附件材料说明

# 附件1：实验数据

Step1：论文实验原数据（数据1、2）收集的是胆系癌症QQ群在线健康社区真实聊天语料。通过两台设备进行的数据收集，数据收集过程中有数据遗漏，因此在进行论文实验时，分别对两个数据文件分别进行处理，经数据处理后，将两者进行合并、去重，最后成标准数据文件（数据3），数据日期是从2020.07.01-2020.12.31日。

Step2：针对（数据3）中的聊天语料进行情感极性分析和健康信息主题分析。

Step3：针对（数据3）中用户的source(id)、target(id)分别提取节点文件（数据4）、边文件（数据5）。

Step4：结合（数据3）中的情感值和健康信息主题和（数据4、数据5）中的节点和边，完成节点指标统计（数据6）。

Step5：根据（数据6）完成节点属性指标的计算（数据7）。

Step6：根据（数据7）完成AttriRank算法的节点文件（数据8）和边文件（数据9）的构建。

Step7：根据（数据8、9）完成AttriRank算法，最后得到d=0.85时，节点的得分文件（数据10）。

Step8：根据（数据10）将节点按照得分大小进行降序排列， 并根据论文中的实验结果（图片）进行数据分析和汇总（数据11），如表1所示。

表1实验结果分析数据汇总

| 序号 | 图片编号 | 统计指标 |
| --- | --- | --- |
| 1 | 图2 | 网络效率 |
| 2 | 图3 | 交互率、互惠率 |
| 3 | 图4 | 覆盖率 |
| 4 | 图5 | 移除节点后情绪倾向度 |
| 5 | 图6 | 活跃度、活跃度/活跃度均值；互惠度、互惠度/互惠度均值 |
| 6 | 图7 | 活跃度、互惠度 |
| 7 | 图8 | 多样性、多样性/多样性均值；均质性、均质性/均质性均值 |
| 8 | 图9 | 多样性、均质性 |
| 9 | 图10 | 倾向度、倾向度/倾向度均值 |

# 附件2：核心程序

1.AttriRank.py是AttriRank算法的核心程序。

2.main.py是AttriRank算法的调用程序。

3.LDA模型.py是健康信息主题识别程序。

4.困惑度.py是判断健康信息主题个数的程序。

5.情感极性分析.py是计算健康信息情感值的程序。

6.网络效率.py是计算关键用户个数的程序。

7.核心程序及源代码文档。

**程序代码建议用python编译器、Notepad++打开，或者查看word文档。**

**详细工作请查看论文《多维属性融合视角下的在线健康社区关键用户识别研究》**