

# 第五届数学中国数学建模网络挑战赛

地址：内蒙古数学学会  
电话：0471-4969085

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2012@tzmcm.cn

---

## 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：张韬

队员 2：杨渭滨

队员 3：刘嘉妮

参赛队教练员（签名）：

参赛队伍组别：本科组

# 第五届数学中国数学建模网络挑战赛

地址：内蒙古数学学会  
电话：0471-4969085

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2012@tzmcm.cn

---

## 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）：

1036

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）：

# 第五届数学中国数学建模网络挑战赛

地址：内蒙古数学学会  
电话：0471-4969085

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2012@tzmcm.cn

## 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

关 键 词 常微分方程 灵敏度 优化模型 基于 SNS 网络的信息传输模型

### 摘 要：

奥运会及奥运会前期的信息传播对于企业利益有着直接的关系。而在网络高速发展的时代，社交网络已成为信息传播的一种途径。本文以企业利用社交网络在奥运会期间进行企业宣传为背景，对社交网络中信息传播人数以及所需费用时间等问题进行了研究。

问题一：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。本文首先将社交网络中的节点分为传播节点、未接触节点及免疫节点三类，并制定社交网络信息传播动态规则。其次，以奥运会开始前 100 天为起始发布消息时间，引入各类节点的比重  $i(t)$ ， $s(t)$ ， $r(t)$  及节点免疫概率  $\mu$ ，建立常微分方程，利用 Matlab 对微分方程进行求解。得出结论：在基数为 1.2 亿的社交网络成员中，大约有 0.35 亿人可看到信息，比例约为 35%。最后，对可变参数  $\mu$  进行灵敏度的讨论。不同的  $\mu$  值下，各节点比重的变化趋势相似，但  $\mu$  值较小时，各比重的变化较为明显。

问题二：本问题分为两个部分。

第一部分要求雇用多名专业信息传播人员对信息进行传播，可看做问题一的深化。本文以花费最小为目标函数，传播覆盖率达到 40% 为约束条件，对第一问的微分方程进行修改从而建立优化模型。利用 Matlab 对模型中的位置参数进行遍历，最终得到结果为：在基数为 9.4 亿的社交网络下，148 名专业传播人士在 16 天的时间内，一共花费 118.4 万人民币，便可将信息传播率达到 40%。

第二部分要求专业传播人员于兼职传播人员共同对信息进行传播。与第一部分相类似，建立优化模型，并用 Matlab 对参数遍历，最终得到结果为：在基数为 9.4 亿的社交网络下，64 名专业传播人员以及 92 名兼职传播人员在 26 天得时间内，一共花费 95 万人民币，便可将信息传播率达到 40%。

在此基础上，经过对问题一模型的分析后发现常微分模型中个别参数存在人为设定的情况，不够精确。因此对模型进行改进，建立基于 SNS 网络的信息传播模型，从而对模型进行了完善。



# 第五届数学中国数学建模网络挑战赛

地址：内蒙古数学学会  
电话：0471-4969085

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)  
Email: 2012@tzmcm.cn

参赛队号 1036

所选题目 C

参赛密码 \_\_\_\_\_  
(由组委会填写)

## 英文摘要（选填）

（此摘要非论文必须部分，选填可加分，加分不超过论文总分的 5%）

Dissemination of information has a direct relationship between the business interests during the Olympic Games. In the era of high-speed network development, the social network has become a means of information dissemination. This paper study on the issue of information dissemination by social network.

Solve the first problem in three steps.

First of all, develop the Social network information dissemination dynamic rules and social network node is divided into three categories, namely dissemination node, no contact node, Immune node.

Then introduce the proportion of various types of nodes and the variable  $\mu$  of node immune probability. We establish the ordinary differential equation and solve the model by using Matlab, the result is approximately 35 million people, 35% of the 120 million people, can see information.

Finally discuss the sensitivity of the variable  $\mu$ . Under different value of  $\mu$ , each node has a similar trend of changes in the proportion, while the proportion of change is more obvious in a smaller value of  $\mu$ .

The paper divide the second problem into two parts.

In the first part, more than one professional personnel are used disseminate information. Base on the first problem, establish optimization model to spend a minimum as the objective function which use coverage as constraints. The result is , in 16 days, 148 professional person can manage the requirement of 40% coverage with 1.184 million yuan.

In the second part, modify the last optimization model and iterate over all integers with Matlab. The result is, in 26 days, 64 professional person and 92 part-time person could finish the requirement of 40% coverage with 950,000 yuan.

On this basis, the individual parameters of ordinary differential equations model exist the situation of man-made, is not precise enough. Therefore, to improve the model, we build SNS network-based information dissemination model.

Keywords: ordinary differential equation optimization model

# 第五届数学中国数学建模网络挑战赛

地址：内蒙古数学学会

电话：0471-4969085

邮编：010021

网址：[www.tzmcm.cn](http://www.tzmcm.cn)

Email: 2012@tzmcm.cn

---

## 1. 问题重述

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。有一个例子：

2012 年 1 月 26 日，一个名为 Jamie Beck 的 Tumblr 博主发布了一张“海怪号 (Mar Mostro)”帆船在沃尔沃环球帆船赛上乘风破浪的照片，随后他收到了 2.5 万条互动信息，其中 60% 是转发这张照片。Jamie Beck 是这艘船的赞助商 Puma 聘请的推广作者，Puma 预计，鉴于 Beck 有 200 万粉丝，这张照片最终可能获得 600 万至 700 万品牌印象度 (Impressions, 衡量到达率的指标之一)，而 Instagram 上会达到 4000 万。在整个沃尔沃帆船赛中，Puma 一共派出 10 位这样的作者去比赛地点阿布扎比，他们在 Twitter, Instagram 和 Tumblr 上更有针对性地发布与 Puma, “海怪号”相关的内容。尽管 Puma 还没有发布它们的奥运广告计划，但 Puma 数字营销负责人 Remi Carlioz 有类似的计划：“我们不是奥运的官方合作伙伴，但我们会想别的办法和我们的受众一起参与到这个话题里来。”

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。

### 第一阶段问题

**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 35 人，考虑到成本，请给企业制定一份合理的用人方案。

附件中的数据是 Twitter 社交网站用户之间的链接关系 (follow 关系) 数据，用于发现用户组，以及分析 Twitter 用户的链接分布。

## ● 2. 问题分析

### 2.1 问题背景

题目着眼于社交网络对信息传播这一现况，解决奥运会信息传递问题。社会性网络服务 (social network service, SNS) 是 Web2.0 体系下的一个技术应用架构。SNS 通过网络聊天、博客、播客和社区共享等途径，实现个体社交圈的逐步扩大，最终形成一

个联结“熟人的熟人”的大型网络社交圈，充分反映出人类社会的六度分离理论（Six Degrees of Separation）[1]，即你和任何一个陌生人之间所间隔的人不会超过五个，也就是说，最多通过五个人你就能够认识任何一个陌生人。根据这个理论，你和世界上的任何一个人之间只隔着五个人，不管对方在哪个国家，属哪类人种，是哪种肤色。

根据以上背景，对问题进行如下分析。

## 2.2 问题一

分析问题可知，第一问是对最终信息传播人数进行求解。在奥运会开幕之前的 100 天，一个专业推广者不断地重复的发布消息，并建立传播网络。将社交网络中的个体看做节点，将节点分为三类：传播节点、未接触节点、免疫节点。分析问题，在建模时，考虑每日新增粉丝数，粉丝看见消息的概率，三种节点所占概率的变化等一系列问题，利用常微分方程对问题进行求解。

## 2.3 问题二

问题二分为两个部分，第二部分是第一部分的扩展。题目中给出要求，最终信息传播需覆盖潜在用户的 40%，而对于现在用户这一群体来说，可以看做是随机分布于整个社交网络中，因此，题目中的要求可看做信息的传播最终达到整个网络的 40%。

第一部分针对多个专业消息传播者进行问题设置。此部分可看做问题一的加深。根据问题一的模型，将其改建，建立规划模型，以花费最少为目标函数，以信息传播覆盖率达到 40%为条件建立优化模型。

第二部分在第一部分的基础上继续深入，在雇用专业宣传人员的同时也雇用兼职网络宣传人员对信息进行宣传。根据上一部分中优化模型的建立，此处，对模型进行修改，从而求得最小花费以及信息传播所需的天数。

# ● 3. 数据预处理

## 3.1 转发率 $\lambda$ 的确定

根据题目背景中对 Jamie Beck 案例的介绍，在总共 600 万到 700 万的品牌印象度中，有 2.5 万条互动信息，其中 60%是转发信息。从而可确定转发率  $\lambda$  为：

$$\lambda = \frac{2.5 \times 60\%}{200} = 0.75\%$$

本文中所有转发率均应用背景所给的转发率。

## 3.2 数据基本处理

表：按粉丝数归类统计表

粉丝数	人数	粉丝数	人数
0-50	356	251-300	74
51-100	166	301-350	57
101-150	139	351-400	74
151-200	104	401-450	229
201-250	77	451-500	1230

由表中可以观察到，在所给 2506 个用户粉丝统计表中，粉丝数量在 451 到 500 之间的人数达到了 1230 人，为比例最大的部分。对粉丝数进行最值均值分别为：500 及 333.4 个。而且第一列所给用户和第二列所给用户之间重叠比例几乎为 0，这就显示出目前 Twitter 用户组的拓扑图还处于初步扩展阶段，即广告还未开始全面传播。所以基于以下假设已知信息的人数占总人数的比重与粉丝重复率相同是合理的。

## ● 4. 模型假设

1. 假设社交网络用户总数量  $N$  不变
2. 假设传播期间，用户都有同等机会接触社交网络。
3. 假设只考虑传播节点对其他节点的影响。
4. 假设社交网络中人群分为三种：传播者、免疫者和未接触者
5. 假设免疫者得到广告后不再分享，而且不会继续被“传染”为传播者，即具有很强免疫力。
6. 假设 2 亿潜在用户在整体社交网络用户中为随机分布。
7. 假设已知信息的人数占总人数的比重与粉丝重复率相同。
8. 假设第二问中兼职宣传人员每日新增的 35 个粉丝与专业宣传者的 500 个粉丝相同，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。

## ● 5. 符号说明

转发率	$\lambda$
传播节点比重	$s(t)$
未接触节点比重	$i(t)$
免疫节点比重	$r(t)$
社交网络中节点个数的基数	$N$
接触后成为免疫节点的概率	$\mu$
专业宣传人员	$x_1$
兼职宣传人员	$x_2$
节点 $j$ 在 $[t, t + \Delta t]$ 时段内处于未感染状态的概率	$p_{ii}^j$
$t$ 时刻节点 $j$ 的邻居中传播节点的数量	$g$
度分布函数	$P(k')$
网络节点平均度	$\bar{k}$

## ● 6. 模型的建立与求解

### 6.1 问题一

分析问题可知，第一问是对最终信息传播人数进行求解。在奥运会开幕之前的 100 天，一个专业推广者不断地重复的发布消息，并建立传播网络。

#### 6.1.1 模型的建立

在 SNS 网络中，一个人发布的消息会被其好友看到，并以一定的概率分享、传播。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 SNS 网络上的用户定义为节点，个体之间的好友关系则可以抽象地用节点之间的边来表示，信息只沿着边传播。根据信息在 SNS 网络中的传播规律，我们把网络中的节点分为三类：传播节点、未接触节点、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，并有机会接受信息。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。

下面针对该问题，利用传染病动力学与复杂网络理论相结合的算法[2][3]，对在线社交网络中的信息传播进行详细的理论建模，有助于更深刻的理解发生社交网络中的传播行为，更加符合网络广告信息的传播机理。

现在我们制定信息传播的动态规则：广告传播节点*i*遇到了节点*j*，如果节点*j*是未感染节点，那么它以 $\lambda$ 的概率成为传播节点；如果节点*j*是传播节点或者免疫节点，那么他就以 $\mu$ 的概率成为免疫节点。

定义*i(t)*，*s(t)*，*r(t)*分别为未接触节点、传播节点和免疫节点在整个系统里所占的比重。显然有*i(t)+s(t)+r(t)=1*。进一步研究发现，在整个网络企业广告传播过程中，*i(t)>0*，而*r(t)<1*，也就是说一部分人从来就没有听说过这个广告，这也和实际情况相符合。

考虑到传播网络建立的过程中，粉丝存在重叠现象，及部分节点在接触信息之后成为免疫节点。根据传播机理，并且将以上两种情况去除，建立平均场常微分方程[4]：

$$\begin{cases} N \frac{di(t)}{dt} = -20\lambda[Ns(t)]i(t) - 500i(t) \\ N \frac{ds(t)}{dt} = 20\lambda[Ns(t)]i(t) - \mu[Ns(t)][s(t) + r(t)] + 500i(t) \\ N \frac{dr(t)}{dt} = \mu[Ns(t)][s(t) + r(t)] \end{cases}$$

其中， $\lambda$ 为转发率， $\lambda=0.75\%$ 。*N*为社交网络中节点个数的基数。根据市场研究公司 InSites Consulting 的数据显示，全球 72%的网民至少已经成为社交网站的用户，总人数达到 9.4 亿，活跃人数为 1.2 亿，故*N*取 1.2。

*N s(t)* 是指社交网络系统中的传播节点；由于一个普通传播节点每天可以新增 20 个粉丝，所以  $20\lambda N s(t)$  是指某个传播节点接触的未接触节点群中可能成为传播节点的个数。

由于假设已知信息的人数占总人数的比重与粉丝重复率相同，所以  $20\lambda N s(t)$  乘 *i(t)* 即考虑了重复率的问题。

而一个专业推广者新增的粉丝都会把信息分享给自己的粉丝，但是分享信息的粉丝中存在以一定概率重叠的现象，所以  $500i(t)$  是指专业推广者平均一天新增的“纯”有效粉丝个数，即不重叠粉丝的个数。

$[s(t)+r(t)]$  是 *t* 时刻未接触过该网广告的比例， $\mu[Ns(t)][s(t)+r(t)]$  是由于不分享广告或者不及时上线等因素而移出传播者成为免疫者的个数，即为第三个方程所代表的含义。

第一个方程的意思是：未接触节点减少个数分为普通传播者和专业推广者让其变为已知者的人数，即有  $20\lambda N s(t) + 500i(t)$  转为传播者。

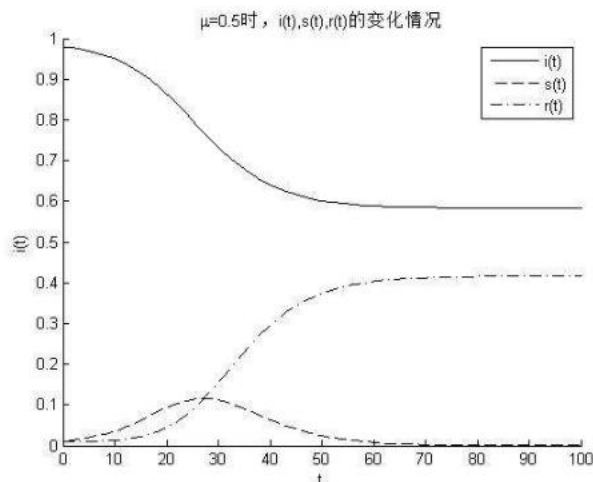
第二个方程的意思是：传播节点增加的个数分别为未接触点转移到普通传播点的个数  $20\lambda N s(t) + 500i(t)$  和由于种种原因移出传播节点范围而成为免疫者所减少的个数  $\mu[Ns(t)][s(t)+r(t)]$ 。

由上述分析可知，三个方程相加可得  $\frac{ds(t)}{dt} + \frac{dr(t)}{dt} + \frac{di(t)}{dt} = 0$ ，这与  $i(t) + s(t) + r(t) = 1$

相一致，所以所建模型较为合理。

### 6.1.2 模型的求解

利用 *Matlab* 求解常微分方程，对信息经过各节点传播过程进行求解。取  $i(t)$ ,  $s(t)$ ,  $r(t)$  的初值分别为 (0.99998, 0.00001, 0.00001)。模型中含有未知参数  $\mu$ ，结合现实并且考虑其意义，对其进行赋值，取  $\mu$  为 0.5 进行模型求解，结果如下图所示：



对上图进行分析可得整个信息传播的过程，简单的概括如下：

首先系统里只有少量的传播节点  $s(t)$ ，即专业传播人员以及每日增加的 500 粉丝。其它都为未接触节点，免疫节点的数量为 0。

随着传播节点开始散播信息，未接触节点的比重  $i(t)$  很快减少，传播节点的比重  $s(t)$  急剧增加。随着信息的进一步扩散，免疫节点的比重  $r(t)$  开始增加而传播节点的数量达到一个峰值以后开始下降。最后，传播节点的数量变为了 0，而网络里就只剩下免疫节点和未接触节点。

当  $\mu$  为 0.5 时，求得最终免疫节点比重  $r(t)$  的稳定值为  $r(t) = 0.35$ ，从而可得奥运会开始后信息传播人数为：

$$n = N \times r(t) = 1.2 \times 0.35 = 0.3526 \text{ 亿}$$

即，100 天后，在基数为 1.2 亿的社交网络成员中，大约有 0.35 亿人可看到信息，比例约为 35%。

### 6.1.3 对 $\mu$ 值灵敏度的讨论

考虑到在求解模型的过程中对  $\mu$  进行了人为赋值，故在此对不同  $\mu$  值得灵敏度进行讨论，结果如下：

表：不同  $\mu$  值对应各项节点比重值表

$\mu$	$s(t)$ 最大值	$r(t)$ 稳定值	$i(t)$ 稳定值
0.1	0.39	0.86	0.16
0.3	0.19	0.58	0.42
0.5	0.12	0.42	0.58
0.7	0.09	0.33	0.67
0.9	0.07	0.27	0.73

由上表可以明确看出，当  $\mu=0.3$  时，传播节点比重的最大值为  $s(t) = 0.19$ ，而免疫节点比重的稳定值为  $r(t) = 0.58$ ，未接触节点比重的稳定值为 0.42。

不同  $\mu$  值对于  $i(t)$ ,  $s(t)$ ,  $r(t)$  的影响如下图所示:

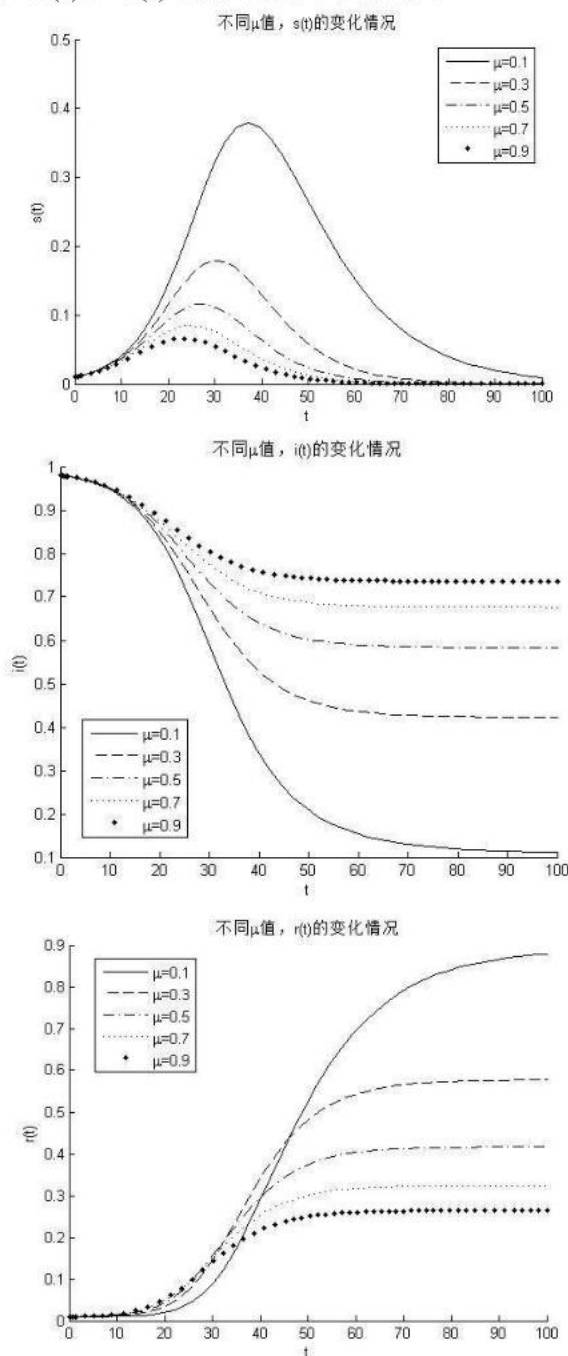


图: 不同  $\mu$  值对应各项节点比重示意图

以传播节点的比重  $s(t)$  为例, 对上图进行分析可看出,  $\mu$  的取值对  $s(t)$  的变化趋势没有直接影响。 $s(t)$  的峰值表示在整个信息传播的过程中出现的最大传播节点的数目, 它在一定程度上反映了信息传播所造成的最大影响。

结合实际, 信息传播的初期, 传播节点的数量以较高的速率增加, 随着时间的推移, 越来越多的节点变为免疫节点, 从而使得传播节点的比重下降, 最终信息达到饱和。

对于不同的  $\mu$  值进行对比, 可看出,  $\mu$  的值越大,  $s(t)$  的最大比重越大, 同时,  $\mu$  值较小时,  $s(t)$  的变化较为明显, 例如,  $\mu=0.1$  时与  $\mu=0.3$  进行对比,  $s(t)$  的最大比重变化明显于  $\mu$  趋近于 1 时的变化。

同理，可对不同  $\mu$  值情况下的免疫节点比重以及未接触节点比重进行分析。

综上所述，在对信息传播人数进行预测的具体案例中，需根据当下情况，选取较为合适的  $\mu$  值对总人数进行求解。

## 6.2 问题二

本问分为两部分，第二部分是对第一部分问题的扩展延伸。根据前一问的结果可知，专业人士粉丝增长速度是由快到慢的。因为信息开始传播时，大部分的用户是未接触节点，而随着节点重叠和免疫节点的增加，造成增长缓慢。

所以，增加相同粉丝数离“0时刻”越早，所需的时间越少。因此，若要雇用多名专业人士并且费用尽可能的少，则需在开始时一同雇用多名专业人士。即不考虑雇用专业人士之间的时间差，均为  $t$  的函数。

由分析可知，本问建立优化模型对问题进行求解。

### 6.2.1 模型的建立

#### (一) 第一部分

假设雇用专业人士的数量为  $x$  人，则在  $t$  时刻到  $t + \Delta t$  时刻的传播过程中，可将传播节点分为两类，分别为：直接与专业人士接触的新粉丝以及与非专业人士接触的粉丝。

当  $t$  以天为单位，并且  $\Delta t$  为最小单位一天时，与专业人士接触的新粉丝数为： $500i(t)x$ 。其中， $i(t)$  为未接触节点的比重。

#### (1) 约束条件

题目要求广告宣传覆盖潜在用户的40%。根据假设，潜在用户随机分布于所有社交网络的用户中。因此，可认为广告宣传覆盖所有社交网络用户中的40%时，宣传工作完成。即为：

$$s(t) + r(t) \geq 40\%$$

其中， $s(t)$  为传播节点的比重， $r(t)$  为免疫节点的比重。

#### (2) 目标函数：

我们的目标是花费最少，即为：

$$\min M = 500x \bullet t$$

其中， $t$  为传播达到40%时的总时间

#### (3) 优化模型[5-7]

根据第一问中信息传播动态规则，以及约束条件，目标函数，建立如下优化模型：



$$\begin{aligned}
 &obj: \min M = 500x \bullet t \\
 &s.t \begin{cases} s(t) + r(t) \geq 40\% \\ N \frac{di(t)}{dt} = -20\lambda Ns(t)i(t) - 500i(t)x \\ N \frac{ds(t)}{dt} = 20\lambda[Ns(t)]i(t) - \mu[Ns(t)][s(t) + r(t)] + 500i(t)x \\ N \frac{dr(t)}{dt} = \mu[Ns(t)][s(t) + r(t)] \\ t > 0, x > 0 \text{ 且为整数} \end{cases}
 \end{aligned}$$

其中,  $\lambda$  为转发率,  $\lambda = 0.75\%$ 。  $N$  为社交网络中节点个数的基数。

由于只有直接和专业推广者接触的粉丝才具有被专业者影响的机会, 所以需  $500i(t)$  要乘以  $x$ , 而  $20\lambda Ns(t)i(t)$  是基于现有的整个传播者影响的粉丝, 所有不用乘以  $x$ 。简单的说, 即传播者与接收者之间所存在的上下关系不同引起的。

## (二) 第二部分

第二部分就第一部分比较来说, 增加了兼职宣传人员, 将问题复杂化, 增加了灵活性。同样建立优化模型对问题进行求解。

假设  $x_1$  为专业宣传人员的人数,  $x_2$  为兼职宣传人员的人数。与第一部分的思路类似, 也是将某一时刻的传播过程看成两部分粉丝的集合进行考虑。

### (1) 目标函数

由于新增兼职宣传人员, 则以最小花费为目标函数, 即为:

$$\min M = 500x_1t + 50x_2t$$

其中,  $t$  为传播达到 40% 时的总时间

以信息传播动态规则以及与第一部分相同的约束条件建立优化模型, 具体模型如下所示:

$$\begin{aligned}
 &obj: \min M = 500x_1t + 50x_2t \\
 &s.t \begin{cases} s(t) + r(t) \geq 40\% \\ N \frac{di(t)}{dt} = -20\lambda Ns(t)i(t) - i(t)(500x_1 + 35x_2) \\ N \frac{ds(t)}{dt} = 20\lambda[Ns(t)]i(t) - \mu[Ns(t)][s(t) + r(t)] + i(t)(500x_1 + 35x_2) \\ N \frac{dr(t)}{dt} = \mu[Ns(t)][s(t) + r(t)] \\ t > 0, x_1 \geq 0, x_2 \geq 0 \text{ 且为整数} \end{cases}
 \end{aligned}$$

其中,  $\lambda$  为转发率,  $\lambda = 0.75\%$ 。  $N$  为社交网络中节点个数的基数。

## 6.2.2 模型的求解

### (一) 第一部分

利用 Matlab, 对模型进行求解。考虑专业宣传者人数  $x$  以及天数  $t$  均为正整数, 故

对其 1 到 10000 进行遍历，找出最优解，从而求得传播最快的时间以及最小的花费，具体结果如下所示：

表：多个专业宣传者进行广告宣传细则

价格/元	人数	天数	$i(t)$	$s(t)$	$r(t)$
1184000	148	16	0.999592	0.000307	0.0001

## (二) 第二部分

同理，分别对专业推广者人数  $x_1$ 、兼职宣传者人数  $x_2$  以及天数  $t$  均为正整数，故对其进行遍历，找出费用最优解，如下表：

表：多个专业宣传者和兼职宣传者进行广告宣传细则

价格	人数 $x_1$	人数 $x_2$	天数	$i(t)$	$s(t)$	$r(t)$
951600	64	92	26	0.999601	0.000299	0.0001

根据所得结果进行分析：专业推广者的理论传播“性价比”为 1 元/人，而兼职推广者的理论传播性价比为 1.4 元/人，所以按这个理论，总是很惯性的认为，不需要兼职推广者，即  $x_2 = 0$ ，但是结果却截然不同，究其原因是因为重叠率和性价比的问题，虽然起初专业推广者的信息传播速度很快，但是随着时间的增加，其传播重叠率也增加较快，造成传播“性价比”下降较多。所以起初传播节点的度可以影响传播速率，也可以影响覆盖率。

## ● 7. 模型评价

### 7.1 模型优点：

- (1) 利用传染病动力学与复杂网络理论相结合的算法，更加符合社交网络中，广告信息的传播机理。
- (2) 应用平均场常微分方程，巧妙地消除了粉丝存在重叠现象。
- (3) 通过对  $\mu$  值灵敏度的讨论，得到了  $i(t)$ ， $s(t)$ ， $r(t)$  关于不同  $\mu$  值的变化规律。

### 7.2 模型缺点：

在微分方程中，个别参数存在人为设定的情况，不够精确

## ● 8. 模型推广

### 8.1 改进微分方程模型

由于问题一中常微分方程模型以上模型是建立在传播行为具有某种规律的基础上的，但是 SNS 网络打破了传统的传播方法，它利用人与人之间的关系改变人与信息的关系，反过来又用人与信息的关系影响人与人之间的关系。所以，下面针对该问题，利用传染病动力学与复杂网络理论相结合的算法，对在线社交网络中的信息传播进行详细的理论建模，有助于更深刻的理解发生社交网络中的传播行为，更加符合网络广告信息的传播机理。故对微分方程模型进行改进。

根据前文对 SNS 的介绍，可知节点在传播状态、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，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，则未感染节点会以概率  $p_1$  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, 会以一定的速度停止传播而变为免疫节点, 且无需与其他节点接触。

以下建立基于 SNS 网络的信息传播模型[8]

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  $j$  从未感染状态转移到传播状态的概率且  $p_{is}^j = 1 - p_{ii}^j$ , 其中

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

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,  $g$  是具有如下二项分布的随机变量:

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

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。 $\omega(k, t)$  可写为如下形式:

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

式中,  $p(k' | k)$  为度相关函数, 表示度为  $k$  的节点与度为  $k'$  的节点邻接的条件概率;  $p(s_{k'} | i_k)$  表示一个拥有  $k'$  条边的节点在它连接到一个度为  $k$  的未感染节点的条件, 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:

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

将  $\omega(k, t)$  代入上式可得:

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

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  $p_{ss}^j = (1 - \Delta t p_2)^g (1 - v \Delta t)$

于是, 得到度为  $k$  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:

$$\begin{aligned} \overline{p_{ss}}(k, t) &= (1 - \Delta t p_2 \sum_k p(k' | k) \rho^r(k', t))^k \\ &\quad \times (1 - v \Delta t) \end{aligned}$$

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  $I(k, t)$ ,  $S(k, t)$ ,  $R(k, t)$  分别为在  $t$  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:

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

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:

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

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:

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

当  $\Delta t \rightarrow 0$  时, 对上式右侧进行泰勒展开得到:

$$\frac{\partial p^i(k, t)}{\partial t} = -k p_1 \rho^i(k, t) \sum_k p(k' | k) \rho^s(k', t)$$

同理, 可以得到

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

由以上三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关系, 传播动力学过程同时受到网络拓扑结构和传播机理的影响。

网络中节点的度相关性被用来描述不同网络结构之间的差异。节点的度相关性, 也称为网络选型连接性, 指的是网络中与高度数(或低度数)节点相连接的节点的度数偏向于高还是低。若连接度大的节点趋向于和其他连接度大的节点连接, 则认为网络呈现协调混合; 若连接度大的节点趋向于和其他连接度小的节点连接, 则认为网络非呈现协调混合。

实际的网络的选型连接性有一些呈现协调混合, 一些呈现非协调混合。如社会网络(电影演员合作网络、科学家合作网络)中节点具有正的度的相关性, 其他类型的网络(信息网络、技术网络、生物网络)则相反。

根据 SNS 网络的度相关性[9]一般小于零, 所以, 本文的度相关函数可以写为:



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

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

首先根据数据所给的 follow 关系, 生成一个社交网络, 并进行拓扑网络每天更新。得到该网络的基本数据: 节点总数, 平均度, 最大度, 聚类系数, 同配系数, 并根据初始度, 设定参数  $p_1$ ,  $p_2$ ,  $v$ , 进行迭代, 最终得到  $s(t)$ ,  $i(t)$ ,  $r(t)$  的变化, 进而估计出一条奥运广告可以被看到的人数。

## ● 9. 参考文献

- [1] 郭海霞, 新型社交网络信息传播特点和模型分析, 现代情报, 第 1 期 第 32 卷: 56-59, 2012. 1
- [2] 张芳, 司光亚, 罗批, 谣言传播模型研究综述, 国防大学信息作战与指挥训练中心报, 第 2 期 第 13 卷: 332-343, 2011. 2
- [3] 周涛, 王达, 曾燕, 复杂网络上传播动力学研究综述, 自然科学进展, 第 5 期 第 15 卷: 513-518, 2005. 5
- [4] 潘灶烽, 汪小帆, 李翔, 可变聚类系数无标度网络上的谣言传播仿真研究, 系统仿真学报, 第 8 期 第 18 卷: 2346-2348, 2006. 8
- [5] 黄忠霖, 黄京, 基本科学计算的 MATLAB 实现, 北京: 国防工业出版社, 2011. 3
- [6] 姜启源, 谢金星, 叶俊, 数学模型 (第三版), 北京: 高等教育出版社 2003. 2
- [7] 吴鹏, Matlab 高效编程技巧与应用, 北京: 北京航空航天大学出版社, 2010. 5
- [8] 张彦超, 刘云, 张海峰, 程辉, 熊飞, 基于在线社交网络的信息传播模型, 物理学报, 第 60 期, 第 5 卷: 72-83, 2011. 5
- [9] 金永生, 王睿, 陈祥兵, 企业微博营销效果和粉丝数量的短期互动模型, 管理科学, 第 4 期 第 24 卷: 72-83, 2011. 8

## ● 附件

Fun1

```
%f=@(N,mu)@(t,i)(20*(N*i-1)*(1-i)-mu*N*i+500*(1-i))/N;
f=@(mu,lamda,N)@(t,y)[-20*lamda*y(2)*y(1)-500*y(1)*lamda/N;

(20*lamda*N*y(2)*y(1)-mu*N*y(2)*(y(2)+y(3))+500*y(1)*lamda)/N;
mu*y(2)*(y(2)+y(3))];

tspan=[0,100];
y0=[0.998 0.001 0.001];

%N=10*100000000;
%mu=0.8;

color='k';
dian={'-','--','-.',':','. '};
mu=[0.1 0.3 0.5 0.7 0.9];
smax=zeros(5,1);
```

```

rwen=zeros(5,1);
for k=1:5;
lamda=0.0075;
N=5*100000000;
[tt yy]=ode45(f(mu(k), lamda, N), tspan, y0);

smax(k)=max(yy(:,2));
rwen(k)=yy(end,3);

figure(1)
hold on
plot(tt, yy(:,1), strcat(dian{k}, color));
figure(2)
hold on
plot(tt, yy(:,2), strcat(dian{k}, color));
figure(3)
hold on
plot(tt, yy(:,3), strcat(dian{k}, color));
end
figure(1)
legend(' \mu=0.1', ' \mu=0.3', ' \mu=0.5', ' \mu=0.7', ' \mu=0.9');
title(' ^2 » Í¬\muÖµ£¬i(t) µÄ±ä»¯Çéìö');
xlabel(' t');
ylabel(' i(t)');

figure(2)
legend(' \mu=0.1', ' \mu=0.3', ' \mu=0.5', ' \mu=0.7', ' \mu=0.9');
title(' ^2 » Í¬\muÖµ£¬s(t) µÄ±ä»¯Çéìö');
xlabel(' t');
ylabel(' s(t)');

figure(3)
legend(' \mu=0.1', ' \mu=0.3', ' \mu=0.5', ' \mu=0.7', ' \mu=0.9');
title(' ^2 » Í¬\muÖµ£¬r(t) µÄ±ä»¯Çéìö');
xlabel(' t');
ylabel(' r(t)');

fun2
f=@(mu, lamda, N)@(t, y)[-20*lamda*y(2)*y(1)-500*y(1)*lamda/N;

(20*lamda*N*y(2)*y(1)-mu*N*y(2)*(y(2)+y(3))+500*y(1)*lamda)/N;
mu*y(2)*(y(2)+y(3))];

```

```

tspan=[0, 100];
y0=[0.998 0.001 0.001];

%N=10*100000000;
%mu=0.8;

color='k';
dian={'-', '--', '-.', ':', '.'};

mu=0.5;
lamda=0.0075;
N=940000000;
[tt yy]=ode45(f(mu, lamda, N), tspan, y0);
figure(1)
hold on
plot(tt, yy(:, 1), strcat(dian{1}, color));
plot(tt, yy(:, 2), strcat(dian{2}, color));
plot(tt, yy(:, 3), strcat(dian{3}, color));
legend('i(t)', 's(t)', 'r(t)');
title('\mu=0.5\pm\epsilon i(t), s(t), r(t)\mu\pm\epsilon\gg\epsilon\epsilon');
xlabel('t');

Fun2_1
h=yy(:, 2)+yy(:, 3);
t=round(tt);
n=length(t);
g=1;
min=100000000;
for k=1:n
    for x=1:1000;
        if t(k)>0&x*h(k)>=0.4
            count(1)=500*t(k)*x;
            count(2)=x;
            count(3)=t(k);
            count(4)=yy(k, 1);
            count(5)=yy(k, 2);
            count(6)=yy(k, 3);
            %g=g+1;
            if count(1)<min
                min=count;
                zcount=count;
            end
        end
    end
end

```

```

end
end

fun2_2
f=@(mu, lamda, N, x)@(t, y) [-20*lamda*y(2)*y(1)-500*y(1)/N*x;

(20*lamda*N*y(2)*y(1)-mu*N*y(2)*(y(2)+y(3))+500*y(1)*x)/N;
mu*y(2)*(y(2)+y(3))];

tspan=[0, 100];
y0=[0.9998 0.0001 0.0001];

mu=0.5;
lamda=0.0075;
N=940000000;

min=100000000;
for x=1:1000;
    [tt yy]=ode45(f(mu, lamda, N, x), tspan, y0);
    t=round(tt);
    n=length(tt);
    h=yy(:, 2)+yy(:, 3);
    for k=1:n
        if t(k)>0&h(k)>=0.4
            count(1)=500*t(k)*x;
            count(2)=x;
            count(3)=t(k);
            count(4)=yy(k, 1);
            count(5)=yy(k, 2);
            count(6)=yy(k, 3);

            if count(1)<min
                min=count;
                zcount=count;
            end
        end
    end
end
end

fun2_3
f=@(mu, lamda, N, x1, x2)@(t, y) [-20*lamda*y(2)*y(1)-(500*x1+35*x2)*y(1)/N;

(20*lamda*N*y(2)*y(1)-mu*N*y(2)*(y(2)+y(3))+(500*x1+35*x2)*y(1))/N;

```



```

mu*y(2)*(y(2)+y(3))];

tspan=[0,100];
y0=[0.9998 0.0001 0.0001];

mu=0.5;
lamda=0.0075;
N=940000000;

min=100000000;
for x1=1:100
    for x2=1:100;
        [tt yy]=ode45(f(mu, lamda, N, x1, x2), tspan, y0);
        t=round(tt);
        n=length(tt);
        h=yy(:,2)+yy(:,3);
        for k=1:n
            if t(k)>0&h(k)>=0.4
                count(1)=500*t(k)*x1+50*t(k)*x2;
                count(2)=x1;
                count(3)=x2;
                count(4)=t(k);
                count(5)=yy(k,1);
                count(6)=yy(k,2);
                count(7)=yy(k,3);

                if count(1)<min
                    min=count;
                    zcount=count;
                end
            end
        end
    end
end
end
end

```