

## 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：杨和林

队员 2：陈霞

队员 3：袁质方

参赛队教练员（签名）：

参赛队伍组别：本科组

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）：  
2310

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 C 题：碎片化趋势下的奥运会商业模式的数学分析

关 键 词 社交网络传播模型 预测 整数规划 分支定界算法

## 摘 要：

本文旨在奥运会背景下, 针对社交网络的信息传播规律进行研究, 首先建立了传染病  $SI$  传播模型, 并通过优化进一步建立社交网络的信息传播模型 (SNS), 解决了观看含有企业广告的奥运新闻的人数预测问题, 然后立足于前面的模型建立了单目标整数规划模型, 接着采用分支定界算法解得企业成本最低情况下为企业制定出了最合理的用人方案, 最终较全面地解决了本文的所有问题。

对于问题一, 首先建立传染病  $SI$  模型对问题进行简化讨论, 但通过分析发现, 本题信息传播网络为无标度网络, 网络信息传播有效率  $P$  也不是一个定值, 它与节点周围“感染者”的数目有关。由此, 本文再次建立在线社交网络信息传播模型, 同时借鉴 Matlab 工具箱得出看到信息总用户量计算公式  $y = N * (1 - \exp(-0.001 * n_1 * t))$ , 其中  $N$  表示市场潜在总用户量,  $n_1$  表示专业推广者人数。假设  $N = 2$  亿、 $t = 100$  天的前提下,  $n_1$  取为特定的数值, 得出看到信息总用户量如下表所示:

| 专业推广者人数 $n_1$ (人) | 1      | 5      | 10     | 15     |
|-------------------|--------|--------|--------|--------|
| 看到信息总用户量 $y$ (亿人) | 0.1904 | 0.7870 | 1.2642 | 1.5538 |

对于第二问, 在第一个问题模型的基础之上, 运用 Matlab 软件逐步缩小搜索范围, 寻求符合有效传播数量要求的专业人员雇佣数量至少为 5 名。接着根据题目兼职推广者的信息, 考虑市场潜在总用户量、时间和有效传播人数三个指标, 建立了以成本最低为单目标的整数规划模型, 并采用分支定界算法, 为企业制定了最合理的用人方案为: 在距离奥运会 100 天时雇佣 2 名专业推广者和 6 名兼职推广者, 再到第 87 天的时候雇佣一个 1 名兼职推广者, 在此方案下, 企业所需最小成本为 13.065 万元。

本文的特色在于基于合理的社交网络传播模型, 巧妙地结合了网络信息传播有效率, 构造了对应的有效总用户量计算公式, 并采用相关的算法很好地预测了信息用户量及为企业制定了最优的用人计划。

参赛队号 2310

所选题目 C 题

参赛密码 \_\_\_\_\_  
(由组委会填写)

## 英文摘要

This article aims to the background of the Olympic Games , the social network for the dissemination of information law , first established the model of the spread of infectious diseases , and by optimizing the dissemination of information to further establish the social networking model (SNS) to solve the Olympic watching with corporate advertising news number of prediction problem, then based at the front of the model to establish a single objective integer programming model , and then using the branch and bound algorithm for solution of the lowest cost of doing business for companies to work out the most reasonable employer , and ultimately a more comprehensive solution to this article problem.

For problem one, First established an infectious disease model to a brief discussion of the problem, but The analysis shows that this question of information dissemination network for scale-free network type, and the network information transmission efficiency  $P$  is not a fixed value, which is related to the number of nodes around the "infection". As a result, this paper re-establish the information dissemination model of online social networking, learn from the Matlab toolbox calculated the formula of seeing information total users is

$y = N * (1 - \exp(-0.001 * n_1 * t))$ , where  $N$  is the total potential market of users,  $n_1$  is the number of professional promotion, assume that  $N = 2$  and  $t = 100$ ,  $n_1$  are some specific numerical values, Obtained the seeing information total users in the table below:

| The number of professional promotion $n_1$ ( people)     | 1      | 5      | 10     | 15     |
|--|--------|--------|--------|--------|
| the seeing information total users $y$ ( million people) | 0.1904 | 0.7870 | 1.2642 | 1.5538 |

For the second question, on the basis of the first question's model, We use Matlab software gradually reduce the search space for effective communication with the number of the professional promoters for employment number at least five. Then according to the subject information, the company can hire part-time propaganda from network. Considering the market potential in total users, time and effective spreading of the number, we set up the lowest cost of single objective integer programming model, at the same time, we use the branch bound algorithm for enterprises to decide the most reasonable scheme: the enterprise can hiring two professional promoters and six part-time promoters in 100 days until the Olympics, and then to 87 days of time hire a part-time promoter. Under the plan, The enterprise pay for the minimum fee :130650 yuan.

This article features is based on the reasonable social network propagation model, ingeniously combines the network information transmission efficiency, the corresponding structure of the total effective calculation formula of users, and the related algorithm are very good for predicting the information users for enterprises to formulate and the optimal plan of choosing the promoters.

## 一、 问题重述

1984 年之前，奥运会是个“非卖品”；之后，奥运与电视结盟，通过发布赞助商广告模式攫取巨大商业利益，这就是“Top 赞助商”的前身。经过 28 年的发展，这种模式已经是商业社会里最重要的公司展示舞台。

品牌选择奥运会的理由，是因为这里凝聚了观众的大量时间。他们希望在观众关注比赛的同时也注意到自己的品牌和产品，而 Top 赞助商们，则可以获得在电视奥运频道里排除行业里其他竞争对手广告的特权。

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，观众不在只关注电视，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。

现有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。

## 二、 问题分析

本题的第一个问题是预测一条含有企业广告的奥运会新闻在离奥运会开幕前 100 天的传播过程中可以被多少人观看到。题目给出可以聘请专业的推广者，一个社交网络的专业推广者平均每天可以新增 500 个粉丝，普通网络用户平均每天可以新增 20 个粉丝。作为互联网的普通用户，当接受到来自网络的信息时，根据信息吸引力对于自身的影响程度的大小，普通用户粉丝层可以自己决定是否转发该信息成为该信息的转播者或免疫者。总的来看，信息在整个转播过程会出现一定的重复转播概率，计算时需要注意对于用户组连接重复的剔除，且考虑普通网络用户以概率接受和转播信息。

本题的第二个问题是计算至少需要雇佣几名专业社交网络推广者才能实现有效推广者达到 8000 万，利用临界值求法，逐步缩小搜索范围寻求符合要求的雇佣数量点；考虑还可以从网络上雇佣兼职宣传者，考虑到成本，请给企业制定一份合理的用人方案，针对这个最合理的用人方案，寻找能达到时间、数量和成本要求的最优“专业推广者”和“兼职推广者”的组合。

### 三、 模型假设

1. 由于社交网络与互联网环境紧密相关，因此，我们假设在奥运会期间，互联网环境保持稳定不变，互联网用户保持不变。
2. 假设有足够的专业网络推广者和兼职宣传者能够雇佣，雇佣的专业社交网络推广者和兼职宣传者都能尽职尽责保质保量的刚好完成规定任务，不超额完成也不缺额完成工作任务。
3. 假设该企业在奥运会期间经营状况良好，不会出现财务恶化、企业形象丑化、管理不善等负面情况。
4. 假设宣传推广途径仅限于社交网络，排除电视、纸质媒体等传播推广途径。
5. 假设网络用户分为信息未感染者、已感染者、免疫者三类。

### 四、 符号说明

|                           |   |
|---------------------------|---|
| $p_1$                     | 未感染节点成为传播节点的概率                                    |
| $p_1$                     | 未感染节点成为传播节点的概率                                    |
| $p_2$                     | 未感染节点成为传播节点的概率                                    |
| $p_{ii}^j$                | 节点 $j$ 在 $[t, t + \Delta t]$ 时段内处于未感染状态的概率        |
| $p_{is}^j$                | 节点 $j$ 从未感染状态转移到传播状态的概率                           |
| $p_{sr}^j$                | 节点 $j$ 在 $[t, t + \Delta t]$ 时段内从传播状态转移到免疫状态的概率   |
| $\omega(k, t)$            | 在 $t$ 时刻从具有 $k$ 条边的未感染节点连接到一个传播节点的概率              |
| $N(k, t)$                 | 在 $t$ 时刻网络中度为 $k$ 的节点总数量                          |
| $I(k, t)$                 | 在 $t$ 时刻网络中度为 $k$ 的未感染节点数量                        |
| $S(k, t)$                 | 在 $t$ 时刻网络中度为 $k$ 的传播节点的数量                        |
| $R(k, t)$                 | 在 $t$ 时刻网络中度为 $k$ 的免疫节点的数量                        |
| $p(k k')$                 | 度相关函数，表示度为 $k$ 的节点与度为 $k'$ 的节点临界的条件概率             |
| $\rho^s(k', t)$           | 在 $t$ 时刻度为 $k'$ 的传播节点的密度                          |
| $\overline{p_{ii}}(k, t)$ | 度为 $k$ 的节点在 $[t, t + \Delta t]$ 时段内处于未感染状态的平均转移概率 |

## 五、 模型建立与求解

### 5.1 问题一：观看广告人数模型的建立与求解

#### 5.1.1 问题一模型的建立

##### (1) 原 SI 模型

SI 模型<sup>[1]</sup>指出普通网络用户粉丝一旦接受某条传播信息，将从此成为该条信息的“感染者”，以  $S$  表示未被该信息“感染”的人数，以  $I$  表示已被该信息“感染”的人数；人员流动趋向为：

$$S \rightarrow I$$

信息传播方式大致抽象为星型传播<sup>[2]</sup>，如图 1 所示。

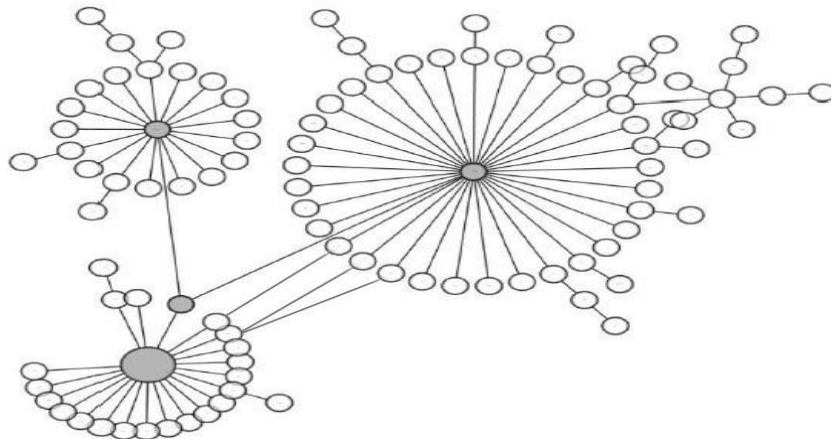


图 1 星型传播模式

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。

假设：

1. 每个专业传播者在单位时间内有效传播的人数为常数  $k_0$ ，概率为  $P$ 。
2. 推广者在社交网络发布和奥运会相关的所有信息传播的期限为 100 天。

记时刻  $t$  “感染者”的人数为  $I(t)$ ，开始时有  $I(1)$  个传播者，则在  $\Delta t$  时间内增加的病人数为

$$I(t + \Delta t) - I(t) = k_0 I(t) \Delta t \quad (1)$$

于是得：

$$\begin{cases} \frac{dI(t)}{dt} = k_0 I(t) \\ I(1) = I_1 \end{cases} \quad (2)$$

其解为：  $I(t) = I_1 e^{k_0 t}$ 。

**模型分析与解释：**这个结果与传播初期比较吻合，但它表明被该信息“感染”人数将按指数规律无限增加，显然与实际不符。事实上，在信息传播期间，一个信息传播者单位时间内能传染的人数  $k_0$  则是在改变的。在初期， $k_0$  较大，随着

该信息的“感染者”增多，未被感染者减少，被传染机会也将减少，于是 $k_0$ 就会变小。

## (2) SI 模型改进

记时刻 $t$ 的未被感染者人数为 $S(t)$ ， $C(t)$ 为 $t$ 时刻传播人数的重复人数， $\alpha(t)$ 为 $t$ 时刻传播人数的重复率：

重复率初值 $\alpha(1) = \frac{C(1)}{I(1)}$ ，由附件数据即可得该初值。

分析信息被传播人数增长到一定数量后增长率下降的主要原因，我们注意到，通过网络关系随机性的传播会对感染者人数总量的增长起着阻滞作用，并且随着感染者人数的增加，阻滞作用越来越大。所谓 SI 模型改进就是考虑这个因素，认为随着时间网络传播信息重复率增加将会阻滞整体染者人数总量的增长。

经论文验证，重复概率 $\alpha(t) = 1 - k_0(1 - \frac{I(t)}{I_m})$ 。当 $I_m = I(t)$ 时，感染者人数总量不再增长，即重复率 $\alpha(t)$ 达到最大值。

则假设：

1. 总人数为常数 $N$ ，且 $I(t) + S(t) = N$ 。
2. 单位时间内一个传播者能传染的人数与当时未被“感染者”人数成正比，比例系数为 $P$ （传播率）， $P$ 的值其实是一个变量，它每天的值都在发生变化，在此模型中，我们认为传播率 $P$ 在数值上与“感染者”数量的增长率是相等的，其中

$$PS(t) = \frac{I(t+1) - I(t)}{I(t)} \quad (3)$$

题目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。即，

$$PS(t) = \frac{[(1 - \alpha(t)) \sum_{l=1}^t (500I(1) * 20^l) - (1 - \alpha(t-1)) \sum_{l=1}^{t-1} (500I(1) * 20^l)]}{(1 - \alpha(t-1)) \sum_{l=1}^{t-1} (500I(1) * 20^l)} \quad (4)$$

$$S(t+1) = S(t) - \Delta I(t) = S(t) - [I(t+1) - I(t)] \quad (5)$$

由 (4) (5) 可得方程：

$$\begin{cases} \frac{dI(t)}{dt} = PS(t)I(t) \\ I(1) = I(1) \end{cases}$$

解得：

$$I(t) = \frac{N}{1 + (\frac{N}{I_0} - 1)e^{-PNt}}。$$



**模型分析：**可以解得  $\frac{dI(t)}{dt}$  的极大值点为： $t_1 = \frac{\ln(\frac{N}{I_0} - 1)}{PN}$ 。这可以表示传播高峰时刻。当传播强度  $P$  增加时， $t_1$  将变小，即传播高峰来得快，这与实际情况吻合。本题此处即需求得  $I(t)$  的最大值  $I_m$ 。

### (3) SI 模型的求解

由被该信息“感染”的总人数  $I(t) = \frac{N}{1 + (\frac{N}{I_0} - 1)e^{-PNt}}$ ， $N$  为世界上所有的网络

用户总量， $I_1$  为网络专业传播者数量的初值， $P$  为单位个体单位时间的传播率， $t$  为信息传播时间。

参数：（1） $t=100$ 天；

（2）传播重复率  $\alpha(t)$  随时间的变化趋势如图 2。

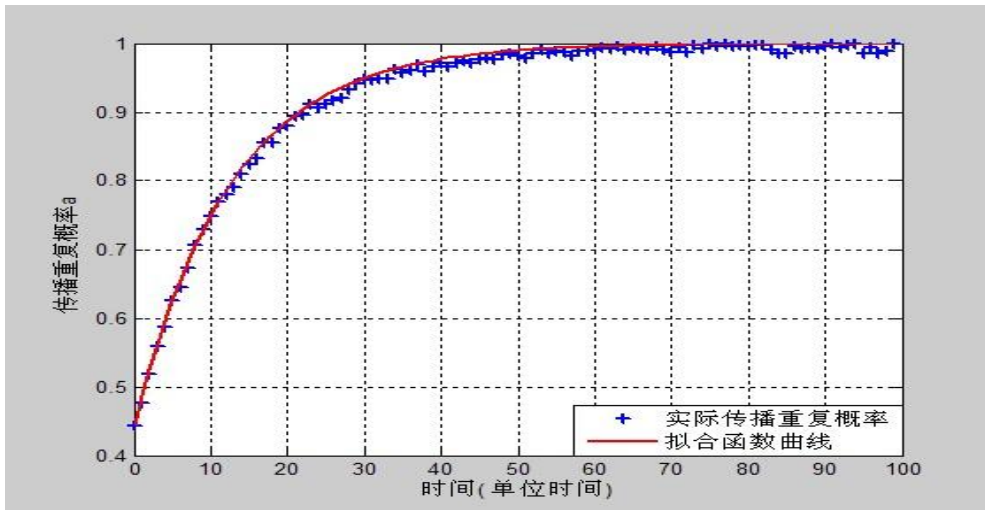


图 2 传播重复率  $\alpha(t)$  随时间的变化趋势

网络信息传播重复率与时间的关系式：

$$\alpha = 1 - 0.87 * \exp(-0.015 * t) - 0.015 \quad (6)$$

可以发现，随着传播时间的延续，在传播过程中的重复概率会逐渐增加，直到信息网络市场达到饱和，实现最大的传播人数。

（2）根据  $P = 1 - \alpha$ ，传播有效率  $P$  随时间的变化趋势如图 3。

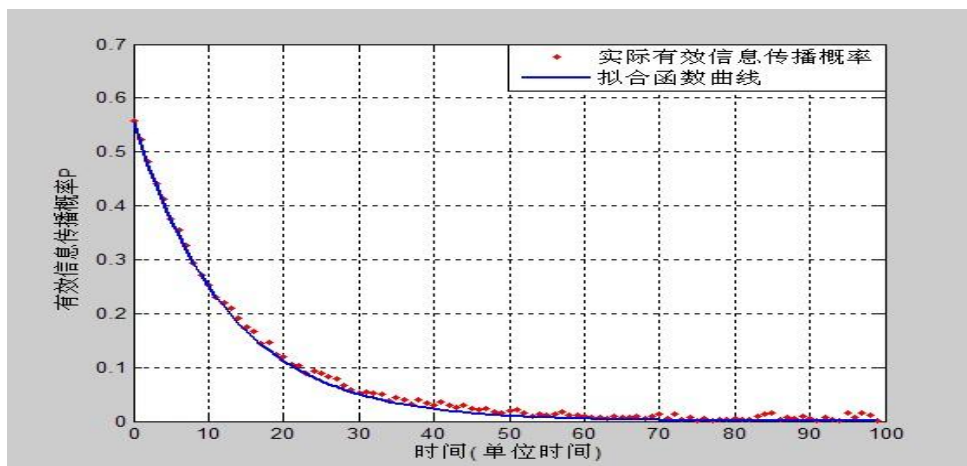


图 3 传播有效率  $P$  随时间的变化趋势

网络信息传播有效率与时间的关系式:

$$P = 0.87 * \exp(-0.015 * t) + 0.015 \quad (7)$$

可以发现,随着传播时间的延续,在传播过程中的信息传播有效率概率会逐渐降低,因为被传播者的重复传播现象。

#### (4) SI 模型缺陷

1、这个模型没有考虑人际网络的结构。它假设任何两个人之间有联系,而且在这两个模型中,节点周围只要有“感染者”,其就以一定概率被传染。概率值是定值,与节点周围“感染者”的数目无关。这也与信息传播的实际情况不符。信息传播中,受传者有从众心理,其受周围人影响的程度显然与其中接受信息的人的比例有关系。个体周围接受信息的人越多,则个体越可能接受信息。

2、该传播模型大都是基于规则网络的,但是在线社交网络上的传播行为与规则网络相比确实存在根本上的不同,对网络个体处理过于简单。

#### (5) SI 模型改进——在线社交网络的信息传播模型

##### 1、信息吸引力讨论

信息扩散面积与信息吸引力<sup>[3]</sup>及小世界性质的关系论证。

在  $100 \times 100$  的方格阵上构造了一个二维的小世界图,  $p$  的值是为方格构造随机边的概率。当  $p = 0$  时,这是一个规则的方格阵。将  $|A^{(n)}|$  定义为扩散面积,于是发现信息扩散面积与信息吸引力有关系: ①扩散面积随信息吸引力的增加而增加; ②方格阵信息吸引力存在一个大于 0.6 的临界点。

在这个临界点以下,信息几乎不会传播开,只有大于这个临界点,信息才会传播开。而且在大于临界点的附近,信息传播面积会随着信息吸引力的增加快速增加。随机边的加入可以使这个临界值减小,但不会小于 0.2。随机边的加入,对信息吸引力在 0.2 到 0.8 之间影响最大,随机边加入的概率越大,扩散面积越大。但随着信息吸引力的增加,随机边的加入对信息扩散面积的影响越来越小,当信息吸引力大于 0.8 时,就几乎没有影响了。如图 4 所示。

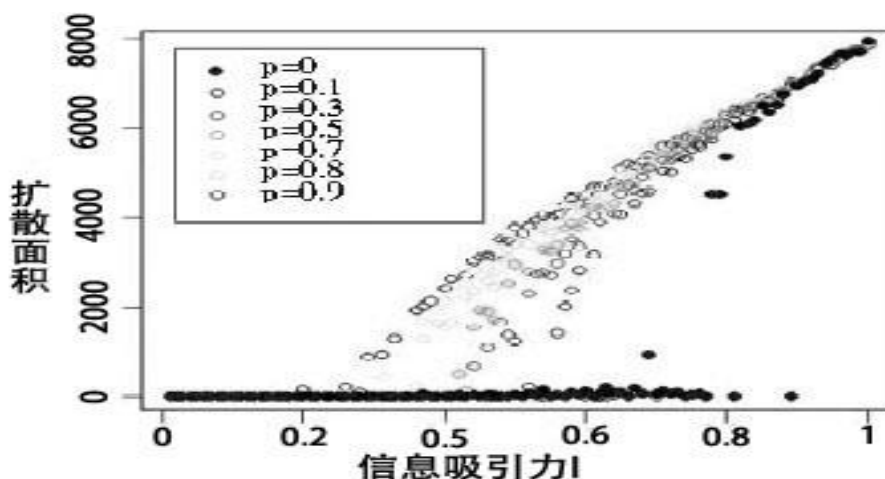


图 4 信息扩散随信息吸引力  $I$  的变化趋势

从图 4 可以看出,在小世界网络和无标度网络上,信息吸引力小于 0.2 时的信息扩散耗时与大于 0.2 时的情况有很明显的不同,这也印证了在这些图上信息吸引力临界值的存在性。

所以,首先存在信息传播临界值。当信息吸引力大于临界值时信息才能传播开,否则几乎不能传播。其次,小世界网络上,当信息吸引力在临界值附近时,随机边的添加会使传播面积增大,从而导致实际的传播临界值减小;但当信息吸引力足够大时,图结构的变化对信息传播面积影响不大。第三,当信息吸引力在临界值附近时,信息传播耗时对网络规模很敏感,但随着信息吸引力的增加,其值趋于稳定。

#### ① 吸引力小于 0.2 或大于 0.8

当信息吸引力小于 0.2 时,该广告信息将不会被广泛传播,传播时间短、受众少;当广告吸引力大于 0.8 时,该广告信息会迅速爆红于网络,传播速度极快。在两种情况之下,该广告信息的传播规律都无法准确寻找,从而无法利用数学工具建模预测出其最终受众。

#### ② 吸引力在 0.2 到 0.8 之间

根据信息在 Twitter 社交网站用户之间的传播规律,我们把网络中的节点分为三类:传播节点、未感染节点、免疫节点。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;未感染节点表示该节点没有接受过来自其邻居节点的信息,并有可能接受信息,即有概率被感染。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,节点在传播状态、未感染状态和免疫状态之间的转移不仅依赖于节点自身的状态,还与它的邻居节点的状态相关<sup>[4]</sup>,网络传播示意图如图 5。

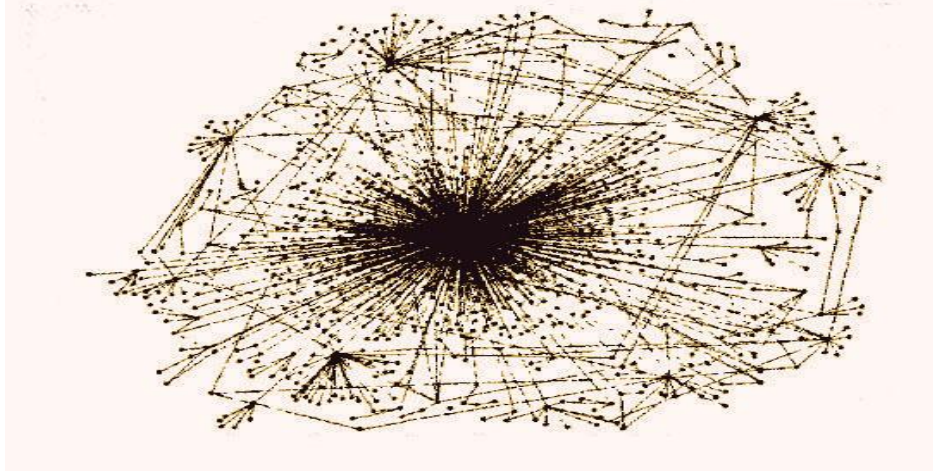


图 5 网络传播示意图

定义以下传播规则：

- 1) 如果一个传播节点与一个未感染节点接触；则未感染节点会以  $p_1$  概率成为传播节点；
- 2) 如果一个传播节点与一个免疫节点接触，则传播节点会以概率  $p_2$  成为免疫节点；
- 3) 传播节点不会无休止地传播下去，会以一定的速度  $v$  停止传播而变为免疫节点，且无需与其他节点接触。

此处假设一个节点  $j$  在  $t$  时刻处于未感染状态， $p_{ii}^j$  表示节点  $j$  在  $[t, t + \Delta t]$  时段内处于未感染状态的概率， $p_{is}^j$  表示节点  $j$  从未感染节点状态转移到传播状态的概率且  $p_{is}^j = 1 - p_{ii}^j$ ，其中

$$p_{ii}^j = (1 - \Delta t p_1)^g \quad (8)$$

式中， $g = g(t)$  表示在  $t$  时刻节点  $j$  的邻居中传播节点的数量。

假设节点  $j$  含有  $k$  条边， $g$  是具体如下二项分布的随机变量：

$$\prod(g, t) = \binom{k}{g} \omega(k, t)^g (1 - \omega(k, t))^{k-g} \quad (9)$$

式中， $\omega(k, t)$  表示在  $t$  时刻从具有  $k$  条边的未感染节点连接到一个传播节点的概率。

$$\begin{aligned} \omega(k, t) &= \sum_{k'} p(k' | k) p(s_{k'} | i_k) \\ &\approx \sum_{k'} p(k' | k) \rho^s(k', t) \end{aligned} \quad (10)$$

式中， $p(k' | k)$  为度相关系数，表示度为  $k$  的节点与度为  $k'$  的节点邻接的条件概率；度相关系数函数可以写为

$$p(k' | k) = \frac{k' P(k')}{\bar{k}} \quad (11)$$

其中， $P(k')$  为度分布函数， $\bar{k}$  为网络节点平均度。

$p(s_{k'} | i_k)$  表示一个拥有  $k'$  条边的节点在它连接到一个度为  $k$  的未感染节点条件下，处于传播状态的概率； $\rho^s(k', t)$  表示在  $t$  时刻度为  $k'$  的传播节点的密度。

所以，得到度为  $k$  的节点在  $[t, t + \Delta t]$  时段内处于未感染状态的平均转移概率

$\overline{p_{ii}}(k,t)$  如下:

$$\begin{aligned}\overline{p_{ii}}(k,t) &= \sum_{g=0}^k \binom{k}{g} (1-p_1\Delta t)^g \omega(k,t)^g \times (1-\omega(k,t))^{k-g} \\ &= (1-p_1\Delta t\omega(k,t))^k\end{aligned}\quad (12)$$

将 (11) 式代入上式得到

$$\overline{p_{ii}}(k,t) = (1-p_1\Delta t \sum_{k'} p(k'|k) \rho^s(k',t))^k \quad (13)$$

同理, 假设  $p_{sr}^j$  表示节点  $j$  在  $[t, t+\Delta t]$  时段内从传播状态转移到免疫状态的概率,  $p_{ss}^j$  表示节点  $j$  保持传播状态的概率且  $p_{ss}^j = (1-p_2\Delta t)^g (1-v\Delta t)$ 。

于是, 得到度  $k$  的节点在  $[t, t+\Delta t]$  时段内处于感染状态的平均转移概率  $\overline{p_{ss}}(k,t)$  如下:

$$\overline{p_{ss}}(k,t) = ((1-p_2\Delta t \sum_k p(k'/k) \rho^r(k',t))^k \times (1-v\Delta t)) \quad (14)$$

则节点从传播状态转移到免疫状态的转移概率为  $\overline{p_{sr}}(k,t) = 1 - \overline{p_{ss}}(k,t)$ 。

假设  $N(k,t)$  为在  $t$  时刻网络中度为  $k$  的节点总数量,  $I(k,t)$ 、 $S(k,t)$ 、 $R(k,t)$  分别在  $t$  时刻网络中度为  $k$  的未感染节点、传播节点和免疫节点的数量, 则

$$I(k,t) + S(k,t) + R(k,t) = N(k,t) \quad (15)$$

于是, 网络中度为  $k$  的未感染节点的数量在  $[t, t+\Delta t]$  时段内的变化情况如下:

$$\begin{aligned}I(k,t+\Delta t) &= I(k,t) - I(k,t)(1-\overline{p_{ii}}(k,t)) \\ &= I(k,t) - I(k,t) \times [1 - (1-p_1\Delta t \sum_{k'} p(k'|k) \rho^s(k',t))^k] \quad (16)\end{aligned}$$

同理, 可相应得到度为  $k$  的传播节点和免疫节点的数量在  $[t, t+\Delta t]$  时段内的变化情况, 如下:

$$\begin{aligned}S(k,t+\Delta t) &= S(k,t) + I(k,t)(1-\overline{p_{ii}}(k,t)) - S(k,t)(1-\overline{p_{ss}}(k,t)) \\ &= S(k,t) + I(k,t) \times \left[ 1 - (1-p_1\Delta t \times \sum_{k'} \rho^s(k',t) p(k'|k))^k \right] \\ &\quad - S(k,t) \left[ 1 - (1-p_2\Delta t \times \sum_{k'} \rho^r(k',t) p(k'|k))^k (1-v\Delta t) \right]\end{aligned}\quad (17)$$

$$\begin{aligned}R(k,t+\Delta t) &= R(k,t) + S(k,t)(1-\overline{p_{ss}}(k,t)) \\ &= S(k,t) \left[ 1 - (1-p_2\Delta t \times \sum_{k'} \rho^r(k',t) p(k'|k))^k (1-v\Delta t) \right] + R(k,t)\end{aligned}\quad (18)$$

由 (15) (16) 式可以得到

$$\frac{I(k,t+\Delta t) - I(k,t)}{N(k,t)\Delta t} = -\frac{I(k,t)}{N(k,t)\Delta t} \left[ 1 - (1-p_1\Delta t \times \sum_{k'} \rho^s(k',t) p(k'|k))^k \right] \quad (19)$$

当  $\Delta t \rightarrow 0$  时, 对 (19) 式右侧进行泰勒展开得到

$$\frac{\partial \rho^i(k,t)}{\partial t} = -kp_1\rho^i(k,t) \sum_{k'} \rho^s(k',t) p(k'|k) \quad (20)$$

同理, 由 (17) 式可以得到

$$\begin{aligned} \frac{\partial \rho^s(k,t)}{\partial t} = & k p_1 \rho^i(k,t) \times \sum_{k'} \rho^s(k',t) p(k'|k) - k p_2 \rho^s(k,t) \\ & \times \sum_{k'} \rho^r(k',t) p(k'|k) - v \rho^s(k,t) \end{aligned} \quad (21)$$

由 (18) 式可以得到

$$\frac{\partial \rho^r(k,t)}{\partial t} = k p_2 \rho^s(k,t) \sum_{k'} \rho^r(k',t) p(k'|k) + v \rho^s(k,t) \quad (22)$$

由 (20), (21), (22) 式联立得到信息传播的动力学演化方程组, 用于刻画传播节点、未感染节点和免疫节点的密度随着时间的变化, 传播动力学过程同时受到网络拓扑结构的传播机理的影响。

### 5.1.2 问题一信息传播模型求解

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分享给自己的粉丝们, 普通网络用户平均每天可以新增 20 粉丝。经上文求得, 网络信息传播有效率与时间的关系式:

$$k = 0.87 * \exp(-0.015 * t) + 0.015 \quad (23)$$

可以发现, 因为被传播者的重复传播现象, 随着传播时间的延续, 在传播过程中的信息传播有效率概率会逐渐降低。

此处将分两种情况对网名总人数和专业推广者人数取特值检验模型的准备性。

(1) 此处假设互联网网名总人数为 2 亿人, 最初只有 1 个专业推广者负责该信息的网络推广, 网络信息节点随时间的转换情况如图 7。

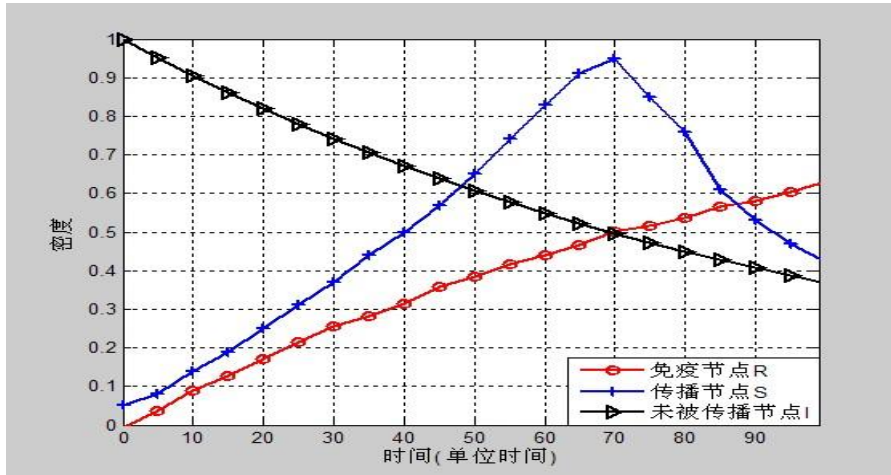


图 7 1 个专业推广者时网络信息节点随时间的转换情况

企业奥运会新闻受众等于免疫节点与传播节点的数量之和, 即  $M = R + S$ 。

(2) 此处假设互联网网名总人数为 2 亿人, 最初有 5 个专业推广者负责该信息的网络推广, 网络信息节点随时间的转换情况如图 8。



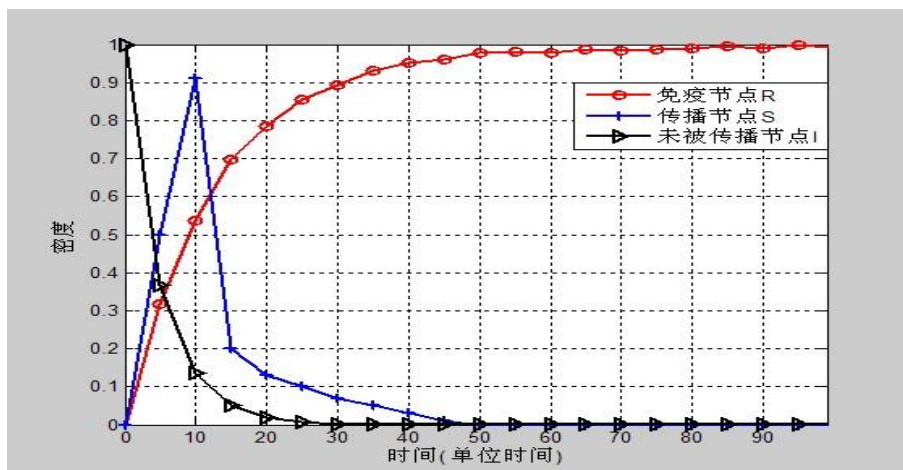


图 8 5 个专业推广者时网络信息节点随时间的转换情况

根据图3，在开始度为  $k = 500$  和以后平均度约  $k' = 20$ ，我们借助matlab工具箱cftool对其拟合，当专业推广者开始为  $n_1$  名及市场潜在用户量为  $N$  时，在时间  $t$  时刻接收到此信息的市场客户累计人数  $y$  为：

$$y = N * (1 - \exp(-0.001 * n_1 * t)) \quad (24)$$

当  $N = 2$  亿人潜在用户， $n_1 = 1$  名专业推广者时，在  $t = 100$  天时由公式得预计市场客户累计人数  $y = 0.1904$  亿人；

当  $N = 2$  亿人潜在用户， $n_1 = 5$  名专业推广者时，在  $t = 100$  天时由公式得预计市场客户累计人数  $y = 0.7870$  亿人；

当  $N = 2$  亿人潜在用户， $n_1 = 10$  名专业推广者时，在  $t = 100$  天时由公式得预计市场客户累计人数  $y = 1.2642$  亿人；

当  $N = 2$  亿人潜在用户， $n_1 = 15$  名专业推广者时，在  $t = 100$  天时由公式得预计市场客户累计人数  $y = 1.5538$  亿人。

## 5.2 问题二：最合理用人方案模型的建立与求解

### 5.2.1 问题二模型的建立

(1) 题目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，即要求雇佣多名推广者在奥运会开幕之前的 100 天之内，使得这期间看到该广告信息的人群数量大于 8000 万。

通过计算，利用 Matlab 编程得出不同的专业推广者完成有效推广数量 8000 万，所需的时间如图 9 所示。

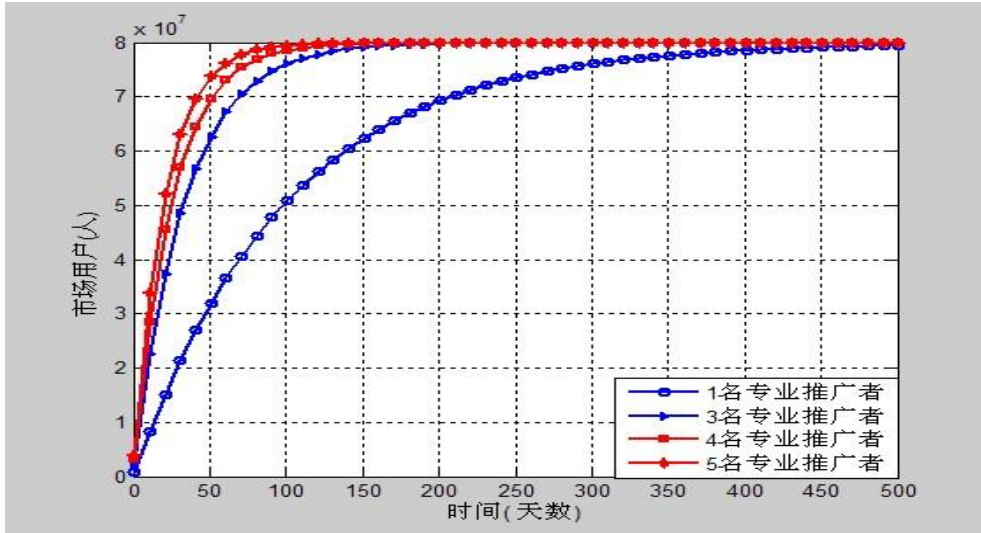


图 9 不同的专业推广者完成有效推广数量 8000 万所需时间

通过图像分析出,至少要求雇佣 5 名推广者在奥运会开幕之前的 100 天之内,使得这期间看到该广告信息的人群数量大于 8000 万。

(2) 因为还可以从网络上雇佣兼职宣传者, 综合目前的推广人员: 专业与兼职。

#### 1、两类推广者按比例组合

假设看到该广告信息的人群数量大于 8000 万, 1 个专业推广者、1 个兼职推广者以及他们同时工作完成有效推广数量的时间发展图如图 9。

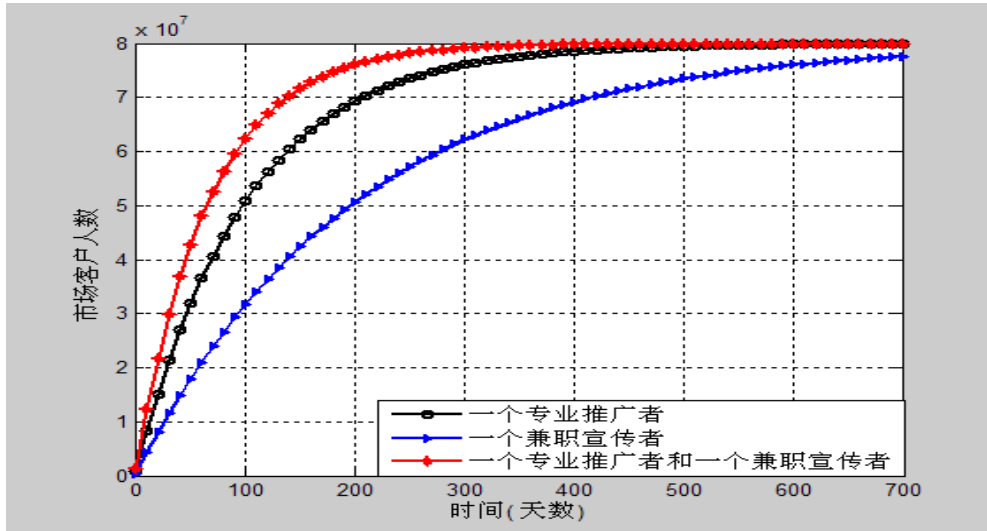


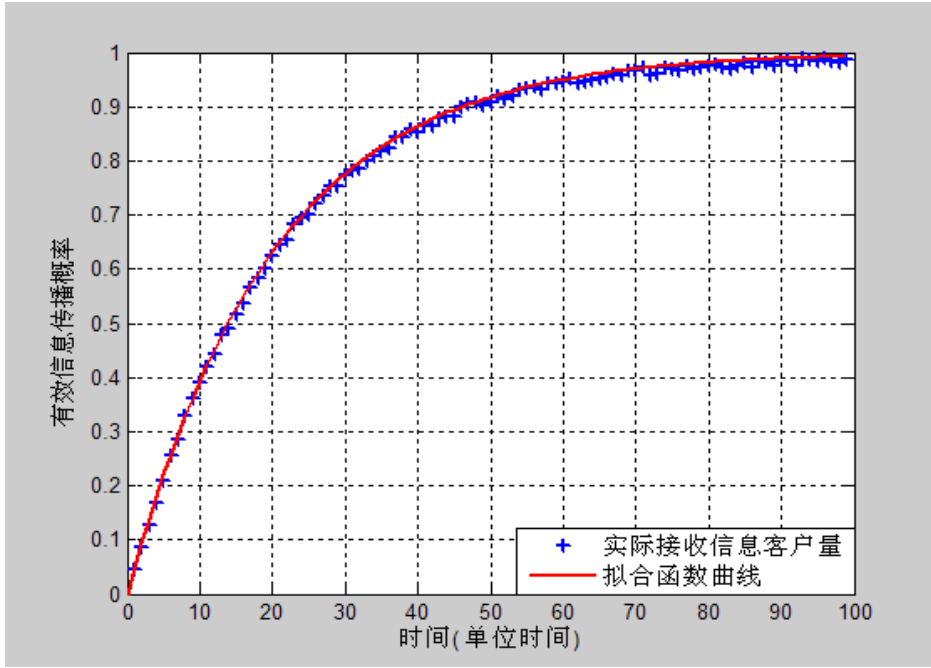
图 10 两类推广者按比例组合工作完成有效推广数量

由上图发现, 一个专业推广者完成 8000 万有效推广数量的时间为 400 天; 一个兼职推广者完成 8000 万有效推广数量的时间为 700 天; 而两者同时工作时, 他们完成 8000 万有效推广数量的时间为 300 天。

2、首先我们先根据问题一涉及的理论知识, 根据实际有效信息传播概率分布图, 在开始度为  $k = 35$  和以后平均度约  $k' = 20$ , 我们借助matlab工具箱 cftool对其拟合, 当兼职推广者开始为  $n_2$  名及市场潜在的潜在用户量为  $N$  时, 在时间  $t$  时刻接收到此信息的市场客户累计人数  $y$  为:

$$y = N * (1 - \exp(-0.0005 * n_2 * t)) \quad (25)$$





考虑时间的限制为 100 天，结合讨论成本因素，最优化两类推广者的比例，制定合理的用人方案。

建立以成本  $Q$  为目标函数及传播数量为约束条件的方程如下：

$$\min Q = 500 \times 100x_1 + 50 \times 100x_2 \quad (26)$$

$$st. \begin{cases} x_1 \geq 0 \\ x_2 \geq 0 \\ 2 \times 10^8 \times (1 - e^{(-0.001 \times 100x_1 - 0.0005 \times 100x_2)}) \geq 8 \times 10^7 \end{cases} \quad (27)$$

由于模型属于整数规划模型，因此可以采用分支定界法对其进行求解。分支定界法<sup>[5]</sup>的基本思路是先求解整数规划的线性规划问题。

### 5.2.2 问题二模型的求解

分支定界法求解整数规划问题的步骤(以求最大化的整数规划为例)：

step1: 确定与整数规划问题（记为问题一 A）对应的松弛线性规划问题记为问题 B

step2: 求解问题 B，如果问题 B 没有可行解，则问题 A 也没有可行解，计算停止；如果问题 B 有最优解，并符合问题 A 的整数条件，则问题 B 的最优解就是问题 A 的最优解，计算停止。若问题 B 有最优解，但不符合问题 A 的整数条件，记问题 B 的目标函数值为  $\underline{Z}$

step3: 用观察法找问题 A 的一个整数解，一般可取  $x_i = 0$  ( $i = 1, 2, \dots, n$ ) 进行试探，

求得其目标函数值，并记作  $\underline{Z}$  以  $\bar{Z}^*$  表示问题 A 的最优目标函数值；这时有： $\underline{Z} < \bar{Z}^* < \bar{Z}$ 。

step4: 在 B 的最优解中任选一个不符合整数条件的变量进行分支，设  $x_i = \bar{b}_i, [\bar{b}_i]$

表示不超过  $\bar{b}_i$  的最大整数，构造两个约束条件  $x_i \leq [\bar{b}_i]$  和  $x_i \geq [\bar{b}_i] + 1$ ，将这两个约束条件分别加入问题 B 中，得到两个后继子问题 B1 和 B2. 不考虑整数条件求解这两个后继子问题。

step5: 以每个后继子问题为一分支标明求解的结果，与其他后继子问题的解的结果进行比较，找出最优目标函数值最大者作为新的上界；从已符合整数条件的各分支中，找出目标函数值最大者作为新的下界  $\underline{z}$ ；若后继子问题的解不符合整数条件，仍然保持原下界值。

step6: 各分支的最优目标函数中若有小于  $\underline{z}$  者，则剪掉这支，即以后不再考虑了；若大于  $\underline{z}$  且不符合整数条件，则重复 step4。一直到最后  $\underline{z} = \bar{z}$  为止，从而得到整数规划问题（问题 A）的最优整数解

算法流程图如下所示：

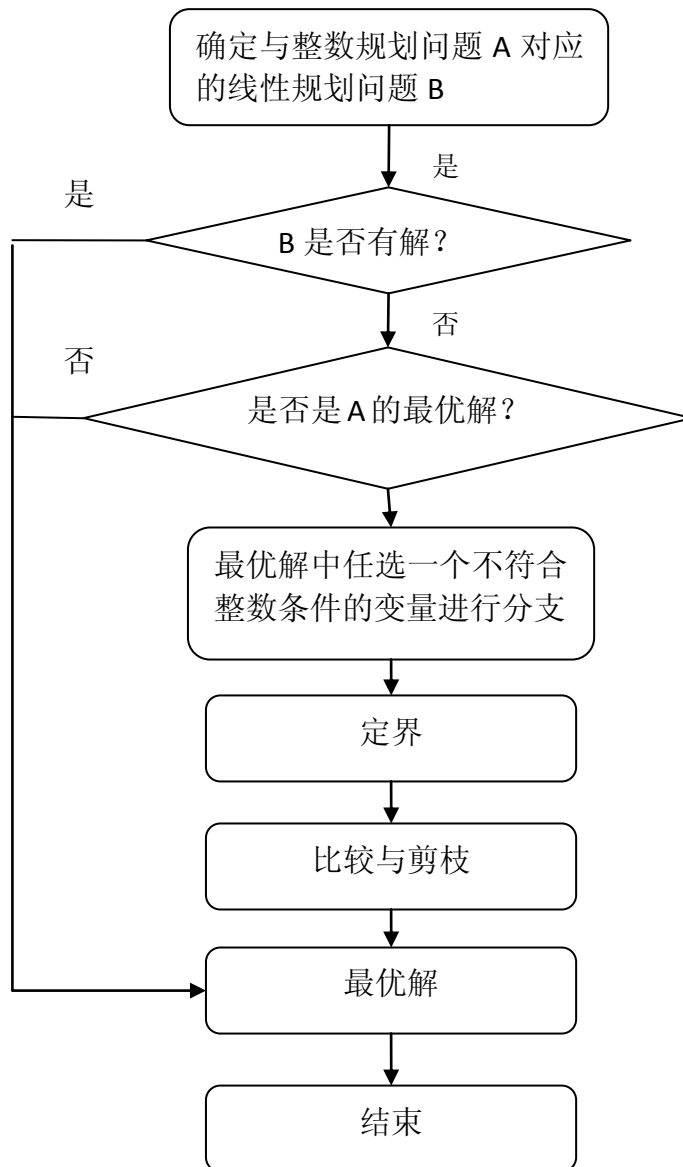


图 11 分支定界算法流程图

利用 matlab 软件分支界定法求解：求得使成本最小，且满足有效传播数量 8000 万的专业推广者与兼职推广者的最优组合为 (2, 7)，即需要雇佣 2 名专业

推广者和 7 名兼职推广者负责该企业的信息网络传播，其解如下图：

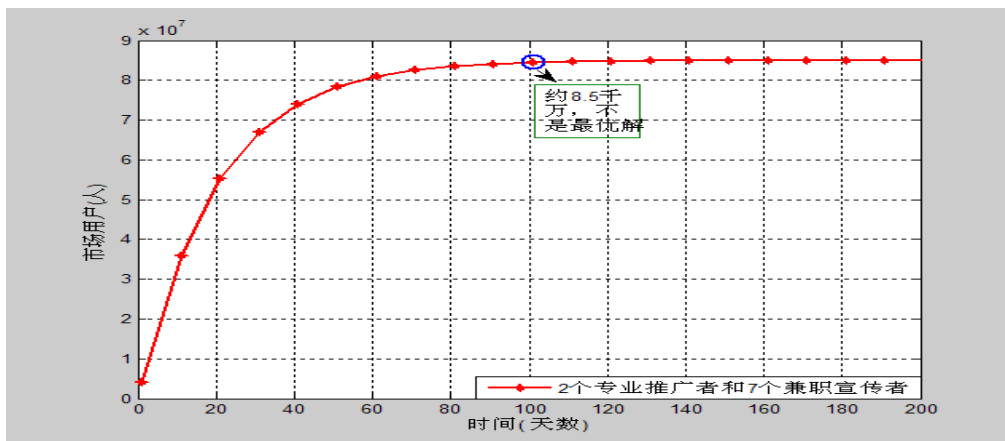


图 12 matlab 软件分支界定法求解

但是在 100 天时所计算的客户约为 8.5 千万人，所发费的成本为  $Q=100*(500*2+50*7)=13.5$  万元，不是最优解。所以我么可以在开始的第一天雇佣 2 名专业推广者和 6 名兼职推广者，再到第 87 天的时候雇佣一个 1 名兼职推广者，计算得到在 100 天时所计算的客户约为 8.0089 千万人，所发费的成本为  $Q=100*(500*2+50*6)+ (100-87)*50=13.065$  万元，为最优解。用 matlab 仿真可得到如图 13 所示的结果。

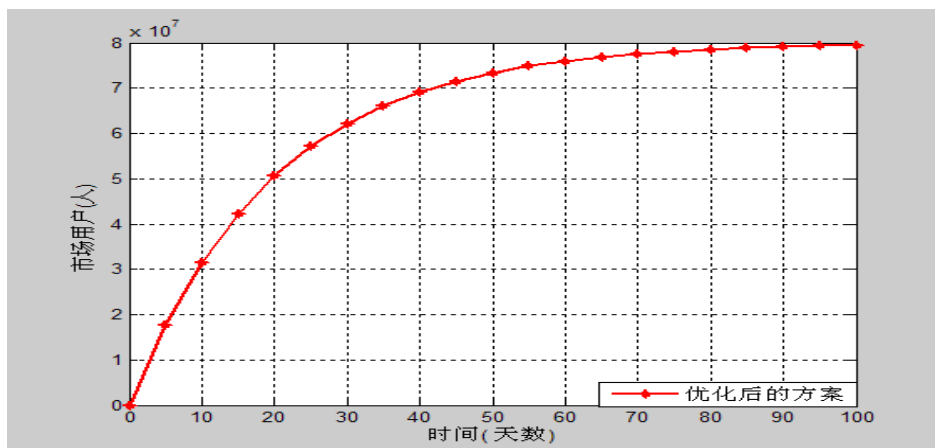


图 13 matlab 仿真求解最优解

综上所述,某个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,在满足以上的要求的同时,考虑到成本,我们给企业制定一份合理的用人方案:在距离奥运会100天时雇佣2名专业推广者和6名兼职推广者,再到第87天的时候雇佣一个 1名兼职推广者,所发费的最低成本为 $Q=13.065$ 万元。

## 六、 模型评价

### 模型的优点

1. 基于 SNS 网络的改进的 SI 信息传播模型定义了三个传播规则,考虑了节点度的影响,结合复杂网络和传染病动力学理论,进一步建立了微分方程组,使建立的模型更符合 SNS 网络的特点;

2. 模型采用了数值计算, 图形观察与理论分析相结合的方法, 可以看作计算机技术与建模方法的巧妙配合;
3. 比较全面地达到了建模的目的, 即描述传播过程、分析感染人数的变化规律, 度量信息传播蔓延的程度;
4. 采用的模型较为人知, 容易让人理解, 可供企业采用, 方便企业决策。

#### 模型的缺点

1. 由于数据来源渠道较少, 使得验证程度有限;
2. 模型假设中使一些条件理想化, 实际网络更加复杂, 使得模型结果与实际会有一定出入。

## 七、 模型推广

本题所涉及的网络类型是连接度大的节点趋向于和其他连接度小的节点连接, 网络非呈现协调混合、无标度类型网络, 基于该网络类型建立的信息传播模型为如何在动态网络、小世界网络中对传播行为进行建模奠定了基础。此模型也可以为传染病、谣言、舆论网络传播等模型的建立提供参考。

## 八、 参考文献

- [1] 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[J], 物理学报, 60 (5): 1-5, 2011。
- [2] 郭海霞, 新型社交网络信息传播特点和模型分析[J], 现代情报, 32(1): 56-59, 2012。
- [3] 孙庆川, 山石, 兰田田, 一种新的信息传播模型及其模拟[J], 图书情报工作, 54 (6): 52-56, 79, 2010。
- [4] 唐四慧, 杨建梅, 两种交互式信息传播网络的传播模型比较研究[J], 科学学研究, 26 (3): 476-479, 2008。
- [5] 周成义, 马垣. 二次整数规划问题的分支定界法. 鞍山钢铁学院学报, 25 (5), 375-378, 2002。
- [6] Hu H B, Han D Y, Wang X F 2010 Physica A 389 1065。
- [7] Hu H B, Wang X F 2010 Phys.Lett. A 37 1105。
- [8] Vázquez A, Weigt M 2003 Phys. Rev. E 67 027101。
- [9] 葛哲学, 孙志强. MATLAB R2007 编程运用[M], 北京: 电子工业出版社, 2007。

## 九、 附录

### 程序一 C++

```
#include <iostream>
#include <string>
#include <fstream>
#include <map>
using namespace std;
int main()
{
void display(const map<string,int> &);
ifstream in_file( "E:\\asd.txt ");
string word;
map<string, int> words;
while (in_file >> word)
{
words[word]++;
}
display(words);
int i;cin>> i;
return 0;
}
void display(const map<string, int>&wordmap)
{
int p=0, q=0;
map<string, int> ::const_iterator it=wordmap.begin();
for (; it!=wordmap.end();it++)
{
cout<<it->first<< "- " <<it-> second << ' ';
q++;
cout<<q;
p=q/835541;
cout<<p;
}
}
```

### 程序二

```
close all;
clear;
clc;
load d.txt;
```

```

x=0:1:99;
cftool;
y1=0.5566*exp(-0.08*x)
y2=1-y1;

figure
plot(x,y1,'r*','LineWidth',1.25,'MarkerSize',3);
hold on
plot(x,y,'-','LineWidth',2);
legend('实际有效信息传播概率','拟合函数曲线',3);
xlabel('时间( 单位时间)','FontSize',10);
ylabel('有效信息传播概率 k','FontSize',10);
grid
figure
y2=1-y1;

plot(x,y2,'+','LineWidth',2);
hold on;
plot(x,1-y,'r-','LineWidth',2);
legend('实际传播重复概率','拟合函数曲线',3);
xlabel('时间( 单位时间)','FontSize',10);
ylabel('传播重复概率 p','FontSize',10);
grid

```

### 程序三

```

close all;
clear;
clc;
n1=1;
n2=1;
x=1:10:1000;
y1=1*exp(-0.01*x)
r1=1-y1;
r2=1-1*exp(-0.005*x);

he=1-1*exp(-(0.005*n2+0.01*n1)*x);

figure

plot(x,r1,'kO-','LineWidth',2,'MarkerSize',4.5);
hold on
plot(x,r2,'b->','LineWidth',2,'MarkerSize',3.5);
hold on

```

```
plot(x,he,'r-','LineWidth',2);
axis([0 700 0 1])

legend('一个专业推广者','一个兼职宣传者','一个专业推广者和一个兼职宣传者',3);
xlabel('时间( 天数)','FontSize',10);
ylabel('', 'FontSize',10);
grid
```

#### 程序四

```
close all;
clear;
clc;
n1=1;
n2=1;
x=1:10:1000;
r1=1-1*exp(-0.01*x);
r3=1-1*exp(-0.03*x);
r4=1-1*exp(-0.04*x);
r5=1-1*exp(-0.05*x);
r1=r1*80000000;
r3=r3*80000000;
r4=r4*80000000;
r5=r5*80000000;
figure
plot(x,r1,'O-','LineWidth',2,'MarkerSize',4.5);
hold on
plot(x,r3,'b->','LineWidth',2,'MarkerSize',3.5);
hold on
plot(x,r4,'r-s','LineWidth',2,'MarkerSize',3.5);
hold on
plot(x,r5,'r-*','LineWidth',2);
axis([0 500 0 80000000])
legend('1 名专业推广者','3 名专业推广者','4 名专业推广者','5 名专业推广者',3);
xlabel('时间( 天数)','FontSize',10);
ylabel('市场用户(人)','FontSize',10);
grid
```

#### 程序五:

```

Function
[newx, newfval, status, newbound]=branchbound(f, A, B, I, x, fval, bound, Aeq, Beq, lb, ub, e)
% 分支定界法求解整数规划
% f, A, B, Aeq, Beq, lb, ub 与线性规划相同
% I 为整数限制变量的向量
% x 为初始解, fval 为初始值
options = optimset('display','off');
[x0, fval0, status0]=linprog(f, A, B, Aeq, Beq, lb, ub, [], options);
%递归中的最终退出条件
%无解或者解比现有上界大则返回原解
if status0 <= 0 || fval0 >= bound
    newx = x;
    newfval = fval;
    newbound = bound;
    status = status0;
    return;
end

%是否为整数解, 如果是整数解则返回
intindex = find(abs(x0(I) - round(x0(I))) > e);
if isempty(intindex)
    newx(I) = round(x0(I));
    newfval = fval0;
    newbound = fval0;
    status = 1;
    return;
end

%当有非整可行解时, 则进行分支求解
%此时必定会有整数解或空解
%找到第一个不满足整数要求的变量
n = I(intindex(1));
addA = zeros(1, length(f));
addA(n) = 1;
%构造第一个分支  $x \leq \text{floor}(x(n))$ 
A = [A; addA];
B = [B, floor(x(n))];
[x1, fval1, status1, bound1] =
branchbound(f, A, B, I, x0, fval0, bound, Aeq, Beq, lb, ub, e);
A(end, :) = [];
B(:, end) = [];
%解得第一个分支, 若为更优解则替换, 若不是则保持原状
status = status1;

```



```

if status1 > 0 && bound1 < bound
    newx = x1;
    newfval = fval1;
    bound = fval1;
    newbound = bound1;
else
    newx = x0;
    newfval = fval0;
    newbound = bound;
end

%构造第二分支
A = [A;-addA];
B = [B,-ceil(x(n))];
[x2,fval2,status2,bound2]
=branchbound(f,A,B,I,x0,fval0,bound,Aeq,Beq,lb,ub,e);
A(end,:) = [];
B(:,end) = [];

%解得第二分支，并与第一分支做比较，如果更优则替换
if status2 > 0 && bound2 < bound
    status = status2;
    newx = x2;
    newfval = fval2;
    newbound = bound2;
end

```