

# 2017 年第十届华中地区大学生数学建模邀请赛经典赛题目

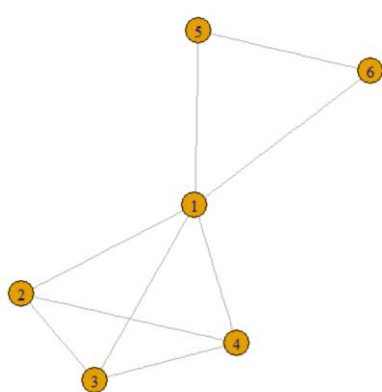
(请先阅读“第十届华中地区大学生数学建模邀请赛论文格式规范”)

## B 题 基于通讯数据的社群聚类

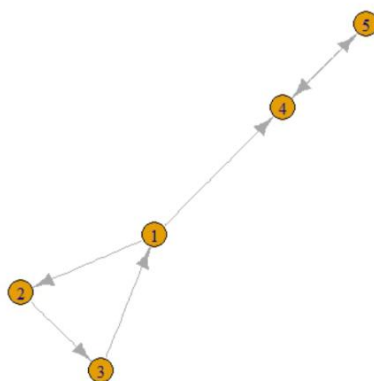
大数据时代的来临使得许多不可能成为了现实。数据分析和数据挖掘技术成功地在多个重大领域取得了巨大成功。现有部分人群的通讯数据，试对人群进行社群分类和相关识别。

社群分类的研究中，往往先将数据表示为图的形式。

1、图一中，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。图一为无向图。试就图一信息，区分个体差异，对个体进行初步识别，实现分群。



图一



图二

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、附件一中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  $10\% \sim 50\%$  的终止传播概率，试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。

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。

特别地，请附上相应的参考文献。