

# 大数据分析

2019 秋

## Homework #3

数据 yelp.inject.2000 评论中随机选了 2000 个账号作为受控账号，对 200 个目标饭店进行了虚假好评的“攻击”。请用**至少两种**课上讲过的大图挖掘的方法（允许其中一个是未讲过的算法）检测目标用户和目标饭店，可以利用网上公开代码和资源。

**数据地址：**

[yelp.inject.2000] <https://github.com/shenghua-liu/HoloScope/blob/master/testdata/yelp.inject.2000.edgelist.gz>

格式: userid objid 1

1 表示有一次好评连接

**提交需求：**【注意格式要求】

1. 提交 **gz 压缩**的用户文件、目标文件，

命名规则：[学号].[算法].userlist.gz, [学号].[算法].objlist.gz

输出文件的格式是：id list 用换行符隔开

提示：1) 如果利用 HoloScope 的话，可以将 objid 输出的可疑分值按照从大到小排列，取前 200 个。

2) 由于评分是利用程序评分，因此务必输出格式正确。至少一种方法检测的 F1 score 在 0.9 以上的作业方可获得满分。

2. 单独压缩文件提交代码实现（只提交自己实现代码部分，包括包引用，调用公开代码部分不提交。如果是完全自己实现，则提交全部代码）。

**参考资源** [若好用记得给星]：

[1] SVDs homework#1

[2] HoloScope: <https://github.com/shenghua-liu/HoloScope>

[3] Fraudar: <https://bhooi.github.io/projects/fraudar/index.html>

[4] SparTAN2: <https://github.com/shenghua-liu/spartan2>