

# 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：

参赛队员（签名）：

队员 1：

队员 2：

队员 3：

参赛队教练员（签名）：

参赛队伍组别：本科

# 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）：

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 碎片化趋势下的奥运会商业模式

关 键 词 SIR 模型 BA 无标度网络 信息传播 计算机仿真

## 摘 要：

本文主要研究了信息在社会网络中的传播过程。首先，基于现实的复杂网络特征，构建与之相似的 BA 无标度网络；其次，利用构建的无标度网络对研究的问题建立模型；最后，利用计算机仿真得出最终结果。

针对问题 1 中 SNS 网络中企业广告的传播，建立经典 SIR 传播模型进行求解，迭代 100 次后，全网节点均被感染，与现实不符合。因此对模型做出改进，由 BA 无标度复杂网络构造  $1000 \times 1000$  的 0-1 矩阵类比于真实的 SNS 网络。采用一定的节点转换规则，对 BA 网络的节点进行 100 次迭代。迭代过后全网有近 63.5% 的节点被感染，与实际较符。因此，可预测最终的信息接收人数为  $N \times 63.5\%$

针对问题 2，主要采用问题 1 的模型，构建一个包含 100 个节点的无标度网络，以此代替整个社会网络的中心网络，以此小型网络为中心向外扩散传播。通过合理假设并忽略对信息接收人数影响较小的因素，建立了最终的信息接收人数的模型为：

$$M(n) = [s \times 500 \times 20 \times n + h \times 35 \times 20 \times n + N(n) \times 20 \times n] \times \lambda(n),$$

并且构建目标函数为：

$$F = s \times 500 \times 100 + h \times 50 \times 100。$$

通过搜索算法则得出使得目标函数最小的安排方案为专业推广者雇佣 12 人，兼职推广者雇佣 20 人，所花费成本为 70 万元。

最后，对问题 1、问题 2 的模型进行改进，加入更加符合实际的影响因素，使得仿真结果更加符合现实。

参赛队号 1090

所选题目 C 题

参赛密码 \_\_\_\_\_

(由组委会填写)

## 1 问题的重述

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，它与电视结盟，在运动员入场仪式、颁奖仪式、热门赛事、金牌榜发布等受关注的时刻发布赞助商广告，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。2008 年，北京奥运会全球合作伙伴最低赞助为 6000 万美元，2012 年伦敦奥运会就变成 8000 万美元。这种模式被奥运会主办方发挥到了极致，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。

现在是一个机会，电视正在受到冲击，法国科技公司源讯 (Atos Origin) 2011 年 10 月便公布了一份《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。他们可以自己决定看什么，定制内容，并可以通过社交网络和志同道合者相互吐槽。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。

**问题一：** 请建立数学模型，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？

**问题二：** 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。

## 2 模型假设及符号说明

### 2.1 符号说明

符号	符号说明
$N$	系统总人数
$s(t)$	广告传播者占总人数 $N$ 的比例
$i(t)$	广告未被感染者占总人数 $N$ 的比例
$r(t)$	广告免疫者占总人数 $N$ 的比例
$k$	网络节点的度
$a_{ij}$	由 $BA$ 无标度网络形成 $0-1$ 矩阵中的元素
$F$	雇佣推广者的成本费用
$M$	信息传播的最终接收人数
$s$	雇佣的专业推广者人数
$h$	雇佣的兼职推广者人数

### 2.2 模型假设

- 1、真实  $SNS$  网络可由  $BA$  无标度网络构造出  $1000 \times 1000$  的  $0-1$  矩阵进行替代
- 2、 $BA$  网络仿真传播中拟定新增节点在每次迭代以恒定数目加入。
- 3、在  $BA$  网络  $0-1$  矩阵中节点有且仅有三种状态

## 3 问题分析

针对问题 1 的分析,由于信息传播的多样性, $SNS$  网络已成为重要的广告媒介。 $SNS$  网络上企业广告传播问题,就其广告传播过程中,存在三类  $SNS$  用户,没有接受到广告的用户;接收到但未传播;以及接收到并传播的用户。其传播规律十分类似于传染病模型,我们参考经典传播模型  $SIR$ , 进行分析求解。其求解结果与实际不符合。对此,进行模型改进。

由大量的实证结果表明,许多真实网络结构可归于无标度复杂网络,经数据分析可知, $SNS$  网络的度服从无标度网络的幂律分布,因此,由可采用  $BA$  无标度网络构造出

1000×1000的0-1矩阵对SNS进行仿真，结果更为准确。

针对问题2的分析，我们利用问题1的模型，生成了一个包含100位用户的无标度网络，用此网络来代替整个社会网络的中心网络，并以此小型网络为基础，向外进行信息传播，采取合理的假设，忽略传播过程中对人数影响较小的因素，建立最终信息接收人数的模型，再构建使成本费用最低的目标函数，将问题转化为优化问题，利用搜索算法，即可求得使成本费用最低，并且满足信息接收人数约束条件的最优解。

## 4 模型的建立

### 4.1 问题1的模型

根据题目要求，先将附件中的数据进行整合、筛选等预处理。利用mysql数据库软件对数据做统计分析，分析其信息的传播过程，可以将本问题近似为传染病传播问题，利用传染病模型中的SIR模型，修正相应的系数，即可预测出信息接收者的最终人数。

#### 4.1.1 SIR模型的建立

##### 4.1.1.1 信息传播机理

在信息传播网络中，一个人发布的信息会被其好友看到，并且对信息进行传播、分享，同时，若其好友对其内容不感兴趣则成为“免疫者”且不会传播，本文把传播网络中的用户定义为疾病传播中的个体，则信息在用户网络中的传播可抽象为传染病的传播过程。

根据信息在社交网络中的传播规律，我们把网络中的用户分为三类：传播者，未被感染者，免疫者。传播者表示该用户向其粉丝传播信息，即传染病模型中的感染者。未被感染者表示该用户没有接受信息，即传染病模型中的健康者。免疫者表示该用户已经接受到了信息，但不具备传播信息的能力。即传染病模型中已经被治愈好的病人。

初始的传播者为随机选取的节点，结合传染病模型的传播机理则可建立网络中的信息传播模型。

##### 4.1.1.2 模型的初步建立

根据假设可知，总人数 $N$ 不变，用户分为了传播者、未被感染者、免疫者三类，三类用户在总人数 $N$ 中所占的比例分别记作 $s(t)$ 、 $i(t)$ 和 $r(t)$ 。将所有用户每天新增加的粉丝数定义为传播者每天的接触人数 $\lambda$ ， $p$ 为未被感染者接触传播者过后变为传播者的概率，则传播者的日有效接触率为 $\lambda p$ 。本文中，假设初始传播者即为专业推广者，并且不会随着时间的推移变为免疫者，则其有效接触率为 $\lambda' p (\lambda' = 500)$ ，其他非专业传播

者的日有效接触率为  $\lambda''p(\lambda'' = 20)$ 。每天传播者变为免疫者的人数占传播者总数的比例为  $u$  [1]。

由以上的假设，显然有

$$s(t) + i(t) + r(t) = 1 \quad (1)$$

对于免疫者而言应有

$$N \frac{dr}{dt} = uNi \quad (2)$$

记初始时刻的传播者和未被感染者的比例分别是  $s_0(s_0 > 0)$  和  $i_0(i_0 > 0)$ ，免疫者的初始值为  $r_0 = 0$ ，则可建立 *SIR* 模型的方程为：

$$\begin{cases} \frac{di}{dt} = p\lambda's(t)i_0 + p\lambda''s(t)[i(t) - i_0] - u[i(t) - i_0] & i(0) = i_0 \\ \frac{ds}{dt} = -p\lambda's(t)i_0 - p\lambda''s(t)[i(t) - i_0] & s(0) = s_0 \end{cases} \quad (3)$$

利用 matlab 编程作数值计算，则可得到  $i(t)$  与  $s(t)$  的变化趋势，即可预测 100 天后接受到信息的人数的比例为  $1 - s(100)$ 。

### 4.1.2 BA 网络模型的建立

利用传染病模型无法精确的描述信息在网络中的传播过程，因此我们利用无标度网络对模型作进一步的改进。

#### 4.1.2.1 BA 网络

度 (Degree) 是评估复杂网络节点重要性最简单直接的指标，通常对于人们最直观的认识就是网络中度最大的节点就是网络中最重要的节点。对于一个节点的度  $d_i$ ，是指与该节点相连接的边数，它是这样被定义的：

$$d_i = \sum_{j \in R} \delta_j^i, \delta_j^i = \begin{cases} 1 \\ 0 \end{cases} \quad (4)$$

$i$  与  $j$  之间有边直接相连， $i$  与  $j$  之间无边直接相连，则  $d_i = 0$ 。节点的度数体现了该节点与周边节点的直接联系，一定程度上反映了该节点的重要性，将网络中节点的度数进行排序，可以发掘重要节点。

企业广告新闻在 SNS 网络中的推广，可看成是遵从于某一规律的网络传播行为。如果人们把 SNS 网络中的每一用户看作一个节点的话，那么 SNS 网络就可当作一个复杂网

络问题来研究。企业广告在 SNS 中的传播就可看作是在复杂网络中的传播行为。

在考虑网络特征的时，常使用两个特征来衡量网络[2]：

特征路径长度 (characteristic path length)：在网络中，任选两个节点， $i$  和  $j$ ，它们之间的距离  $d_{ij}$  定义为连接着两个节点的最短路径上的边数。网络中任意两个节点之间的距离的最大值称为网络的直径 (diameter)，记为  $D$ ，即  $D = \max d_{ij}$ ，网络的平均路径长度  $L$  定义为任意两个节点之间的距离的平均值，即

$$L = \left( \frac{1}{\frac{1}{2} N(N+1)} \right) \sum_{i \geq j} d_{ij} \quad (5)$$

其中  $N$  为网络节点数，这是网络的全局特征。

聚类系数 (clustering coefficient)：假设某个节点有  $k$  个边，则这  $k$  条边连接的节点 ( $k$  个) 之间最多可能存在的边的个数为  $\frac{k(k-1)}{2}$ ，用实际存在的边数除以最多可能存在的边数得到的分数值，定义为这个节点的聚合系数。所有节点的聚类系数的均值定义为网络的聚类系数。聚类系数是网络的局部特征，反映了相邻两个人之间朋友圈子的重合度，即该节点的朋友之间也是朋友的程度。

规则网络的特征路径长度大，聚类系数高；随机网络特征路径长度小，聚类系数低。现实中网络大多既非规则，也非随机网络。

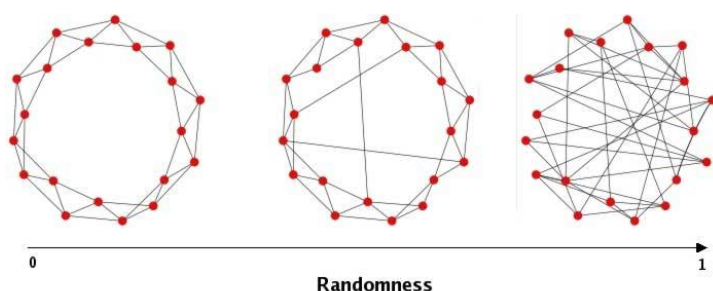


图 1 从规则网络到随机网络的演化

大量的实证结果表明，许多真实复杂网络结构往往呈现小世界特性，即特征路径长度小，聚类系数高。并且具有标度不变性，及无标度网络，节点的度服从幂率分布，即

$$p(k) \propto k^{-\gamma} \quad (6)$$

其中， $\gamma$  为幂率指数。

现用 mysql 对原始数据进行统计分析，去除冗余数据，求出每位 twitter 用户的连接关系数，也及复杂网络中节点的度。

导入 matlab，绘出原始数据度分布图，如下：



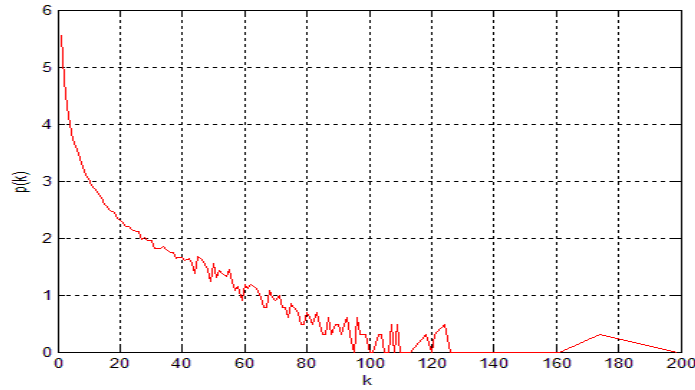


图 2 原始数据度分布图

其中， $k$  为每个节点所对应的度， $p(k)$  为对度为  $k$  的节点数目取 10 为底的对数。

据图像分析可知，twitter 用户的度分布大体上满足负指数分布。及体现了原始数据呈无标度复杂网络的特性。

#### 4.1.2.2 计算机网络仿真

由于原始数据量大，处理冗余数据后，节点数仍有  $4 \times 10^5$ ，难以在原有数据基础上进行复杂网络分析，因此，本文基于其无标度复杂网络的特性，构造与原数据网络结构相似的无标度 BA 网络进行研究。

无标度网络即为服从幂律分布的网络。BA 网络模型是第一个无标度网络演化模型。其构造算法如下：

1) 增长：在初始状态，假定系统中已有少量  $m_0$  个节点，在以后的每一时间间隔中，新增度为  $m$  的点 ( $m \leq m_0$ )，并将  $m$  条边连接到网络中已存在的  $m$  个不同节点上。

2) 择优连接：当于网络中选择节点与新增节点连接时，假定被选择的节点  $v$  与新节点连接的概率  $\Pi(k_v)$  和节点  $v$  的度成正比，即  $\Pi(k_v) = \frac{k_v}{\sum_j k_j}$ 。经过  $t$  个时间间隔后，便会

形成一个有  $N = m_0 + t$  个节点、 $mt$  条边的网络。该网络最终演化为标度不变状态，即几点具有度  $k$  的概率服从度指数  $\gamma = 3$  的幂律分布。

运用 matlab 实现算法，构造  $1000 \times 1000$  的 BA 网络模型，采用 0-1 矩阵表示。

$$a = \begin{bmatrix} 0 & 0 & \dots & \dots & 1 & \dots & 0 \\ 0 & 0 & \dots & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & 0 & \dots & 1 & \dots \\ \dots & 1 & 0 & 1 & 1 & \dots & 1 \\ 1 & \dots & \dots & 1 & \dots & 0 & \dots \\ \dots & 0 & 1 & \dots & 0 & \dots & \dots \\ 0 & 0 & \dots & 1 & \dots & \dots & 0 \end{bmatrix}$$

其中,  $a_{ij} = \begin{cases} 0 \\ 1 \end{cases}$ , 当  $a_{ij} = 0$  时表示该节点无连接, 当  $a_{ij} = 1$  时表示该节点存在连接。

我们认为矩阵中的节点能够处于 3 类状态。

状态 0: 节点  $i$  还未接受其它节点传来的信息。

状态 1: 节点  $i$  接收到其他节点传来的信息, 还未转发。

状态 2: 节点  $i$  转发信息到其邻近节点, 而后成信息免疫者。

状态为 0 的节点如果没有接受到其它节点传来的信息, 其状态仍然是 0, 如果接受到了其他节点传来的一条信息, 其状态变为 1; 处于状态 1 的节点, 如果没有转发信息, 其状态仍为 1, 如果转发信息将进入状态 0; 状态 2 的节点考虑其已经接受到了信息并已经向它的邻接节点转发了信息, 所以它不会再转发信息, 保持状态 2 不变。

设  $BA$  系统初始时仅有一个节点处于状态 1。她将去搜索 0-1 矩阵中与其相连的节点。并将其信息以一定的概率传发给其邻接节点。

本文参考二八原理, 拟定节点的接收概率与转发概率均为 20%。

对于新粉丝的联入, 我们按 0-1 矩阵的每行进行加入新粉丝操作, 每步仅加入一个新粉丝。在第一行中, 随机抽取第  $j$  列, 如果第  $j$  列为 0, 则将其 0 变为 1, 若不为 0, 则再在此行进行随机抽取列, 直至此列数为 0, 并将其 0 置 1。同样, 第二次添加新粉丝时, 在第二行中重复进行在第一行中的操作。以此类推, 在一轮迭代中添加 1000 个新粉丝, 进行 100 次的迭代。

进行 matlab 仿真, 系统在 100 次迭代之前, 将会趋于一个稳态, 并最终求得一个稳定比例。

## 4.2 问题 2 的模型

问题 (1) 中, 我们已经建立了模拟社会网络的小型网络, 其特性可以反映社会网络的部分复杂性质, 因此我们可以根据建立的小型网络来分析信息的传播过程, 通过建立相应的假设, 并增加约束条件与目标函数, 再据搜索算法, 搜索出使成本费用最低的初始专业推广者与兼职推广者。

## 4.2.1 问题分析

在(1)问中建立的无标度小型网络模型中，我们生成了包含 100 位用户的无标度网络，通过(1)问的分析，此网络的特性与题中所给的复杂网络具有相当高的耦合性，因此我们可以以此小型网络来代替社会网络的中心网络，即小型网络中的用户对信息传播具有相当大的影响力。

在信息的传播过程中，我们把信息的传播分为了两个部分，即在小型网络内的传播与小型网络外的传播，在小型网络内的传播，我们可以根据(1)问的模型来模拟信息的传播过程，但是其传播过程有以下几点区别：在本问的信息传播过程中，我们假设了初始传播用户（即被雇佣者）一直传播下去，而不会在中途被终止雇佣。

在小型网络外的传播，我们假设推广者新增的粉丝即为信息的接收者，并且这些粉丝会进一步转发其信息给他本身的粉丝，到这一步为止，信息不再进一步往下层传播。而非推广者的粉丝不会再往下层传播。同时，设定非重复率，消除重复的粉丝对信息传播的影响。

建立信息接收人数的约束函数与成本费用最低的目标函数，利用搜索算法，则可搜索出使成本费用最低的初始推广者人数。

## 4.2.2 模型建立

### 4.2.2.1 约束函数建立

我们所雇佣的推广者来自于所建立的小型网络，假设雇佣的初始专业推广者人数为  $s$ ，雇佣的兼职推广者为  $h$ ，因为小型网络内的用户相对较少，则在计算信息接收人数的模型中，可以忽略网络内的接收人数， $N(t)$  为每个时刻小型网络内的已接受信息并正在传播的用户人数（不包括初始推广者）。设定  $\lambda(t) = \frac{6}{t}$ ，随着时间的增长，则  $\lambda(t)$  会逐渐减小。

在传播第一天，信息的接收人数为：

$$M(1) = [s \times 500 \times 20 + h \times 35 \times 20 + N(1) \times 20] \times \lambda(1)$$

在传播的第二天，信息的接收人数为：

$$M(2) = [s \times 500 \times 20 \times 2 + h \times 35 \times 20 \times 2 + N(2) \times 20 \times 2] \times \lambda(2)$$

在传播的第  $n$  天，信息的接收人数为：

$$M(n) = \left[ \underbrace{s \times 500 \times 20 \times n}_{\text{专业推广者的信息传播}} + \underbrace{h \times 35 \times 20 \times n}_{\text{兼职推广者的信息传播}} + \underbrace{N(n) \times 20 \times n}_{\text{小型网络内其他用户的信息传播}} \right] \times \lambda(n) \quad (7)$$

因此，100 天过后，信息接收的总人数为每天接收人数的累加，即为：

$$M = M(1) + M(2) + \cdots + M(n) \quad (n = 1, 2, 3 \cdots 100) \quad (8)$$

### 4.2.2.2 优化模型建立

雇佣推广者需要付出一定的雇佣费用，本问中，我们已经假设了雇佣的推广者不会在中途被终止雇佣，因此其所需要的雇佣费用为：

$$F = s \times 500 \times 100 + h \times 50 \times 100 \quad (9)$$

因此建立的优化模型为：

$$\begin{aligned} \min F &= s \times 500 \times 100 + h \times 50 \times 100 \\ \text{st.} \quad &\begin{cases} M = M(1) + M(2) + \cdots M(n) \geq 80000000 \\ s, h \text{ 为正整数} \end{cases} \end{aligned} \quad (10)$$

利用 matlab 编写搜索算法程序，则可搜索出最佳的推广者安排方案。

## 5 模型的求解

### 5.1 问题 1 的求解

#### 5.1.1 SIR 模型的求解

根据所建立的 SIR 模型，并利用 matlab 作图，得到三类人群随时间变化的比例变化，如下图：

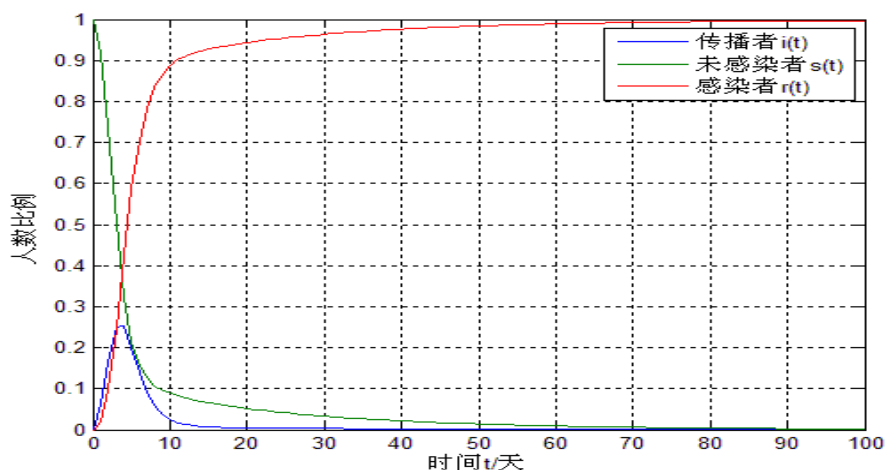


图 3 各类人群人数比例的变化

其未感染者与传播者的关系如下图所示：

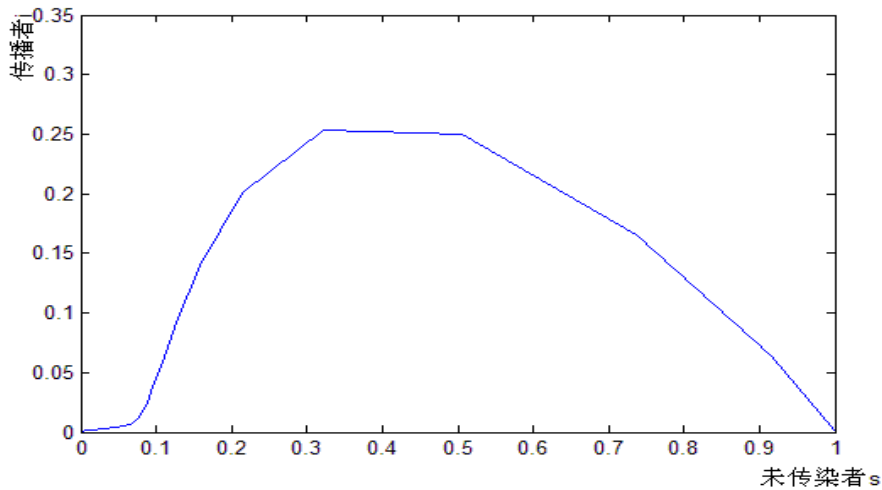


图 4 未感染者与传播者的关系

利用传染病模型，则最后的感染者即接收信息的人数覆盖了全体人数，因此传染病模型不适用本问题的求解

### 5.1.2 BA 网络模型的求解

利用 matlab 作数值迭代，迭代 100 此过后，统计其最终到达稳态时的状态 2 节点数，如表 1。

表 1 看到信息的人数随迭代次数的演变

18	31	61	143	168	207	252	308	336	367	400	413	443	475	498	501
509	535	545	558	568	576	602	635	635	635	635	635	635	635	635	635
635	635	635	635	635	635	635	635	635	635	635	635	635	635	635	635
635	635	635	635	635	635	635	635	635	635	635	635	635	635	635	635
635	635	635	635	635	635	635	635	635	635	635	635	635	635	635	635
635	635	635	635	635	635	635	635	635	635	635	635	635	635	635	635
635	635	635	635												

其占总体节点的比例图：

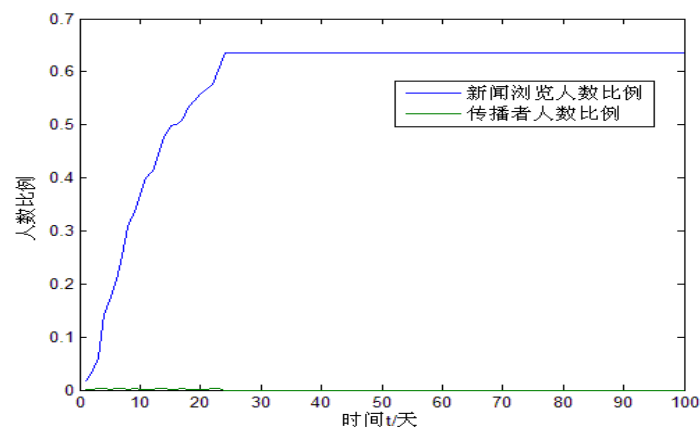


图 5 接收信息人数占总体的比例

因此，信息传播网络中，最终会趋于 0.635 的比例，较好的符合了真实世界的传播趋势，在含有  $N$  个节点的网络中，其最终接受信息的总人数为  $N \times 0.635$ 。

## 5.2 问题 2 的求解

利用 matlab 编写搜索算法程序，将初始的专业推广者控制在 [1,30] 的范围之内，初始的兼职推广者控制在 [5,60] 的范围之内。进行搜索，得到最终的优化安排方案为：专业推广者的雇佣人数为 12 人，兼职推广者的雇佣人数为 20 人，其花费的总费用为：70 万元。

# 5 模型的优缺点与改进

## 5.1 问题 1 模型的进一步改进

针对模型一，在 BA 网络中传播广告新闻时，初始只设了一个人来传发广告新闻，转发者转发后将变为免疫者，不会再进行传播，网络中每节点传播一次后将不会进行转发。且转发给邻近节点的概率恒为 0.2。与真实广告传播出入较大。

对此，在模拟的初始时刻，首先让网络中每一节点按固定概率取 1，即让每一节点按一定的概率接受到广告新闻。假定系统总的节点数目为  $N$ 。 $N_1(0)$  为初始时刻系统中状态为 1 的节点数。定义

$$\rho_1(0) = \frac{N_1(0)}{N} \quad (11)$$

为系统初始时刻接受到该短信息的节点密度，称为初始接受密度。

由于节点与节点之间相互传递短信息，系统节点的状态会发生改变，如果  $t$  时刻，系统中有  $N_2(t)$  个节点是转发了短信息的节点，定义此时系统的转发密度定义为：

$$\rho_2(t) = \frac{N_2(t)}{N} \quad (12)$$

节点  $k_i$  的接收概率应与其度和邻接节点是否为转发者有关，暂仅考虑节点度对接受概率的影响，就一般而言，与其连接的节点越多则它接收到广告的概率越大，及其度越大接收的概率越大，定义接收概率为：

$$p_i = \frac{k_i}{N} \quad (13)$$

节点与节点之间会相互转发信息，在  $t$  时刻，节点  $i$  给其邻接节点  $j$  转发广告只存在两种状态，要么给  $j$  传，则状态为 1，要么不传，则状态为 2。其概率应为二项分布，对所有的节点求取其概率分布，及对二项分布求极限，及为泊松布。

但节点与节点之间不会一直都处于相互转发状态，她们的转发概率会逐渐变缓直至整个系统演化到一个不随时间变化的定态，我们称之为系统的终态。在系统的终态，系统的转发密度、接受密度和广告转发量都有确定的值，不再随时间变化，转发概率将为 0。

设经过时间  $t_s$  后达到网络系统的终态，改进节点转发概率分布：

$$p_i(t) = \frac{\lambda^{k_i} e^{-\lambda}}{k_i!} (t_s - t) \quad (14)$$

将系统的接受密度  $\rho_1(0)$ 、转发密度  $\rho_2(t)$ ，转发概率  $p_i(t)$  与接收概率  $p_i$  代入 BA 网络中，设为其网络传播的一组状态参量。进行计算仿真，其结果将更为真实可靠。

## 5.2 问题 2 模型的进一步改进

针对问题 2 的模型，为了解决问题的方便，我们作了相应的假设，比如，在雇佣推广者的问题上，我们假设了推广者一直被雇佣而不会出现中途被解除雇佣的情况，但这种情况往往与实际问题不相符合，因此针对这种情况，对模型作相应的改进。

我们将考虑每一天的专业推广者人数和兼职宣传者人数，假设第  $i$  天雇佣的专业推广者人数为  $s_i$ ，雇佣兼职推广人数为  $h_i$ ，则在传播的第  $i$  天，信息的接收人数为：

$$M(i) = [500 \times 20s_i + 35 \times 20h_i + 20N(i)] \times \lambda(i) \quad (14)$$

其中， $N(i)$  为每个时刻小型网络内的已接受信息并正在传播的用户人数（不包括初始推广者）

因此，我们可以建立如下改进模型

$$\begin{aligned} \min \quad & z = 500 \sum_{i=1}^{100} s_i + 50 \sum_{i=1}^{100} h_i \\ & \begin{cases} \sum_{i=1}^{100} M(i) > 80000000 \\ s_i \geq 0; h_i \geq 0; i = 1, \dots, 100 \end{cases} \end{aligned} \quad (15)$$

利用计算机搜索算法可以找到模型的最优解。该模型更加接近于现实，具有一定的实用价值。

## 5.2 模型的优缺点

优点：我们采用了经典的 SIR 传染性模型，该模型将人群分成了三类进行讨论，这和社交网络信息传播非常相似，SIR 模型能够正确地反映社交网络信息传播。此外，依据题目所给的要求，我们通过 BA 模型生成一个  $1000 \times 1000$  的网络矩阵，并在该网络中进

行社交网络信息传播的模拟，这能够清晰地反映网络信息的传播过程。

缺点：在计算机模拟中，一些初始数据由我们自己所设定的，这可能造成模拟的数据和现实数据相差甚远。在第二问中，我们假设了推广者一直被雇佣而不会出现中途被解除雇佣的情况，但这种情况往往与实际问题的实际情况不相符合，不能完全地反映现实情况。因此还需要做进一步的模型改进。



## 参考文献

- [1]姜启源 谢金星 叶俊 数学模型 高等教育出版社
- [2]魏婧 刘业政 刑小云 在线社会网络的口碑传播研究 情报杂志 4 期 2011 年
- [3]平亮 宗利永 基于社会网络中心性分析的微博信息传播研究 上海理工大学管理学院 03 期 2009 年
- [4]张发 李璐 宣慧玉 传染病模型综述 系统工程理论与实践 9 期 2001 年
- [5]Albert R,Barabasi A L.Statistical mechanics of complex networks.Rev Mod Phys,2002,74:47-97
- [6]Newman M E J, The structure and function of complex networks.SIAM Review,2003.45:167-256

## 附录 1——SIR 模型求解

```
%微分方程 M 文件
function y=ill(t,x)
p=30;u=0.3;i0=0.02;
y=[(p+500*t)*x(1)*x(2)+(p+20*t)*x(1)*(x(2)-i0)-u*(x(2)-i0),(p+500*t)*x(1)
)*x(2)+(p+20*t)*x(1)*(x(2)-i0)];
%SIR 模型求解
ts=0:100;
ba=0.02;
x0=[ba,1-ba];
[t,x]=ode23('ill',ts,x0);
plot(t,x(:,1),t,x(:,2),t,1-x(:,1)-x(:,2));
xlabel('t')
grid on
```

## 附录 2——BA 网络矩阵

```
% BA 网络矩阵生成函数
function [Nodes,Cii]=BA(N,m)
Nodes=zeros(N);
Cii=zeros(1,N);
t=zeros(1,N);
for i=1:m
    Nodes(i,m+1)=1;
    Nodes(m+1,i)=1;
    list(i)=i;
end
for i=m+1:2*m
    list(i)=m+1;
end
for n=m+2:N
    t=2*m*(n-m-1);
    for i=1:m
        list(t+i)=n;
    end
    k=1;
    while k<m+1
        p(k)=round((t+1)*rand(1));
        if p(k)>0&p(k)<(t+1)
            if Nodes(n,list(p(k)))==0
                list(t+m+k)=list(p(k));
                Nodes(n,list(p(k)))=1;
                Nodes(list(p(k)),n)=1;
                k=k+1;
            end
        end
    end
end
```

```

        end
    end
end
for i=1:N
    Cii(i)=sum(Nodes(:, i));
    t(i)=list(i);
end

```

### 附录3——在线网络信息传播模拟

%在线网络信息传播模拟

```

clear;
a=BA(1000, 10);%生成一个 0-1 网络矩阵
T=100;%时间
N=1000;
M=1;
count=0;
sum=zeros(1, N);
flag=zeros(1, N);
flag(1)=1;
p1=0.2;
p2=0.8;
for t=1:T
    sum=zeros(1, N);
    time=0;
    counttime=0;
    for k=1:N
        if flag(k)==1
            time=time+flag(k);
        end
    end
    for i=1:N
        %获得第 i 个节点的度
        for k=1:N
            sum(i)=sum(i)+a(i, k);
        end
        if counttime>=time
            break;
        end
        if flag(i)==1
            flag(i)=2;%标记该节点为免疫节点
            counttime=counttime+1;
            temp=p1*sum(i);
            temp=floor(temp);%浏览人数

```

---

```

        templ=p2*temp;
        templ=floor(templ);%转发人数
        temp=temp-templ;
    for j=1:N %转化为传播节点
        if count>=temp
            break;
        end
        rand=randint(1,1,[1,N]);
        if a(i,rand)==1
            flag(rand)=1;
            count=count+1;
        end
    end
    count=0;
    for j=1:N %转化为免疫节点
        if count>=templ
            break;
        end
        rand=randint(1,1,[1,N]);
        if a(i,rand)==1
            flag(rand)=2;
            count=count+1;
        end
    end
end
end
for i=1:N %增加粉丝数目
    count=0;
    for j=1:N
        if count<M
            break;
        end
        rand=randint(1,1,[1,N]);
        if a(i,rand)==0
            a(i,rand)=1;
            count=count+1;
        end
    end
end
end
as=0;xx=0;yy=0;
for i=1:N %数据统计
    if flag(i)==2
        as=as+1;
        xx=xx+1;
    end
end

```

```

        end
        if flag(i)==1
            as=as+1;
            yy=yy+1;
        end
    end
    xx1(t)=xx;
    yy1(t)=yy;
    topl(t)=as;
end
t=1:T;
plot(t, topl/1000, t, yy1/1000);

```

## 附录 4——信息传播搜索算法

%搜索算法源代码

```

col=1;
for N=1:30
    for M=1:50
        a=BA(100, 10); %生成 BA 网络矩阵
        NUM=100;
        TIME=100;
        number=0;
        flag=zeros(1, NUM);
        for t=1:TIME
            count=0;
            for j=1:NUM %随机生成 N 个专业推广者
                if count>=N
                    break;
                end
                rand=randint(1, 1, [1, NUM]);
                if flag(rand)==0
                    flag(rand)=1;
                    count=count+1;
                end
            end
            count=0;
            for j=1:NUM %随机生成 M 个普通推广者
                if count>=M
                    break;
                end
                rand=randint(1, 1, [1, NUM]);
                if flag(rand)==0
                    flag(rand)=2;
                    count=count+1;
                end
            end
        end
    end
end

```

---

```

        end
    end
    tnum=0;%推广人数
    pnum=0;%普通推广人数
    ynum=0;%普通人员
    for i=1:NUM
        if flag(i)==1
            tnum=tnum+1;
        end
        if flag(i)==2
            pnum=pnum+1;
        end
        if flag(i)==3
            ynum=ynum+1;
        end
    end
    end
    number=number+(tnum*500*20*t+pnum*35*20*t+ynum*20*t)*(6/t);
    for i=1:NUM %网络传播
        if flag(i)==1||flag(i)==2||flag(i)==3
            if flag(i)==3
                for j=1:NUM
                    if a(i,j)==1&&flag(j)==0
                        flag(j)=3;
                    end
                end
                flag(i)=4;
            end
            for j=1:NUM
                if a(i,j)==1&&flag(j)==0
                    flag(j)=3;
                end
            end
        end
    end
    end
    end
    if number>800000000 %存储专业推广者人数和普通推广者人数
        result(col,1)=N;
        result(col,2)=M;
        col=col+1;
    end
end
end
end

```