

# 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它公开的资料（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：           146          

参赛队员（签名）：

队员 1:           覃楚珺          

队员 2:           徐  博          

队员 3:           张凌霄          

---

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：         146        

---

(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：基于通讯数据的社群聚类

## 【摘要】

本文针对基于通讯数据的社群聚类问题，对社群网络进行了分类，提出了该问题的有效算法；在满足信息覆盖率最大和成本最小的前提下，给出了合理的投放方案，并在此基础上考虑多个因素的影响，对题中所给通讯网络重新进行了分类。

问题 1、2、3：问题 1、2 分别针对无向图和有向图个体识别及分类，利用邻接矩阵对网络图中节点连接情况进行初步描述；问题 3 利用 Excel 对表格中的数据进行统计分析，根据各个用户之间的通讯情况构建了以每个用户为节点的通讯网络。对这 3 个问题均以模块度为分类依据，采用在 G-N 算法的基础上进行改进的 Fast-Newman 算法进行 MATLAB 编程。在问题 1 中将 1、5、6 分为一类，2、3、4 分为一类；在问题 2 中将 1、2、3 分为一类，4、5 分为一类；问题 3 的分类结果见文中表 6-6。

问题 4：利用 Gephi 使问题 3 中的分类结果可视化，并得出了信息传播的最长路径为 5。在假定一定的有效信息接受率为 15%之后，基于信息传播终止概率的极大极小两种情况分别进行讨论，在此基础上得出了信息传播终止率介于 10%与 50%之间的投放方案。在传播终止率介于 10%-31.6%的情况下，在任意节点处投放信息即可最大可能地遍历整个网络；在传播终止率为 50%的情况下，信息最多只能传播两次。在这种情况下，针对通讯网络的中心性指标特性提出了介数中心性模型，利用问题 3 的邻接矩阵对信息的传播进行了讨论，合理地选择了投放节点。综合考虑信息传播覆盖率和投放成本，对投放点的选择做出优化。在传播终止率为 31.6%-50%的情况下得到可行的投放方案为在用户毕婕婧和蔡月处投放信息，此时成本为 2m, 信息传播覆盖率的期望值最大。

问题 5：针对社群网络中除拓扑结构外的其它有用信息，对用户的三类其它信息（通话时间、通话频率、通话地点）进行主成分分析，利用用户主成分信息构成网络权值，并将网络权值引入社群网络，完成对网络的进一步刻画。以信息流为分类依据，采用 InfoMap 算法对网络进行重新分类，并利用网站 [www.mapequation.org](http://www.mapequation.org) 所提供的信息图可视化工具 Map Generator 可视化信息图。问题 5 的分类结果见文中表 6-14。

最后给出了模型的评价、推广与改进，针对模型的优缺点进行了分析，并将通讯网络的模型推广于社交网络、电力网络等多个网络，同时从重叠社群发现和动态网络两个方面，对模型进行了一定的改进。

**【关键词】** 社群探测；介数中心性模型；Fast-Newman 算法；InfoMap 算法；主成分分析

## 1. 问题重述

### 1.1 问题背景

大数据时代的来临使得许多不可能成为了现实。数据分析和数据挖掘技术成功地在多个重大领域取得了巨大成功。现有部分人群的通讯数据，试对人群进行社群分类和相关识别。

### 1.2 需要解决的问题

(1) 图 1-1 中，六个点分别表示六个个体。点与点的边表示个体之间的联系。图一是无向图。试就图 1-1 信息，区分个体差异，对个体进行初步识别，实现分群。

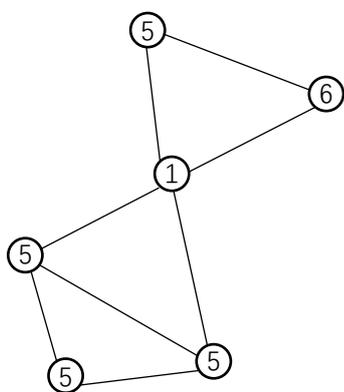


图 1-1：无向图

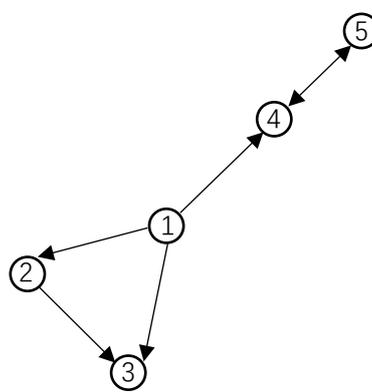


图 1-2：有向图

(2) 图二中，同样五个点表示五个个体，此时为有向图。在识别个体的同时，希望能进一步刻画任意两点之间的相似关系，从而达到“物以类聚，人以群分”的目的。

(3) 根据题目给出某营业部近三个月的内部通讯记录。结合并完善问题 1, 2 中的数学模型，对其进行分析和识别，实现个体的分群，并得出相应的结论。

(4) 社群聚类的应用中，我们通常可尝试定向的信息传播以达到较佳的传播效果，比如可以定向对某几个节点进行信息投放(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，可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，都有 10%~50%的终止传播概率，试以题目中所给记录，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。

(5) 上述问题中，实际上忽略了很多有用的信息，如通讯的位置、时间、通话频率等。那么，你能否考虑位置、时间等多种因素建立综合的数学模型，挖掘更多的信息。如有需要，可以自行补充数据，验证自己的模型，得出相应结论。

## 2. 问题假设

- (1) 题目中提供的数据真实有效，
- (2) 问题中涉及到的社群为非重叠社群，即一个节点仅属于一个社群，不存在同时属于多个社群的节点。
- (3) 问题 2 中两个节点间的距离不作为节点连接的属性。

(4) 问题 3 中仅考虑主叫用户与被叫用户是否存在联系，不考虑其它附加信息，如通话时长、通讯位置等。

(5) 问题 4 中信息传播存在有效性，当传播概念低于某个阈值时，认为信息不能被正确有效地接收，即信息传播完全终止。

(6) 问题 5 中同一信息经过传播节点，信息终止传播概率相同。

(7) 问题中研究的网络为静态网络，节点间的属性不随时间变化。

### 3. 符号说明

符号	含义
$G$	网络图
$C$	网络的邻接矩阵
$V$	网络图中的节点构成的集合
$A$	网络图中的边构成的集合
$Q$	社群中某个节点的模块度
$k_i$	节点 $i$ 的度, $i = 1, 2, \dots, n$
$e_{ij}$	网络中连接两个不同社群中的节点的边在所有边中占的比例
$E_d$	由 $e_{ij}$ 构成的 $d \times d$ 维的对称矩
$a_i$	$E_d$ 中每行各元素之和
$k_i^{in}$	节点 $i$ 的入度, $i = 1, 2, \dots, n$
$k_i^{out}$	节点 $i$ 的出度, $i = 1, 2, \dots, n$
$N$	网络中节点的总数
$M$	网络中边的总数
$C_D(v)$	节点 $v$ 的度中心性归一化值
$C_B(v)$	节点 $v$ 的介数中心性
$\sigma_{st}(v)$	节点 $s$ 到节点 $t$ 的所有最短路径中经过节点 $v$ 的数量
$R$	有效信息接收率
$s$	终止传播率
$E$	信息第一次传播到的节点构成的集合
$m$	在一个节点投放信息的成本

## 4. 问题分析

在很长的一段时间里，通讯网络、电力网络、生物网络等，分别是通讯科学、电力科学、生命科学等不同学科的主要研究对象。社群结构是复杂网络一个极其重要的特性，复杂网络通常会呈现出社群结构特性。如何在实际网络中高效地发现社群结构是近年来复杂网络的研究特点之一。

通讯数据的社群聚类分析和信息传播的研究，涉及通讯次数、位置、时间、通话频率等诸多因素，这些因素内在联系较为细微，各种因素难以精确分析，处理起来较为复杂，而且通讯社群的节点众多，数据处理的繁琐性进一步增大了问题本身的难度。

### 4.1 问题 1、2、3 的分析

问题 1 所给出的图一为无向图，要求对图中的个体进行初步识别并分类，可根据图中每个个体的度进行初步识别，并用  $N \times N$  邻接矩阵表示出来个体之间的连接情况（即两节点之间有连接邻接矩阵元素置 1，两节点之间无连接邻接矩阵元素置 0）。

问题 2 要求对有向图进行个体识别并刻画两点间的相似关系，每个个体的的度存在出度和入度的差别，通过对个体出度和入度的统计，更具体的刻画了任意两点之间的相似关系，同样也可构建有向图的邻接矩阵。

问题 3 中题目给出了大量的数据，对具体通话情况数据的搜寻相对容易，但附表有隐含的信息需要自己挖掘。对题目进行仔细分析后，我们对题目所给数据进行了整理，并分析了用户之间有无通讯的通讯情况。为了便于分析，我们对用户名进行了编号，并针对用户之间有无通讯的情况，构造了反映用户之间通讯关系的邻接矩阵（用户之间有通讯发生邻接矩阵元素置 1，用户之间无通讯发生邻接矩阵元素置 0）。

由以上分析可知，从问题 1、2、3 所给出的信息出发，我们可在个体构成的网络中依据节点间的连接关系构造出邻接矩阵，邻接矩阵可很好的反映网络的拓扑结构。因此以上三个问题都可从邻接矩阵入手建立模型求解。

### 4.2 问题 4 的分析

问题 4 以问题 3 为基础，要求在 36 个用户的分群结果下定点投放信息，需要考虑信息覆盖率和成本等因素，这两个因素限制了投放点的多少和所在的位置。首先，确定信息的传播与网络节点之间的连接有关，即如果网络节点间有连接，则表明信息可以在这两节点间传播。因此，在此问题中我们仅仅考虑网络的拓扑结构进行问题求解与优化。由于存在终止传播概率，信息传播次数越多则信息越不可达，信息覆盖率越低。如果增加投放点，可以加大信息传播的覆盖率，但会增加成本（极端情况下，给所有节点投放信息可达最大信息覆盖率，但此方案显然不可行）；如果减少投放点，可以减小成本，但信息传播覆盖率会随之下降。由以上分析可知，本问题的核心是确定信息传播的有效次数。

根据以上分析：我们由不同的终止传播概率入手，对在终止概率为 10%和终止概率为 50%的两种极大和极小概率展开讨论，得出两种概率条件下信息的有效传播次数。结合信息覆盖率和成本因素，在此基础上建立了基于通讯网络的介数中心性指标模型。

下面的思路流程图是我们文章结构的一个缩影，它完整而形象地反映了我们文章地建模思路。

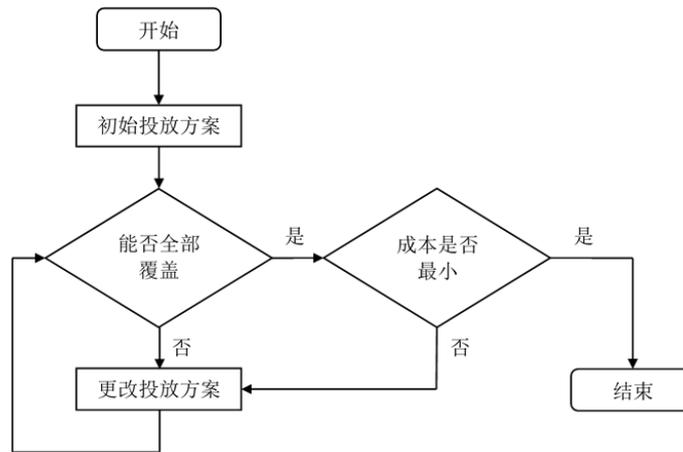


图 4-1：思维流程图

### 4.3 问题 5 的分析

问题 1-4 仅仅依据网络的拓扑结构进行社群分类，而对网络其它信息有诸多忽略，在本问题中需要综合考虑网络拓扑结构和网络节点其它有用信息（如通讯位置、通讯时间、通讯频率）建立综合数学模型进行社群分类和相关识别。为了把网络节点其它有用信息引入到网络中，我们对通讯位置、通讯时间和通讯频率进行主成分分析从而得出网络权值，并把上述问题中的无权网络化为有权网络。此时邻接矩阵已经不足以表示网络所有信息，这里我们将网络信息表示为{源点节点(主叫)，目标节点(被叫)，网络权值(主成分)}的格式，并改进针对网络拓扑结构的分类方法，利用信息熵的概念，从网络信息流分析方面进行有权有向社群网络的分类求解。

## 5. 建模前的准备

在建立模型之前我们对该社群问题概念以及相关基本算法做了研究。

基于网络社群拓扑结构的传统算法可分为两大类，第一类是基于图论的算法，如 K-L 算法<sup>[1]</sup>、谱平分法<sup>[2]</sup>等；第二类是层次聚类算法，如基于边介数的分裂算法<sup>[3]</sup>、基于相似度量度量的凝聚算法<sup>[4]</sup>、Kernighan-LinLin 算法<sup>[5-7]</sup>、GN 算法<sup>[8]</sup>、FN 算法<sup>[9]</sup>、CNM 算法等。

### 5.1 GN 算法

Girvan 和 Newman 于 2001 年提出了一种基于边介数的分裂算法，简称 GN 算法，这种算法根据网络中社群内部高内聚、社群之间低内聚的特点，逐步去除社群之间的边，取得相对内聚的社群结构。算法利用边介数概念来探测边的位置，某边的边介数定义为网络上所有顶点之间的最短路径通过该边的次数，其准确度较高，但其算法复杂度比较大，仅限于研究中等规模的复杂网络。

Newman 还并提出了模块度的概念。它是基于同型混合来定义的。考虑某种划分形式，它将网络划分为  $d$  个社群。定义一个  $d \times d$  维的对称矩  $E_d = \{e_{ij}, i=1,2,\dots,d, j=1,2,\dots,d\}$ ，其中元素  $e_{i,j}$  表示网络中连接两个不同社群中的节点的边在所有边中占的比例，这两个结果分别位于第  $i$  个社群和第  $j$  个社群。其中模块度的衡量标准是利用完整网络计算的。设矩阵中对角线上各元素之和为

$$Tre = \sum_i e_{ii} \quad 5-1$$

它给出了网络中连接某一个社群内部各节点的边在所有边中所占的比例。定义每行中各元素之和为

$$a_i = \sum_j e_{ij}, \quad 5-2$$

它表示于第  $i$  个社群中的节点相连的边在所有边中所占的比例。在以上陈述的基础上，用下式定义模块度

$$Q = \sum_i (e_{ii} - a_i^2) \quad 5-3$$

上式的物理意义是：网络中连接两个同种类型的节点的边（即社群内部边）的比例减去在同样的社群结构下任意连接这两个节点的边的比例的期望值。如果社群内部边的比例不大于任意连接时的期望值，则有  $Q=0$ 。一般认为  $Q$  的最大值对应的社群结构就是网络的社群结构。 $Q$  的上限为  $Q=1$ ， $Q$  越接近这个值，就说明网络的社群结构越明显，实际网络中， $Q$  值通常位于  $0.3-0.7$  之间<sup>[10]</sup>。模块度越大则表明社群划分效果越好。

## 5.2 Fast - Newman 算法

2004 年，Newman 提出了在 GN 算法上改进优化的基于贪婪算法思想的一种凝聚算法—Fast - Newman 算法，该算法可以用于分析节点数达 100 万的复杂网络。这种算法是在使模块度不断增加的基础上进行，即每次合并沿着使模块度增大最大和减小最小的方向进行。其算法如下：

Step1: 初始化网络为  $N$  个社群，即每个节点为一个独立社群。初始的  $e_{ij}$  和  $a_i$  满足

$$e_{ij} = \begin{cases} 1/2M, & \text{如果结点 } i, j \text{ 之间有边相连} \\ 0, & \text{其他} \end{cases}, \quad 5-4$$

$$a_i = k_i / 2M$$

其中  $k_i$  为节点  $i$  的度， $M$  为网络中总的边数。

Step2: 依次合并有边相连的社群对，并计算合并后的模块度增量

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ij} - a_i a_j) \quad 5-5$$

根据贪婪算法的原理，每次合并应该沿着  $Q$  增大最多或者减小最少的方向进行，其算法复杂度为  $O(M)$ 。每次合并后，对相应的元素  $e_{ij}$  更新，并将  $i, j$  社群相关的行和列相加，其时间复杂度为  $O(N)$ ，因此步骤 2 总的时间复杂度为  $O((M+N)N)$ ，对于稀疏网络则为  $O(N^2)$ ，其中  $N$  为网络中节点的个数。

整个算法完成后可以得到一个社群结构分解的树状图。再通过选择在不同位置断开可以得到不同的网络社群结构。在这些社群结构中，选择一个对应着局部最大  $Q$  值的，就得到最好的网络社群结构。

在本题中，我们选用了 Fast-Newman 算法进行求解。

## 6. 模型的建立与求解

在下面数据结构的讨论中，我们首先假设图  $G=(V,A)$ ， $|V|=n,|A|=m$ ，并假设  $V$  中的顶点用自然数  $1, 2, \dots, n$  表示或编号， $A(G)=\{(i_1,j_1),(i_2,j_2),\dots,(i_M,j_M)\}$  代表边集合。如果  $(i,j)=(j,i)$ ，则为无向图，否则为有向图。将该图采用邻接矩阵表示法进行表达。邻接矩阵表示法是将图以邻接矩阵 (Adjacency Matrix) 的形式存储在计算机中。图  $G=(V,A)$  的邻接矩阵是如下定义的： $C$  是一个  $n \times n$  的 0-1 矩阵，即：

$$C = (c_{ij})_{n \times n} \in \{0,1\}^{n \times n} \quad 6-1$$

$$c_{ij} = \begin{cases} 1, (i,j) \in A \\ 0, (i,j) \notin A \end{cases} \quad 6-2$$

当网络是一个无向无权网络时，邻接矩阵  $C=\{c_{ij}\}$  是一个 0-1 对称矩阵，当网络是有向图时，邻接矩阵  $C$  是非对称的 0-1 矩阵，当网络是加权网络时， $C$  中的非元素代表边的权重。图中节点的度表示与该点连接的边数，如果用  $k_i$  表示节点  $i$  的度，则可以用邻接矩阵  $C$  来表示为：

$$k_i = \sum_{j \in N} c_{ij} \quad 6-3$$

如果是有向图，点的度数则包括两部分：出度与入度，即该点所指向其它点的边数和指向该点的边数，如果用  $k_i^{out}$ 、 $k_i^{in}$  来表示节点的出度与入度，则：

$$k_i = k_i^{in} + k_i^{out} \quad 6-4$$

问题一中的图一是无向图，其邻接矩阵为 0-1 对称矩阵，对于问题二、问题三的信息可以得到有向图，其节点的连接信息也可以用邻接矩阵表示，因此无论是问题 1，还是问题 2、3 都可以用邻接矩阵来刻画，因此对于问题 1、2、3 的数据形式，采用均 Fast-Newman 算法。

### 6.1 问题 1 模型的建立与求解

图 1 是一个无向图，将其中的每个个体看成一个节点，可以得到各个节点的度，具体数据如下表所示：

表 6-1：问题 1 节点度数

节点编号	1	2	3	4	5	6
度数	5	3	3	3	2	2

我们可以将表 1 的信息转化成邻接矩阵，得到的 0-1 对称矩阵如下：

$$\begin{pmatrix} 0 & 1 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}$$

我们根据 Fast-Newman 算法编写了 MATLAB 程序，见附录 1。运行程序，经过聚类

分析得到图 1 中六个个体的分群结构图：

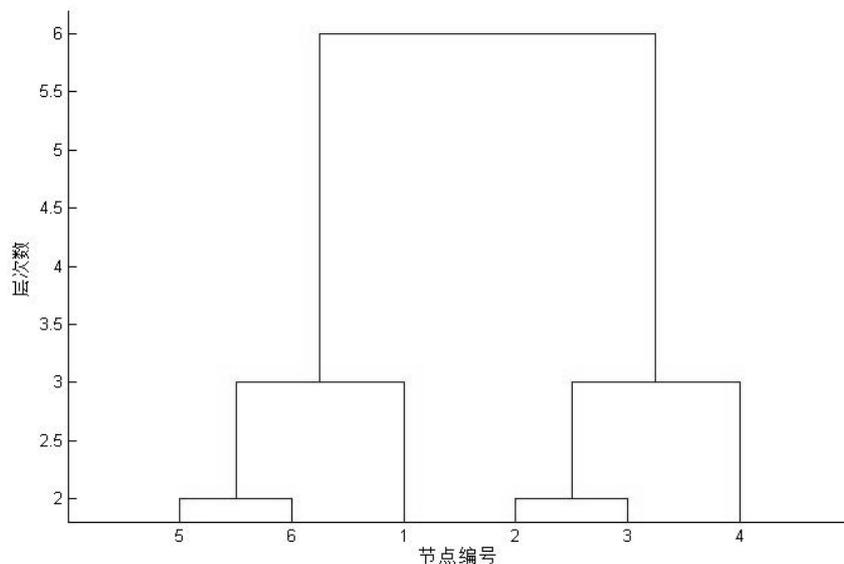


图 6-1： 图 1 的结构分解树状图

根据以上结构分解树状图，我们将 1、5、6 分为一类，2、3、4 分为一类，分类情况如下表：

表 6-2： 问题 1 分类结果

类别	节点编号
1	1、5、6
2	2、3、4

通过比对分类结果与图 1 中直观信息可知，1、5、6 节点之间的模块性和 2、3、4 节点之间的模块性较好，因此认为分两类结果合理。

## 6.2 问题 2 模型的建立与求解

图 2 为一个有向图，把图中的每个个体看成一个节点，可以得到各节点的出度和入度，具体数据如下表所示：

表 6-3： 问题 2 节点出入度

节点编号	1	2	3	4	5
出度	2	1	1	1	1
入度	1	1	1	2	1

我们同样可以将图二的信息转化成邻接矩阵，得到的非对称的 0-1 矩阵如下：

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

在 MATLAB 程序中对数据进行更换进行聚类分析，得到的结构分解树状图如下：

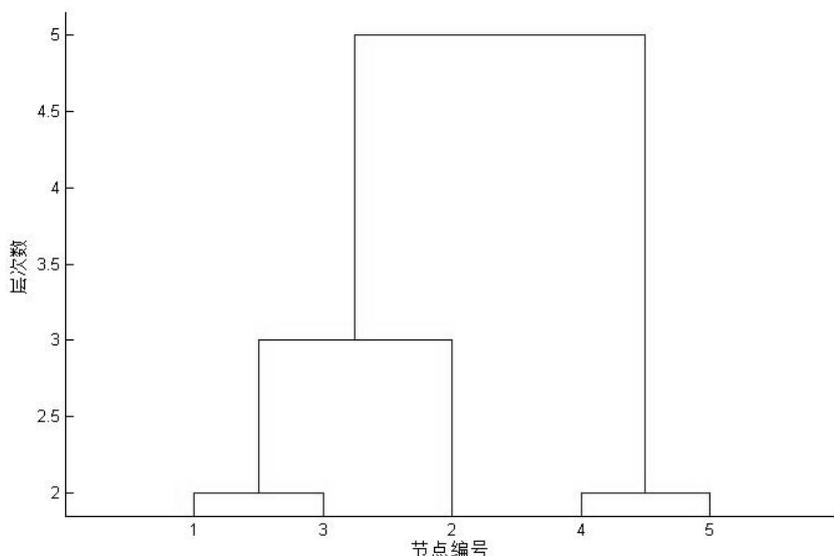


图 6-2： 问题 2 的结构分解树状图

根据以上结构分解树状图，可看出 1、2、3 的相似度较高，4、5 的相似度较高，因此我们可以将 1、2、3 分为一组，4、5 分为一组，分类情况如下表：

表 6-4： 问题 2 分类结果

类别	编号
1	1, 2, 3
2	4, 5

通过比对分类结果与图 2 中直观信息可知，1、2、3 节点之间的模块性和 4、5 节点之间的模块性较好，因此也认为分两类结果合理。即，1、2、3 节点之间的相似性较高，4、5 节点之间的相似性较高。

### 6.3 问题 3 模型的建立与求解

题目附件中一共给出了 36 个用户之间的通讯记录，在不考虑通讯的位置、时间、通话频率的情况下，对题目所给数据进行整理，如果将每个用户看成一个节点，对这 36 个用户进行编号，如下表所示：

表 6-5： 用户编号表

用户	毕婕靖	蔡月	曾帅	陈斓	高淼	康芸晴
编号	1	2	3	4	5	6
用户	柯雅芸	李熹俊	梁茵	廖颜翠	林礼琴	凌慧雯
编号	7	8	9	10	11	12
用户	刘菁芸	刘松荷	柳谓	陆盈	骆娴	孟芳
编号	13	14	15	16	17	18
用户	潘澜巧	潘立	彭荃	孙锦	孙怡毓	孙翼茜
编号	19	20	21	22	23	24
用户	童豫	涂蕴知	王蕴姣	文芝	吴霄	吴宇晓
编号	25	26	27	28	29	30

用户	谢斑尚	易贞	张荆	张培芸	张庭琪	钟倩
编号	31	32	33	34	35	36

这 36 个节点构成了一个网络，我们利用 Gephi 对每个点进行标记，得到如图 6 所示的有向网络图：

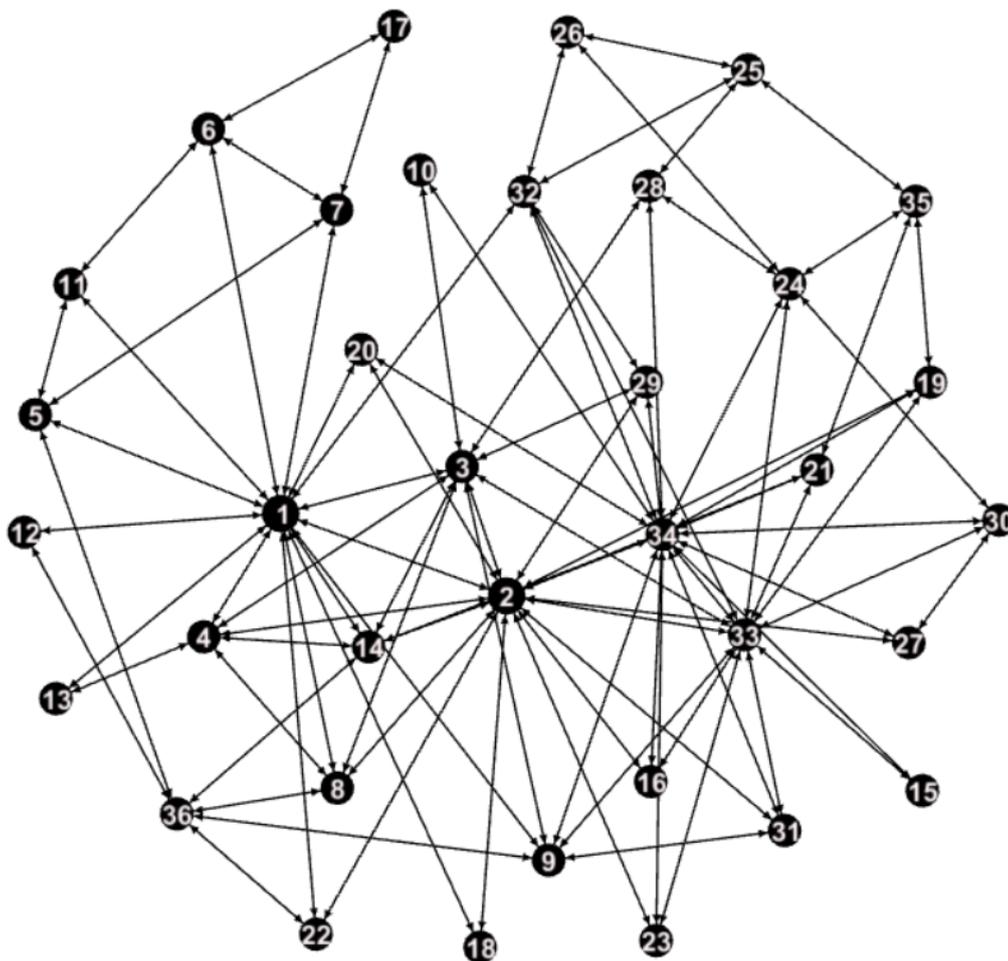


图 6-3：初始网络图

将两个用户之间有通话视为两个节点之间有边。若主叫用户对被叫用户进行呼叫，则该边的方向为从主叫用户代表的节点指向被叫用户代表的节点。由此得到的邻接矩阵如附录 2 所示，在 MATLAB 程序中更换相应数据进行聚类分析，得到的结构分解图如图 6-4，可视化网络聚类图如图 6-5。

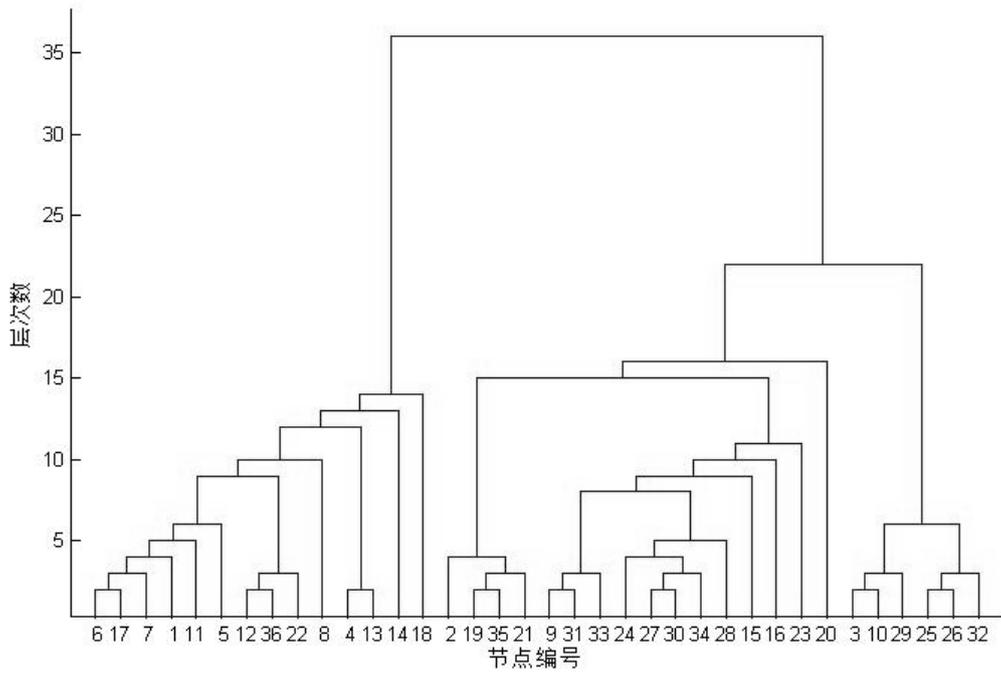


图 6-4：问题 3 的结构分解树状图

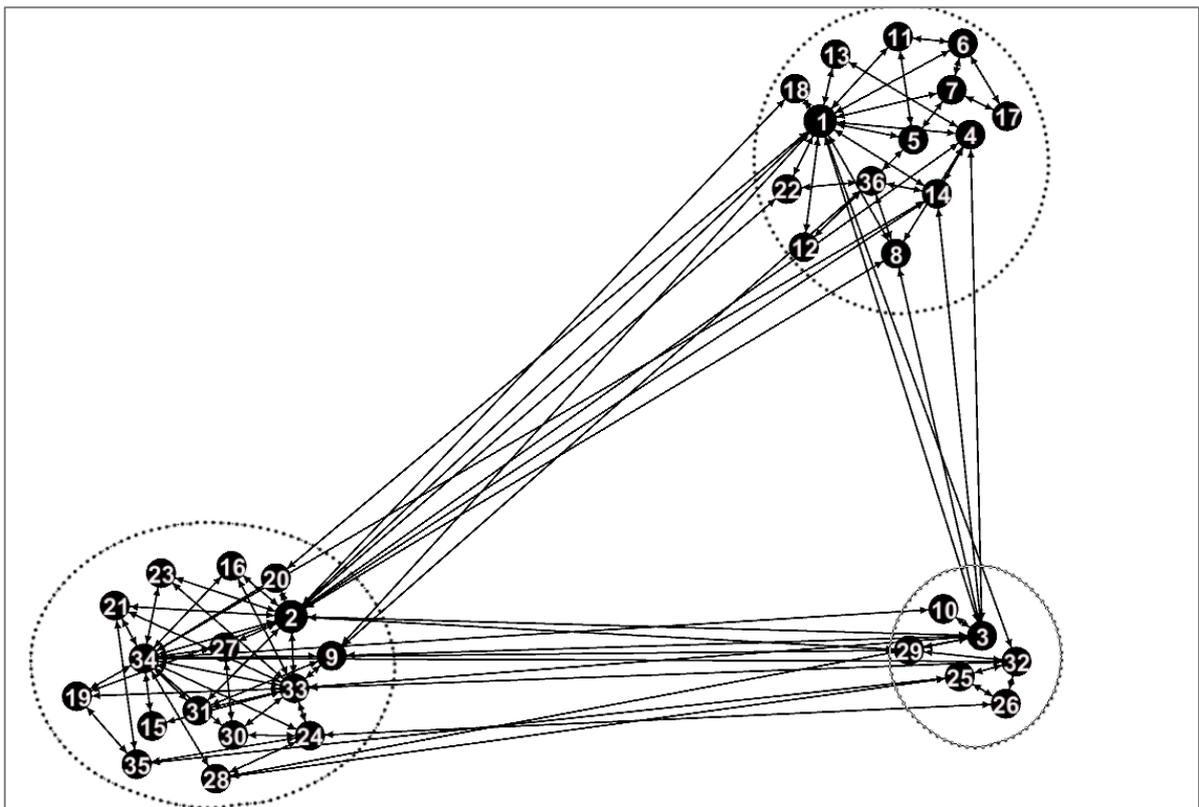


图 6-5：社群聚类后网络图

根据以上结构分解树状图，可以将 36 个节点分为三类：

表 6-6: 社群聚类结果

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

## 6.4 问题 4 模型的建立与求解

本问要求选择最佳投放点，根据题目的要求，要达到较高的信息传播率同时控制成本。具体化到这 36 个用户构成的通讯网络，要求在最少的节点投放达到最大的信息传播率。

在下面的讨论中，引入中心性 (Centrality) 的概念进行分析。

### 6.4.1 中心性指标介绍

寻找网络中的关键节点一直都是网络科学的重要研究内容之一，在社会网络分析中，节点的重要性排序常用“中心性”来表示。衡量节点中心性的几个常用指标包括度中心性、紧密中心性、介数中心性和 k-核中心性。下面从有向网络的角度来介绍节点的中心性指标。

#### (1) 度中心性

衡量网络节点中心性最直接的指标就是度中心性 (Degree centrality)。它认为一个节点的度越大意味着与越多节点相连接，那么这个节点也就越重要。在有向网络中，定义消息的传播方向为出度方向，则用出度表示节点的度中心性以衡量节点影响力。在一个  $N$  个节点的网络中，每个节点可能的最大度值为  $N - 1$ ，则出度为  $k_{out}$  的节点的度中心性归一化值为：

$$C_D(v) = \frac{k_{out}}{N-1} \quad 6-5$$

#### (2) 紧密中心性

紧密中心性 (Closeness centrality) 显示了一个节点与其它节点连接的紧密程度。它的值是由用户到其它节点最短距离之和的倒数决定，一般情况下计算紧密中心性时认为节点到其它所有的节点都是可到达的<sup>[11]</sup>。

#### (3) 介数中心性

衡量网络中心性特征的另一重要指标就是介数中心性 (Betweenness centrality)，它认为网络中心的节点就是通过该节点的最短路径数目越多的节点，其计算公式为：

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}(N-1)(N-2)} \quad 6-6$$

其中， $\sigma_{st}(v)$  代表节点  $s$  到节点  $t$  所有最短路径中经过节点  $v$  的数量， $\sigma_{st}$  为  $s$  到  $t$  所有的最短路径数， $(N-1)(N-2)$  表示有向网络中最大可能的介数值。

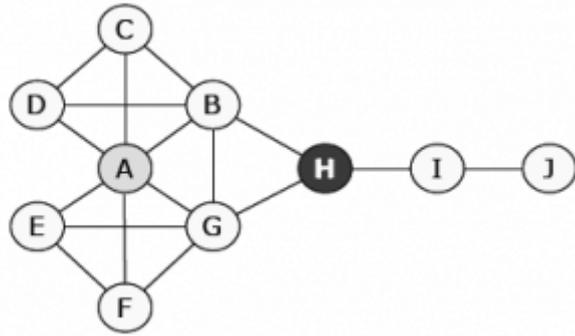


图 6-6 网络社群

例如，对于图 6-6 网络节点 A 的度中心性高于节点 H，因为节点 H 的直接连接更少。而节点 H 有更高的中介中心性，它在这个网络中扮演经纪人的角色，它处在网络内许多节点交往的路径上，因此，它具有控制其它人交往的能力。

#### (4) k-核中心性

k-核中心性(也称 k-shell 或者 k-core )<sup>[12]</sup>揭示了网络拓扑的层次特性，节点的核值  $k_c$  越大就越接近网络的中心。k-核分解方法是通过递归地移去网络中所有度小于  $k$  的节点及节点的边，直到网络中所有的节点度都不小于  $k$ 。

### 6.4.2 中心性分析

对于第三问所构造的社群聚类网络，我们采用 Gephi 对其中心性分析，结果如下：

表 6-7：节点中心性

ID	Label	Eccentricity	Closness Centrality	Harmonic Closness Centrality	Betweenness Centrality
编号	姓名	离心率	紧密中心度	调和紧密中心度	介数中心度
1	毕婕靖	3	0.603	0.710	430.163
2	蔡月	3	0.625	0.729	289.779
34	张培芸	4	0.583	0.721	235.636
32	易贞	3	0.530	0.576	110.681
33	张荆	4	0.538	0.650	103.462
3	曾帅	3	0.556	0.629	92.711
9	梁茵	3	0.515	0.567	46.986
7	柯雅芸	4	0.402	0.467	34.917
6	康芸晴	4	0.398	0.462	32.667
24	孙翼茜	5	0.412	0.506	31.571
14	刘松荷	3	0.515	0.567	27.404
19	潘澜巧	4	0.461	0.512	20.605
21	彭荃	4	0.461	0.512	20.605
28	文芝	4	0.455	0.507	19.883
36	钟倩	4	0.402	0.483	18.956
4	陈斓	3	0.493	0.552	17.312
25	童豫	4	0.385	0.443	12.686

5	高淼	4	0.402	0.467	8.936
35	张庭琪	5	0.354	0.422	8.667
8	李熹俊	3	0.486	0.538	8.199
20	潘立	3	0.486	0.519	7.216
22	孙锦	3	0.461	0.500	5.090
27	王蕴姣	4	0.443	0.488	4.474
26	涂蕴知	4	0.372	0.419	2.986
30	吴宇晓	5	0.398	0.472	2.667
29	吴霄	4	0.461	0.512	2.369
12	凌慧雯	4	0.389	0.433	1.769
11	林礼琴	4	0.398	0.452	0.917
31	谢斑尚	4	0.455	0.507	0.400
10	廖颜翠	4	0.427	0.464	0.286
15	柳谓	5	0.380	0.434	0.000
16	陆盈	4	0.443	0.488	0.000
23	孙怡毓	4	0.443	0.488	0.000
17	骆娴	5	0.294	0.340	0.000
13	刘菁芸	4	0.385	0.429	0.000
18	孟芳	3	0.449	0.481	0.000

注：<sup>[16]</sup>

1. 在 Gephi 中互相连接的节点的图距离为 1，
2. 离心率 (Eccentricity) 定义为从一个给定起始节点到距离其最远节点的距离
3. 介数中心度 (Betweenness Centrality) 定义为一个节点出现在网络中最短路径上的频率
4. 紧密中心度 (Closness Centrality) 定义为从一个给定起始节点到所有其它节点的平均距离。

### 6.4.3 信息传播转发特征

以下分析求解基于通讯网络的信息传播转发特征：

通讯用户呼叫图是一个社群网络。如果假定信息转发概率恒定，则通过持续转发可以得到较高的信息转发长度，但实际上，转发路径的长度通常较短：根据 Kwak<sup>[13]</sup>等的分析的传播路径表明，97.6%的转发路径都小于 6，且最长距离也不会超过 11。沈珂轶等<sup>[14]</sup>认为这主要是因为随着信息传播层数的增加，人们对信息的兴趣是递减的。Eytan · Baks 等<sup>[15]</sup>同样在研究中发现信息的最大转发长度只能达到 9，并且随着转发长度的增加，出现频率急剧下降。

### 6.4.4 信息传播分析

#### (1) 最小终止传播率传播

本题中所给的终止传播概率为 10%–50%，首先考虑终止传播概率为 10%的情况。

在这种情况下，从一个节点传播信息到另一个节点的有效信息接收率为 90%。我们假定有效信息接收率低于 15%即视为不能有效传播。

观察表 6-7 Gephi 给出的离心率，得出从投放起始节点到距其最远节点的距离为

5, 即信息传播六次之后即可循环回到投放起始点, 就是说在起始点投放过后最多经过五次就可以遍历整个网络。

在经过五次传播后有效信息接收率为  $R = (1-10\%)^5 = 0.59049 > 0.15$

上式是在从起始点投放, 信息沿最长传播路径传播的情况下得出, 此时能有效传播, 即从起始点投放到达任意节点的有效率始终大于 59%。因此可以得出, 在 10% 的终止传播概率下, 在任意一个节点投放, 即可最大可能地遍历整个通讯网络, 同时也能满足成本最小的条件。

## (2) 最大终止传播率传播

下面考虑终止传播概率为 50% 的情况, 我们建立了基于社交网络中的介数中心性指标模型。该模型主要是结合社交网络中节点的介数中心性指标来进行选取种子节点。根据介数中心性指标来判断一个用户的影响力。如果该用户的联系人越多, 他所发出的消息就可以被很多人接收, 消息被人转发的可能性更大, 间接说明他可能影响到的人越多。所以介数中心性大的节点影响力大。根据介数中心性对每一类的节点进行排序, 每类中介数中心性最大的节点, 即通过该节点的最短路径数目越多的节点, 也就是网络中心的节点。由于此种方法可能导致种子节点之间都是邻居或者种子节点之间的邻居存在很多重叠的节点, 最终选中的种子节点集合效果不稳定。这种模型的缺点是不稳定, 效果可能会随网络结构的变化而变化。例如, 当每个节点的介数中心性都差不多时, 选出的节点集合和随机选出的节点集合效果差不多。其优点是方法简单, 时间复杂度小, 对于选取种子节点个数少的情况下比较适用。

在信息传播终止率为 50% 的情况下, 信息经过两次传播后, 有效到达第三个节点的概率为  $R = (1-50\%)^2 = 0.25 > 0.15$ , 认为经过两次传播后能成功到达第三个节点。信息经过三次传播后, 有效到达第四个节点的概率为  $R = (1-50\%)^3 = 0.125 < 0.15$ , 认为经过三次传播后不能成功到达第四个节点。

观察表 6-7 的介数中心性指标结合第三题中对 36 个用户组成的通讯社群网络的分类, 可以得到, 在第一、二、三类中介数中心性最大的用户分别是节点编号为 1 的用户毕婕靖、节点编号为 2 的用户蔡月、节点编号为 32 的用户易贞。

下面从选定这三个用户对应的节点为种子节点, 在第三题中根据用户信息构造的邻接矩阵中对信息投放方案进行讨论。

设下图的邻接矩阵中  $C$  的元素为  $c_{ij}$ , 其值表示两个用户之间的通讯情况

$$c_{ij} = \begin{cases} 1, & \text{用户 } i(\text{主叫用户}) \text{ 与用户 } j(\text{被叫用户}) \text{ 有通话记录} \\ 0, & \text{用户 } i(\text{主叫用户}) \text{ 与用户 } j(\text{被叫用户}) \text{ 无通话记录} \end{cases}$$

① 在节点 1 处投放一个信息, 信息传播过程如下:

Step1: 在第一信息传播层中, 信息沿如图 6-7 所示的横向箭头方向在邻接矩阵中横向传播至所有与节点 1 直接联系的标为灰色的节点  $j$ , 其中  $j$  满足  $c_{1j}=1$ , 记  $j \in E$ 。

Step2: 在第二信息传播层中, 信息沿如图 6-7 所示的纵向箭头方向在邻接矩阵中纵向传播至所有与第一信息传播层中直接联系的标为灰色的节点  $k$ 。其中  $k$  满足为

$$c_{kj} = 1 \quad (k = 2, 3, \dots, 36, j \in E)。$$

观察下图可知，在节点 1 处投放信息后，可以看出编号为 15、30 等的节点处未接收到信息，因此只在节点 1 处投放达不到最大的信息传播覆盖率，此时成本为  $m$ 。

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36					
编号	姓名	毕婕靖	蔡月	曾帅	陈澜	高淼	康芸晴	柯雅芸	李熹俊	梁茵	廖颜翠	林礼琴	凌慧雯	刘菁芸	刘松荷	柳渭	陆盈	骆娴	孟芳	潘澜巧	潘立	彭荃	孙锦	孙怡毓	孙翼茜	童豫	涂蕴知	王蕴姣	文芝	吴霄	吴宇晓	谢斑尚	易贞	张荆	张培芸	张庭琪	钟倩					
1	毕婕靖	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0			
2	蔡月	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0			
3	曾帅	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0			
4	陈澜	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
5	高淼	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
6	康芸晴	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0		
8	李熹俊	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1		
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	
10	廖颜翠	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	刘松荷	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
15	柳渭	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0		
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
17	骆娴	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	孟芳	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
20	潘立	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	0	
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	1	0	0	1	0	
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	0		
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
31	谢斑尚	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
33	张荆	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	0	1	1	1	0	0	1	1	1	0	1	0	0	1	1	1	1	1	1	1	1	1	1	0	0	0	0	
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	

图 6-7：在节点 1 处投放信息传播层

②在节点 2 处投放一个信息，信息传播规律与节点 1 相同。观察下图可知，在节点 2 处投放信息经过两次传播后可以看见有节点 25 等未接收到信息，因此只在节点 2 处投放达不到最大信息传播覆盖率，此时成本为  $m$ 。

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
1	毕婕靖	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
2	蔡月	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0	1	1	0	0	
3	曾帅	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	
4	陈澜	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	高淼	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
6	康芸晴	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	李熹俊	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1	
10	廖颜翠	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	刘松荷	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1
15	柳谓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	骆娴	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	孟芳	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
20	潘立	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	
31	谢斑尚	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	1	0
33	张荆	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0	0
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

图 6-8：在节点 2 处投放信息传播层

③在节点 32 处投放一个信息，信息传播规律与节点 1 相同。观察下图可知，在节点 32 处投放信息经过两次传播后有节点 4 等未接收到信息，因此只在节点 32 处投放达不到最大信息传播覆盖率，此时成本为  $m$ 。

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36		
1	毕婕靖	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0		
2	蔡月	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0	1	1	0	0	0	
3	曾帅	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	
4	陈澜	1	1	1	0	0	0	0	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	高淼	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
6	康芸晴	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	李熹俊	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1		
10	廖颜翠	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	刘松荷	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1		
15	柳谓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0		
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0		
17	骆娴	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	孟芳	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0		
20	潘立	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0		
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0		
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0		
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0	1	0	0	1		
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0		
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1		
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0		
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0		
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0	
31	谢斑尚	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0		
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0	
33	张荆	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	1	1	0	0	0	
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

图 6-9：在节点 32 投放信息传播层

④在节点 1、2、32 处各投放一个信息，信息传播规律与节点 1 相同。观察下图可知：在节点 1、2、32 处投放信息经两次传播后，可以覆盖整个网络。

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	
	毕婕靖	1																																				
1	毕婕靖	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	
2	蔡月	1	0	1	1	0	0	0	1	0	0	0	0	0	1	1	0	1	1	1	1	1	1	1	1	0	0	1	0	1	0	1	1	1	1	0	0	
3	曾帅	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	
4	陈澜	1	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	高淼	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
6	康芸晴	1	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	李熹俊	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1		
10	廖颜翠	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	刘松荷	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	
15	柳渭	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
17	骆娴	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	孟芳	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
20	潘立	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	
31	谢斑尚	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0
33	张荆	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	1	1	1	0	1	0	0
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	1	0	0	0
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

图 6-10：在节点 1，2，32 投放信息传播层

在 4 的讨论中，我们已经找到了一种达到使信息传播达到最大覆盖率的投放方案，此时的成本为  $3m$ 。为了进一步减少成本，可以尝试减少投放点。由问题 3 可知第三类中所包含的节点数最少，且第 32 个节点的介数中心性在 1、2、32 这三个节点中最小，所以我们尝试着不在第 32 个节点处投放，只考虑在 1、2 节点处投放。

⑤在节点 1、2 处投放，信息传播规律与节点 1 的相同。观察下图可知：在节点 1、2 处投放信息经两次传播后，可以覆盖整个网络，此时的成本为  $2m$ 。

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
1	毕婕婧	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	
2	蔡月	1	0	1	0	0	0	1	0	0	0	0	0	0	1	0	1	0	1	1	1	1	1	0	0	0	0	1	0	1	0	1	1	0	0	0	
3	曾帅	1	1	0	0	0	0	1	1	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	
4	陈澜	1	1	0	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	高淼	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
6	康芸晴	1	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	李熹俊	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1		
10	廖颜翠	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	刘松荷	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	
15	柳渭	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
17	骆娟	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	孟芳	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
20	潘立	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	1	0	0	0	0	
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	1	0	0	
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	1	1	0	0
31	谢斑尚	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0
33	张荆	0	1	1	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	1	0	1	0	1	0	0	0	0	0	1	1	1	0	1	0	0	
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	0	0	0
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0	0	0	0	0	0	0
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0

图 6-11：在节点 1，2 投放信息传播层

由上述①②③④⑤的讨论可知，仅在 1、2、32 中的一个节点处投放，虽然成本较低，但是达不到最大信息传播覆盖率；在 1、2、32 三个节点处投放，虽然可达到最大信息传播覆盖率，但是成本较高；而在 1、2 两个节点处投放既能满足信息传播覆盖率最大，其成本也适中。综合考虑信息传播覆盖率和成本等因素，我们选择了在 1、2 节点处投放这一可行的方案，即仅在用户毕婕婧和蔡月处投放信息。

### (3) 其余终止传播率传播

当终止传播率介于 10%-50%时，在最长传播路径的情况下，若满足有效信息接受率大于设定阈值 15%，则认为在任意节点投放信息都能最大可能遍历整个网络，记终止传播率为  $s$ ， $s$  应满足：

$$(1 - s)^5 < 0.15$$

解得  $s < 0.316$ 。故当  $10\% < s < 31.6\%$  时，选用任意节点便能使信息传播覆盖率的期望值达到最大，此时投放成本为  $m$ 。当  $31.6\% < s < 50\%$  时，为使信息传播覆盖率的期望值达到最大，应选用节点 1（用户毕婕婧）和节点 2（用户蔡月）处投放信息，此时投放成本为  $2m$ 。

## 6.5 问题 5 模型的建立与求解

### 6.5.1 网络权值选取

在上述问题的节点分类中，节点属性仅考虑了到了与其它节点的连接关系，而在实际应用中，一个网络的节点具有多重属性，例如在一个公路网络中，一个节点与另一个节点间的距离，社交网络中两个节点的联系紧密程度都是十分重要的属性，因此我们需要引入其它量来描述这类属性。这里我们引入权的概念，将权定义为两个节点的联系紧密程度，从而将问题 3 中的有相无权网络转换成有相有权网络。

在通讯网络中，两个用户之间包含的信息包括通讯的位置、时间、通话频率等。在本问题中我们选用通讯位置，通话时间和通话频率作为两个节点的属性值。其中，通话频率 =  $\frac{\text{某用户通话次数}}{\text{某用户通话天数}}$ ，通话时长采用附件中原始数据。而由于附件数据中通讯位置（通话地点）全为武汉，位置信息对于各网络节点影响等效，故在此问题中将位置属性值忽略，仅考虑通话时间和通话频率。

下面采用中主程序分析法考虑通话时间和通话频率属性，为网络建立权值。主成分分析法是一类对非随机变量的分析方法，主要目的是希望用较少的变量去解释原有的大部分变异，将许多相关性很高的变量转化成彼此相互独立或不相关的变量，通常是选出比原始变量个数少，能解释大部分变异的几个新变量，即所谓主成分，并用以解释资料的综合性指标，即主成分分析实际上一种降维方法。

为分析主成分，我们首先分析通话时间和通话频率的变量相关性，在这里我们采用皮尔逊积矩相关系数来对变量相关性进行描述。皮尔逊积矩相关系数定义为：

$$\rho_{X,Y} = \frac{\text{cov}(X,Y)}{\sigma_X \sigma_Y} = \frac{E((X - \mu_X)(Y - \mu_Y))}{\sigma_X \sigma_Y} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - E^2(X)}\sqrt{E(Y^2) - E^2(Y)}} \quad 6-7$$

该系数越接近 ±1，表明变量 X 和变量 Y 的相关性越大。

在此问题中，令 X 表示通话时间，Y 表示通话频率，并进行双变量线相关性分析，所得结果如表 6-8。

表 6-8： 通话时间和通话频率相关性结果

		通话时长	通话频率
通话时长	皮尔逊相关性	1	0.999
	显著性（单尾）		0.000
	个案数	192	192
通话频率	皮尔逊相关性	0.999	1
	显著性（单尾）	0.000	
	个案数	192	192

\*在 0.01 级别（单尾），相关性显著。

通过表 6-8 可知，通话时长与通话频率的相关性为 0.999，相关性显著，故可任选其一作为主成分，这里我们选通话频率作为主成分，从而得到了各个节点间的权值。

## 6.5.2 社群分类模型优化

社群检测算法除了利用网络拓扑分析，还可以利用流分析。其中，拓扑分析主要用于无权网络，而流分析主要应用于有权网络，思路是发现在网络的某种流动（物质、能量、信息）中形成的社群结构。

上述问题中，FN 算法是基于网络拓扑结构，利用模块度最大的原则进行社群分类，这样的分类方法没有考虑到网络权值对于社群分类的影响。在此问题中，为了考虑到网路权值对于社群分类的影响，我们引入信息流的概念，采用流分析对于社群网络重分类。

在流分析中，Infomap 算法被认为是目前社群划分效果最优的社群检测算法之一，本问题采用 Infomap 算法进行社群分类。

Infomap 以图中随机行走路径的编码长度作为要优化的目标函数，把网络中的社群检测问题转化成了信息压缩编码问题。如果对每个网络中每个节点进行霍夫曼编码表示，那么就可以将网络中的任意一条路径用一定的码字来表示。首先对整个图进行社群划分，对每个社群都会进行社群级别的霍夫曼编码，后在每个社群内部，对处于同一个社群内的节点进行节点级别的霍夫曼编码，所以每个节点可以由它所处的社群霍夫曼编码和社群内的具体节点的霍夫曼编码联合确定。对于网络中社群结构的检测就转换为使得编码长度最短的编码压缩问题。在这个编码体系中，包括三类码表（code book）。第一个是主码表（master code book），决定每个社群的编码；第二个是传送门码表（exit code book），决定每个社群的出口连边的编码；第三个是节点码表（node code book），决定每个社群内的节点的编码。一旦对网络的社群划分 P

（partition）给定，就存在一个社群码表，一个传送门码表，和  $m$  个节点码表，其中  $m$  是社群的个数。最后，社群划分的目标就是要最小化所有码表的总长<sup>[17]</sup>。由于目标函数为网络中所有随机行走路径的总编码长度，所以在优化目标函数时，算法会倾向于把联系紧密的节点归并到同一个社群中去，这样，最优的社群划分方式会获得最大的编码压缩量<sup>[18]</sup>。

M. Rosvall 和 D. Axelsson 利用数学公式 (the map equation) 表示了上述问题<sup>[19]</sup>。

$$L(M) = q_{\sim} H(Q) + \sum_{i=1}^m p_{\sim}^i H(P^i) \quad 6-8$$

$$H(Q) = - \sum_{i=1}^m \frac{q_{i\sim}}{\sum_{j=1}^m q_{j\sim}} \log \left( \frac{q_{i\sim}}{\sum_{j=1}^m q_{j\sim}} \right) \quad 6-9$$

$$H(P^i) = - \frac{q_{i\sim}}{q_{i\sim} + \sum_{\beta \in i} p_{\beta}} \log \left( \frac{q_{i\sim}}{q_{i\sim} + \sum_{\beta \in i} p_{\beta}} \right) - \sum_{\alpha \in i} \frac{p_{\alpha}}{q_{i\sim} + \sum_{\beta \in i} p_{\beta}} \log \left( \frac{p_{\alpha}}{q_{i\sim} + \sum_{\beta \in i} p_{\beta}} \right) \quad 6-10$$

其中  $M$  即对网络的某种社群划分得到  $m$  个社群。L 是该划分所对应的信息描述长度。 $q_i$  代表进出第  $i$  个社群的概率（先考虑无向网络），因此  $q$  代表社群间跳转的总概率。 $H(Q)$  是社群间跳转行为的香农信息量。 $P_i$  代表第  $i$  个社群内节点间跳转的总概率， $H(P_i)$  是第  $i$  个社群内节点间跳转行为的香农信息量。

在公式的两个部分，信息量都用其被使用的频率进行加权。经过展开化简，可以得到简化形式如下：

$$L(M) = w_{\infty} \log(w_{\infty}) - 2 \sum_{i=1}^m w_{i_{\infty}} \log(w_{i_{\infty}}) - \sum_{\alpha=1}^n w_{\alpha} \log(w_{\alpha}) + \sum_{i=1}^m (w_{i_{\infty}} + w_i) \log(w_{i_{\infty}} + w_i)$$

6-11

通过 Info Map 算法 (C++代码在附件中给出) 我们可以得出社群分类如下：

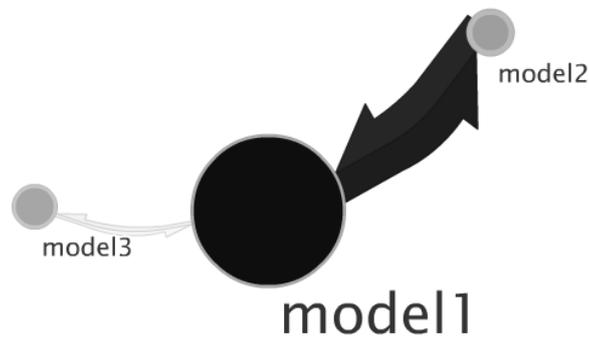


图 6-12 : Info Map 社群聚类结果图

总共分为三个模块，其中模块中节点信息如下表，模块间信息流信息如表，其中节点流量比例 =  $\frac{\text{节点流量}}{\text{网络总流量}}$ 。

表 6-9: 模块 1 内部节点信息

节点名称	节点流量比例 (%)
2	13.8
34	10.7
1	9.98
20	9.86
33	5.15
3	4.15
9	2.48
4	2.42
36	2.42
32	2.36
14	2.3
8	1.98
30	1.69
21	1.59
31	1.53
29	1.51

19	1.5
16	1.27
27	1.23
23	1.21
22	1.09
13	0.884
18	0.873
10	0.865
12	0.85
15	0.762
总流量	84.454

表 6-10: 模块 2 内部节点信息

节点名称	节点流量比例 (%)
24	2.61
28	1.66
35	1.58
25	1.55
26	1.24
总流量	8.64

表 6-11: 模块 3 内部节点信息

节点名称	节点流量比例 (%)
7	1.64
5	1.61
6	1.58
11	1.25
17	0.853
总流量	6.933

表 6-12: 模块间信息流信息

信息流出模块	信息流出模块	流量比 (%)
1	2	3.709
1	3	2.104
2	1	3.594
3	1	2.077

## 6.6 结论

表 6-13: 无权通讯网络分群结果 (基于 FN 算法)

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

表 6-14: 有权通讯网络分群结果 (基于 Info Map 算法)

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

通过比较无权通讯网络分群结果和有权通讯网络分群结果可知, 此问题中权值 (通话时间或通话频率) 对于分类结果影响较大。有权通讯网络分类模块 1 作为中心模块对网络影响较大, 模块 2 与模块 3 通过模块 1 进行信息交流; 模块 1 中节点数目增加, 有效的降低模块之间的耦合性, 同时提高了模块内部紧密型。因此, 在现实的社群网络分类中, 简单地根据网络拓扑结构, 则忽略了用户的主观能动性, 得到的分类结果往往忽略了很多重要信息。

## 7. 模型的推广

本文基于通讯社群进行了三个问题的讨论, 可以推广到实际生活中的诸多网络, 如社交网络、信息传送网络、电力网络等, 这些网络也是由诸多节点构成的相互联系的复杂社群, 可以基于本文提到的分类方法进行分类

## 8. 模型的评价

### 8.1 模型的优点

(1) 在 1、2、3 问中根据题目所给的信息, 本团队将题目中的信息都转化为邻接矩阵表示, 使前三问在形式上统一, 便于采用一种算法求解。

(2) 针对前三问本团队采用了在 G-N 算法上进行改进优化得出的 Fast-Newman 算法进行求解, 这种算法可用于分析节点较多的复杂网络, 思路清晰, 具有很好的普适性和实用性。

(3) 在解决问题的过程中, 本文运用了多种软件极大地简化了分析和求解问题地难度。如采用 Excel 工作表上作业方式, 进行数据处理减少数据的导入、导出; 采用 SPSS 统计软件进行主成分分析, 得到各个因素间的相关系数; 采用 Gephi 进行复杂网络可视化及各种指标分析; 利用 MapEquation 在线可视化软件 Map Generator 对 InfoMap 算法给出的 .map 文件进行可视化。

(4) 模型具有较好的普适性和可移植性, 不仅可用于通讯网络, 还可用于电力系统网络、社交网络等多种网络的分析。

## 8.2 模型存在的不足

(1) 文中所建立的模型都是基于非重叠社群，没有考虑出现重叠区的情况。

(2) 在问题 3 中对于有效信息接收概率阈值的设定具有一定的主观性，且在考虑终止传播概率时，认为每次传播的终止传播概率恒定，而在实际网络中每次传播的终止传播概率可能逐级递减。

(3) 在对问题 5 的多因素进行主成分分析时，我们仅利用了题中所给的信息，没有补充更多信息来验证模型的可行性。

(4) 问题 5 以静态网络为基础进行分析，而实际的网络可能是具有较大可变性的动态网络。

## 9. 模型的改进

本文采用的是基于非重叠社群发现的算法，但本文中的网络可能存在重叠社群，因此可以采用基于重叠社群发现的算法，比如基于二阶段聚类的重叠区发现 (Overlapping Community Detection Based on Single-Pass and Fuzzy Clustering, SPFC)，基于派系过滤的派系过滤算法 (Clique Percolation Method, CPM)<sup>[20]</sup>对本文中的网络进行社群探测。

本问题中的网络为静态网络，网络中的节点不随时间变化。可以通过改进算法或采用其它算法来研究动态网络，例如采用动态网络社群结构探测算法 (Community Structure Detection Algorithm for Dynamic Networks, CDD)。该算法依据节点加入引起模块性指标变化的情况，对网络节点进行社群划分，从而可以发现网络社群结构随时间的动态变化过程。同样也可以实现静态网络的社群结构探测；与 NG 算法和 AN 算法等社群结构探测算法相比，算法的速度快，所获得的模块性指标也基本相当<sup>[21]</sup>。

## 参考文献

- [1]Kernighan B W ,Lin S. An efficient heuristic procedure for portioning graphs[J]. Bell System Techni cal Journal, 1970, 49; 291-307
- [2]Fiedler M.Algebraic connectivity of graphs[J].Czechoslovak Mathematical Journal,1973, 23(98); 298-305.
- [3]Girvan M , Newman M E J . Community structure in social and biological networks[J]. PNAS, 2001, 99(12); 7821-7826
- [4]Newman M E J.Fast algorithm for detecting community structure in networks[J].Physical Review E, 2004, 69(6); 066133
- [5]KERNIGHAN B W,LIN S.An efficient heuristic procedure for partitioning graphs[J].Bell System Technical Journal,1970,49(2):291 - 307.
- [6]Pothena, Simonhd, Lioukp.Partitioning sparse matrices with eigenvectors of Graphs [J]. SIAM J. Matrix Anal. Appl., 1990, 11(3):430 - 452.
- [7]Fiedlerm. Algebraic connectivity of graphs[J].Czechoslovak Mathematical Journal, 1973, 23(2):298 - 305.
- [8]Radicchi F, Castellano C, Cecconi F, et al. Defining and identifying communities in net-works[J].Proc. of the National Academy of Science, 2004, 101(9):2658-2663.
- [9]NEWMAN M E J, GIRVAN M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2):026113.
- [10]Newman M E J, Grivan M. Finding and evaluating community structure in networks[J]. Physical Review E, 2004, 69(2): 026113.
- [11]NEWMAN M. The mathematics of networks[J]. The New Palgrave Encyclopedia of Economics, 2008, 2:1-12.
- [12]DOROGOVTSEV SN, GOLTSEV A V, MENDES J F. k-Core organization of complex networks[J]. Phys Rev Lett, 2006, 96(4):040608.
- [13]Kwak H. Lee C, Park H. Eet al. What is Twitter, a social network or a new media?[C]||Proceedings of the 19<sup>th</sup> international conference on World Wide Web. ACM, 2010:591-600.
- [14]沈珂轶. 社会网络的社群发现与动态特性研究 [ D ]. 上海: 上海交通大学, 2011.
- [15]Bakshy E, Hofman J M, Mason W A, et al. Everyone' s an Influencer:quantifying Influence on Twitter[C]||Proceedings of the Fourth ACM International Conference on Web Search and Data Mining. ACM, 2011:65-74.
- [16]Ulrik Brandes, A Faster Algorithm for Betweenness Centrality, in Journal of Mathematical Sociology 25(2):163-177, (2001).
- [17]<http://wiki.swarma.net/index.php/网络社群划分算法>.
- [18]高华, 姜建国, 社交网络中的团检测算法研究 社交网络中的团检测算法研究 社交网络中的团检测算法研究, 二〇一三年四月十九日, 西安电子科技大学。
- [19]M. Rosvall and D. Axelsson, The map equation, Department of Physics, Umea University, SE-901 87 Ume a, Sweden, February 2, 2014.

- [20] 蒋盛益, 杨博泓, 李敏敏, 吴美玲, 王连喜. 基于二阶段聚类的重叠社群发现算法[J]. 模式识别与人工智能, 2015, (11):983-991.
- [21] DU Haifeng, YUE Zhongshan, LI Shuzhuo, 等. Community structure detecting algorithm for dynamic networks based on modularity 基于模块性指标的动态网络社群结构探测方法[J]. 系统工程理论与实践, 2009, 29(3):162-171.

## 附录

### 附录 1: Fast Newman 算法 MATLAB 程序

```
function [Z H] = FastNewman(adjacent_matrix)
% FastNewman 算法实现社群发现
% 该算法参见文献《Fast algorithm for detecting community structure in networks》
(2003)
% 输入
% adjacent_matrix - 邻接矩阵
% 输出
% Z - n-1*3 矩阵, 第 i 行表示第 i 次合并, 第 1 列和第 2 列表示合并的社群标号, 第 3 列是
合并后的模块度
% H - 聚类树图的句柄

n = size(adjacent_matrix,1); % 节点数目
max_id = n;
Z = [];
clusters = [1:n; zeros(1,n); 1:n]; % 初始划分, 第 1 行是节点标号, 第 2 行是社群标号
的变换, 第 3 行是社群标号
step = 1;
while numel(unique(clusters(3,:))) ~= 1
    [Q e a clusters] = GetModularity(adjacent_matrix, clusters);
    k = size(e,1); % 社群数目
    DeltaQs = [];
    for i = 1:size(e,1)
        for j = 1:size(e,1)
            if i ~= j
                DeltaQ = 2*(e(i,j)-a(i)*a(j));
                DeltaQs = [DeltaQs [i;j;DeltaQ]];
            end
        end
    end
    [maxDeltaQ, id] = max(DeltaQs(3,:));
    id = id(1);
    i = DeltaQs(1,id);
    j = DeltaQs(2,id); % 找出 DeltaQ 最大的社群对的序号
    max_id = max_id + 1;
    c_id1 = find(clusters(2,:) == i);
    c_id2 = find(clusters(2,:) == j);
    id1 = unique(clusters(3,c_id1));
    id2 = unique(clusters(3,c_id2));
    clusters(3,c_id1) = max_id;
    clusters(3,c_id2) = max_id;
    Z = [Z; [id1 id2 length([c_id1 c_id2])]];
    step = step + 1;
end
if nargin == 2
    H = dendrogram(Z,0);
end
end

function [Q e a out_clusters] = GetModularity(adjacent_matrix, clusters)
%% 获取无向无权网络的某种划分的模块度
%% 模块度的提出和计算方法见如下文献:
% 《Finding and evaluating community structure in network》(2004)
% 输入
```

```

% adjacent_matrix - 网络的邻接矩阵
% clusters - 网络的社群划分, 第1行是节点标号, 第2行是社群标号的变换, 第3行是社群标号
% 输出
% Q - 划分的模块度
% e - k*k 矩阵, k 是社群数目, 其中元素的定义见上述文献
% a - 意义见上述定义
% out_clusters - 与输入含义相同
uni = unique(clusters(3,:));
for i = 1:length(uni)
    idices = find(clusters(3,:) == uni(i));
    clusters(2,idices) = i;
end
m = sum(sum(adjacent_matrix))/2; % 网络的边的数目
Q = 0 ;
k = numel(unique(clusters(2,:))); % 社群数目
e = zeros(k);
for i=1:k
    idx = find(clusters(2,)==i);
    labels_i = clusters(1,idx);
    for j=1:k
        idx = find(clusters(2,)==j);
        labels_j = clusters(1,idx);
        for ii=1:length(labels_i)
            vi = labels_i(ii);
            for jj=1:length(labels_j)
                vj = labels_j(jj);
                e(i,j) = e(i,j)+adjacent_matrix(vi,vj);
            end
        end
    end
end
end
e = e/2;
e = e/m;
a = [];
for i=1:k
    ai = sum(e(i,:));
    a = [a; ai];
    Q = Q + e(i,i)-ai^2;
end
out_clusters = clusters;
end

clear all
clc
adjacent_matrix = load('.\dataset\test.dat');
[Z, H] = FastNewman(adjacent_matrix);

```

## 附录 2：问题 3 邻接矩阵

编号	姓名	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36		
1	毕婕靖	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0	0		
2	蔡月	1	0	1	1	0	0	0	1	0	0	0	0	0	1	0	1	0	1	1	1	1	1	1	0	0	0	1	0	1	0	1	0	1	1	1	0	0	
3	曾帅	1	1	0	1	0	0	0	1	1	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	
4	陈澜	1	1	1	0	0	0	1	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
5	高淼	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
6	康芸晴	1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
7	柯雅芸	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
8	李嘉俊	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
9	梁茵	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0	1		
10	廖颜翠	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0		
11	林礼琴	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
12	凌慧雯	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
13	刘菁芸	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
14	刘松荷	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1		
15	柳渭	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	
16	陆盈	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
17	骆娴	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
18	孟芳	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
19	潘澜巧	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
20	潘立	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
21	彭荃	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0	
22	孙锦	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
23	孙怡毓	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
24	孙翼茜	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	1	0	0	1	1	1	0	0	
25	童豫	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1	0	0	
26	涂蕴知	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0	0	0	0	0	
27	王蕴姣	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	
28	文芝	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	1	0	0	
29	吴霄	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
30	吴宇晓	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	1	1	0	0	
31	谢斑尚	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0
32	易贞	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	1	0	0	0	1	1	0	0	0	
33	张荆	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0
34	张培芸	0	1	0	0	0	0	0	0	1	1	0	0	0	1	1	1	0	0	1	1	1	0	1	1	0	0	1	1	1	1	1	1	1	1	0	0	0	0
35	张庭琪	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0
36	钟倩	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0