

# 第十届华中地区大学生数学建模邀请赛

## 承 诺 书

我们仔细阅读了第十届华中地区大学生数学建模邀请赛的竞赛细则。

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。

我们的参赛报名号为： 97

参赛队员（签名）：

队员 1：           

队员 2：           

队员 3：           

---

---

湖北省工业与应用数学学会

第十届华中地区大学生数学建模邀请赛组委会

# 第十届华中地区大学生数学建模邀请赛

## 编号专用页

选择的题号：     B    

参赛的编号：     97    

---

(以下内容参赛队伍不需要填写)

竞赛评阅编号：

# 第十届华中地区大学生数学建模邀请赛

题目： 基于通讯数据的社区聚类与信息传播研究

## 【摘 要】

本文针对复杂网络的社群聚类与影响力最大化等问题，建立了相应的数学模型，并对于多因素共同作用的复杂网络结构建立了一个综合的评价模型以获取更符合实际情况的社群分组。

我们将前三个问题归结于社群聚类问题。首先基于图论构建邻接矩阵来描述无向图和有向图网络，然后基于广度优先搜索（BFS）算法构造网络的相似矩阵来定量描述网络中任意两点之间的相似关系；对于复杂通讯网络的聚类，将其类比有向图构建邻接矩阵，然后采用基于优化  $Q$  函数的快速复杂网络聚类（FN）模型对其聚类分析，共分为 4 组，绘出其网络结构图并根据无重叠的意见领袖模型讨论各组的意见领袖；最后对  $Q$  函数关于分类簇数进行灵敏性分析，得出  $Q$  函数随之起伏较大的结论，并证实该模型分类簇数的正确性。

针对问题四，这是一个基于信息传播的影响力最大化问题，第一步通过基于 *Random Search* 思想改进的爬山贪婪（*Hill-Climbing Greedy*）算法计算各节点影响力，得出节点影响力排序表；第二步采用 FN 算法进行  $n$  簇分类并选出每类影响力最大节点；第三步确定信息传播机制建立线性阈值模型对该通讯网络进行仿真模拟。经测试，我们的方案对于给定的分类簇数能够自动选取最优投放点在三步之内实现 90% 以上的信息覆盖率。

针对问题五，我们对社群聚类问题做了更深一步的分析，提出适用于多因素共同作用网络的基于主成分分析（PCA）的聚类模型。首先对通讯网络的特征进行主成分分析得到综合评价价值，作为权值赋以网络中的边集得到带权邻接矩阵；然后通过 FN 算法对附件中的通讯网络重新聚类分析，绘出其网络结构图并与问题三结果相对比，结果表明基于主成分分析的聚类模型能够更好地表现出实际复杂网络系统中各节点的内部功能。

最后对以上问题的模型进行了评价和推广，对处理相关实际问题有一定的参考价值！

关键词：邻接矩阵 广度优先搜索算法 基于优化  $Q$  函数快速复杂网络聚类模型  
改进爬山贪婪算法 线性阈值模型 主成分分析法

## 1、问题重述

### 1.1 问题背景

马克思说过“人是一切社会关系的总和”。现代生活中，随着信息时代的快速发展，我们和他人的连接日趋紧密。我们每个人都处在各种各样的社交网络中，如何从纷繁的复杂网络中去发掘更多有价值的信息，成了备受关注的研究问题。

社群聚类与信息挖掘即是这类复杂网络研究中的一个热点。通过对具有相似特性的节点进行聚类，可以识别具有相似喜好的用户社区。这既有助于为用户提供更有针对性的推荐服务，也可以使广告的投放成本大幅降低，效果大幅提高，具有巨大的研究与商业价值！

### 1.2 问题描述

#### 问题一：基于图论的聚类

(1) 给定六个点代表六个个体，由个体之间的关系构建**无向图**。要求根据图中信息区分个体之间的差异，对其进行分群。

(2) 给定五个点代表五个个体，由个体之间的关系构建**有向图**。要求在识别个体的同时，进一步**刻画任意两点之间的相似关系**，以达到“物以类聚”的目的。

(3) 结合完善(1)、(2)中的数学模型，根据附件一中某营业部近三个月的内部通讯记录对个体进行分群聚类，并分析类群特征规律以得出相应结论。

#### 问题二：确定定向信息传播的最佳传播覆盖率

在社群聚类应用中，可以定向对某些节点进行信息投放（如投放广告或制造商业谣言），通过现有通讯网络的传播，达到较高的信息传播率。

现假设每投放一个节点的信息成本为 $m$ ，信息在传播过程中每经过一个节点都会有10% ~ 50%的终止传播概率，要求以附件一中所给记录计算最少要在哪些节点投放信息就可以达到最大的信息传播覆盖率。

#### 问题三：综合考虑多种因素建立综合模型挖掘潜在信息

上述问题的建模过程中或多或少忽略了许多有用的信息，比如通讯的位置、时长、频次等等，要求考虑多种因素建立综合数学模型，挖掘更多的信息。

可自行补充数据验证建立的综合模型，得出相应结论。

## 2、问题分析

### 2.1 问题一的分析

问题一要求先对所给定的无向图和有向图分别建立聚类模型，实现对图中所显示个体的分群聚类，其中在对有向图分析时，需要刻画出任意两点之间的相似关系。然后结合并完善两个模型，根据具体通讯记录确定出衡量个体之间连接关系强弱的参数，根据这一参数使用具体算法将其聚类为不同的簇，由此对个体分群聚类并得出相应结论。

### 2.2 问题二的分析

问题二要求对该营业部群体网络在综合考虑成本、终止传播概率情况下确定最少在哪些节点（个体）投放信息可达到最大的信息传播覆盖率。这是一个影响力最大化传播问题，首先根据传播覆盖率确定节点影响力的评价标准，然后根据问题一的方法确定不同类别中最具影响力的节点，对其投放信息，并确立信息传播机制。最终通过模拟得出信息传播覆盖率。

### 2.3 问题三的分析

问题三要求考虑位置、时长、通话频率等多种因素修正综合的数学模型，挖掘出更多的信息，得出相关结论。并且可以自行补充数据，验证自己的模型，对比得出相关结论。

## 3、模型假设与符号系统

### 3.1 模型假设

针对本问题，建立如下假设：

- （1） 参数准确性假设：附件中按照通话记录所取参数值均准确无误；
- （2） 信息湮没假设：节点的信息湮没概率与通话总强度成正相关；
- （3） 转移概率假设：信息传播过程中的转移概率与该节点与其他节点的联系强度成线性正相关。

### 3.2 符号系统

符号	意义
$V = \{v_i\}_{i=1}^n$	顶点集

$E = \{(v_i, v_j)   v_i, v_j \in V\}$	边集
$W(A, B, C)$	邻接矩阵
$W_{\otimes}$	相似矩阵
$G(V, E)$	网络结构图
$Q$	$Q$ 函数
$k_i$	节点 $i$ 的度
$m$	网络中总的边数
$C$	网络簇内实际的连接数目
$X$	随机的连接情况下簇内期望的连接数目
$\theta_v$	阈值
$b_{v\omega}$	节点 $v$ 的激活概率
$AN_v$	已接受信息节点集
$\delta(AN_v)$	节点能够影响成功的节点数目
$\xi$	信息覆盖率
$P$	总结点个数
$(X_1, X_2)$	两个指标：平均时长、通话频率
$Z_k$	主成分
$O_k$	主成分 $Z_k$ 的贡献率
$\Delta$	综合评价值
$\mu_j$	第 $j$ 个指标的样本均值
$s_j$	第 $j$ 个指标的样本标准差
$R$	相关系数矩阵

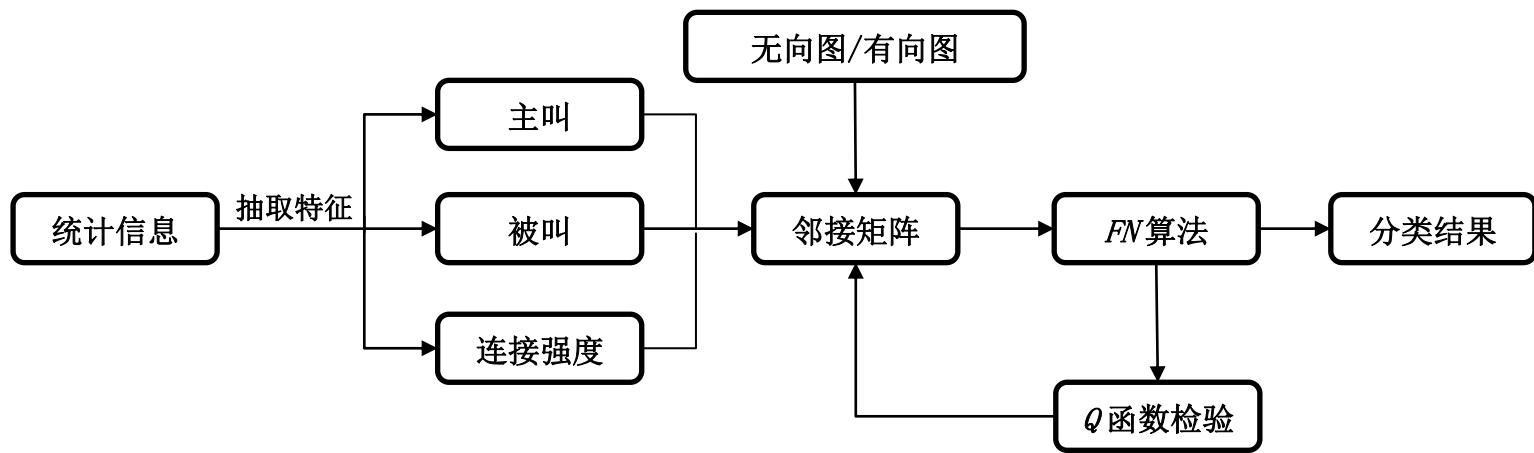
• 注：未列出的符号以及重复符号以出现处为准。

## 4、问题一的建模与求解

### 4.1 问题一的分析

在问题一中，首先明确无向图和有向图各节点之间的相似关系，根据表征相似关系的邻接矩阵<sup>[1]</sup>采用基于局部搜索的快速复杂网络聚类算法<sup>[2]</sup>（*FN*）对有向图和无向图进行聚类；对于实际通讯记录的社群划分问题，首先获取其统计信息，根据个体之间的互联频次确定该网络的邻接矩阵，然后采用 *FN* 算法对其聚类并采用网络模块性评价函数（*Q* 函数）<sup>[3]</sup>定量地刻画网络簇结构的优劣。

问题一的建模流程图如下：



### 4.2 建模准备

#### 4.2.1 网络结构图

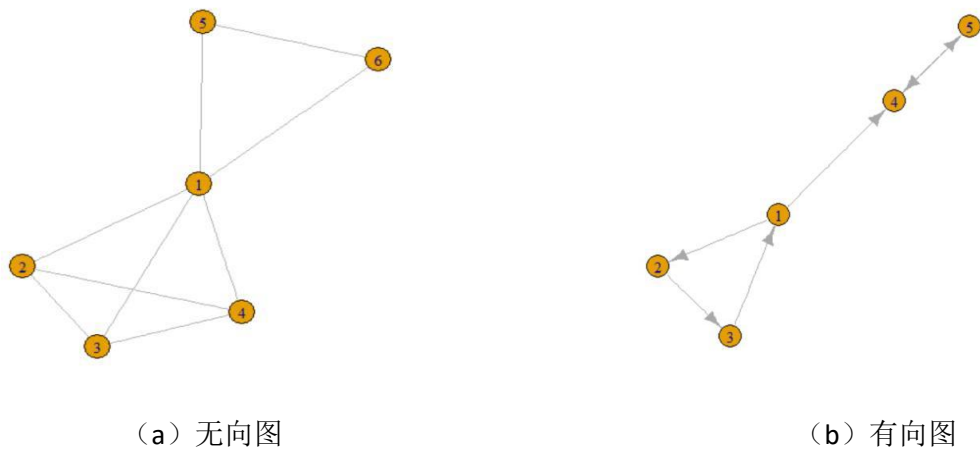


图 4.1 网络结构图

对于网络  $G(V, E)$ ，有顶点集  $V = \{v_i\}_{i=1}^n$  和边集  $E = \{(v_i, v_j) | v_i, v_j \in V\}$ ， $W_{ij}$  是网络的邻接矩阵，它代表数据之间的关系，且当  $(v_i, v_j) \in E$  时， $W_{ij} = 1$ ，否则  $W_{ij} = 0$ 。对于无向图而言，它是实对称矩阵，即  $W_{ij} = W_{ji}$ ；而对于有向图而言，其未必关于对角线对称。

#### 4.2.2 数据信息统计

由于原有附件中营业部内部通话记录数据庞大且重复性高，故将其信息统计整理。其统计信息如表 4.1 所示：

表 4.1 通讯记录统计表

名称	个数
主叫人	36 人
被叫人	36 人
有效通话线路	192 条
通话次数	10713 次
单人单向通话频次	见附录一

对个体绘制其网络结构图如图 4.2 所示：

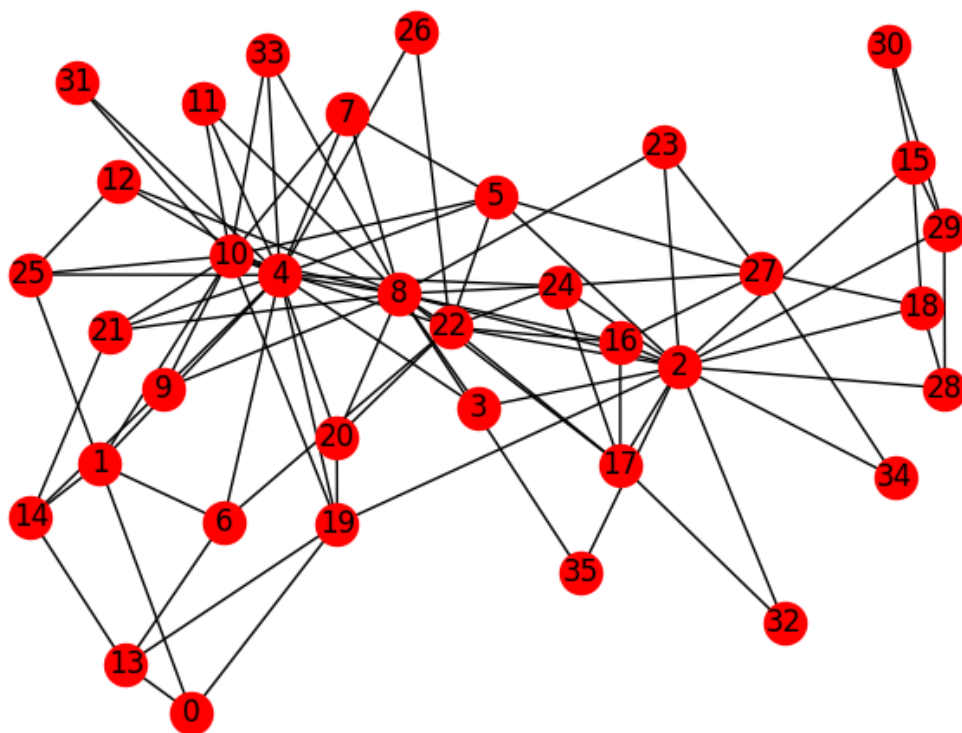


图 4.2 通讯网络结构图



图中共有 36 个节点，相互连接的边共 192 条，每条边上的权值可由单人单向通话频次拟定，符合表 4.1 中的统计信息。

### 4.3 问题一模型建立

在对网络结构的分类过程中，刻画任意两点之间的相似关系是影响分类精度的关键因素。

对于无向图和有向图而言，两点之间的相似关系可由两点之间是否连接、连通路程条数、每条路径所经过的边数确定。

对于具体的通讯网络而言，其数学本质就是一个有向图的复杂网络，我们可以类比有向图的方式来定义两点之间的相似关系，从而表达网络中两个通信节点间的相似关系。

这样我们就可以通过量化上述因素来刻画任意两节点之间的相似关系。从而通过具体的网络聚类算法实现网络结构的分类。

#### 4.3.1 无向图邻接矩阵

无向图的特性是其每条边代表两个节点互联，我们知道图中两点之间相似关系是由两点之间是否连接、连通路程条数、每条路径所经过的边数确定的，若要将这些因素量化，则需要引入邻接矩阵来反应任意两点之间的连通关系。邻接矩阵可以完全保留图中的信息，实现量化过程。 $WA_{ij} = 1$  代表  $V(i) \leftrightarrow V(j)$  相互连通，此时矩阵关于对角线对称。

问题（1）中无向图的邻接矩阵如下：

$$WA = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (4.1)$$

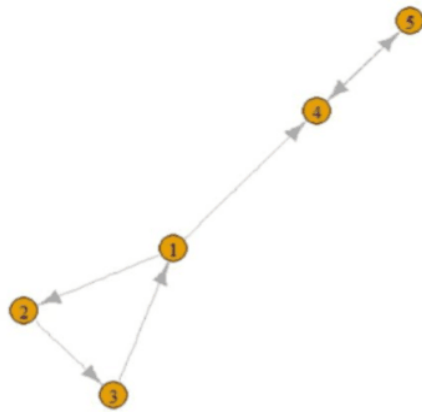
#### 4.3.2 有向图邻接矩阵

有向图的特性是其每条边具有方向，代表一个节点指向另一个节点。此时，两点之间的连接具有方向性，其邻接矩阵没有关于对角线对称的性质，同样可以用来反应有向图中任意两点之间的连通关系， $WB_{ij} = 1$  代表  $V(i) \rightarrow V(j)$  连通。

问题（2）中有向图的邻接矩阵如下：

$$WB = \begin{pmatrix} 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix} \quad (4.2)$$

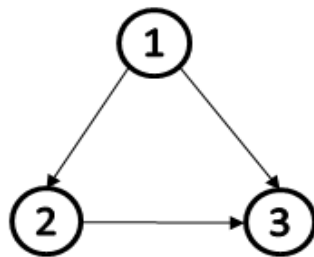
对于有向图任意两个节点之间的相似关系，使用广度优先搜索算法<sup>[1]</sup>，以下图为例：



假设每条边的连接强度均为 0.5，若以 1 开始搜索，目标为 5，则第一步可以到达 2、4，第二步可以到达 3、5，找到目标！

记第一步所到达的节点 2、4 与源节点 1 的相似强度为 0.5；第二步所到达的节点 3、5 与源节点 1 的相似强度为 0.25。

在优化算法中，对于下图：



若同时存在两条路径到达目的节点，比如从源节点 1 到目标节点 3：此时连接强度的计算公式为：

$$\tau = 1 - \left[ (1 - 0.5) \times (1 - 0.5 \times 0.5) \right] = 0.625 \quad (4.3)$$

由此规则我们可以求出有向网络的相似矩阵，矩阵中每一元素对应着出发节点和目的节点的相似度。

### 4.3.3 通讯网络的邻接矩阵

对于通讯网络而言，存在节点和连接节点的边（信道），而且通信过程具有方向性，所以通讯网络可以对应于有向图的建模方式： $WB_{ij} = 1$  代表  $V(i) \rightarrow V(j)$  连通。其邻接矩阵可表述为：

$$WC_{36 \times 36} = \begin{pmatrix} 0 & 1 & \dots & 0 & 0 \\ 1 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (4.4)$$

此时通讯网络为有向图，邻接矩阵不一定关于对角线对称。

将矩阵可视化， $X, Y$  坐标均为通讯网络中的个体，一个是出发节点，另一个是到达节点，矩阵元素为 1 的部分呈现黑色，为 0 的部分呈现为白色。

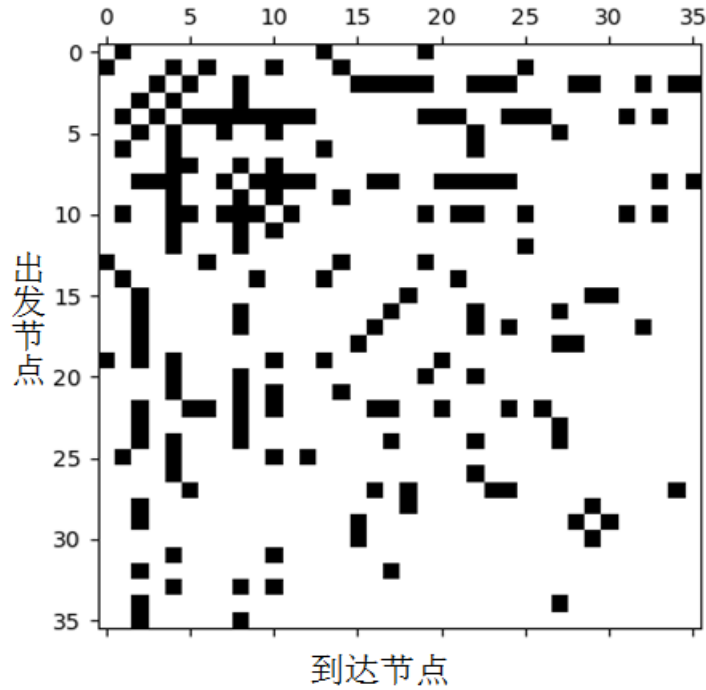


图 4.3 邻接矩阵可视化

### 4.3.4 基于 $FN$ 算法的网络聚类模型

基于优化  $Q$  函数的快速复杂网络聚类（ $FN$ ）算法从局部观点出发，使网络中每个节点独立计算并优化自身的局部函数，通过所有节点协同作用来实现对  $Q$

函数的优化，具有快速、并行的特点。

我们首先对  $Q$  函数来分析，将表征网络结构图的邻接矩阵作为输入传入  $FN$  算法，首先对于无权网络  $G(V, E)$ ，假设点集  $V$  被聚类为若干类簇，网络中任一节点  $i$  的标签为  $V(i)$ ，其所属的类簇为  $C_{V(i)}$ ，定义  $Q$  函数为“网络簇内实际的边数”与“完全随机的连接情况下簇内期望的边数”之差，数学表示为：

$$Q = \frac{1}{2m} \sum_{ij} \left( \left( W_{ij} - \frac{k_i k_j}{2m} \right) \times \delta(V(i), V(j)) \right) \quad (4.5)$$

其中  $W = (W_{ij})_{n \times n}$  表示网络结构图的邻接矩阵，对于函数  $\delta(u, v)$  有

$$\delta(u, v) = \begin{cases} 1, & u = v \\ 0, & otherwise \end{cases} \quad (4.6)$$

$k_i$  表示节点  $i$  的度：

$$k_i = \sum_j W_{ij} \quad (4.7)$$

$m$  表示网络中总的边数：

$$m = \frac{1}{2} \sum_{ij} W_{ij} \quad (4.8)$$

将 (4.4) 拆解得到：

$$Q = \frac{1}{2m} \left\{ \sum_{ij} (W_{ij} \times \delta(V(i), V(j))) - \sum_{ij} \left( \frac{k_i k_j}{2m} \times \delta(V(i), V(j)) \right) \right\} \quad (4.9)$$

可以看出，式 (4.8) 中减号左边表示网络簇内实际的连接数目，减号右边表示完全随机的连接情况下簇内期望的连接数目。显然他们的差值（ $Q$  函数）就可以度量一个网络聚类结果所对应簇结构的优劣。 $FN$  算法的目标就是对网络聚类，使得聚类结果所对应的  $Q$  函数值最大。

确定了聚类簇结构的评价标准  $Q$  函数之后， $FN$  算法可具体描述为：

#### 基于优化 $Q$ 函数的快速网络聚类算法

**输入** 表征网络结构图各节点相似度的邻接矩阵  $W = (W_{ij})_{n \times n}$

**输出** 最优网络簇个数 & 各簇中节点编号

**Step1.** 将每个节点都作为一个单独的网络簇

**Step2.** 对网络中剩余可以连接的边，各自计算其连接后：

$C \leftarrow$  网络簇内实际的连接数目

$X \leftarrow$  随机的连接情况下簇内期望的连接数目

**Step3.** 寻找使  $Q$  函数  $Q = C - X$  增长最快的一条边，将此边连接，该边连接的原有簇将构成新簇；

**Step4.** 对剩余的可连接边继续执行 **Step 2**，若无论连接剩余的哪条边， $Q$  函数均会下降，则网络分类完毕，即找到最优网络簇分类。

### 4.4 问题一模型求解

建立基于  $FN$  算法的网络聚类模型后，采用 Python 语言对  $FN$  算法进行编程求解，传入刻画相关度的邻接矩阵  $WA$ 、 $WB$ 、 $WC$ ，得到输出为分得的簇类数及每簇的节点编号。

#### 4.4.1 无向图模型求解

对 4.3.1 的无向图邻接矩阵，共有 6 个节点，将其编号为 0-5。进行聚类共将其分为两类如表 4.2 所示：

表 4.2 无向图网络聚类结果

簇类别	节点个数	节点编号
1	3	{1, 2, 3}
2	3	{0, 4, 5}

根据表 4.2，做出网络结构图如图 4.4

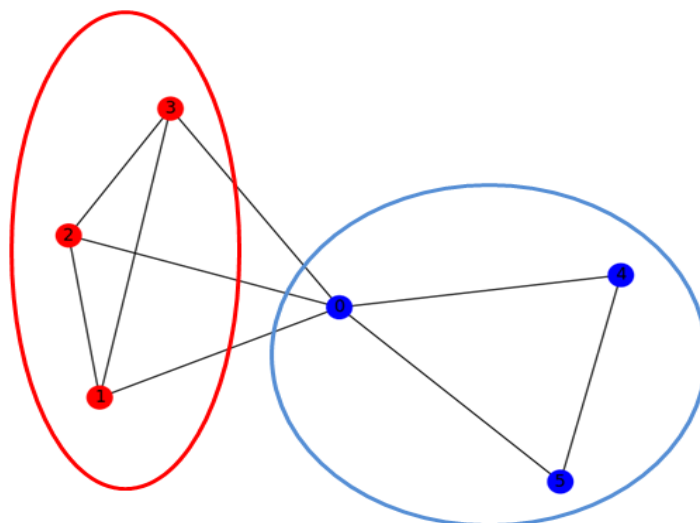


图 4.4 无向图网络聚类结果

#### 4.4.2 有向图模型求解

对 4.3.2 中的有向图邻接矩阵，共有 5 个节点，将其编号为 0-4。进行聚类共将其分为两类如表 4.3 所示：

表 4.3 有向图网络聚类结果

簇类别	节点个数	节点编号
1	3	{0, 1, 2}
2	3	{3, 4}

根据表 4.3，做出网络结构图如图 4.5

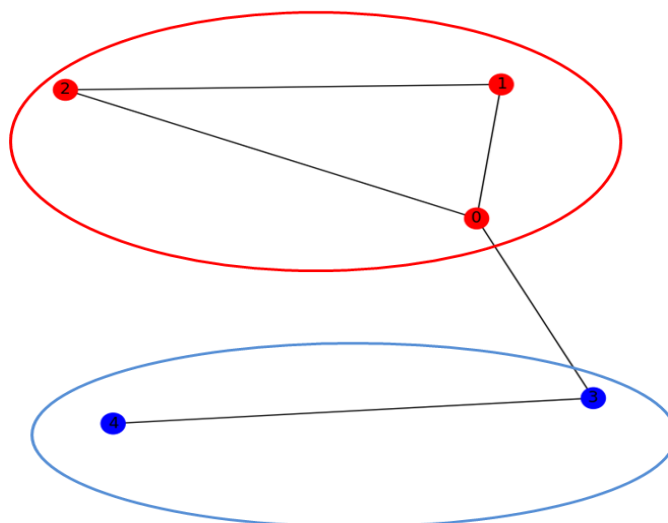


图 4.5 有向图网络聚类结果

按照广度优先算法及其相应计算规则可得到该有向网络的相似矩阵:

$$W_{\otimes} = \begin{pmatrix} 1 & 0.5 & 0.25 & 0.5 & 0.25 \\ 0.25 & 1 & 0.5 & 0.125 & 0.0625 \\ 0.5 & 0.25 & 1 & 0.25 & 0.125 \\ 0 & 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 & 1 \end{pmatrix} \quad (4.10)$$

其中  $W_{\otimes}$  为相似矩阵, 列为出发对象, 行为到达对象。

#### 4.4.3 面向通讯记录的综合模型求解

对 4.3.3 中的通讯网络邻接矩阵, 共有 36 个节点, 将其编号为 0-35。进行聚类共将其分为四类如表 4.4 所示:

表 4.4 通讯网络聚类结果

簇类别	节点个数	节点编号
1	14	{32, 2, 34, 35, 15, 16, 17, 18, 23, 24, 27, 28, 29, 30}
2	16	{1, 33, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 21, 25, 31}
3	3	{0, 19, 13}
4	3	{26, 20, 22}

根据表 4.4, 做出网络结构图如图 4.6

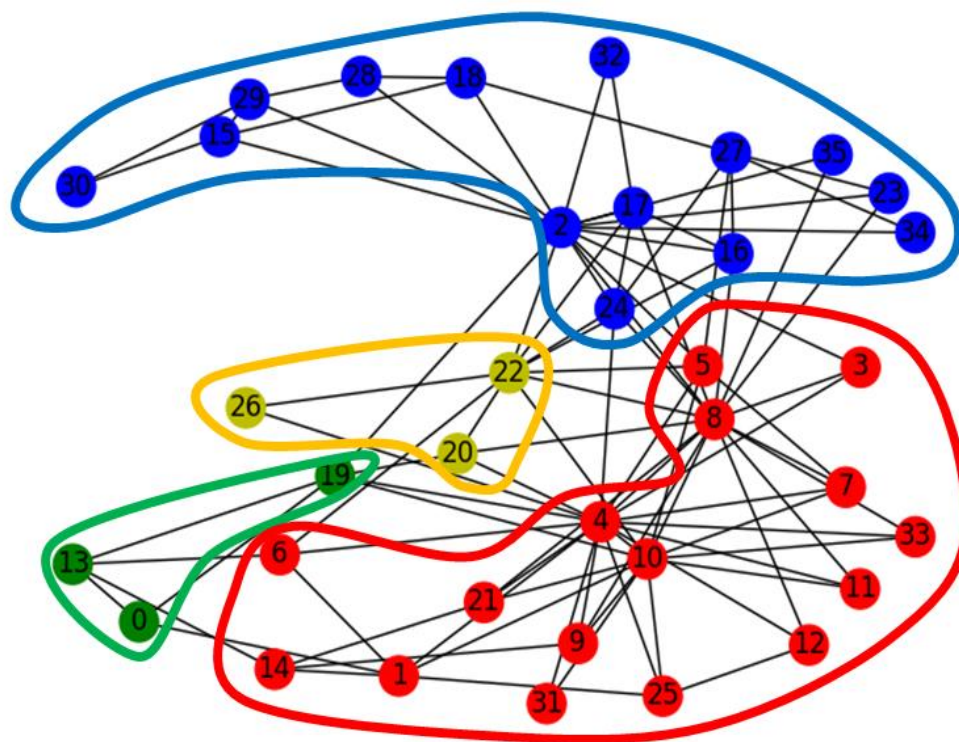


图 4.6 通讯网络聚类结果

(注: 此图为关系图, 图中节点间距离不代表相似度大小)

基于无重叠的意见领袖模型<sup>[4]</sup>，考虑各节点的度（与其相连接的边的个数），绘制出表如表 4.5 所示：

表 4.5 各节点度表

节点	度	节点	度	节点	度	节点	度
0	3	9	4	18	4	27	6
1	6	10	13	19	6	28	3
2	16	11	3	20	4	29	4
3	3	12	3	21	4	30	2
4	18	13	4	22	10	31	2
5	6	14	4	23	3	32	2
6	4	15	4	24	6	33	3
7	4	16	5	25	4	34	2
8	17	17	6	26	2	35	2

根据表 4.4、表 4.5 和图 4.6，可得如下结论：

- （1）该营业部内部人员之间联系紧密，可能分为四个项目组。其中两个项目组人员较多，另外两个较之偏少。
- （2）项目组成员之间联系频繁，组间成员之间也存在相应联系。
- （3）根据意见领袖模型，我们可以得到：

组 1 的意见领袖为 2 号毕婕靖，她是组 1 的组长，通过观察我们也发现她几乎与组 1 所有人都有联系；

组 2 的意见领袖为 4 号张培芸，通过观察可以发现其度很大，和组内成员及组间成员交流密切。

组 3 各节点度的差异很小，我们认为在其中不存在明确的意见领袖，他们的工作关系应该相对平等；

组 4 只有三个节点，22 号曾帅为意见领袖，通过观察我们也发现其与其他各组尤其是组 1、2 的组长有直接联系。

#### 4.5 $Q$ 函数对分类簇数的灵敏性

将  $Q$  函数随分类簇数的变化示意图绘制出来如图 4.7 所示：



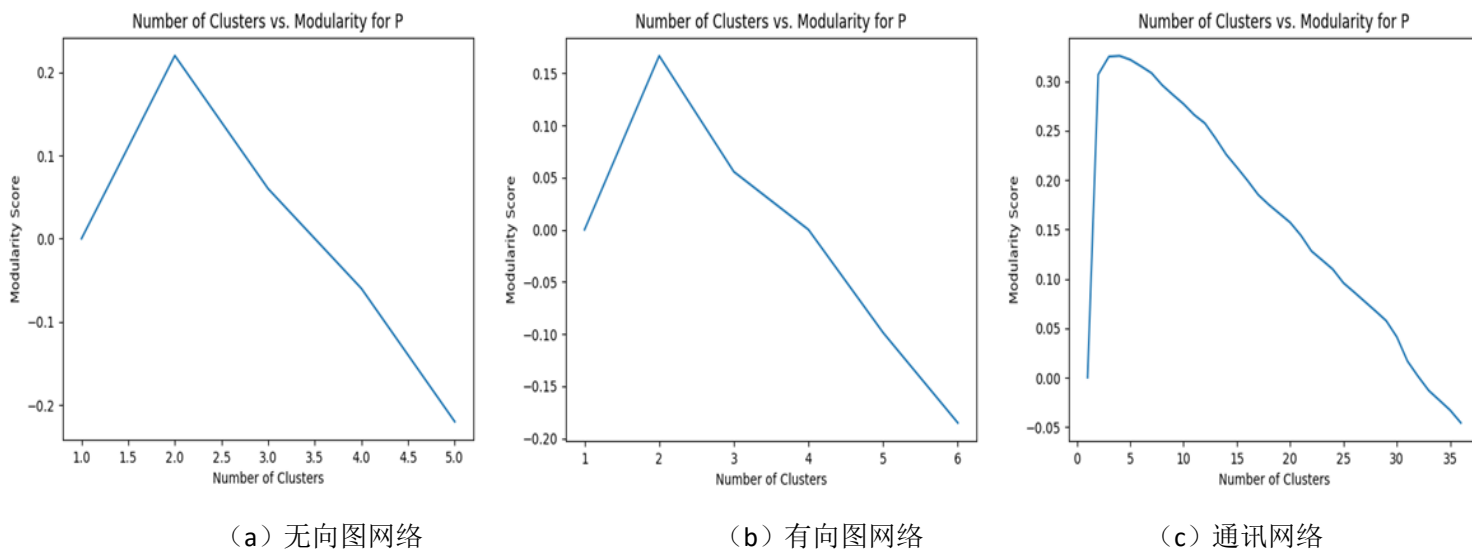


图 4.7  $Q$  函数大小随分类簇数变化图

由图易知， $Q$  函数随着分类簇数的增加先上升后下降。且随分类簇数的改变波动较大，对有向图网络，分类簇数增 1 或减 1， $Q$  函数都会随之减小或增加 0.15，对另外两个网络亦然，故  $Q$  函数对分类簇数比较敏感！

对于 4.3.1 中无向图网络，存在最优分类簇数为 2 类；

对于 4.3.2 中有向图网络，存在最优分类簇数为 2 类；

对于 4.3.3 中的通讯网络，存在最优分类簇数为 4 类。

将其与算法自动划分的簇数相比较，均与之匹配，模型求解正确！

## 5、问题二的建模与求解

### 5.1 问题二的分析

问题二要求对该营业部群体网络在综合考虑成本、终止传播概率情况下确定最少在哪些节点（个体）投放信息可达到最大的信息传播覆盖率。

### 5.2 问题二模型建立

对该影响力最大化传播问题，首先需要确立影响力评价函数来确定各类别中影响力最大的节点，通过改进的爬山贪婪算法<sup>[5]</sup>确定影响力最大节点进行信息传播。

为确立节点之间的信息传播机制，将信息传播过程分解为传播和湮没两步看待：传播过程构建阈值判断下一点是否激活，即成功接收到消息；湮没过程则利用终止传播概率，判断已激活节点是否继续传播，从而建立线性阈值模型<sup>[6]</sup>，进行仿真模拟以获取信息传播覆盖率，并记录此时的节点数目和节点编号。

问题二的建模流程图如图 5.1 所示：

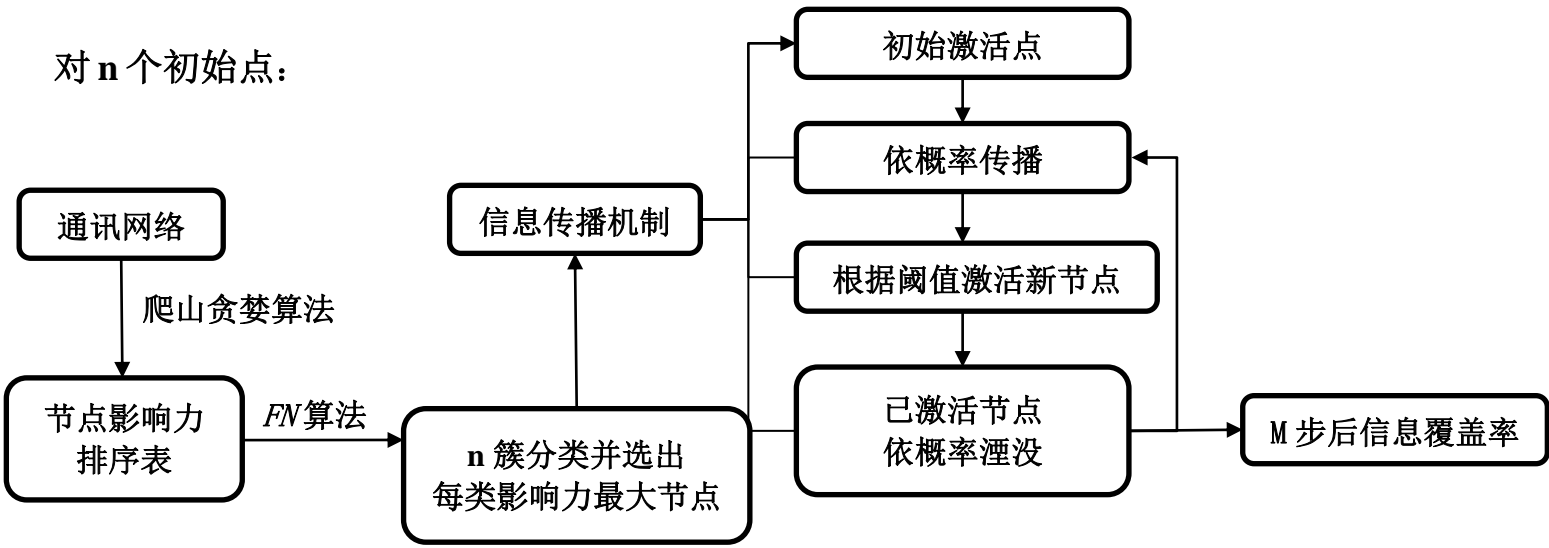


图 5.1 问题二建模流程图

### 5.2.1 确定信息传播机制

线性阈值模型（ $LT$ ）作为一种重要影响力传播模型，它的传播机制描述如下<sup>[8]</sup>：

- 对于网络图  $G(V, E)$ ，每个节点只能处于激活状态或未激活状态，只有自发的信息湮没过程才能使已激活节点变为未激活状态，其余情况激活的节点不能从激活状态变为未激活状态。
- 对于网络中每一个节点  $v$ ，用一个阈值  $\theta_v$ （threshold）表示节点  $v$  被激活的难易程度， $\theta_v$  越大表示节点  $v$  越难被激活，反之节点  $v$  越容易被激活。

根据实验模拟，阈值设置为  $\theta_v = 0.5$  时对该通讯网络有最佳效果。对节

点  $v$  有影响力的节点用集合  $AN_v$  表示，对于该集合中任意节点  $\omega$ ，用符号  $b_{v\omega}$  来表示节点  $v$  的激活概率，即为节点之间相互影响力，对于问题二中的有向图而言， $b_{v\omega}$  就是边的权值。

- 如果  $AN_v$  中所有节点对节点  $v$  的影响力之和大于等于节点  $v$  的阈值  $\theta_v$ ，则该节点被激活，否则节点不被激活。
- 根据问题要求，信息传播过程中每经过一个节点，都有 10% ~ 50% 的终止传播概率，即已激活节点依概率 10% ~ 50% 湮没。
- 传播不断进行，直至图中已无可以被激活的节点为止。

该  $LT$  模型中，共有两个参数且已被确定，即每一个节点的激活阈值  $\theta_v = 0.5$  和节点之间的相互影响力（传播成功概率） $b_{v\omega}$ ，即为通讯有向网络中边的权值。

### 5.2.2 计算节点影响力

给定社会网络  $G(V, E)$ ，对其采用爬山贪婪算法计算其影响力，算法先设置初始种子节点  $v$ ，根据相应传播模型的传播机制找到被影响成功的所有节点集  $AN_v$ 。

实验证明，运用爬山贪婪算法至少可以达到 63% 的近似最优解<sup>[9]</sup>。

计算节点的最大影响力问题可以用数学表达式描述为：

$$\max \left\{ \delta(AN_v) \mid AN_v \subseteq V \right\} \quad (5.1)$$

其中  $\delta(AN_v)$  表示节点能够影响成功的节点数目

在实际算法实现中，采用随机搜索（**Randomization in Search**）<sup>[10]</sup> 的策略，由于该通讯网络规模较小，可以保证得出的解在全局最优解附近  $\pm 5\%$ 。具体算法描述如下：

### 采用 Random Search 思想的改进爬山贪婪算法

输入 影响力传播模型、网络  $G(V, E)$

输出 节点影响力排序表

#### 步骤

Step1. 读取图数据，设定网络的结构和参数（激活阈值  $\theta_v = 0.25$  和传播成功概率  $b_{vw}$ ）

Step2. 依据给定的影响传播模型的传播机制，随机设定初始点，定义其为最大影响力节点

Step3. 对最大影响力节点执行信息传播的模拟，得到被成功传播的下一代节点个数及节点号，进行下次传播，若有子代节点影响力大于初始节点，则最大影响力节点变为影响力最大的子代节点

Step4. 不断执行 step3，直至没有子代节点影响力大于父节点，该节点即为最大影响力节点

Step5. 重复 step2-4 找到第二大影响力节点，第三大影响力节点……

Step6. 重复 step2-5，重新随机生成初始点，获得近似全局最优解的解，得到节点影响力排序表

### 5.2.3 模拟实验得出信息覆盖率

采用广度优先搜索的思路按照信息传播机制对给定节点个数  $n(1 \leq n \leq 36)$  进行仿真，并记录每一步传播过程所对应的信息覆盖率，信息覆盖率的公式如下：

$$\xi = \frac{\delta(AN_v)}{P} \times 100\% \quad (5.2)$$

其中  $\xi$  表示信息覆盖率大小， $\delta(AN_v)$  表示节点能够影响成功的节点数目， $P$  表示总节点个数。

### 5.2.3 综合模型

对于该信息传播问题，我们的目标是：

$$\begin{aligned} Min : m \times n \\ Max : \xi = \frac{\delta(AN_v)}{P} \end{aligned} \quad (5.3)$$

其中  $m$  是投放一次信息的成本， $n$  是信息投放点数， $\xi$  为信息传播覆盖率，由此可以将模型简要描述为以下几个步骤：

- 1) 首先设置投放信息节点个数  $n(1 \leq n \leq 36)$ ，通过采用 Random Search 思想改进的爬山贪婪算法计算得出节点影响力排序表。
- 2) 然后使用基于优化  $Q$  函数的快速复杂网络聚类 (FN) 算法将该通讯网络分为  $n$  簇，根据节点影响力排序表在每簇中找到影响力最大的节点。
- 3) 根据确定的信息传播机制进行模拟实验，得出对给定节点个数  $n(1 \leq n \leq 36)$  每一步基于广度优先思想的传播过程所对应的的信息覆

$$\text{盖率: } \xi = \frac{\delta(AN_v)}{P} \times 100\%$$

- 4) 通过对不同投放节点个数  $n(1 \leq n \leq 36)$  的比较找出可以使信息覆盖率达到最大值的最小投放节点个数，即为问题的解。

### 5.3 问题二的模型求解

#### 5.3.1 节点影响力排序表

通过采用 Random Search 思想改进的爬山贪婪算法计算得出节点 0 ~ 35 所对应的按其影响力大小排序的表，如表 5.1 所示：

表 5.1 节点影响力排序表

名次	节点编号	名次	节点编号	名次	节点编号
1	8	13	21	25	31
2	19	14	9	26	18
3	22	15	7	27	15
4	24	16	23	28	34
5	5	17	33	29	28
6	2	18	11	30	29
7	3	19	6	31	32
8	4	20	12	32	27
9	10	21	35	33	13
10	17	22	1	34	14
11	16	23	25	35	0
12	20	24	26	36	30

### 5.3.2 选出 $n$ 簇中影响力最大的节点

使用基于优化  $Q$  函数的快速复杂网络聚类 (  $FN$  ) 算法将该通讯网络分为  $n$  簇, 根据节点影响力排序表在每簇中找到影响力最大的节点。考虑成本原因, 综合投放信息节点数与传播覆盖率可知, 若使投放成本尽可能小, 信息覆盖率尽可能大, 则  $n$  的范围可粗略划定为  $1 \leq n \leq 6$ 。

对处于该范围的  $n$  执行  $FN$  算法, 找出每簇中影响力最大的点构成点集如表 5.2 所示:

表 5.2  $n$  簇中影响力最大的点集

$n$	影响力最大点集
1	{8}
2	{8,22}
3	{8,19,22}
4	{4,8,19,22}
5	{4,8,19,22,34}
6	{4,8,19,21,22,34}

### 5.3.3 计算信息覆盖率

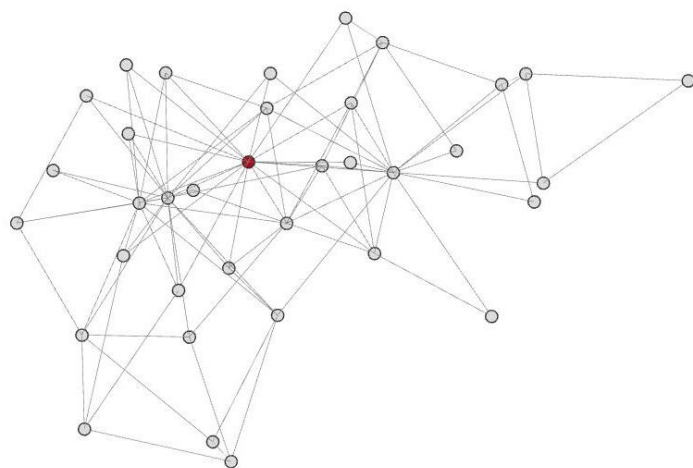
根据确定的信息传播机制使用 MATLAB 软件进行模拟实验, 为了模拟传播过程中信息湮没概率的随机性, 每次传播过程中湮没向量的产生是根据节点发生信息湮没的概率随机生成的。为了消除这种随机性带来的偶然偏差, 我们每次取 1000 次循环取平均值以获得覆盖率的实际期望。最后得出对应确定的  $n$  ( $1 \leq n \leq 6$ ) 个信息投放点的每一步信息覆盖率如表 5.3 所示:

表 5.3 信息传播过程中的信息覆盖率变化表

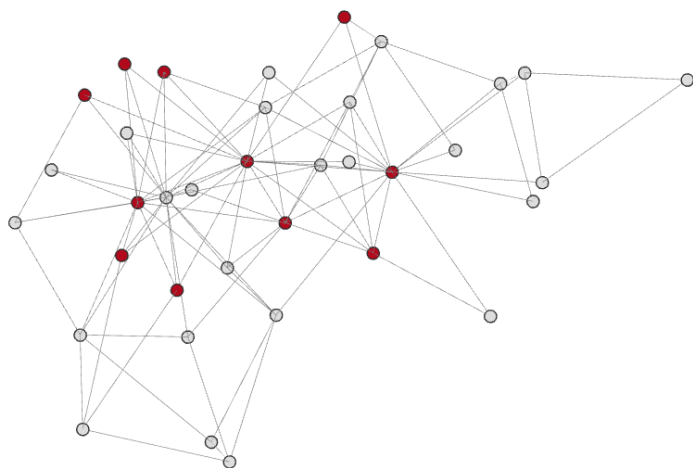
$n$	第一步	第二步	第三步
1	31.62%	72.93%	92.16%
2	36.88%	74.92%	92.19%
3	41.78%	78.51%	92.36%
4	46.64%	78.53%	92.42%
5	49.97%	78.94%	92.50%
6	51.72%	80.06%	92.48%

(注：实验结果为循环 1000 次取平均)

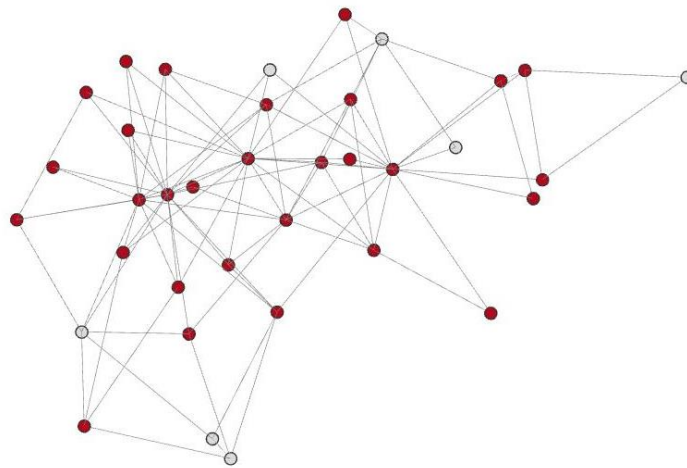
单独投放信息给点 8 号，其对应的节点信息扩散图可绘制如图 5.2:



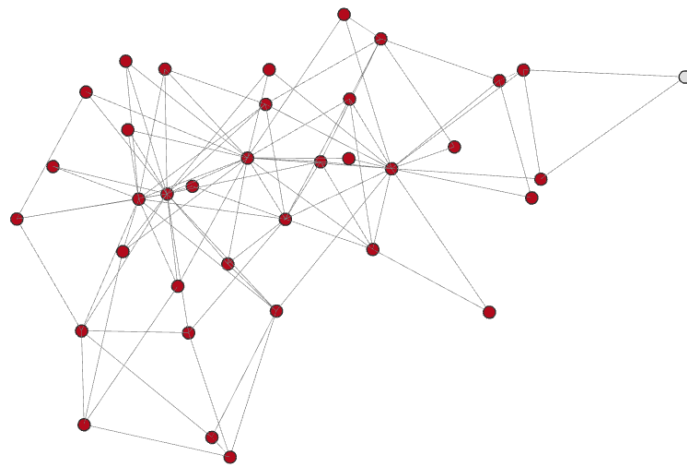
(1) 向节点 8 投放信息



(2) 第一步传播



(3) 第二步传播



(4) 第三步传播

图 5.2 信息扩散示意图

由表 5.1、表 5.2、表 5.3、图 5.2 可以得到如下结论：

- (1) 8 号节点蔡月是影响力最大的节点
- (2) 信息覆盖率随投放节点数目的增多而增大
- (3) 信息覆盖率随传播步数的增加而增大，并趋于饱和
- (4) 节点数目增加所带来的信息覆盖率增量逐步减小
- (5) 当初始投放节点范围是 1 – 6 时，第一步信息覆盖率可达到 30%—50%；第二步信息覆盖率可达到 70%—80%，第三步信息覆盖率可达到 92%—92.5%



## 6、问题三的建模与求解

### 6.1 问题三的分析

问题三要求考虑位置、时长、通话频率等多种因素修正综合的数学模型，挖掘出更多的信息，得出相关结论。并且可以自行补充数据，验证自己的模型，对比得出相关结论。

### 6.2 问题三模型建立

因为问题三要求考虑多种因素的共同作用，故构建基于主成分分析的评价模型，基于此评价模型对通讯网络构建评价体系，得到更加符合实际情况的邻接矩阵，然后采用  $FN$  算法对该通讯网络进行分类，挖掘出更多的信息。

问题三建模流程图如图 6.1 所示：

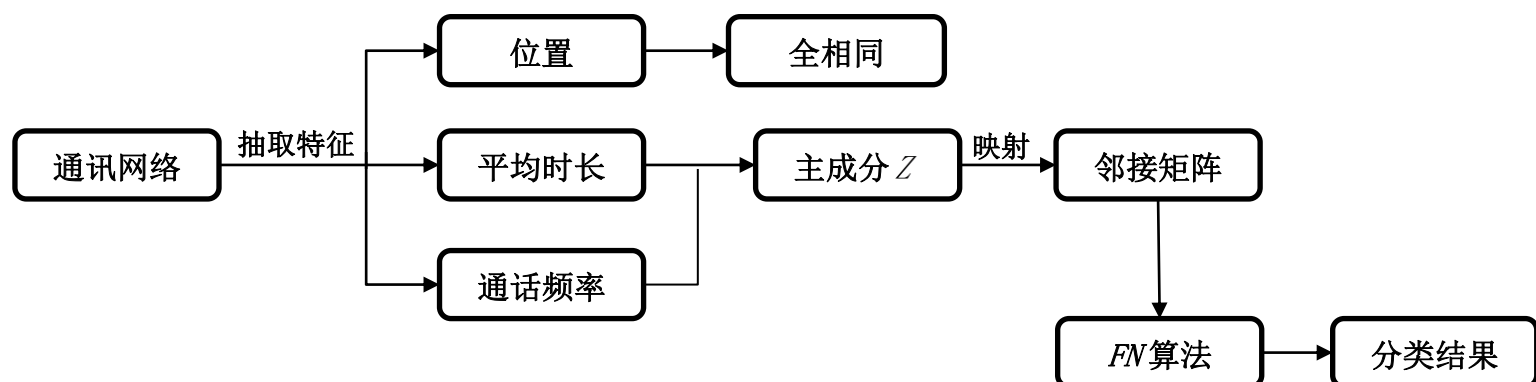


图 6.1 问题三建模流程图

#### 6.2.1 主成分选取

主成分分析（Principal component analysis）<sup>[11]</sup>数学模型是一种正交变换，在几何上就是做一个坐标旋转，因此，主成分分析可以利用其在二维空间中明显的几何意义逐步向  $n$  维空间推广，解决实际问题。在本问题中，该方法对时长和通话频次两个指标映射到  $n$  维空间，进而得到指标构成邻接矩阵，采用  $FN$  算法对反应通讯网络信息的邻接矩阵进行聚类，得到更符合实际的分类结果，挖掘出更深入的信息。

在二维空间中有时间和通话频率两个指标  $(X_1, X_2)$ ，它们大致分布在一个椭圆

内，旋转坐标后，以主成分几乎覆盖原指标信息为优，坐标图如图所示：

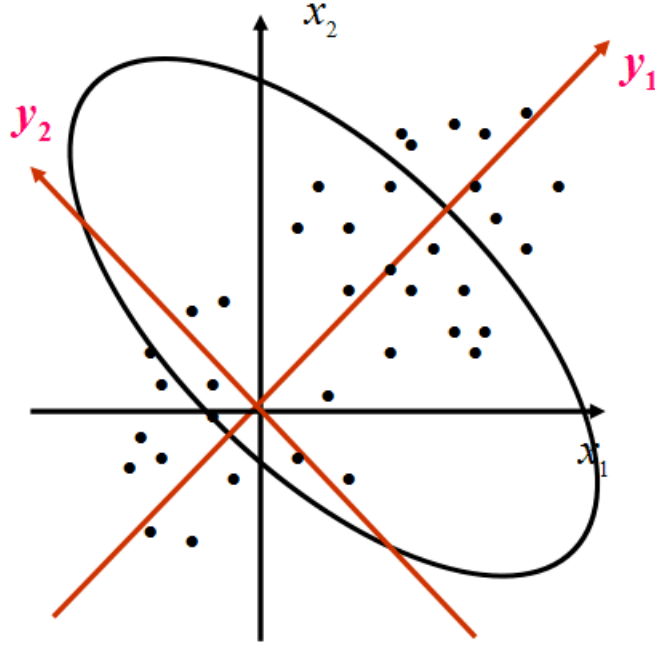


图 6.2 坐标图

旋转公式为：

$$\begin{cases} Y_1 = X_1 \cdot \cos \theta + X_2 \cdot \sin \theta \\ Y_2 = -X_1 \cdot \sin \theta + X_2 \cdot \cos \theta \end{cases} \quad (6.1)$$

新变量  $Y_1$  和  $Y_2$  是原变量  $X_1$  和  $X_2$  的线性组合，它的矩阵表示为：

$$\begin{bmatrix} Y_1 \\ Y_2 \end{bmatrix} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix} \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} = TX \quad (6.2)$$

其中， $T$  为旋转变换矩阵，它是正交矩阵，即  $T'T = I$

推广到  $p$  维随机向量，设  $(X_1, X_2, \dots, X_p)$  是  $p$  维随机向量，考虑线性变换：

$$\begin{cases} Z_1 = a'_1 X = a_{11}X_1 + a_{21}X_2 + \dots + a_{p1}X_p \\ Z_2 = a'_2 X = a_{12}X_1 + a_{22}X_2 + \dots + a_{p2}X_p \\ \dots \dots \dots \dots \dots \dots \\ Z_p = a'_p X = a_{1p}X_1 + a_{2p}X_2 + \dots + a_{pp}X_p \end{cases} \quad (6.3)$$

则有，

$$Var(Z_i) = a'_i \sum a_i (i = 1, \dots, p) \quad (6.4)$$

$$Cov(Z_i, Z_j) = a'_i \sum a_j (i, j = 1, \dots, p)$$

用  $Z_k = a'_k X$  来代替原来  $p$  个向量  $(X_1, X_2, \dots, X_p)$ ，显然， $Var(Z_k)$  越大，表示第

$k$  个主成分  $Z_k$  包含的信息越多,对于本题,只需使方差累计率超过 90%以上即可。

获取主成分  $Z_k$  以后,分别以  $k$  个主成分的贡献率为权重,构建主成分综合评价模型:

$$\Delta = O_1 Z_1 + O_2 Z_2 + \dots + O_k Z_k \quad (6.5)$$

其中  $O_k$  代表主成分  $Z_k$  的贡献率。

### 6.2.2 邻接矩阵构建

由于节点之间位置没有差异,故将其通话时长作为  $X_1$ , 通话频率作为  $X_2$  构成  $(X_1, X_2)$  两个指标。对于通讯网络中的边集  $E_{i \rightarrow j} (i, j = 0, 1, \dots, 35)$  中权值不为 0 的每条边,将边上的权值作为目标  $\Delta$ 。

对两个指标  $(X_1, X_2)$  进行主成分选取,得到主成分  $Z$  及其贡献率,根据公式 (6.5) 得到目标值  $\Delta$ ,即将两个指标映射到边集  $E_{i \rightarrow j} (i, j = 0, 1, \dots, 35)$  构成权重。从而形成新的带权邻接矩阵  $WC_{36 \times 36}$ 。

对该邻接矩阵进行指数变换,将矩阵中为负的元素变为正。即将  $WC_{36 \times 36}$  中的元素  $a_{ij} (i, j = 0, 1, \dots, 35)$  变为:

$$a'_{ij} = \begin{cases} e^{a_{ij}} & \text{if } a_{ij} \neq 0 \\ 0 & \text{if } a_{ij} = 0 \end{cases} \quad (6.6)$$

### 6.2.3 对新邻接矩阵进行聚类分析

此过程与 4.3.4 中基于优化  $Q$  函数的快速复杂网络聚类 (FN) 算法一致。

### 6.2.4 综合模型

针对问题三构建主成分分析模型,将通话时长和通话频率作为两个指标求其

主成分  $Z_k$ ，获取主成分  $Z_k$  以后，分别以  $k$  个主成分的贡献率为权重，构建主成分综合评价模型：

$$\Delta = O_1 Z_1 + O_2 Z_2 + \dots + O_k Z_k$$

将  $\Delta$  映射为邻接矩阵中的权值，对新的邻接矩阵通过  $FN$  算法对其进行聚类，结合通讯网络得出相应结论。

### 6.3 问题三模型求解

#### 6.3.1 主成分分析模型求解

**Step 1:** 先将原始数据平均通话时长和通话频率  $A = a_{ij} = (X_1, X_2)$  进行标准化处理，

将各指标值  $a_{ij}$  转换为标准化指标  $\tilde{a}_{ij}$ ，有：

$$\tilde{a}_{ij} = \frac{a_{ij} - \mu_j}{s_j} \quad (i = 0, 1, \dots, 191; j = 0, 1) \quad (6.7)$$

式中  $\mu_j = \frac{1}{192} \sum_{i=0}^{191} a_{ij}$ ； $s_j = \sqrt{\frac{1}{191} \sum_{i=0}^{191} (a_{ij} - \mu_j)^2}$ ，即  $\mu_j, s_j$  为第  $j$  个指标的样本均值和样本标准差。

**Step 2:** 计算相关系数矩阵  $R = (r_{ij})_{2 \times 2}$ ，有

$$r_{ij} = \frac{\sum_{k=0}^{191} \tilde{a}_{ki} \cdot \tilde{a}_{kj}}{191 - 1} \quad (i, j = 1, 2) \quad (6.8)$$

式中： $r_{ii} = 1$ ； $r_{ij} = r_{ji}$ ， $r_{ij}$  为第  $i$  个指标与第  $j$  个指标的相关系数。

**Step 3:** 利用 MATLAB 软件计算  $R$  的特征值  $\lambda_1, \lambda_2$  及对应的标准化特征向量  $u_1, u_2$ 。

**Step 4:** 计算主成分  $Z_k (k = 1, 2)$  及其信息贡献率，其贡献率可由公式 (6.9) 求得：

$$b_k = \frac{\lambda_k}{\sum_{i=1}^5 \lambda_i} (i = 1, 2) \quad (6.9)$$

**Step 5:** 得到综合评价模型求解结果:

$$\Delta = \sum_{k=1}^2 b_k \cdot Z_k \quad (6.10)$$

**Step 6:** 根据目标值 $\Delta$ 构建新的带权邻接矩阵 $WC_{36 \times 36}$

### 6.3.2 求解结果

(1) 主成分分析结果:

表 6.1 主成分分析结果

主成分	特征根	贡献率	累计贡献率
1	1.0191	50.9565	50.9565
2	0.9809	49.0435	100

(2) 标准化变量前主成分对应的特征向量:

表 6.2 标准化变量前主成分对应的特征向量

	1	2
第一特征向量	0.7071	-0.7071
第二特征向量	0.7071	0.7071

以主成分的贡献率为权重, 构建主成分综合评价模型

$$\Delta = 0.509565 \cdot Z_1$$

(3) 由综合评价值构建邻接矩阵:

$$WC_{36 \times 36} = e^{\begin{pmatrix} 0 & -0.1810 & \dots & 0 & 0 \\ -0.2770 & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix}} = \begin{pmatrix} 0 & e^{-0.1810} & \dots & 0 & 0 \\ e^{-0.2770} & 0 & \dots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 0 & 0 \\ 0 & 0 & \dots & 0 & 0 \end{pmatrix} \quad (6.11)$$

其中对该邻接矩阵求指数定义为对矩阵中的非零元取指数。

### 6.3.3 分类结果

对 6.3.2 中的通讯网络邻接矩阵, 共有 36 个节点, 将其编号为 0-35。进行聚类共将其分为四类如表 6.3 所示:

表 6.3 通讯网络聚类结果

簇类别	节点个数	节点编号
1	15	{1, 33, 4, 5, 6, 7, 9, 10, 11, 12, 14, 21, 25, 26, 31}
2	14	{32, 2, 34, 15, 16, 17, 18, 22, 23, 24, 27, 28, 29, 30}
3	4	{0, 19, 20, 13}
4	3	{8, 3, 35}

根据表 6.3 做出网络聚类图如图 6.3:

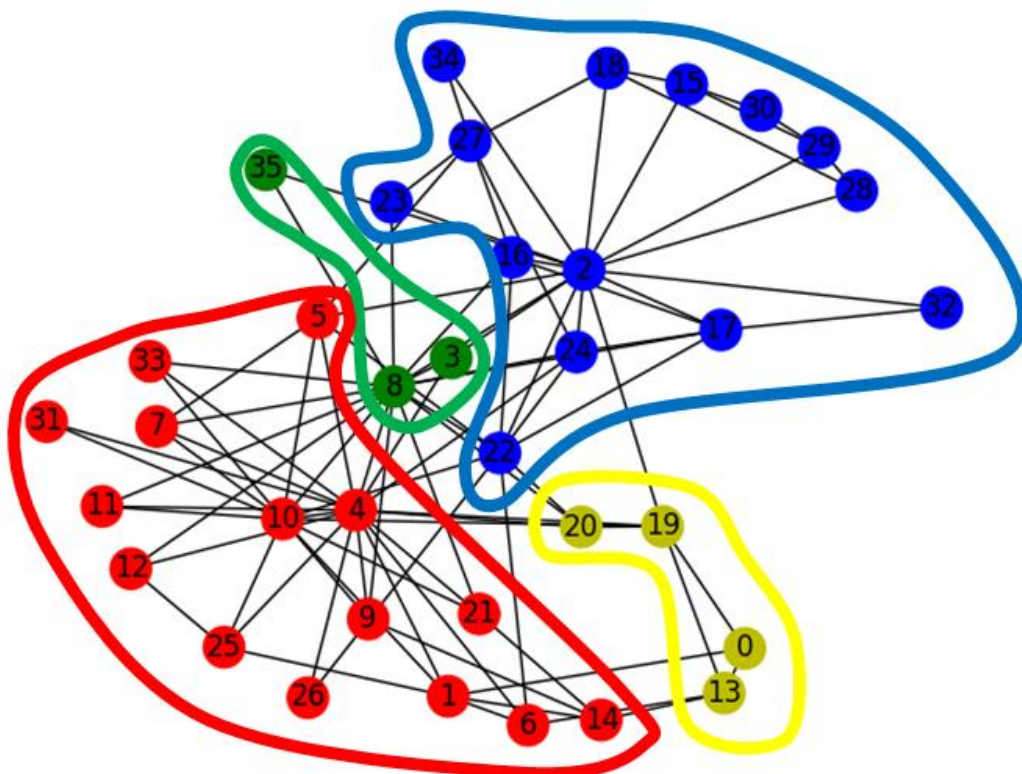


图 6.3 聚类结果图

根据表 4.5、表 6.2、表 6.3 和图 6.3，可得出结论：

- （1）该营业部内部人员之间联系紧密，可能分为四个项目组。其中两个项目组人员较多，另外两个较之偏少。
- （2）项目组内成员之间联系频繁，组间成员之间也存在相应联系。
- （3）根据意见领袖模型，我们可以得到：

组 1 的意见领袖为 4 号张培芸；组 2 的意见领袖为 2 号毕婕靖；

组 3 各节点度的差异很小，我们认为在其中不存在明确的意见领袖；

组 4 只有三个节点，8 号蔡月尚为意见领袖，通过观察我们也发现他与其他各组尤其是组 1、2 的组长有直接联系。又组 4 的 35 号孟芳只与 8 号蔡月及 3 号潘立有联系，故推测组 4 为该营业部的领导组，且 8 号蔡月可能

为营业部业务负责人，35号孟芳有可能为业务部总领导。

- (4) 将该结论与问题一中求解结果相对比，发现意见领袖由 2,4,22 变为了 2,4,8。从表 4.5 中可以看出，8 号的度远高于 22 号，更符合营业部管理层业务领导人的位置；此外，改进后 2 号和 4 号与小组内所有人都相连，更符合小组内直接领导的地位。

#### 6.4 $Q$ 函数对分类簇数的灵敏性

将  $Q$  函数随分类簇数的变化示意图绘制出来如图 6.7 所示：

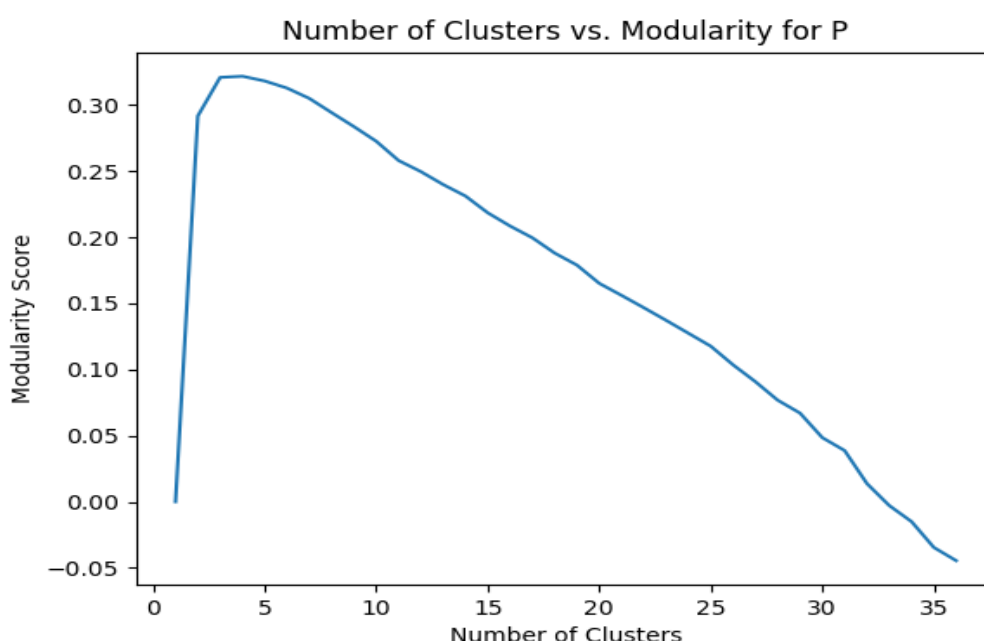


图 6.5  $Q$  函数大小随分类簇数变化图

由图易知， $Q$  函数随着分类簇数的增加先上升后下降。在分类簇数比较低时随分类簇数的改变波动剧大，当分类簇数超过 5，则  $Q$  函数随之线性减小，故  $Q$  函数对分类簇数比较敏感！

对于 6.3.3 中的通讯网络，存在最优分类簇数为 4 类。将其与算法自动划分的簇数相比较，与之匹配，模型求解正确！

## 7、模型的优缺点和推广

### 7.1 模型一的优缺点与推广

#### 模型一的优点：

- (1) 合理利用了网络的连接特性，提出了有效刻画节点间相似程度的方法。
- (2) 利用基于优化 $Q$ 函数的快速复杂网络聚类（ $FN$ ）算法，实现了无监督的高准确度社群聚类。
- (3) 计算速度快，模型通用性强，可以适用于各种有向图无向图网络，对于上千个节点的超大型网络也有很可靠的表现。

#### 模型一有待提高之处：

- (1) 该模型使用邻接矩阵，将无向图的边看为两条有向图相反方向的边，运算时会出现重复，可以进一步优化。

#### 模型一改进方向：

- (1) 对无向图运算时，一次直接连接双向的边，可进一步降低运算成本。

#### 模型一推广方向：

- (1) 可用于对连接密集、没有先验经验的复杂网络进行高效聚类。
- (2) 支持更多节点，超大型的复杂网络。

### 7.2 模型二的优缺点与推广

#### 模型二的优点：

- (1) 能够成功模拟信息传播的时间过程，并设置“信息湮没”过程模拟节点终止传播的行为。
- (2) 通过对于传播成功概率和激活阈值的设计，体现了不同个体在传播过程中的行为差异。



- (3) 不必完全枚举就可以得到近似的全局最优解，具有比较高的效率。

#### **模型二有待提高之处：**

- (1) 对于超大复杂网络，爬山贪婪算法计算成本比较高，而且还可能会困在局部最优解，这回直接对模型最终效果造成影响。

#### **模型二改进方向：**

- (1) 使用其他更先进更复杂的算法得到影响力最大节点，以求在超大网络中提高对全局最优解的逼近准确性。

#### **模型二推广方向：**

- (1) 可适用于连通性更强或连通性更弱的网络。
- (2) 对于相互不连接的网络簇，该模型也可以找到合适的最优投放点

### **7.3 模型三的优点与推广**

#### **模型三的优点：**

- (1) 评价模型基于数据而非基于经验，避免了主观设置参数的不确定性
- (2) 可以发掘因素间的内在联系，找出主要影响因素，可适用于多属性复杂网络。

#### **模型三有待提高之处：**

- (1) 题目数据因素比较少，不能最大程度体现主成份分析法的威力。

#### **模型三改进方向：**

- (1) 寻找数据因素更多的现实数据，用于体现主成份分析法找出主要影响因素的作用。

#### **模型三推广方向：**

- (1) 能够更好地适用于多种因素共同作用的复杂网络。

## 8、参考文献

- [1] 李明哲、金俊、石端英, 图论及其算法, 北京: 机械工业出版社, 2010. 10
- [2] Newman MEJ. Fast algorithm for detecting community structure in networks. *Physical Review (E)*, 2004, 69(6):066133
- [3] Newman MEJ. Girvan M. Finding and evaluating community structure in networks. *Physical Review E*, 2004, 69(2):026113
- [4] Katz E. The two-step flow of communication: An up-to-date report on an hypothesis. *Public Opinion Quarterly*, 1957
- [5] 通用局部搜索算法之爬山法, CSDN 博客, 2017. 4  
[http://blog.163.com/highway\\_xxxw/blog/static/26754625200972884316842/](http://blog.163.com/highway_xxxw/blog/static/26754625200972884316842/)
- [6] 周生福, 线性阈值模型下影响最大化算法研究, 云南大学, 硕士学位论文, 2014. 5
- [7] 兰如钦. 社会网络上的影响力最大化算法研究[D]. 解放军信息工程大学, 2013.
- [8] 冀进朝, 黄岚, 李红明, 李三义. 一种新的机遇社区结构的影响最大化方法[J]. *吉林大学学报 (理学版)*, 2011, 49 (1): 93-97
- [9] David Kempe. Jon Kleinberg. Eva Tardos. Maximizing the Spread of Influence through a Social Network. 2003, ACM 1-58113-737-0/03/0008  
<http://www.cs.cornell.edu/home/kleinber/kdd03-inf.pdf>
- [10] Probabilistic Graphical Models, Stanford University.  
<http://pgm.stanford.edu/Algs/page-1156.pdf>
- [11] 司守奎, 孙兆亮, 数学建模算法与应用 (第二版), 北京: 国防工业出版社, 2016. 1

## 9、附录

### 9.1 附录一：单人单向通话频次统计表

	涂蕴知	孙翼茜	毕婕靖	潘立	张培芸	梁茵	文芝	谢斑尚	蔡月
涂蕴知	0	43	0	0	0	0	0	0	0
孙翼茜	48	0	0	0	39	0	46	0	0
毕婕靖	0	0	0	234	0	48	0	0	159
潘立	0	0	315	0	300	0	0	0	568
张培芸	0	48	0	232	0	45	37	38	183
梁茵	0	0	48	0	35	0	0	40	0
文芝	0	38	0	0	41	0	0	0	0
谢斑尚	0	0	0	0	48	37	0	0	40
蔡月	0	0	186	590	185	0	0	45	0
彭荃	0	0	0	0	37	0	0	0	45
张荆	0	49	0	0	44	48	0	43	43
陆盈	0	0	0	0	50	0	0	0	41
王蕴姣	0	0	0	0	45	0	0	0	45
童豫	38	0	0	0	0	0	39	0	0
张庭琪	0	42	0	0	0	0	0	0	0
柯雅芸	0	0	45	0	0	0	0	0	0
李熹俊	0	0	50	0	0	0	0	0	47
陈澜	0	0	50	0	0	0	0	0	49
高淼	0	0	38	0	0	0	0	0	0
易贞	39	0	40	0	35	0	0	0	0
吴霄	0	0	0	0	46	0	0	0	48
潘澜巧	0	0	0	0	40	0	0	0	46
曾帅	0	0	35	0	0	44	46	0	42
孙锦	0	0	35	0	0	0	0	0	49
刘松荷	0	0	45	0	48	0	0	0	37
吴宇晓	0	48	0	0	46	0	0	0	0
廖颜翠	0	0	0	0	42	0	0	0	0
钟倩	0	0	0	0	0	38	0	0	0
林礼琴	0	0	49	0	0	0	0	0	0
康芸晴	0	0	40	0	0	0	0	0	0
骆娴	0	0	0	0	0	0	0	0	0
柳谓	0	0	0	0	43	0	0	0	0
刘菁芸	0	0	35	0	0	0	0	0	0
孙怡毓	0	0	0	0	46	0	0	0	40
凌慧雯	0	0	44	0	0	0	0	0	0
孟芳	0	0	50	0	0	0	0	0	41

	彭荃	张荊	陆盈	王蕴姣	童豫	张庭琪	柯雅芸	李熹俊	陈澜
涂蕴知	0	0	0	0	43	0	0	0	0
孙翼茜	0	37	0	0	0	41	0	0	0
毕婕靖	0	0	0	0	0	0	45	36	39
潘立	0	0	0	0	0	0	0	0	0
张培芸	41	47	49	48	0	0	0	0	0
梁茵	0	47	0	0	0	0	0	0	0
文芝	0	0	0	0	36	0	0	0	0
谢斑尚	0	39	0	0	0	0	0	0	0
蔡月	49	37	41	43	0	0	0	38	45
彭荃	0	38	0	0	0	38	0	0	0
张荊	44	0	49	0	0	0	0	0	0
陆盈	0	43	0	0	0	0	0	0	0
王蕴姣	0	0	0	0	0	0	0	0	0
童豫	0	0	0	0	0	39	0	0	0
张庭琪	37	0	0	0	37	0	0	0	0
柯雅芸	0	0	0	0	0	0	0	0	0
李熹俊	0	0	0	0	0	0	0	0	47
陈澜	0	0	0	0	0	0	0	43	0
高淼	0	0	0	0	0	0	43	0	0
易贞	0	49	0	0	39	0	0	0	0
吴霄	0	0	0	0	0	0	0	0	0
潘澜巧	0	40	0	0	0	46	0	0	0
曾帅	0	44	0	0	0	0	0	43	35
孙锦	0	0	0	0	0	0	0	0	0
刘松荷	0	0	0	0	0	0	0	0	41
吴宇晓	0	38	0	39	0	0	0	0	0
廖颜翠	0	0	0	0	0	0	0	0	0
钟倩	0	0	0	0	0	0	0	49	0
林礼琴	0	0	0	0	0	0	0	0	0
康芸晴	0	0	0	0	0	0	41	0	0
骆娴	0	0	0	0	0	0	36	0	0
柳谓	0	46	0	0	0	0	0	0	0
刘菁芸	0	0	0	0	0	0	0	0	47
孙怡毓	0	37	0	0	0	0	0	0	0
凌慧雯	0	0	0	0	0	0	0	0	0
孟芳	0	0	0	0	0	0	0	0	0

	高淼	易贞	吴霄	潘澜巧	曾帅	孙锦	刘松荷	吴宇晓	廖颜翠
涂蕴知	0	38	0	0	0	0	0	0	0
孙翼茜	0	0	0	0	0	0	0	39	0
毕婕靖	37	36	0	0	49	37	38	0	0
潘立	0	0	0	0	0	0	0	0	0
张培芸	0	48	39	36	0	0	50	49	42
梁茵	0	0	0	0	48	0	0	0	0
文芝	0	0	0	0	43	0	0	0	0
谢斑尚	0	0	0	0	0	0	0	0	0
蔡月	0	0	37	38	46	41	36	0	0
彭荃	0	0	0	0	0	0	0	0	0
张荆	0	39	0	49	48	0	0	48	0
陆盈	0	0	0	0	0	0	0	0	0
王蕴姣	0	0	0	0	0	0	0	44	0
童豫	0	42	0	0	0	0	0	0	0
张庭琪	0	0	0	38	0	0	0	0	0
柯雅芸	41	0	0	0	0	0	0	0	0
李熹俊	0	0	0	0	40	0	0	0	0
陈斓	0	0	0	0	45	0	42	0	0
高淼	0	0	0	0	0	0	0	0	0
易贞	0	0	40	0	0	0	0	0	0
吴霄	0	49	0	0	41	0	0	0	0
潘澜巧	0	0	0	0	0	0	0	0	0
曾帅	0	0	42	0	0	0	39	0	47
孙锦	0	0	0	0	0	0	0	0	0
刘松荷	0	0	0	0	45	0	0	0	0
吴宇晓	0	0	0	0	0	0	0	0	0
廖颜翠	0	0	0	0	41	0	0	0	0
钟倩	46	0	0	0	0	36	37	0	0
林礼琴	45	0	0	0	0	0	0	0	0
康芸晴	0	0	0	0	0	0	0	0	0
骆娴	0	0	0	0	0	0	0	0	0
柳谓	0	0	0	0	0	0	0	0	0
刘菁芸	0	0	0	0	0	0	0	0	0
孙怡毓	0	0	0	0	0	0	0	0	0
凌慧雯	0	0	0	0	0	0	0	0	0
孟芳	0	0	0	0	0	0	0	0	0

	钟倩	林礼琴	康芸晴	骆娴	柳谓	刘菁芸	孙怡毓	凌慧雯	孟芳
涂蕴知	0	0	0	0	0	0	0	0	0
孙翼茜	0	0	0	0	0	0	0	0	0
毕婕靖	0	43	39	0	0	47	0	43	43
潘立	0	0	0	0	0	0	0	0	0
张培芸	0	0	0	0	42	0	37	0	0
梁茵	41	0	0	0	0	0	0	0	0
文芝	0	0	0	0	0	0	0	0	0
谢斑尚	0	0	0	0	0	0	0	0	0
蔡月	0	0	0	0	0	0	50	0	49
彭荃	0	0	0	0	0	0	0	0	0
张荆	0	0	0	0	41	0	47	0	0
陆盈	0	0	0	0	0	0	0	0	0
王蕴姣	0	0	0	0	0	0	0	0	0
童豫	0	0	0	0	0	0	0	0	0
张庭琪	0	0	0	0	0	0	0	0	0
柯雅芸	0	0	35	50	0	0	0	0	0
李熹俊	46	0	0	0	0	0	0	0	0
陈澜	0	0	0	0	0	46	0	0	0
高淼	50	39	0	0	0	0	0	0	0
易贞	0	0	0	0	0	0	0	0	0
吴霄	0	0	0	0	0	0	0	0	0
潘澜巧	0	0	0	0	0	0	0	0	0
曾帅	0	0	0	0	0	0	0	0	0
孙锦	37	0	0	0	0	0	0	0	0
刘松荷	40	0	0	0	0	0	0	0	0
吴宇晓	0	0	0	0	0	0	0	0	0
廖颜翠	0	0	0	0	0	0	0	0	0
钟倩	0	0	0	0	0	0	0	43	0
林礼琴	0	0	47	0	0	0	0	0	0
康芸晴	0	48	0	40	0	0	0	0	0
骆娴	0	0	39	0	0	0	0	0	0
柳谓	0	0	0	0	0	0	0	0	0
刘菁芸	0	0	0	0	0	0	0	0	0
孙怡毓	0	0	0	0	0	0	0	0	0
凌慧雯	50	0	0	0	0	0	0	0	0
孟芳	0	0	0	0	0	0	0	0	0

## 9.2 附录二：代码说明

项目代码使用 Python 和 Matlab 编写，包含如下文件及文件夹：

Search_Function.py	——	广度优先搜索
Load_data.py	——	数据读取和整理
Plot_network.py	——	绘制网络的连接图
Clustering.py	——	聚类计算
Adj_Mat.txt	——	邻接矩阵
TL_model.m	——	信息传播过程模拟

3D连接图文件夹包含：

figure.js	——	js脚本
index.html	——	互动3D图像文件
Les-Miserables.png	——	例图
plotlyjs-bundle.js	——	js脚本

（注：代码已打包发送）