基于图的防欺诈分析

---- *极意网络 开发总监 洪晓龙*

极意成立

2012

- 我们开始有了团队，核心理念“为世所想，为世所用”正式开创验证安全 2.0 新纪元

极验2.0发布

2012

- 仅凭口碑就已有近3500家中小型网站主

天使投资

2013

- 获得杭州天使湾数十万投资

红杉投资

2015

- 获得 IDG 资本的数百万美元投资

2016

- 由红杉资本领投、IDG 跟投的 2400 万美元 B 轮融资

极验3.0发布

2016

- 懂科技更懂人性，用户量达到 16万

安全平台

2017

- Test-button解决方案
- One pass解决方案
- G-guard解决方案


SIIDC
安畅网络
爱屋吉屋
OKLink
iRanShao
GUERLAIN
PARIS
医界贷
YIJIEDAI
d.cn当乐网
掌阅
石墨文档
very
OHOME 橙家
咪咕阅读
锁板街
人众金服
RENZHONG FINANCIAL
六间房
6.cn
虎嗅
HUXIU.COM
汽车之家
autohome.com.cn
365易贷
59store
问卷星
华住酒店集团
加盟商服务管理平台
West Air
网金社
中山大学
SUN YAT-SEN UNIVERSITY
银联商务
ChinaUMS
战旗TV
agoda
36Kr
艾瑞网
都市丽人
火猫TV
HuoMao TV
PANDA.TV
airbnb
中国東方航空
CHINA EASTERN
Suntime Wealth
朝阳潮富
易企秀
EQXIU.COM
丁香园
WWW.DXY.CN
MEIZU
兑吧
ZCOOL站酷
你我贷
niwodai.com
红岭创投
RONGLING CAPITAL
新浪SHOW
bilibili
www.bilibili.tv
汽车超人
qccr.com
点融网
Dianrong.com
babytree
宝宝树
PPmoney
ZEALER
科技生活第一站
HONGKONG AIRLINES
香港航空
italki
龙珠直播
longzhu.com
CH.com春秋
长江证券
CHANGJIANG SECURITIES
OKCoin 币行
PPTV
微拍堂
欧浦钢网
opsteel.cn
GTOKEN
AIR CHINA
斗鱼
DOUYU.COM

- 5年
- 7亿次
- 22万家

背景

随着互联网的发展，人们的生活越来越方便，但是网络黑产也跟着快速发展，使得大多数互联网企业遭受到各种攻击：

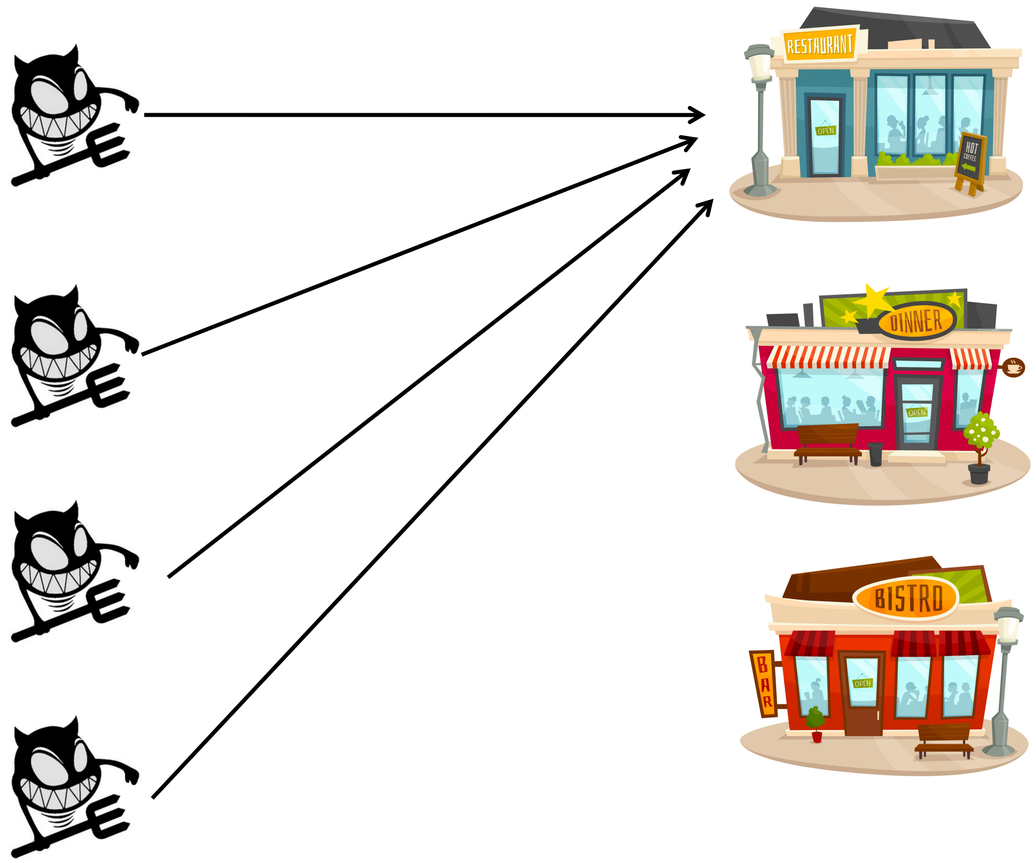
- 羊毛党
- 恶意爬虫
- 恶意刷单
- 等等

如何发现这些攻击者？

- 频繁项挖掘
- 聚类
- 黑名单

攻击者特点

- 目的性强
- 流量大



攻击者成本

- 人力成本
- 技术成本
- 资源成本

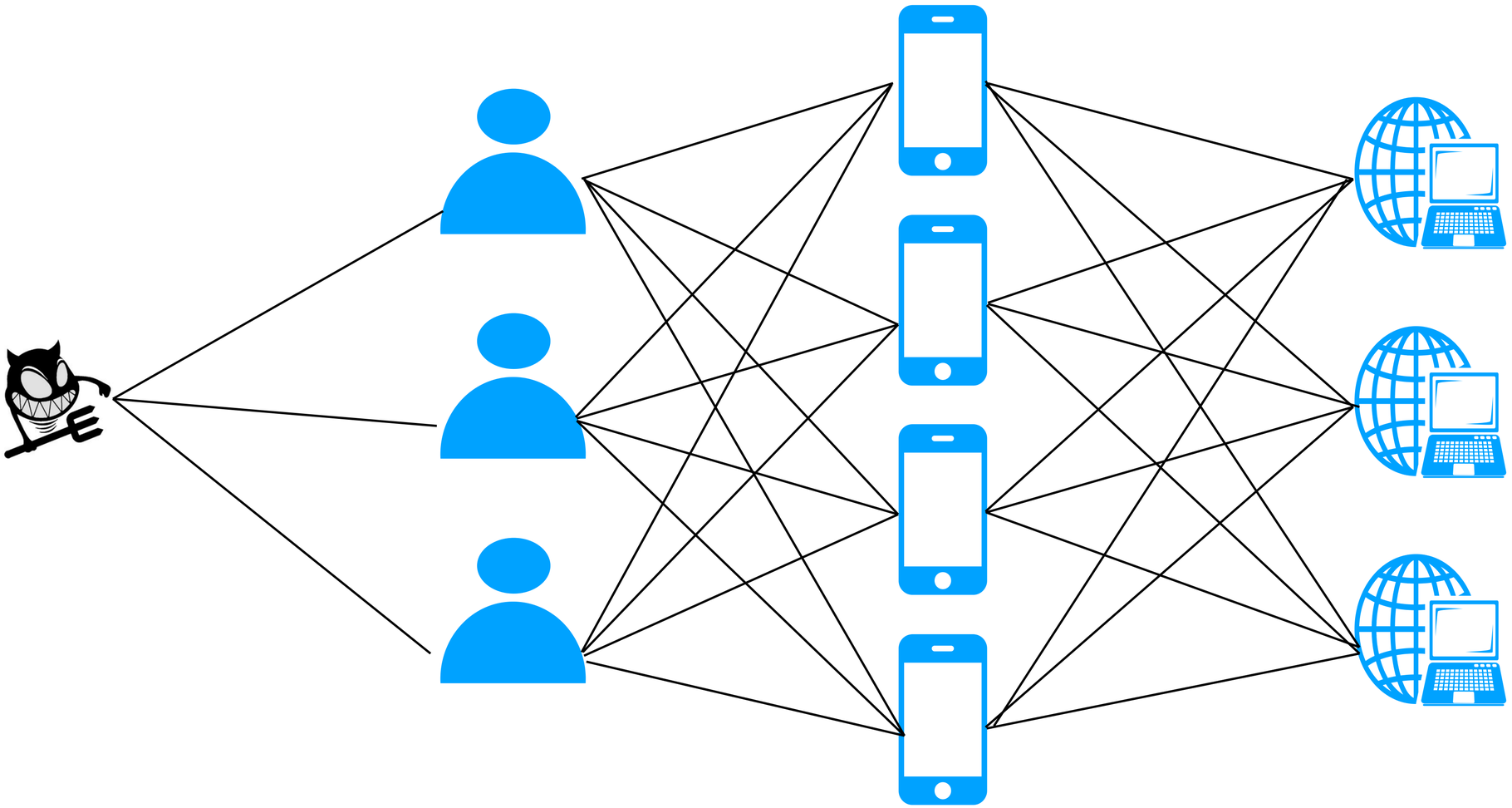
账户/手机号

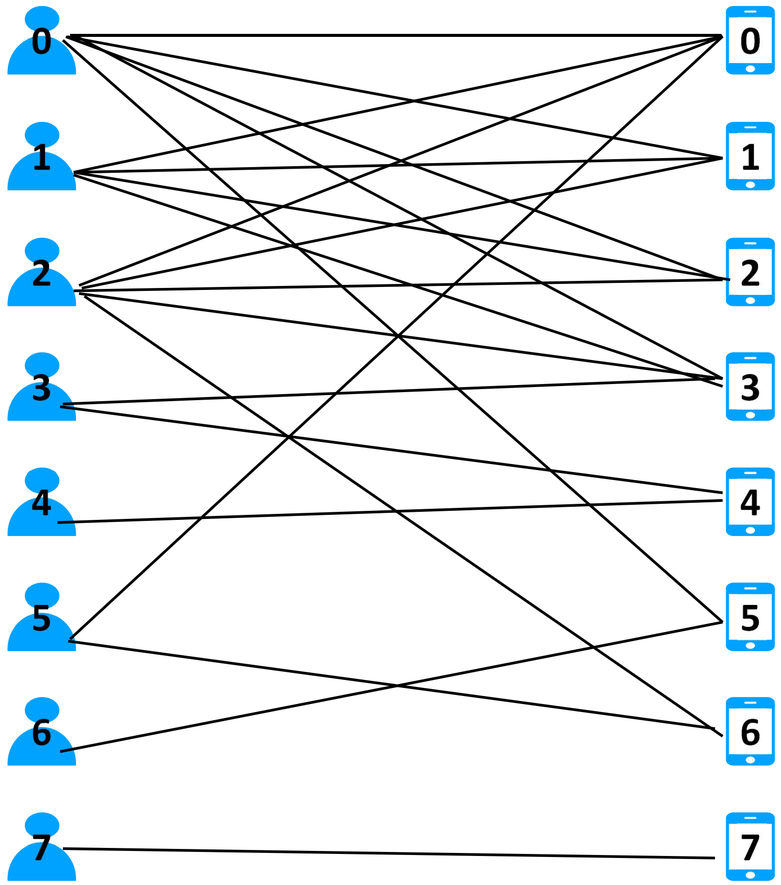
设备

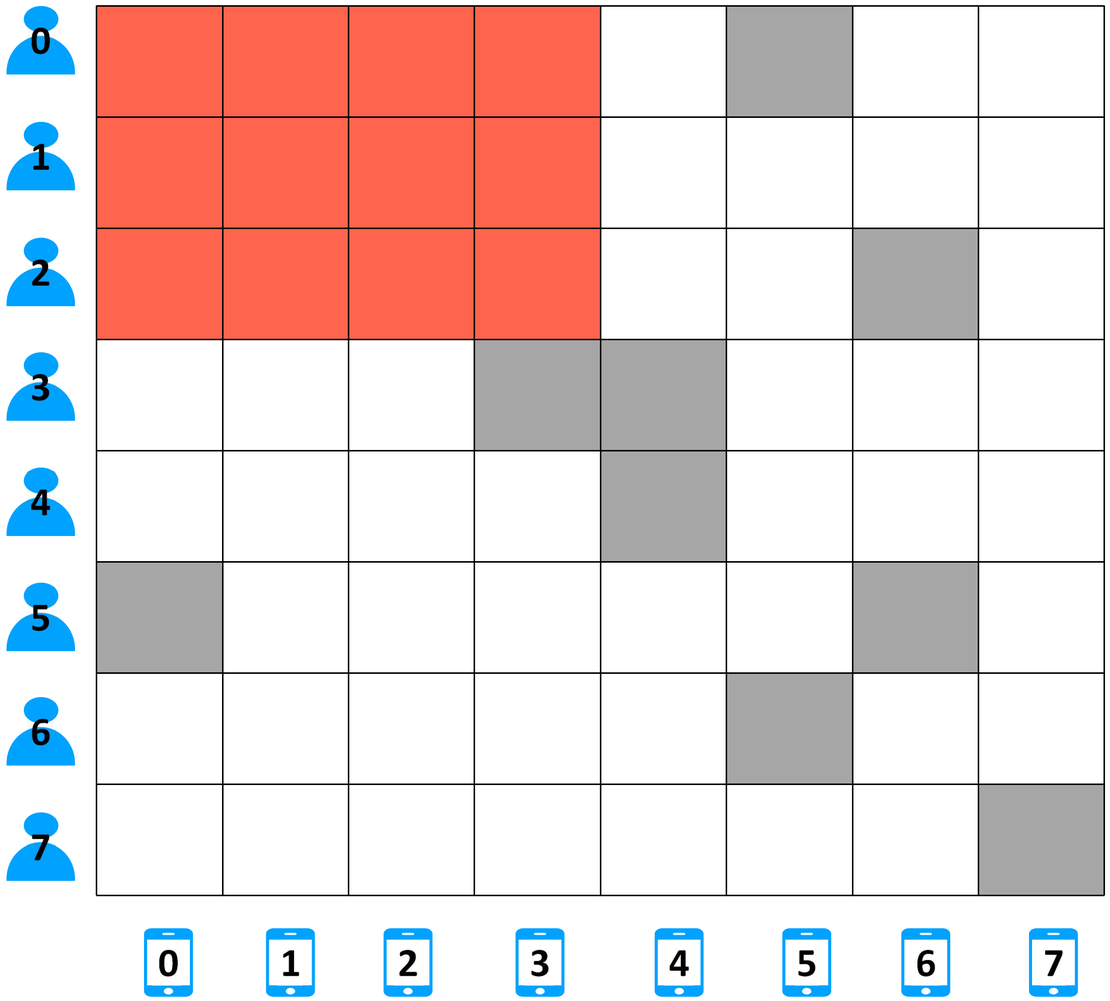
IP

资源成本

攻击者手里的资源（账号/手机号，设备，IP等）有限，必须重复组合地去利用。




0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7


0
1
2
3
4
5
6
7
0
1
2
3
4
5
6
7

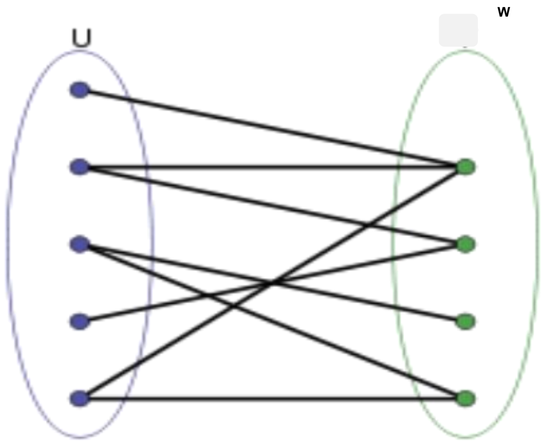
一些尝试

- 聚类
- 社群发现

建模

二部图

- 设G=(*V*,E) 是一个无向图，如果顶点*V*可分割为两个互不相交的子集 (U,W)，并且图中的每条边 (i, j)所关联的两个顶点 i 和 j 分别属于这两个不同的顶点集 (i in U, j in W)，则称图 G 为一个二部图。

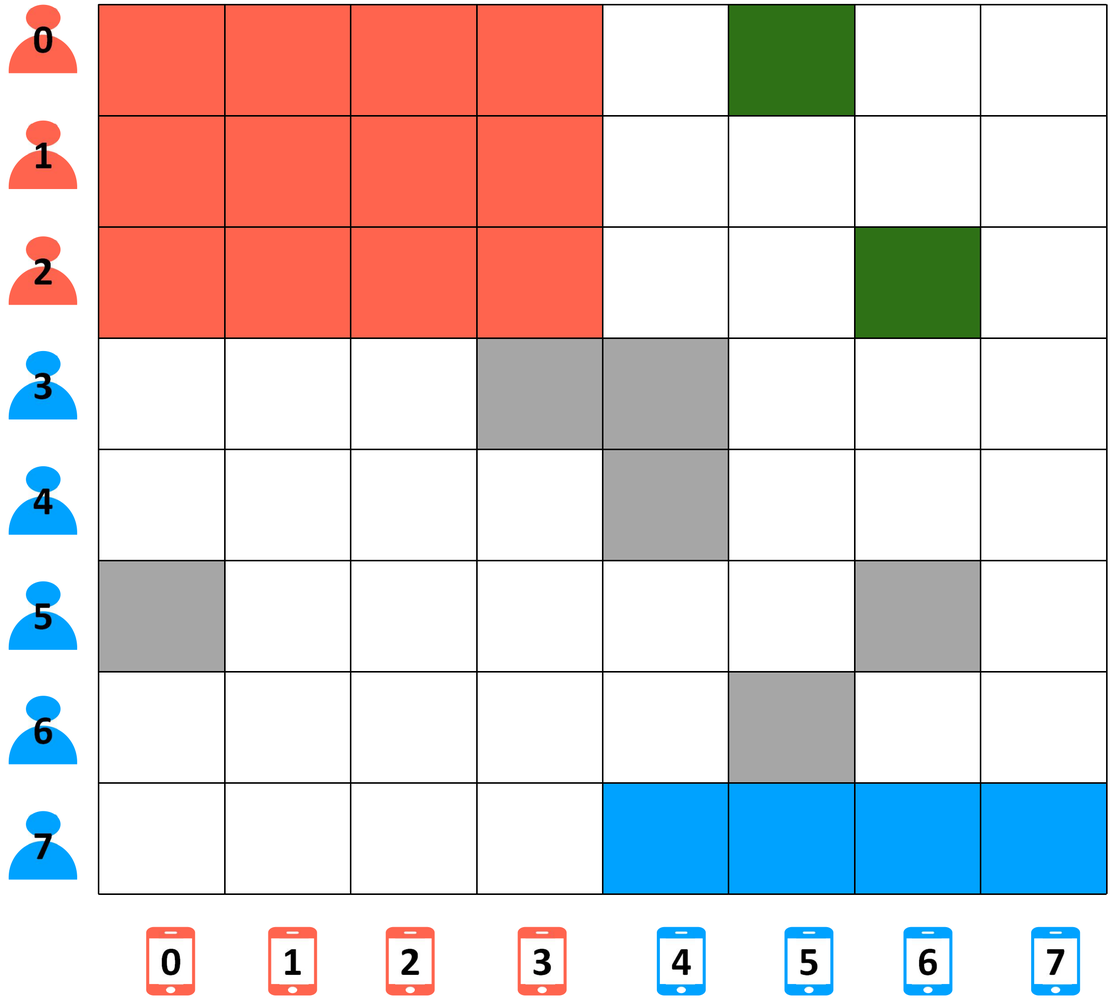


- 在我们的实验中，二部图的节点有两类：一类是ip，另一类是UA。当ip使用某个UA进行验证时，便会在代表ip的节点和代表UA的节点之间构建一条边。

如何求得高密度子图？

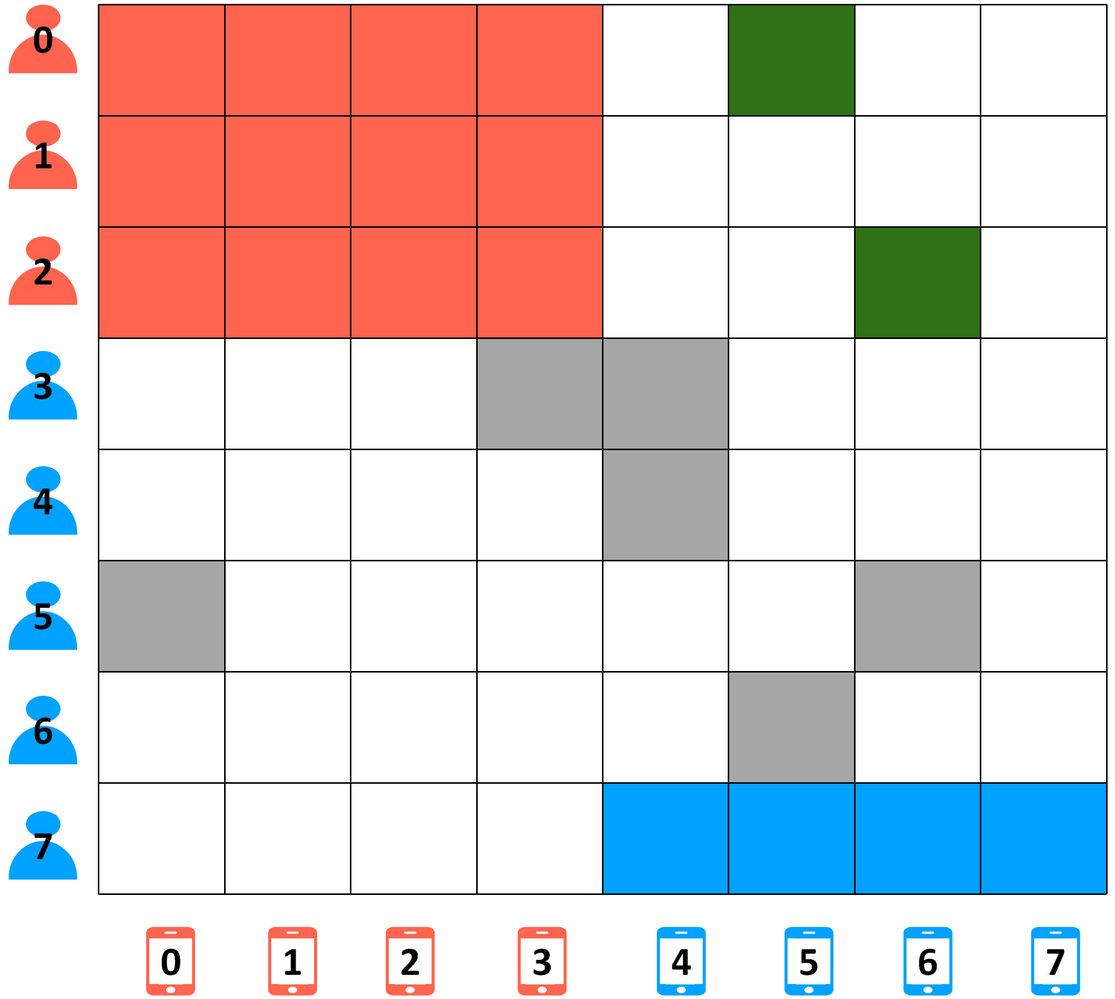
正向求解

- 暴力枚举所有子图
- 遍历联通分量求每个分量密度
- 直接选取度大的点



逆向求解

每次选择度最小的点删除



算法思想

算法的本质就是**贪心算法**查找**高密子图**的过程

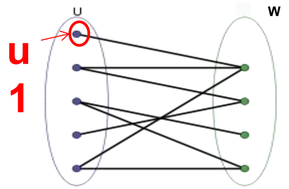
高密子图度量标准：g(S) = f(S)/|S|

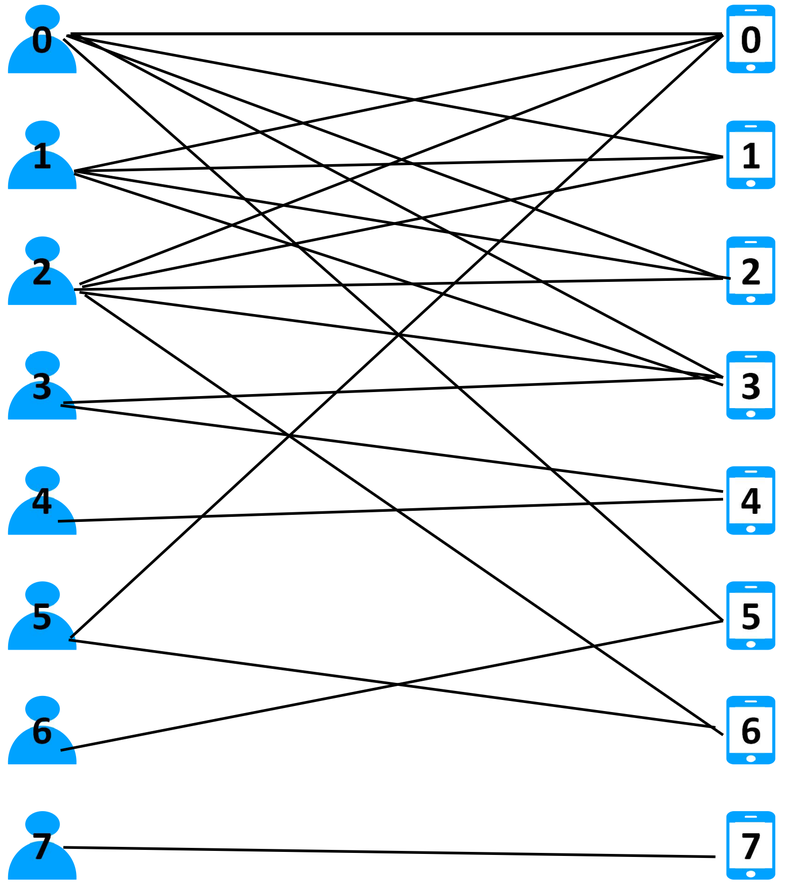
其中fs为子图中边的条数之和，|S| 为子图的顶点集中顶点的个数

贪心策略：选取使得度量函数g(S)最大的子集S

迭代：对顶点集*V*，每次删除度最小的节点ui，对剩下的集合

{V \ ui }计算g({V \ ui })





算法优化

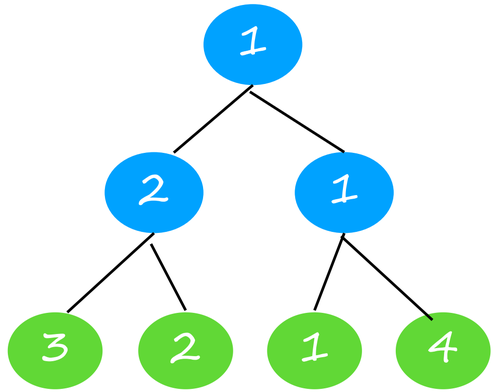
复杂度O(N*N)

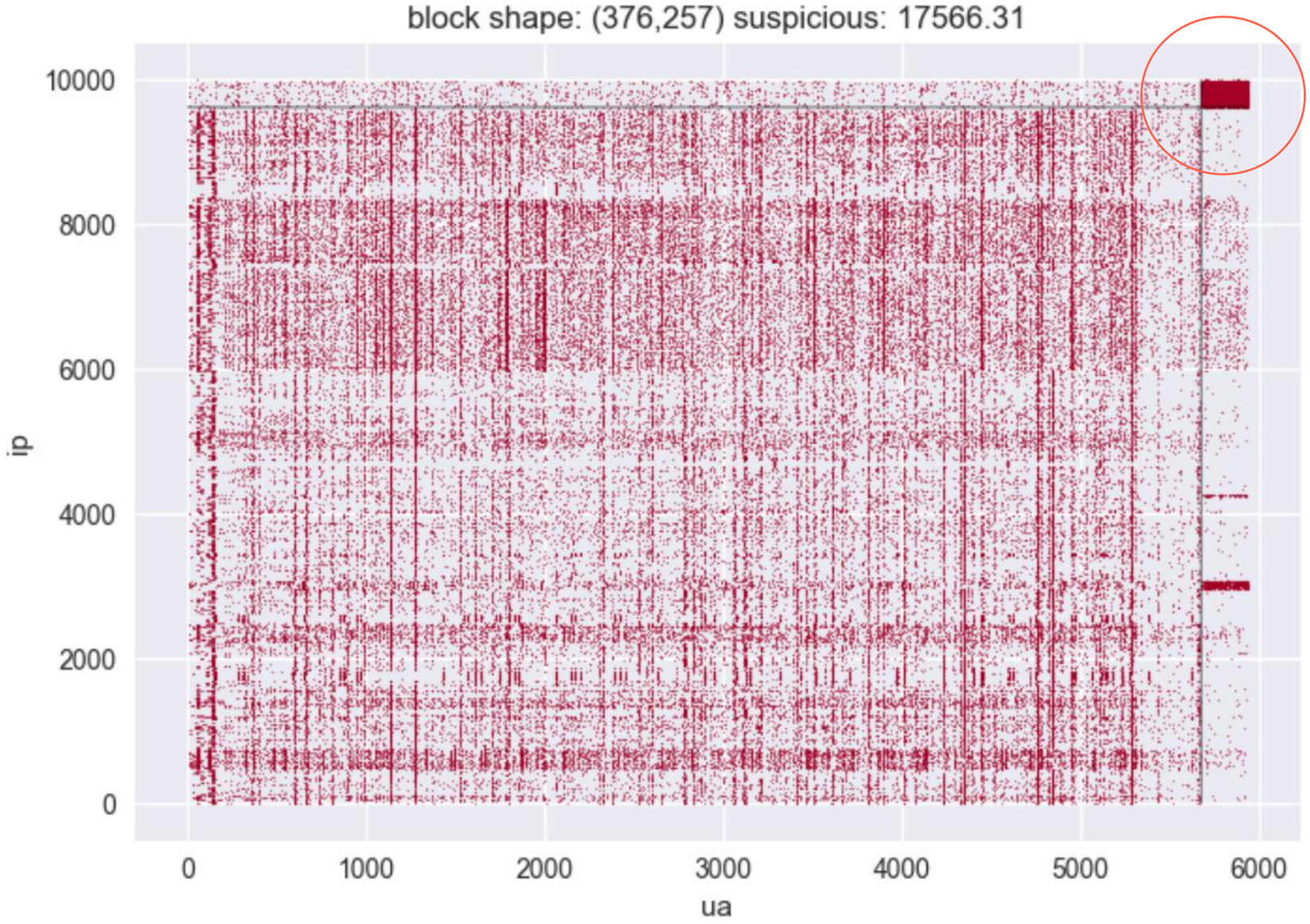
如何去优化？

小根堆优化

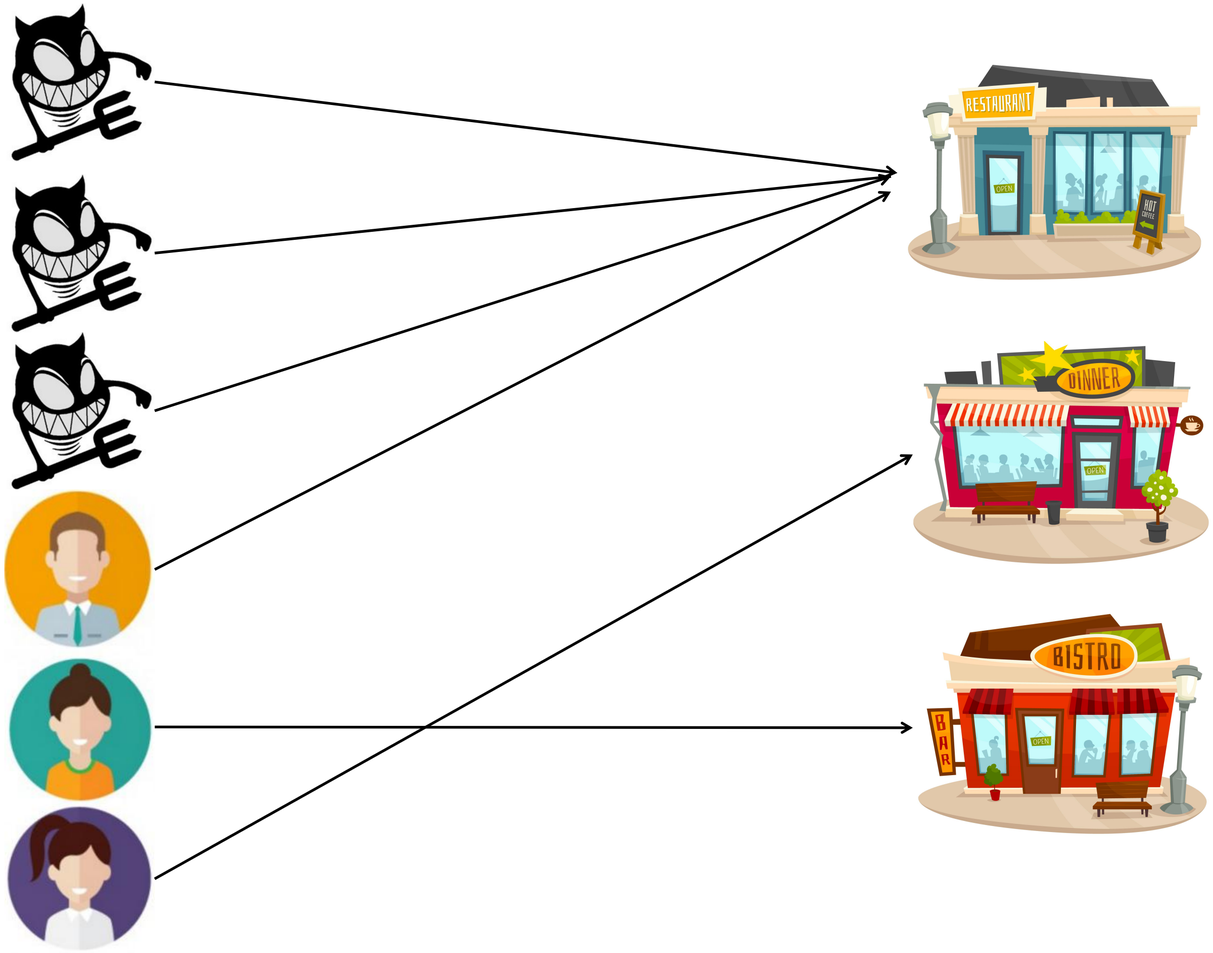
优先树

优化后的复杂度O(N*LOG N)




block shape: (376,257) suspicious: 17566.31
10000
8000
6000
4000
2000
0
ip
0
1000
2000
3000
4000
5000
6000
ua

GeeGuard



F&Q