，其基本思想是从一个初始值 $\omega(0)$ 开始，根据训练数据，每次沿着目标函数的梯度下降方向移动一小步，步长即学习速率 δ ， δ 控制着每次沿梯度下降方向移动的距离，一般设置一个比较小的值，每次迭代都更新 ω 值，直到目标函数收敛。随机梯度下降算法可使目标函数收敛在一个全局或局部最优点，为应对可能出现的目标函数收敛到一个局部最优点，求解过程会尝试从多个不同起始点 $\omega^{(0)}$ 出发

进行迭代优化，最后比较不同收敛点处的函数值。

随机梯度下降算法的迭代形式为

$$\omega^{(t+1)} = \omega^{(t)} - \delta \times \frac{\partial L(\omega^{(t)})}{\partial \omega^{(t)}}$$

具体而言，对于参数 α ：

$$\alpha^{(t+1)} = \alpha^{(t)} - \delta \times \frac{\partial L(\omega^{(t)})}{\partial \alpha^{(t)}}$$

$L(\omega^{(t)})$ 关于 $\alpha^{(t)}$ 的梯度计算如下

$$\frac{\partial L(\omega^{(t)})}{\partial \alpha^{(t)}} = - \sum_{1 \leq k \leq K} \sum_{(v, t_v) \in D^k} \left\{ \frac{\partial \lg p(u, v, c^k)}{\partial \alpha^{(t)}} + \sum_{w \in Q^k(v) \setminus par(v)} \left(\frac{\partial p(w, v, c^k)}{\partial \alpha^{(t)}} \times \frac{-(1 - \exp\{-\tau(w, v, c^k)(t_v - t_w)\})}{1 - p(w, v, c^k)(1 - \exp\{-\tau(w, v, c^k)(t_v - t_w)\})} \right) \right\}$$

参数 β 的求解方法与参数 α 的类似，在此不再给出。

4.2.2 模型的求解

考虑到问题二的实际情况，与某一信息高度关联的信息的出现对该信息造成的影响，所以我们可以不考虑传播客体的影响，仅考虑 S：传播主体和 C：信息。即令特征参数 α 指代传播主体，参数 β 指代具体信息内容。

由于直接编程求解该问题显然难度过大，因此我们利用 MATLAB 中 PDETOOL 工具箱输入已有边界条件对偏微分方程进行求解，并可画出数值解的 3D 图形。

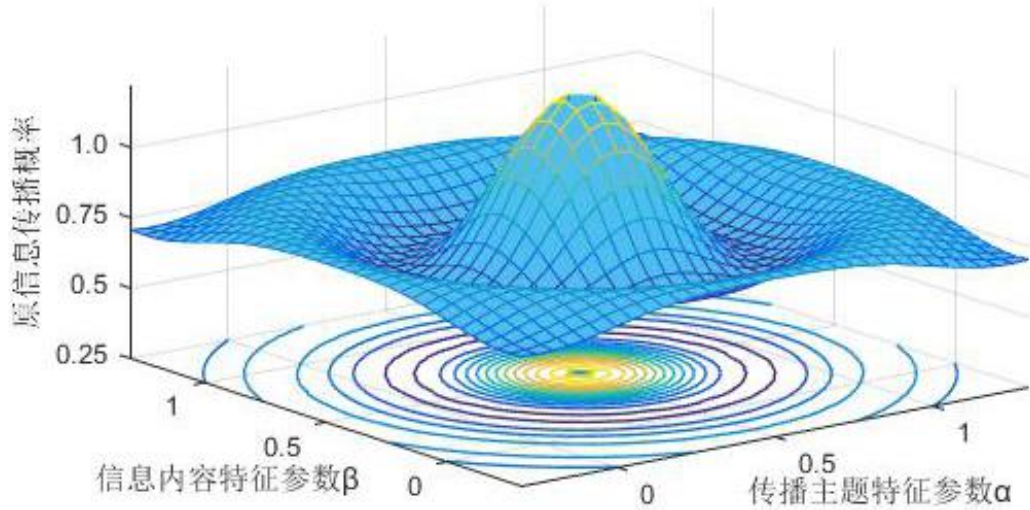


图 4：数值解的 3D 图

问题中的其他未知参数仍沿用问题一给出的参考数据。

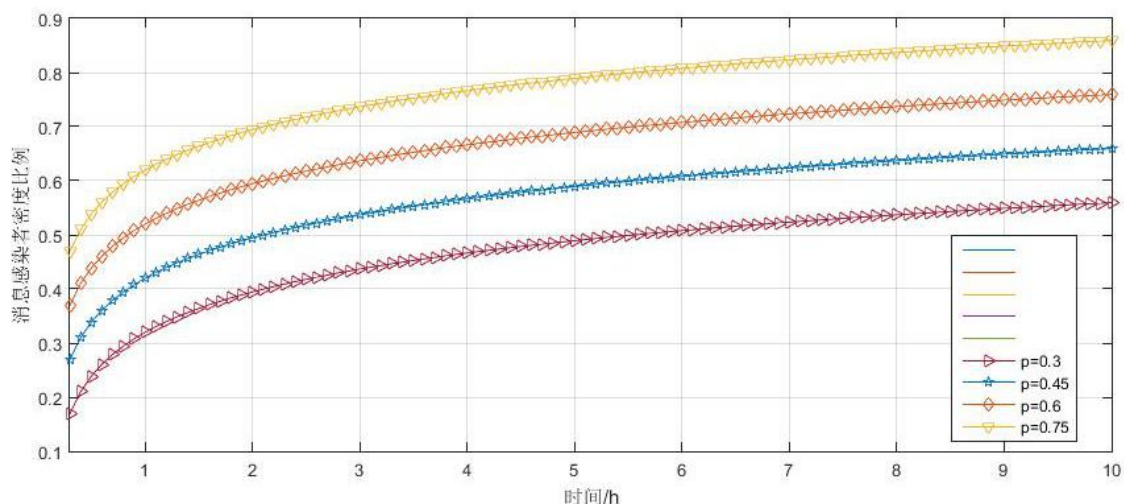


图 5：感染者密度百分比与原信息感染的概率关系图

由求解结果可知当传播主体特征系数 $\alpha=0.64$ ，信息内容特征参数 $\beta=0.412$ 时，原信息最易受到新信息的干扰。此时原信息感染的概率达最大值。实际上我们通过改变问题一中的原信息感染的概率我们发现其感染者密度百分比会随原信息感染的概率增大而增大，如图 5 所示。

所以我们可以得出结论，当某一新的相关信息出现时，当其传播主题和信息内容符合一定条件时，原信息的传播概率会增大，同时会影响着传播范围及原信息传播速率。

4.3 问题三的求解

4.3.1 不同信息传播差异影响模型的建立

针对问题三显然由于信息的不同受众群体的偏好差异，不同类型的信息必然会致使传播过程中的差异；同时鉴于不同网络媒体平台的影响力的差距，同样会致使信息的传播能力的变化。

在本问题中我们不考虑不同类型信息间的竞争，只考虑不同类型信息在传播过程中带来的影响。显然符合用户喜好的信息会具备有更强大的竞争力，在平台间具备更快的传播速度。这里我们沿用问题一采用的思想，为简化数据处理过程，我们使用元胞自动机对实际信息的传播进行仿真，探讨在不同信息传播速度下的，信息传递差异。

元胞自动机模型（CA）是一种用来仿真局部规则和局部联系的方法。典型的元胞自动机是定义在网格上的，每一个点上的网格代表一个元胞与一种有限的状态。变化规则适用于每一个元胞并且同时进行。且具有时间上、空间上和状态上都离散的特性，以局部规则同步演化来反映整个系统的复杂变化，这与信息的传播机制十分相似，在建立模型的过程中用个体的移动性反映了实际中个体的活跃程度，个体活跃程度越高，信息的传播速度越快，越容易信息传递节点密度越容易饱和（图 6 为元胞自动机的仿真过程）。

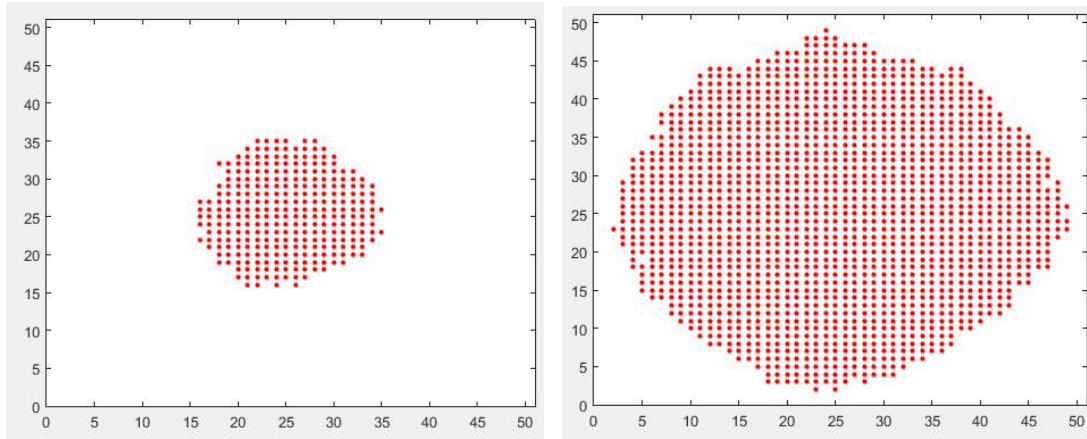


图 6：个体活跃程度越高，信息的传播速度越快，越容易产生饱和

我们分别调整节点间信息传播的速度 v ，以此来区分不同信息的种类及该信息的受欢迎度，显然 v 与信息的受欢迎度是呈正比的，下图中绘制了不同 v 时，传播范围变化率的曲线（图 7）。

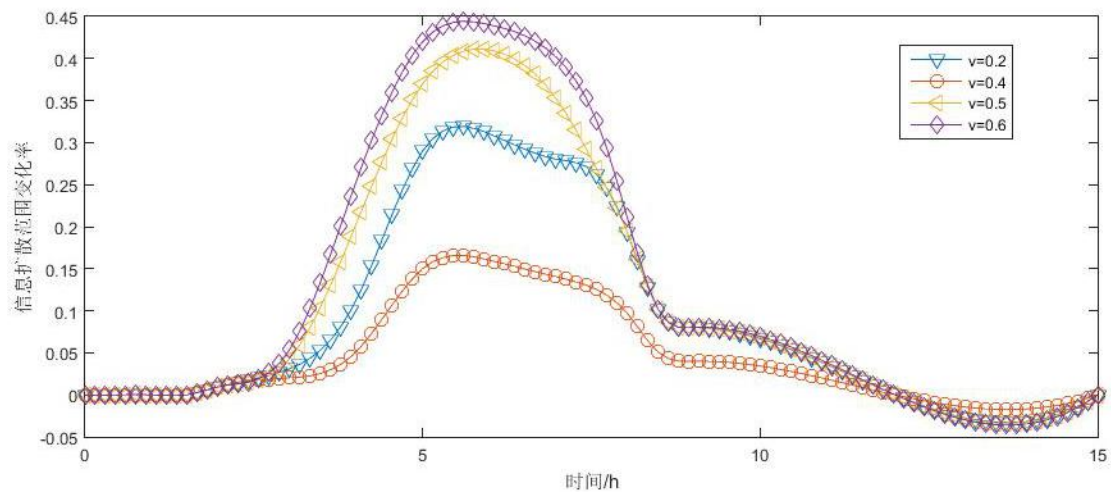


图 7： v 不同时，传播范围变化率曲线

由图表分析可知当信息扩散越快时，节点间的传播速度对其的影响越明显，该部分可对应问题一中的免疫患者或易感染患者密度突变的部分。

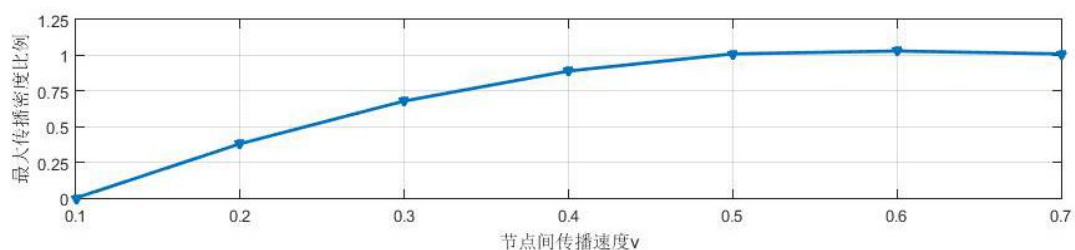


图 8：节点间传播速度与最大传播密度比例关系图

此外，通过统计节点间传播速度及最大传播密度比例可发现两者呈正相关（图 8），且当 v 达到一定值时，该信息所能达到的影响范围不再随其增加而变

化，这是与我们生活常识所相符的。

4.3.2 不同媒体平台传播差异影响模型的建立

与问题一的处理方法相同我们将不同媒体平台视作信息传播节点。为确定识别社交网络节点影响力，我们必须依据一定指标。我们知道在考虑不同节点时不能只使用节点的度作为衡量节点信息传递的能力的。

实际上，真实网络节点的特征远比想象中的要复杂的多，在复杂网络理论中，常见的节点网络拓扑特征有度、紧密度和介数。这三类特征均可以作为度量节点影响力的典型指标，但各自都具有较为明显的劣势。在现实社交网络中，与影响力较大的个体存在较强的好友关系的个体，即使其度值较小，但其具有的潜在影响力却难以忽略。此外，紧密度和介数均为全局性度量指标，其时间复杂度较大，适用范围较为有限。综上所述，我们基于相关参考文献设立了一新型指标并提出基于网络社区叠加的局域型权重指标（Local Weight Index Based on Community Superposition 简称 LWC 指标），且大量的真实数据集中的仿真实验表明，指标具有更加准确的识别性能。

计算公式如下：

$$LWCS(i) = A * s(i) / \max(s(i)) + B * C(i) * \langle k \rangle / \max(k);$$

其参数物理含义如下：

A 和 B 为调节因子,本模型中 A,B 均取 1；

k 为网络社区的平均节点度；

S(i)为节点权重，区分不同媒体平台的信息传播能力；

假设网络中每个节点本身具有一定的权重，此处定义节点的权重为与其相连的同一社区内的所有边的权值之和。节点权重具体表示为：

$$s(i) = \sum w_{ij};$$

其表示所域社区内的邻居节点 w_{ij} 的求和， w_{ij} 表示为节点和之间的边关系权重，可用二者之间的共同好友比例来表示，具体公式可表示为：

$$w_{ij} = O_{ij} / (k_{i-1} + k_{j-1} + O_{ij});$$

其中 O_{ij} 为节点和的共同好友数， k_i 和 k_j 分别表示节点 i 和 j 的度。

$C(i)$ 为社区紧密度，由于本模型不考虑将社交网络分化的情况，取为常数 1。

为验证模型的正确性及相对于其他指标的优越性，我们选取单个节点为起始点以研究其影响力分别计算其对应的 LWCS 值，利用某时刻的节点传播密度和其传播过程中最大的传播密度作为验证模型的指标。

表 1：计算 LWCS 值，节点传播密度，最大的传播密度

序号	介数	紧密度	度	LWCS	t=5 时刻传播密度	最大传播密度
1	125	125	125	169	0.617	0.689
2	169	169	169	338	0.599	0.685
3	338	338	338	125	0.613	0.683
4	335	335	335	206	0.584	0.679

5	295	295	295	295	0.598	0.681
6	206	206	206	335	0.593	0.68
7	91	91	91	158	0.578	0.681
8	195	158	158	213	0.569	0.676
9	58	195	87	114	0.588	0.678
10	158	114	195	195	0.575	0.686
11	87	60	213	87	0.582	0.674
12	97	213	114	60	0.574	0.675
13	248	284	284	284	0.561	0.678

通过观察表中 13 组数据的结果发现，最大传播节点密度差异不大表示信息覆盖范围近乎相同，这也进一步说明了仿真结果符合无标度网络中传播阈值几乎为零的结论。

此外，处理数据发现某一时刻不同传播源点所产生的传播结果是不同的，因此可以作为评价指标，且该时刻的传播密度与 LWCS 呈正相关。这个表明在特定时刻，不同节点信息的传播速度是不同的。同时这充分说明以具有不同指标的节点为传播源点，所造成的信息传播速度和影响范围也不同。同时初始传播节点的 LWCS 越大，初始时刻传播信息的用户的好友越多，SNS 网络中信息的传播速度越快，传播规模的增长越快。因此初始传播节点的 LWCS 的大小会对社交网络中信息传播有很大的影响。所以为能尽快地传播信息，应该选择 LWCS 较大的节点作为初始传播节点。但是，无论初始节点的 LWCS 是多少，收到信息的节点密度百分比最终都会趋于稳定值 1，表示网络中所有的节点最终都将收到该信息。由于在社交网络的高度连通性，信息在网络中传播的门槛几乎为零；初始传播节点的 LWCS 越大信息越容易在网络中迅速传播；中心节点具有较大的社会影响力。

模型的检验：我们利用 MATLAB 中的 corr 函数，分别计算各指标与特定时刻的传播密度的皮尔逊系数。

表 2：各指标与特定时刻的传播密度的皮尔逊系数

	LWCS	度	紧密度	介数
皮尔逊系数	0.8931	0.6016	0.4712	0.6619

显然 LWCS 作为指标明显优于其他指标，且选取 LWCS 作为指标直接的证明的我们的假设同一消息在不同自媒体平台上传播是否存在差异，且依据 LWCS 系数的构成我们可以更好地规划选取网络节点。

综上所述，不同类型的数据其在媒体平台间的传播速度取决于其受公众欢迎程度，并且这会影响到该信息整体的传播扩散的速度并能显著的影响其辐射范围。然而不同类型的媒体平台仅对指定时间点上的传播速度与范围有影响，对信息的最终辐射范围没有显著的影响。

4.4 对自媒体时代消息传播的见解及建议

自媒体时代的消息传播分析

随着自媒体时代的到来,每个人都是记录着和信息传播者,人们借助微博、微信、博客等移动终端进行自我观点的表述与传播。考虑到自媒体拥有的交互性、自主性和双面性,一方面我们应看到它的利处,从而对社会生活产生有益的效应;另一方面,我们要严格控制基于自媒体的社会不良舆论的传播,避免对人民群众造成危害。

对于这个问题,我们队伍首先建立了基于 SEIR 的社会网络信息传播模型,对传统的网络信息传播进行分析,并以“中共中央、国务院决定设立河北雄安新区”这条消息为例,分析其具体的传播过程。结果表明,易接受传播信息的人群在信息出现的起始阶段占了很大比重,而且成为传播信息主力,但随着传播时间的增长最终趋于消失;传播节点在时间历程中,呈现出先增后减的趋势,符合社会规律;而由于一些特殊原因产生的潜伏传播节点也呈现出先增后减的变化。最终,随着传播时间的增加,网络信息传播趋于消逝。通过模型,可以看出在实际的自媒体平台管理中,我们应当控制易接受传播信息人群和传播节点的数量,并注意其变化的趋势,在出现对社会有益的信息时大力借助多方自媒体进行宣扬,而出现危害社会的消息时应减少传播节点。而且,还应对可能存在的传播节点进行监管,防止其爆炸性地出现对社会造成不良影响。

现实生活中,某条信息的传播往往会受到新的与之高度关联的消息的影响,为了分析这种情况下的传播过程变化,我们在传播节点和信息特征的基础上建立了信息传播概率和传播延迟模型。在不考虑传播客体的影响下,发现了当某传播主题和信息内容与已有信息符合传播主体特征系数为 0.64,信息内容特征参数为 0.412 时,原信息的传播受到影响最大,传播概率、范围和速率都发生变化。借鉴到实际的自媒体平台管理上,为了宣传某一事件,我们可以增加与其相关联的信息数量,从而传播效果达到预期最大值。

同时,由于信息的不同受众群体的偏好差异,不同类型的信息必然会致使传播过程中的差异。对此,我们在忽略不同类型信息间竞争的基础上,建立元胞自动机模型进行仿真,结果发现,不同类型信息传播的速度与其所受欢迎度是成正比的。同时信息扩散得越快,信息在节点间的传播速度对信息扩散的影响越大。所以说,在实际自媒体管理上,应主题调控各种信息类型的比重,在吸引大众眼球的同时,注意对社会造成的影响、对重点事件的宣传报道,使其达到一种平衡的状态。

除了信息的类型会对传播过程产生很大的影响,同一信息在不同的自媒体平台上传播也存在差异。针对这种现象,我们将不同的媒体平台视作信息传播节点,并设立了一种基于网络社区叠加的局域型权重指标,而后通过计算公式具体建立模型,最终发现不同类型的自媒体平台仅对指定时间节点上的传播速度与范围有影响,而对最终的传播辐射范围没有显著影响。因此在自媒体平台管理中,应根据不同类型的信息,将不同的信息传播平台和相应的时间节点结合,确保在一定时间范围内,信息传播达到最大化或最小化。

综上所述,我们所建立的多种数学模型对于自媒体平台的管理具有很大的实用价值,在分析传播范围、速度及受众等多方面均有指导性意义,为我国的网络信息传播管理做出一份贡献。

5. 模型评价

5.1 模型的优点

1、问题一的解决中，在传统 SIR 模型的基础上引入了潜伏节点，用来模拟还不能传播信息的节点，同时解释 SNS 的信息传播机理，建立了基于 SNS 的信息 SEIR 传播模型。

2、问题二的解决中，具体分析了节点属性和信息内容等多个因素，同时考虑到信息传播时可能出现的延迟现象，进而对影响自媒体平台信息传播的主要因素进行建模。

3、问题三的解决中，利用元胞自动机动态仿真信息传播的过程，以局部的演化来反映整个系统的复杂变化，这与信息的传播机制十分相似，取得很好的效果。

5.2 模型的缺点

1、在建模过程中，忽略了一些次要因素，例如传播客体的特征，这对于现实生活中的影响是十分巨大的。

2、模型所选用的数据不一定具有代表性，利用这些在一般情况下产生数据而得到的普适性的结论可能与实际有稍微偏差。

参考文献

- [1] 司守奎，孙兆亮等，数学建模算法与应用，北京：国防工业出版社，2016.1
- [2] 姜启源、谢金星、叶俊，数学模型（第三版）[M]．北京：高等教育出版社，2003。
- [3] 张彦超、刘云、张海峰、程辉、熊菲，基于在线社交网络的信息传播模型，物理学报，Vol.60,No.5 (2011)。
- [4] 王超、杨旭颖、徐珂，马建峰，基于 SEIR 的社交网络信息传播模型，电子学报，Vol.42 No.11，Nov.2014。
- [5] 孔素真、赵长伟，多关系类型社交网络信息传播模型，河南师范大学学报（自然科学版），Vol.42 No.5，Sept.2014。