

## 第五届“认证杯”数学中国

### 数学建模网络挑战赛

#### 承 诺 书

我们仔细阅读了第五届“认证杯”数学中国数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站([www.madio.net](http://www.madio.net))公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛队号为：1081

参赛队员（签名）：

队员 1：张扬锐

队员 2：

队员 3：

参赛队教练员（签名）：

参赛队伍组别：研究生组

第五届“认证杯”数学中国

数学建模网络挑战赛

编 号 专 用 页

参赛队伍的参赛队号：（请各个参赛队提前填写好）：  
1081

竞赛统一编号（由竞赛组委会送至评委团前编号）：

---

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

# 2012 年第五届“认证杯”数学中国 数学建模网络挑战赛

题 目      社交网络中信息扩散的综合研究

关 键 词      社交网络 层次分析法 差分方程 信息扩散 元胞自动机

## 摘 要：

本文建立了一个社交网络中有关信息扩散的模型。通过社交网络中用户之间的粉丝关系，我们对所给的问题进行了多方面综合的分析，确定了信息在社交网络中的扩散速度和企业为了达到宣传目的所采用的用人方案。

首先，我们确定了影响社交网络中用户转发信息的三个因素，即活跃度、对信息的兴趣和从众指标，通过层次分析法计算权重得到用户转发信息概率。然后，利用差分方程描述了信息覆盖人数的变化。

通过对模型的进一步分析，我们引入了衰减因子来修正模型。通过元胞自动机的模拟，得到对应宣传时间的衰减因子的值，并用灰色预测模型计算出所有衰减系数。对优化后的模型，我们得出奥运会开始后，一条含有企业广告的奥运会新闻大约可以被  $2.0107 \times 10^8$  人观看到。

对于问题二，我们建立了最优化模型，从只考虑雇佣专业推广员宣传和考虑成本两方面对问题进行分析，利用 0.618 法求解，得到最佳的人员配置方案为 23 个专业社交网络推广员、11 个兼职宣传者。

最后，我们对模型进行了灵敏度分析并作出了合理的评价，同时对该类型问题未来的研究方向做出预测。

参赛队号      1081

参赛密码      \_\_\_\_\_  
(由组委会填写)

所选题目      C

## 英文摘要（选填）

### Summary

This paper propose a general model for the information spread on the social network. The problems are given an extensive analysis to get the diffusion speed of information and the best employing scheme through the follow relationship between the users.

We first describe the probability of information spread with three indexes: vitality, interest and conformity. Then, using the Analytical Hierarchy Process, we obtain the probability list. Additionally, we apply difference equation to identify the changes of the covered users. Clear and easy to understand, our model is efficient to deal with the large amount of data sets.

Furthermore, we introduce attenuation factor to fix our model. Through the simulation of cellular automata and the gray model, the values of the attenuation factor corresponds to the time are obtained. We get the result that there are about  $2.0107 \times 10^8$  users have seen the corporate advertising before the Olympic Games through the modified model.

For question two, we have established the optimal model, and analyze the problem from two sides of only consider employment of professional promoters and consider the cost. Using 0.618 method ,we come to the conclusion that 23 professional social network promoters and 11 part-time advocates is the best employing scheme.

Finally, we analysis the model sensitivity , make a reasonable assessment and predict the future research direction of this type of problem.

**Keywords:** social network   AHP   difference equation   information spread  
cellular automaton

## 一、问题重述

从 1984 年的美国洛杉矶奥运会开始，奥运会就不再成为一个“非卖品”，它在向观众诠释更高更快更强的体育精神的同时，也在攫取着巨大的商业价值。它在每个行业中仅挑选一家奥运全球合作伙伴，这就是“Top 赞助商”的前身。

每届奥运会，Top 赞助商的赞助费用都以 10% 至 20% 的速度在增长。宣传费用的门槛把绝大多数企业排除在了奥运会之外。但是越来越多的企业不甘心错过奥运会这个吸引大众眼球的宣传机会，他们在寻找新的新闻传播渠道。法国科技公司源讯 (Atos Origin) 2011 年 10 月公布了一份《奥运会十大科技事实》清单，其中提到 2012 年伦敦奥运会期间，将有 85 亿台平板、智能手机等移动设备联网。

一切都在数字化，数字化不仅仅打碎了时间，它让传播渠道、受众的注意力、品牌营销方式甚至一切都碎片化了，观众不再只关注电视，他们利用社交网络可以获得更加丰富的比赛信息和网友的评论。这也为更多的企业提供了在奥运期间宣传自己的机会。

一家企业想利用社交网络在奥运会期间进行企业宣传，假设现在距离奥运会开幕还有 100 天，一个社交网络的专业推广者平均每天可以新增 500 个粉丝，这些粉丝会把推广者发布的和奥运会相关的所有信息都分享给自己的粉丝们，普通网络用户平均每天可以新增 20 个粉丝。

我们试图解决下列问题：

(1) 建立数学模型，预测奥运会开始后，一条含有企业广告的奥运会新闻可以被多少人观看到？

(2) 假设企业产品的潜在用户大约有 2 亿人，他们都在使用社交网络，企业希望广告宣传覆盖其中 40% 的人群，至少需要雇佣几名专业社交网络推广者才能实现？假设专业推广者每天的工资是 500 元。还可以从网络上雇佣兼职宣传者，每天仅需要付 50 元的工资，但是他们平均每天新增的粉丝数仅为 35 人，考虑到成本，请给企业制定一份合理的用人方案。

## 二、问题分析

目前社交网络有三种主要形式，传统的博客、类似 facebook 的互动网络、微博，社交网络一个重要的特点便是向外传递信息。

从通信角度来看，博客是类似于广播的单工系统，facebook 是互动的双工系统，微博是移动双工系统，这三种形式各自不同的特点，对于信息的传播也会有不同程度的影响，为了能方便解决问题，我们只关心他们共有的传播特点。

基于传播信息和传播者的特点，信息的传播模式又可分为中心式传播，多关键点传播，链式传播等，不同的传播模式有着不同的传播速率。

社交网络的最大特点是以人为中心，依靠人与人之间的好友、粉丝等关系进行信息的传播。如今，人们在获取信息的时候也更加在乎信息的来源。基于社交网络的这一特点，同时为了模型的可行性，我们着重考虑社交网络的粉丝数量这个属性。

通过对社交网络结构特点的分析，一个典型的社交网络具有方向性、时效性、鲁棒性（一个或者一个关注的缺失，并不影响整个系统的运行）等特点，那么一个社交网络可以抽象成为类似无限传感网的模型来理解。具体来说，某个人所有的粉丝都可看做是传感节点，个体之间的follow关系则可以抽象地用节点之间的边来表示，用户依靠粉丝来传递信息（见图1）。

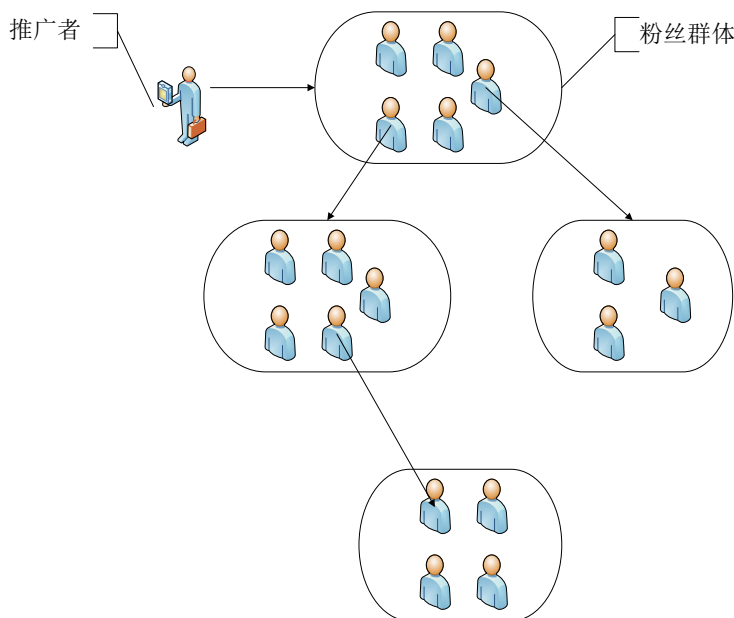


图 1 用户传递信息示意图

对于社交网络中的普通用户，收到信息后是否会分享取决于很多因素，这也直接影响到信息传播的速率。考虑调查研究和数据的可获得性，对于一个特定的社交网络普通用户，如果他的活跃度高，即具有众多粉丝或者喜欢转发信息，那么他分享信息的可能性就大，同理如果用户对某一信息感兴趣或者有从众心理，那么他也会非常乐意转发此消息。由此我们认为用户的活跃度、对信息的兴趣和从众心理是决定分享概率的主要因素。

为了得到信息扩散的特性，我们还需要建立一个传播模型来计算信息扩散的速度来体现受到宣传人数的增长情况。由于消息通过转发这一特性，我们考虑用差分方程来计算宣传人数和制定用人方案。

通过以上分析，我们用如下流程图（见图 2）表示问题解决的步骤。

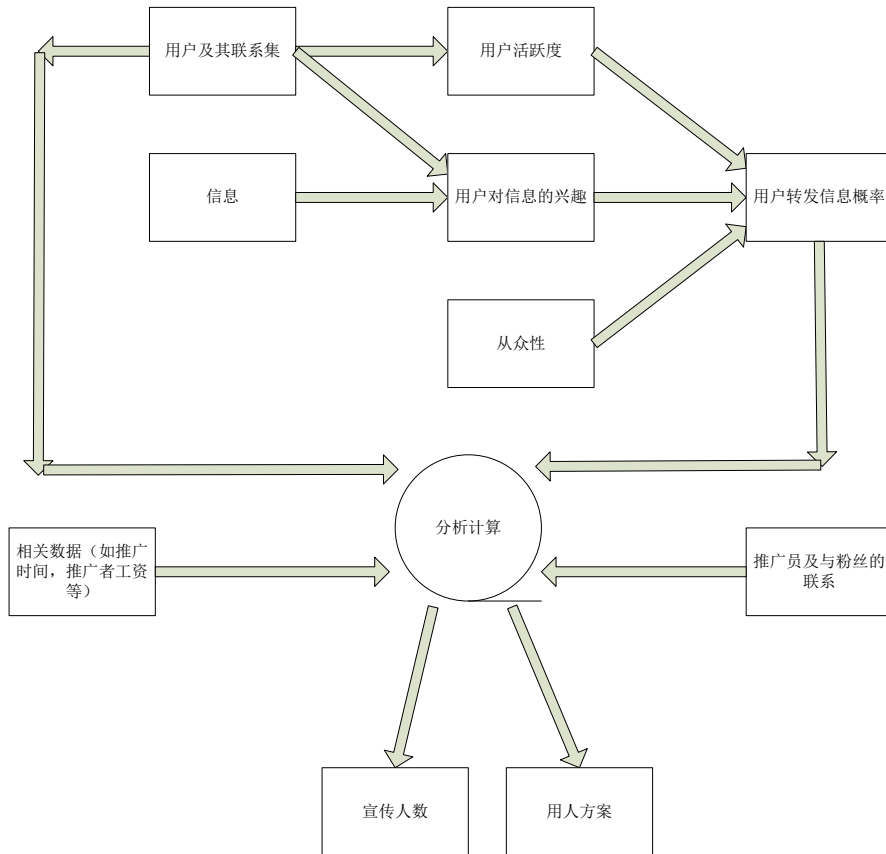


图 2 问题求解的流程图

### 三、模型假设

1. 假设我们所研究的社交网络中信息的扩散方式仅为类似于广播的一对多中心式传播。
2. 假设除了推广者以外，其他用户的社会地位相同，即不具有公众人物的号召影响力。
3. 不考虑一些特定功能对广播式扩散的影响，比如微博@功能的点对点互动。
4. 忽略社会学、经济学等对用户分享信息可能性的影响，认为用户转发信息的概率只与用户活跃度、对信息的兴趣、从众心理有关，且这三种因素之间没有显著相关性。
5. 用户之间的关系只考虑粉丝这一个属性。
6. 从众心理通过信息的转发量体现，当一条信息被大量分享时，用户更倾向于去转发。
7. 假设所研究的社交网络中总体用户的数量在研究时间范围内保持不变。

## 四、模型的建立与求解

### 4.1 问题一

#### 4.1.1 符号约定

$F_i$	用户 $i$ 的粉丝数
$C_i$	用户 $i$ 的活跃度
$I_i$	用户 $i$ 对特定信息（本问题中为奥运新闻）的兴趣
$E$	从众指标
$S_i$	用户 $i$ 转发特定信息的概率
$\{\}$	给定 Twitter 相关数据的集合
$N(t)$	$t$ 时刻已看到信息的人数
$P(t)$	$t$ 时刻所有已看到信息的用户向其未看到信息的粉丝分享的平均概率
$L(t)$	$t$ 时刻已看到信息的用户的还未看到信息的粉丝的数量
$V(t)$	$t$ 时刻粉丝群体间的重叠人数

#### 4.1.2 模型一

##### 1) 确定用户的活跃度、对信息的兴趣、从众指标

用户的活跃度由粉丝的数量直接体现，粉丝数量越多，标明该用户越活跃。三种因素的数值均进行 0 到 1 之间的归一化处理。

$$C_i = \frac{F_i - \text{Min}\{F_i\}}{\text{Max}\{F_i\} - \text{Min}\{F_i\}}$$

考虑到奥运会的特殊性，我们认为所有用户均倾向于对相关信息转发，即用户对奥运信息的兴趣指标为 1 且从众指标也为 1。

$$I_i = 1, E = 1$$

##### 2) 确定用户转发信息的概率

考虑对于用户兴趣和心理的属性相关数据难以获取，我们使用层次分析法来确定三种因素的权重，构造判断矩阵如图 3 所示。



	C	I	E	W
C	1	5	3	0.637
I	1/5	1	1/3	0.105
E	1/3	3	1	0.258

图 3 层次分析法判断矩阵

计算得到一致性指标  $CI = 0.019 < 0.1$ , 我们认为权重的确定结果具有满意的一致性。由权重确定用户转发信息概率

$$S_i = 0.637 \times C_i + 0.105 \times I_i + 0.258 \times E_i$$

对于已知 Twitter 中用户联系数据, 按照次序依次给用户编号, 得到至少有一个粉丝的用户数为 7276。计算每个用户转发奥运信息概率, 得到如下统计图 (见图 4)。

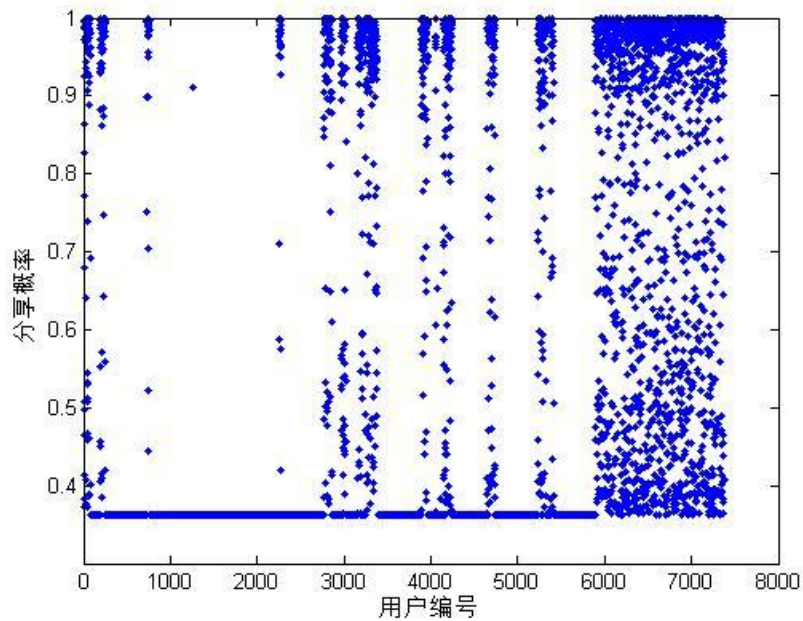


图 4 用户分享信息概率分布图

### 4.1.3 模型二

#### 1) 确定已看到信息人数与时间的关系

为了确定一次消息转发的速度，我们通过查询人人网上相关数据（消息发布时间和回复时间），进行抽样统计调查（如图 5），得到转发消息这一动态过程单位时间 $\Delta t$ 平均值为 2 小时，平均在线时间为 4 小时。

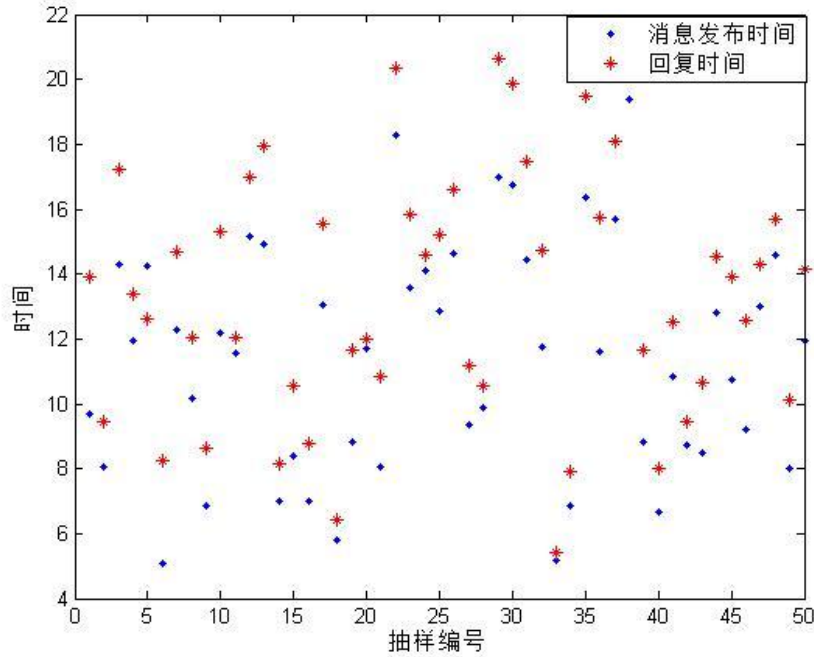


图 5 消息转发数据统计图

在单位时间 $\Delta t$ 内，看到信息所增加的人数为所有已看到信息的用户的粉丝收到分享的人数，其中收到分享的概率即为与之联系的用户分享的平均概率。

$$N(t + \Delta t) - N(t) = P(t)L(t)$$

$P(t)$ 是 $t$ 时刻所有已看到信息的用户向其未看到信息的粉丝分享的平均概率，实际上是与 $t$ 时刻的已看到信息的用户有关，设 $t$ 时刻已看到信息用户的集合为 $M$ ，这些用户的数量记为 $|M| = N(t)$ 。

$$P(t) = \frac{\sum_{i \in M} S_i}{|M|}$$

$L(t)$ 表示 $t$ 时刻已看到信息的用户的还未看到信息的粉丝的人数，即在所有看到信息的用户中，找出他们所有的粉丝，并且排除这些粉丝中已看到分享的一部分，设该部分重叠人数为 $V(t)$ ，剩下的人数为 $L(t)$ 。同时考虑到专业推广者每天新增 500 个粉丝，普通互用每天新增 20 个粉丝，且平均每天进行 2 次消息转发，得到

$$L(t) = \sum_{i \in M} Fi + 10 \times 500 \div 2 + |M| \times 20 \div 2 - V(t)$$

各用户及其粉丝动态变化过程如图 6 所示。

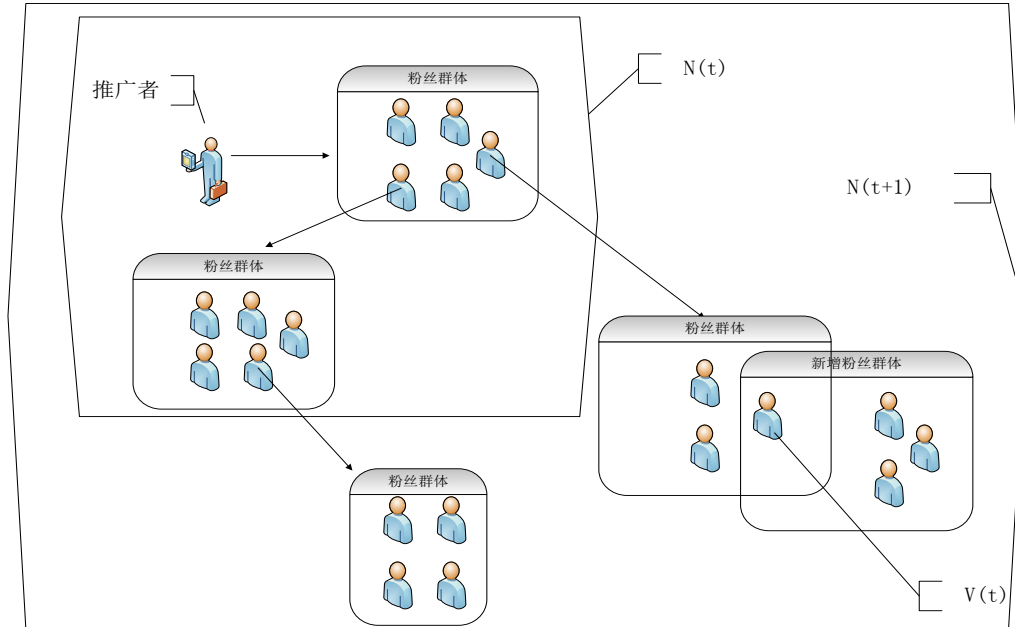


图 6 已收到信息群体的动态变化示意图

确定初始状态各项的值,  $N(0) = 10, P(0) = 1, L(0) = 10 \times 500 \div 2 = 2500$ , 即可得到  $N(t)$  随  $t$  变化的值, 设  $t$  单位为天。

## 2) 模型修正

由于推广员的每日增长粉丝相对于网络中总人数过大, 在所给 Twitter 部分数据做仿真的过程中, 仿真结果并不理想。为了增强模型的可移植性, 适用于一般网络以便研究问题中奥运信息的推广, 对模型做出如下修改和假设。

通过数据查询, 目前社交网络总用户为十二亿, 为了估算 100 天信息的传播量, 粗略认为所有社交网络为一个大的整体网络, 即忽略不同地区不同社交网络的无法互通的特性。同时, 由所给 Twitter 数据库分析, 并由之前已经建立的模型得到对于奥运信息, 用户的平均转发概率为 0.5047, 用户的平均粉丝数为 111.9681, 不考虑个体间差异, 将所有用户 (推广员除外) 一般化。

另一方面在仿真过程中, 如果用户不行添加粉丝, Twitter 数据库中的用户联系能模拟出每次信息传播后粉丝的重叠, 一旦加入每日添加粉丝的考虑, 尤其是对于普通用户, 20 个粉丝的增长量在 100 天后粉丝群体会变得非常大, 其过程中粉丝的重叠将很大程度影响已看到信息的用户数量。

我们提出一种简化思想, 不考虑粉丝的重叠, 得到  $N(t)$  随  $t$  变化的图像 (见图 7)。

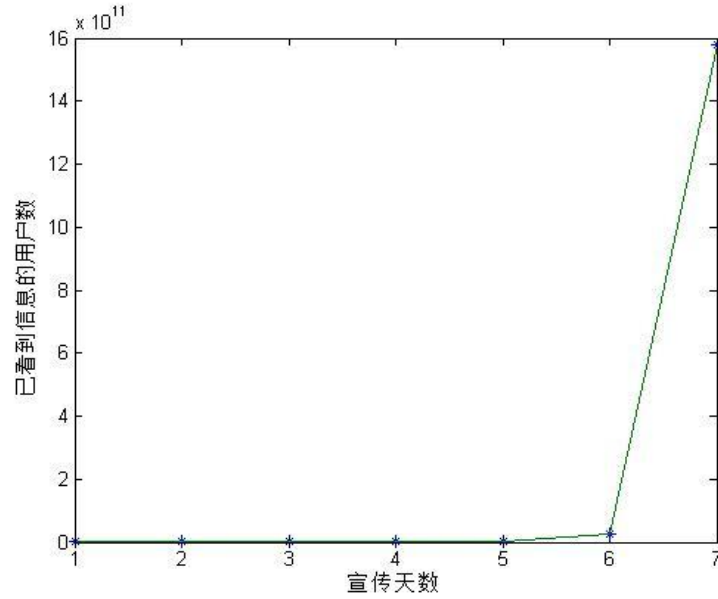


图 7 已看到信息的用户数随宣传天数的变化

由图像结果可知，一条比较重要的信息将在 7 天内覆盖社交网络全部 12 亿用户，这显然是不合理的，因此，用户间粉丝的重叠必须考虑。

通过进一步分析，由于用户粉丝数随天数增长，随着覆盖范围的逐步扩大，重复的粉丝必定越来越多，我们无法通过现有数据来准确计算这个数值。

通过对模型二的进一步分析，我们提出一种简化思想，设定衰减因子 $\beta(t)$ 表示粉丝的重叠对 $N(t)$ 的影响，得到修改后的模型

$$L(t) = \beta(t) \left( \sum_{i \in M} Fi + 10 \times 500 \div 2 + |M| \times 20 \div 2 \right)$$

为了确定衰减因子 $\beta(t)$ ，我们需要建立一个小范围的模拟的社交网络进行仿真，对比实际仿真结果和不考虑衰减因子情况下的 $L(t)$ 。

#### 4.1.4 模型三

我们将二维元胞自动机作为一种模拟信息传播行为的工具，为了能让元胞自动机适用于我们所研究的社交网络，需要对网络的一些参数做出修改。

针对模型一中已求得的数据，用户的平均转发概率为 0.5047，用户的平均粉丝数为 111.9681，将用户平均粉丝数改为 8，为了满足增长比例的相同，假设普通用户的粉丝数平均每天增长 1.64。

定义元胞自动机模型为：

1. 元胞分布于规则划分的二维网格上(100×100)；
2. 元胞具有两种状态：0 表示没有看到信息，1 表示已经看到信息；
3. 元胞的状态由邻居状态决定，邻居为与元胞相邻的8个元胞；
4. 演化规则以及算法：
  - 1) 初始状态：随机100个元胞状态为1，其余为0（假设有一个推广员，他的粉丝数为100）

- 2) 考虑每一个状态为1的元胞的邻居，若邻居状态为0，通过计算机产生随机数，令该邻居以0.5047的概率变为1；若邻居状态为1，不予考虑；
- 3) 每次演变后，记录当前网格中状态为1的数量 $n$ ，并在网格中随机选取 $1.64n$ 个元胞，表示粉丝每天增长的情况。若状态为0，通过计算机产生随机数，令该邻居以0.5047的概率变为1；若状态为1，表示粉丝重叠的情况，不再重复考虑；

仿真结果如下图（见图8、图9）所示，分别表示演变的大体趋势和在100天（即200次单位时间演化）时间点上多次模拟的平均结果。

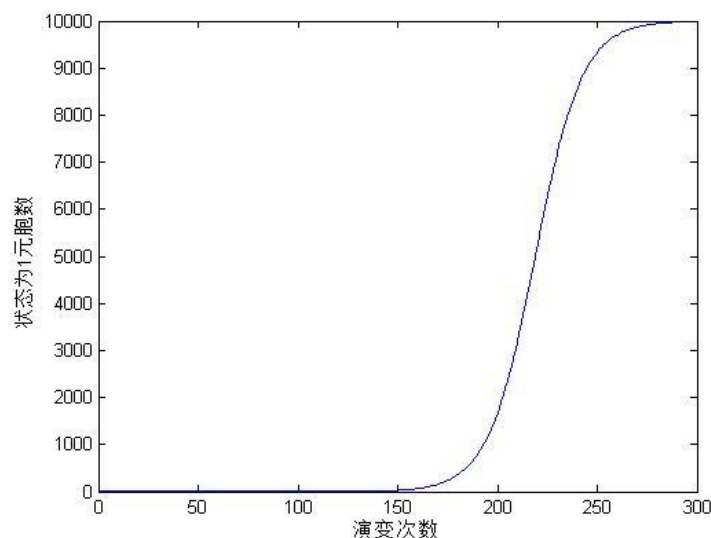


图 8 状态为 1 元胞数随演变次数变化情况

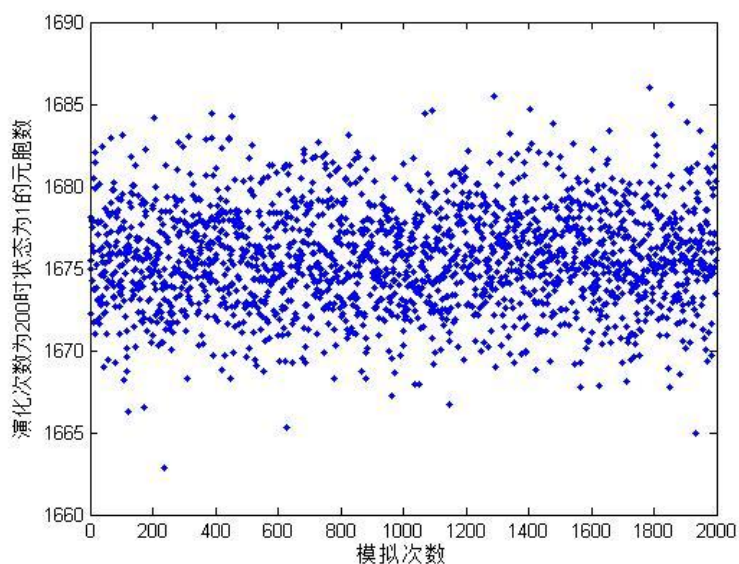


图 9 演化次数为 200 时状态为 1 的元胞数进行 2000 次模拟的结果

图 8 的结果比较好的模拟了信息扩散的实际情况，开始阶段因为基数的数量

小,增长速度随之较慢;随着看到信息的人数越来越多,信息传播的速率也越快;在信息大面积覆盖后,信息传播速率趋于缓和,最后所有人都能收到信息。

我们对之前不考虑衰减因子情况下的模型三做相同模拟,得到一个类似于指数增长的模拟趋势(见图 10)。

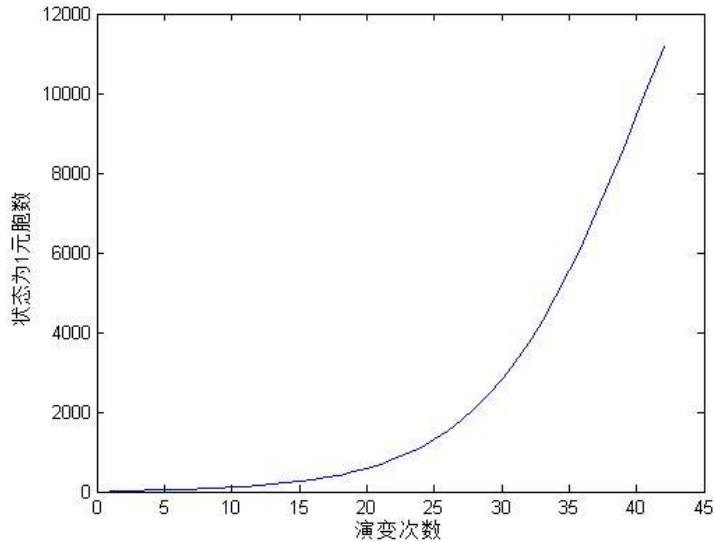


图 10 不考虑衰减因子时状态为 1 元胞数  
随演变次数变化情况

由此得到衰减因子 $\beta(t)$ 随 $t$ 的变化情况(见图 11,  $t$  表示天数, 等同于两个演化次数)。

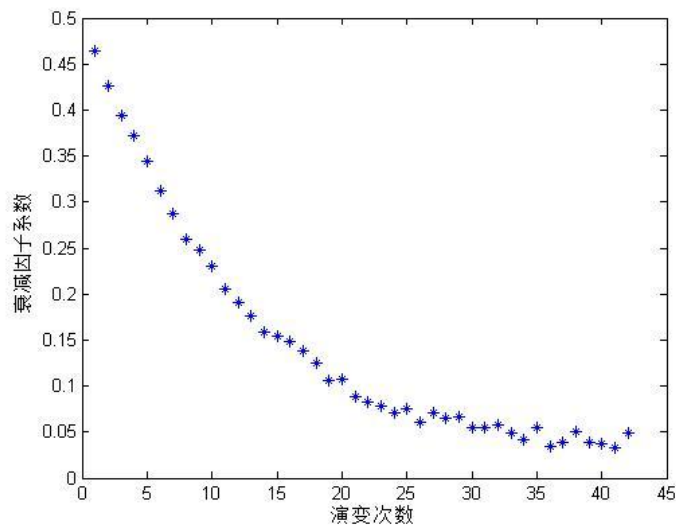


图 11 衰减因子随演变次数变化

由于在图 10 中演变次数到 40 左右, 网格中所有元胞状态均为 1, 而在模型三种我们至少需要得到 200 次演化(100 天)的衰减因子系数, 故需要对衰减因子做预测分析。

#### 4.1.4 模型四

考虑到衰减因子随演变次数变化规律性较强, 具有明显的指数函数递减趋势, 我们采用单序列一阶线性微分方程模型 GM (1, 1) 对数据做预测。

##### 1. 模型建立

设有原始数据列  $x^{(0)} = (x^{(0)}(1), x^{(0)}(2), \dots, x^{(0)}(n))$ ,  $n$  为数据个数。

(1) 原始数据累加以便弱化随机序列的波动性和随机性, 得到新数据序列:

$$x^{(1)} = (x^{(1)}(1), x^{(1)}(2), \dots, x^{(1)}(n))$$

其中,  $x^{(1)}$  中各数据表示对应前几项数据的累加。

$$x^{(1)}(t) = \sum_{k=1}^t x^{(0)}(k), t = 1, 2, \dots, n$$

(2) 将  $x^{(1)}$  视为一连续函数, 对  $x^{(1)}$  建立一阶线性微分方程:

$$\frac{dx^{(1)}}{dt} + ax^{(1)} = u$$

其中,  $a, u$  为待定系数, 分别称为发展系数和灰色作用量,  $a$  的有效区间是  $(-2, 2)$ , 并记  $a, u$  构成的矩阵为

$$\partial = \begin{bmatrix} a \\ u \end{bmatrix}$$

只要求出参数  $a, u$  就能求出  $s(t)$ , 进而求出  $x$  的未来预测值。

(3) 对累加生成数据做均值并进行处理后生成  $B$  与常数项向量  $Y_n$ , 即

$$B = \begin{bmatrix} -0.5 \times (x^{(1)}(1) + x^{(1)}(2)) & 1 \\ -0.5 \times (x^{(1)}(2) + x^{(1)}(3)) & 1 \\ \vdots & \vdots \\ -0.5 \times (x^{(1)}(n-1) + x^{(1)}(n)) & 1 \end{bmatrix}, Y_n = (x^{(0)}(2), x^{(0)}(3), \dots, x^{(0)}(n))^T$$

(4) 用最小二乘法求解灰参数  $\partial$ , 则

$$\partial = \begin{bmatrix} a \\ u \end{bmatrix} = (B^T B)^{-1} B^T Y_n$$

(5) 将灰参数  $\partial$  代入  $\frac{dx^{(1)}}{dt} + ax^{(1)}(t) = u$ , 进行求解, 得

$$\hat{x}^{(1)}(i+1) = \left(x^{(0)}(1) - \frac{u}{a}\right) e^{-ai} + \frac{u}{a}$$

由于  $\partial$  是通过最小二乘法求出的近似值, 所以  $\hat{x}^{(1)}(i+1)$  是一个近似表达式

(即为预测值), 为了与原序列  $x^{(1)}(i+1)$  区分开来, 故记为  $\hat{x}^{(1)}(i+1)$ 。

(6) 对函数表达式  $\hat{x}^{(1)}(i+1)$  及  $\hat{x}^{(1)}(i)$  进行离散, 并将二者做差以便还原  $x$  原序列,

得到近似数据序列  $\hat{x}^{(0)}(i+1)$  如下:  $\hat{x}^{(0)}(i+1) = \hat{x}^{(1)}(i+1) - \hat{x}^{(1)}(i)$

(7) 对建立的灰色模型进行检验, 步骤如下:

① 计算  $x^{(0)}(t)$  与  $\hat{x}^{(0)}(t)$  之间的残差  $e(t)$  和相对误差  $q(t)$ :

$$e(t) = x^{(0)}(t) - \hat{x}^{(0)}(t)$$

$$q(t) = \frac{e(t)}{x^{(0)}(t)}$$

② 求原始数据  $x^{(0)}(t)$  的方差  $s_1$

③ 求残差  $e(t)$  的方差  $s_2$

④ 计算方差比  $C = \frac{s_2}{s_1}$  和  $q(t)$  的平均值  $\bar{q}$

⑤ 灰色模型精度检验如表格2所列

等级	相对误差 $\bar{q}$	方差比
I	$<0.01$	$<0.35$
II	$<0.05$	$<0.50$
III	$<0.10$	$<0.65$
IV	$>0.20$	$>0.80$

## 2. 模型求解

(1) 对数据做一次累加, 得

$x^{(1)} = (0.8812, 0.8095, 0.7437, 0.6832, 0.6277, 0.5768, 0.5300, 0.4870, 0.4476, 0.4114, 0.3782, 0.3478, 0.3199, 0.2943, 0.2708, 0.2493, 0.2296, 0.2115, 0.1950, 0.1799, 0.1662, 0.1536, 0.1421, 0.1317, 0.1223, 0.1137, 0.1059, 0.0990, 0.0927, 0.0871, 0.0822, 0.0778, 0.0740, 0.0708, 0.0681, 0.0658, 0.0640, 0.0627, 0.0618, 0.0614, 0.0614)$

(2) 计算得  $\hat{\theta} = \begin{bmatrix} a \\ u \end{bmatrix} = \begin{bmatrix} 0.0692 \\ 29.5322 \end{bmatrix}$

(3) 为了得到 100 天的宣传天数并说明传播规律, 我们计算了 200 个预测值, 即加上原始数据时间长度一共为 121 天。

(4) 对模型进行检验

$$C = 0.1698$$

经过验证, 该模型的精度较高, 可以用来预测。还原后得到的数据如图所示 (见图12)。



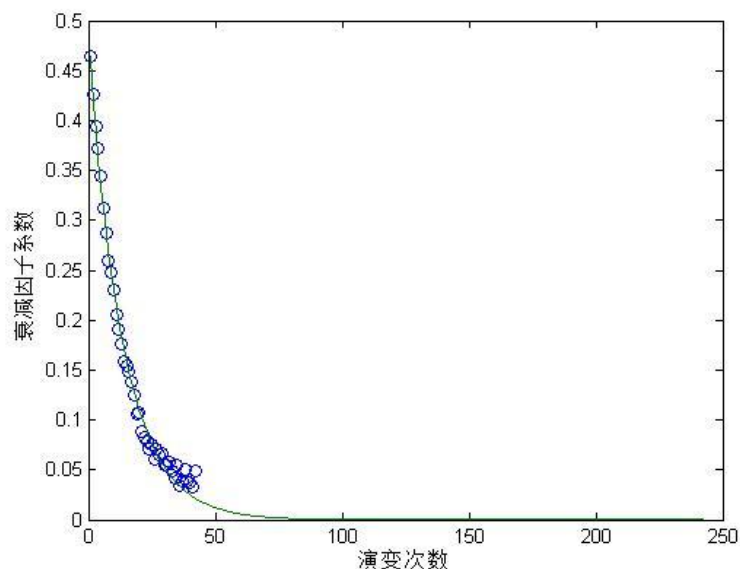


图11 预测演变次数在42次以后衰减因子系数的变化

#### 4.1.5 结果的分析和计算

经过分析，考虑到实际网络中几种数据的值均有改变，演变次数实际是网络中所有状态为 1 的元胞个数的体现，衰减因子实质上是由当前网络中所有已看到信息的用户数直接决定，故对衰减因子进一步按倍数缩小，得到模型三中所需要的衰减因子具体值带入修改后的模型三中，计算结果如下图所示（见图 12）。

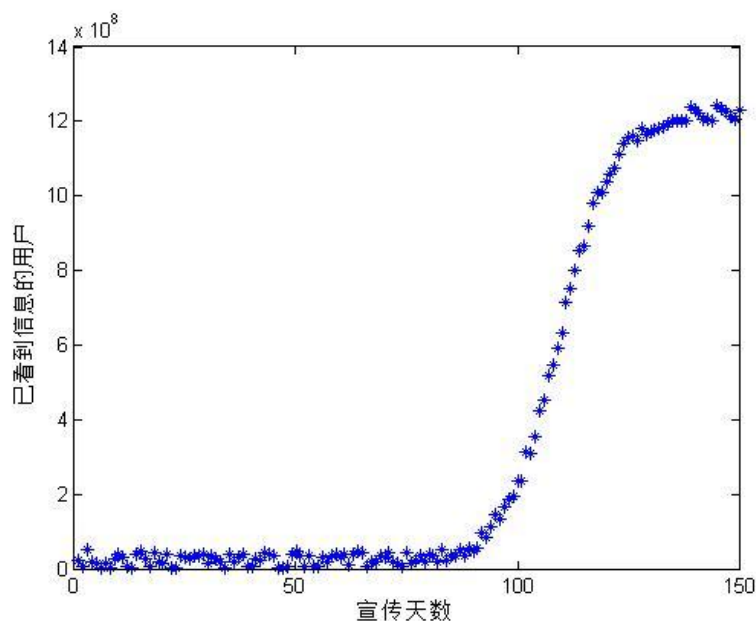


图 12 引入衰减因子后已看到信息的用户随宣传天数的变化

取宣传天数为 100 时已经看到信息的用户数，得到奥运会开始后，一条含有

企业广告的奥运会新闻大约可以被  $2.0107 \times 10^8$  人观看到。

## 4.2 问题二

企业希望广告宣传覆盖 2 亿潜在用户 40% 的人群。我们不妨看做这 2 亿用户是一个单独的网络，由问题一中已建模型得到 100 天后该广告能被 0.3241 亿用户看到，覆盖了 16.21% 的潜在用户。

1. 只考虑雇佣专业推广员宣传

设公司需雇佣专业社交网络推广员数为  $x$ ，需要满足条件

$$N(100) > 2 \times 10^8 \times 0.4$$

$$L(t) = \beta(t) \left( \sum_{i \in M} Fi + 250x + |M| \times 20 \div 2 \right)$$

其余条件和问题一中建立模型相同，公式中  $Fi$  同样受刚开始推广员数量的影响，推广员越多，开始时刻受到宣传的粉丝就越多，因此不能直接由不等式得到结果，需要从初始条件通过差分方程逐步计算。

为了快速找到事宜的推官员数，我们先假设  $x=50$ ，得到结果为 100 天时宣传覆盖率超过了预期的比例 40%，由此以 10 和 50 作为上下界进行 0.618 法依次查找  $N(100)$  的值，得到  $x=26$ 。

2. 考虑成本

由问题所给条件建立使用  $x$  个专业社交网络推广员、 $y$  个兼职宣传者成本指标  $Z(x, y)$ ，得到最优化模型

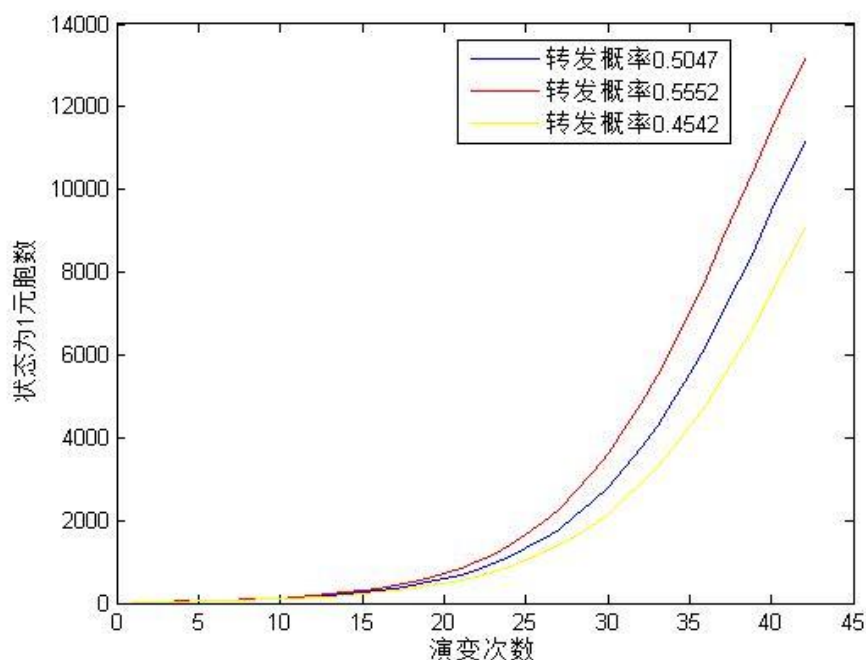
$$\begin{aligned} \min Z(x, y) &= 100 \times (500x + 50y) \\ \text{s.t.} \quad &\begin{cases} N(100) > 2 \times 10^8 \times 0.4 \\ N(0) = x + y \\ N(t+1) - N(t) = 0.5047\beta(t) \left( \sum_{i \in M} Fi + 250x + |M| \times 20 \div 2 \right) \\ \beta(t), M, Fi \text{ 值同之前模型} \\ 26 \geq x \geq 0, y \geq 0, x, y \text{ 均为整数} \end{cases} \end{aligned}$$

为了能用相关软件(lingo)对模型求解，需要对模型进行简化，对于  $Fi$  由于没有具体网络关系可以模拟，我们用平均粉丝数 111.9681 代替，同时  $|M| = N(t)$ ，解得用人方案为  $x=23$ ， $y=11$ ，100 天宣传的成本为 1205000 元。

## 五、模型灵敏度分析

根据第一问的模型，用户对信息转发的概率的确定是决定信息传播速率的重要指标，我们需要讨论转发概率的改变对模型结果的影响。

着重探讨元胞自动机进行模拟时转发概率变化对不考虑衰减因子情况下的模型三结果的改变，将其增加或减少 10%，做相同模拟，得到对比曲线如下：



分析得知,该图所得到数据直接决定了衰减因子的大小,在演变次数较小时,转发概率对信息传播影响较小,而当演变次数增大时,转发概率对传播速率起着决定性影响。所以我们在研究较长时间信息的传播问题时,传播概率的计算和确定尤为重要。

## 六、模型的评价与推广

### 6.1 模型的优点

1. 逐步分析并且改进建立的模型,模型之间层层递进、相互关联,最终得到问题的结果。
2. 通过数据分析,模型得到合理优化。在模型二建立求解过程中得知模型的要求难以实现,经分析后建立模型三并修改模型二,引入衰减因子,不仅通过模型计算出结果,同时也使模型得到了验证。
3. 模型中假设有理有据,并且符合实际情况。模型二处理过程中发现假设与实际情况不符,我们因此做了进一步的分析并提出了修改方案。
4. 模型结构清晰,逻辑明了,数据处理算法简单,对大型数据的处理有较高的实效性。

### 6.2 模型的缺点

1. 模型一中层次分析法带有主观因素。对于人们心里活动的研究难于用数字衡量,如果能对这方面有较精确的指标,那么模型将会更加准确和可信。
2. 一些情况考虑得比较单一,比如用户之间关系、影响用户转发信息因素。

### 6.3 模型未来的研究

随着近年来社交网络的迅猛发展,对社交网络的研究也越来越多,目前的研究也多是从小复杂网络、数据挖掘等角度出发。

实际生活中用户转发信息受很多因素影响,包括心理学、社会学、行为学等方面,对于信息传播的研究,以后的研究应该是从这些角度去考虑,重点在于用户信息的多方面多属性获取并且把这些数据与用户的具体活动和心理联系起来。

## 七、参考文献

- [1] 陆丽华、罗鹏飞、李星等,网络恶意移动代码扩散模型综述[J],计算机应用, 2003, 23(6): 18-22
- [2] 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, 60(5), 2010
- [3] Lin Guoji, Jia Xun, Ouyang Qi, Predict SARS infection with the small world network model, Journal of Peking University (Health Sciences), vol. 35(z1), 2003, 66-69
- [4] Stephen Eubank, Hasan Guclu, V. S. Anil Kumar, et al. Modeling disease outbreaks in realistic urban social networks. Nature, 2004. 5, 429, 180-184
- [5] Grabowski A., Kosinski R. A., Epidemic spreading in a hierarchical social network, Phys. Rev. E, 2004, 70(3): 031908-1~031908-7