

# 华为云GES服务图引擎中的实时推荐算法

参加了华为EI学习训练营，记录在华为云GES中实施推荐算法参数的实验。

## GES实时推荐算法学习

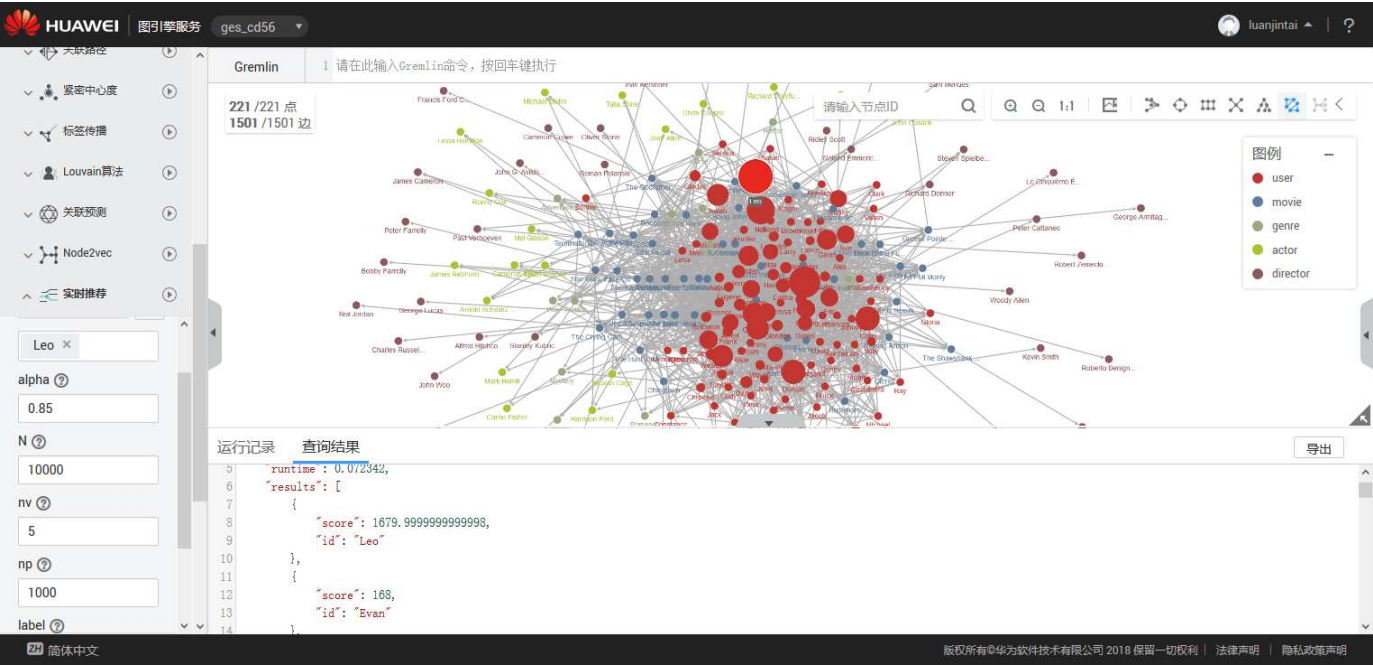
使用提供的training-data数据，其中包括  
算法包括的参数有sources,alpha,N,nv,np,label,directed，下面分参数进行分析（未提到的参数皆设置为默认值）：

### sources

节点ID，最大数量为30，可以选择各种类型，在本例中，例如导演、电影等。

### 实验

以sources=Leo，label=user为例，为Leo推荐用户：

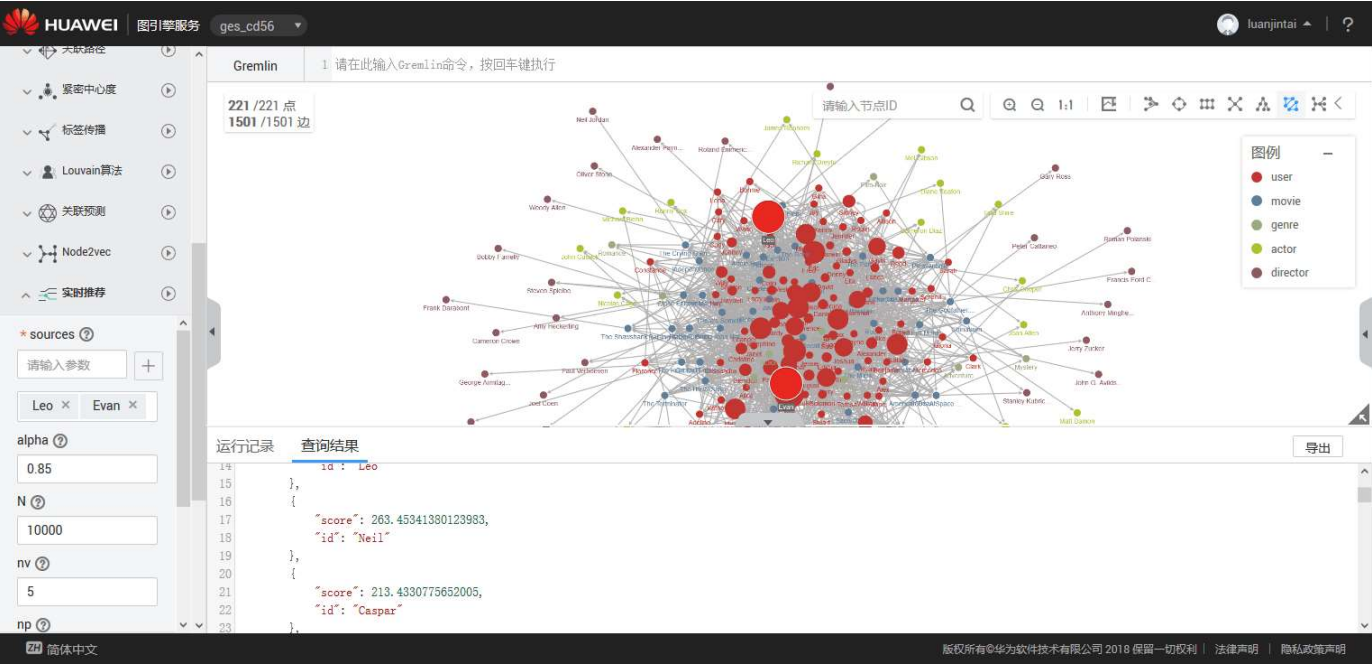


```
{
  "sources": [
    "Leo"
  ],
  "runtime": 0.072342,
  "results": [
    {
      "score": 1679.9999999999998,
      "id": "Leo"
    },
    {
      "score": 168,
      "id": "Evan"
    }
  ]
}
```

```
    },
    {
      "score": 151,
      "id": "Neil"
    },
  ],
}
```

可发现与Leo距离最近的用户为Evan，得分为168，第二为Neil，得分151：

在ID中加入Evan，为Leo和Evan同时推荐用户，运行时间基本没变，Neil的得分增加为263：



结论

- 实时推荐系统基于随机游走算法，score得分的计算可能基于用户的信任关系以及各节点的相似度计算，得分越高说明相关度越高。
- 选择不同的source可以达到不同的推荐效果，sources可以选择类型不同的节点，例如用户、电影名称、导演名称等，可以根据用户喜好和用户路径为其推荐任意label的节点。

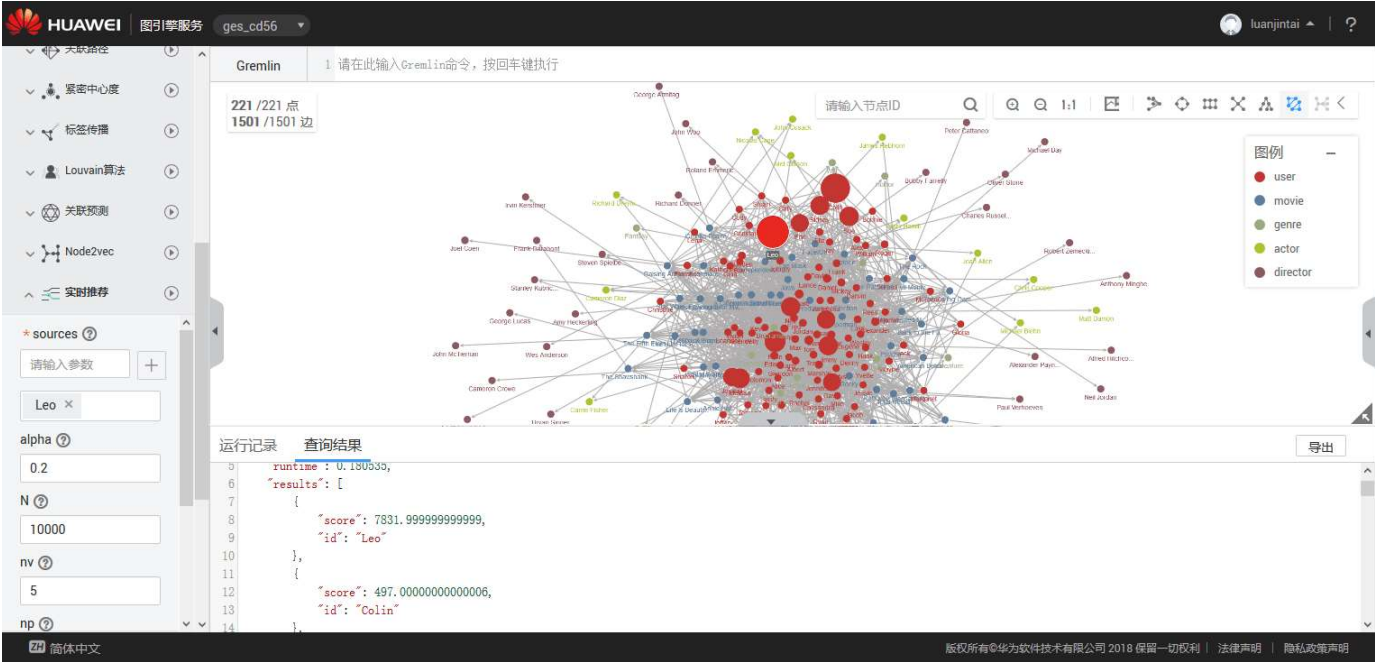
alpha

权重系数，0-1，系数越大，步长越长。

实验

固定其他变量，调整alpha，注意观测运行时间和查询结果。

alpha = 0.2:



结论

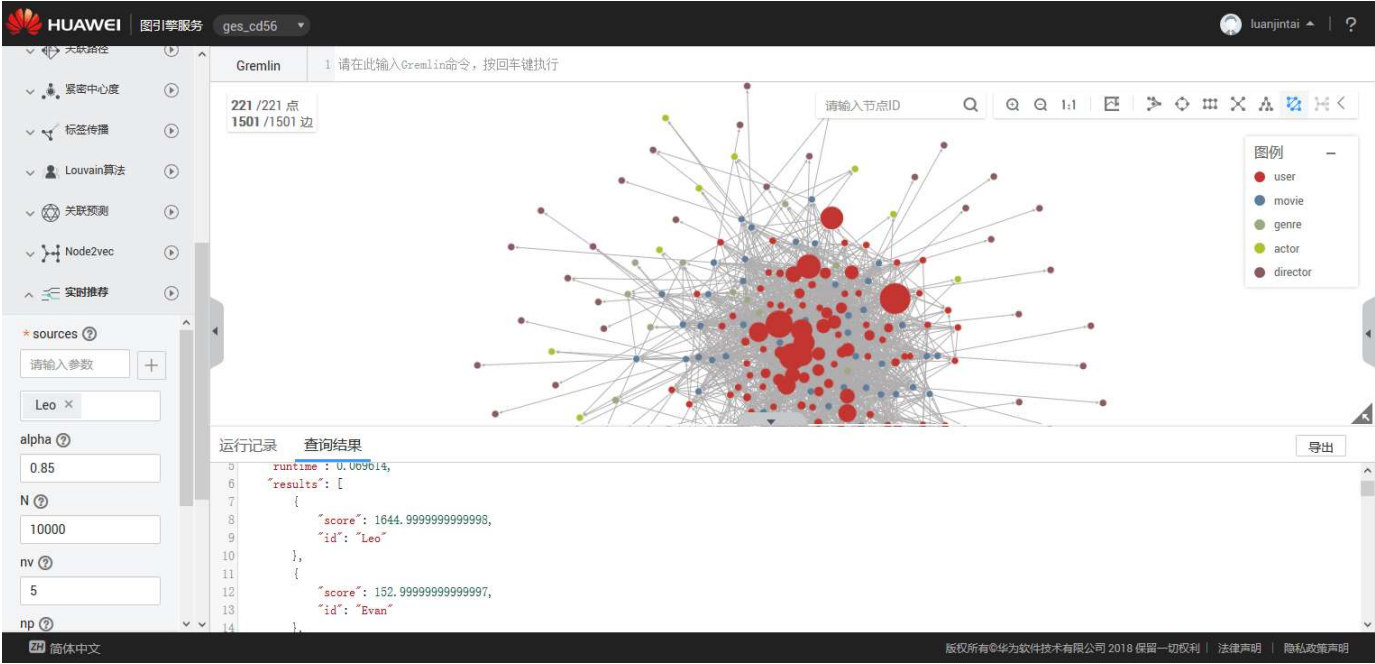
- alpha越大，步长越长，算法运行时间越短，合理的增加步长有利于减少计算资源占用，缩短预测时间；
- alpha越大，Leo的score越低，并且出现不同的推荐结果，理论上，步长越小，结果越精确。

N

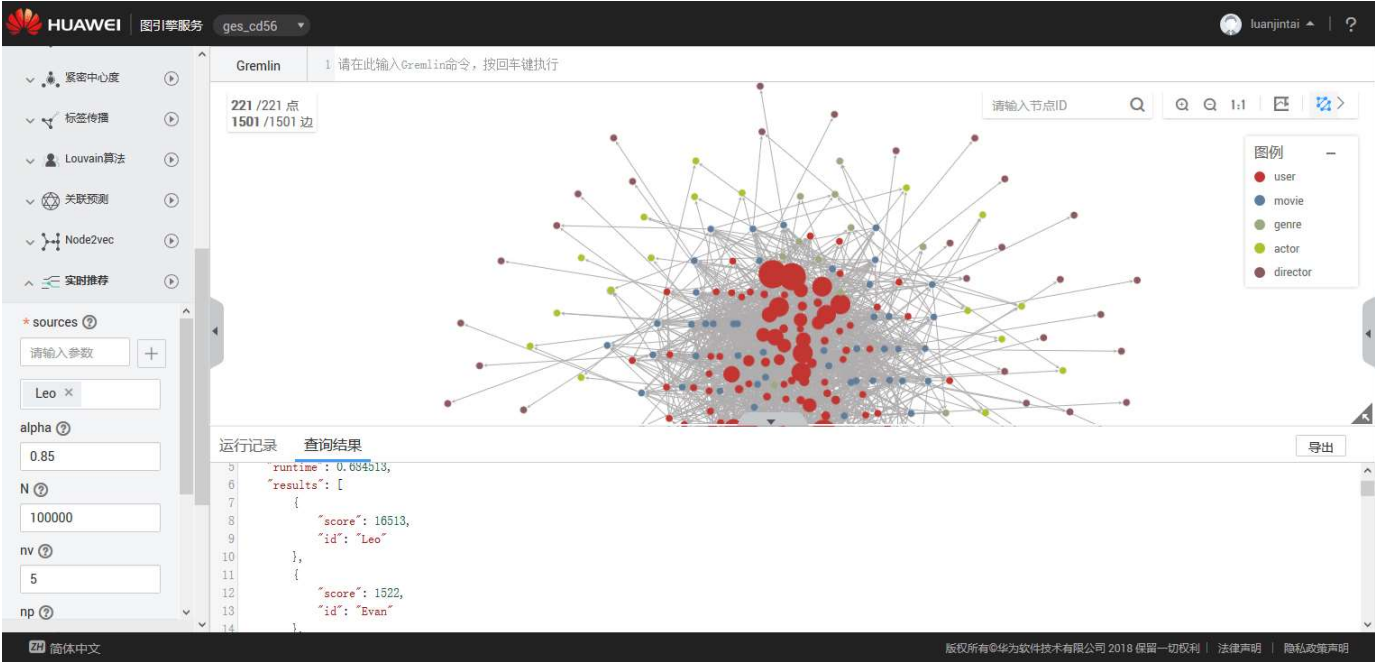
游走步数，[1,200000]。

实验

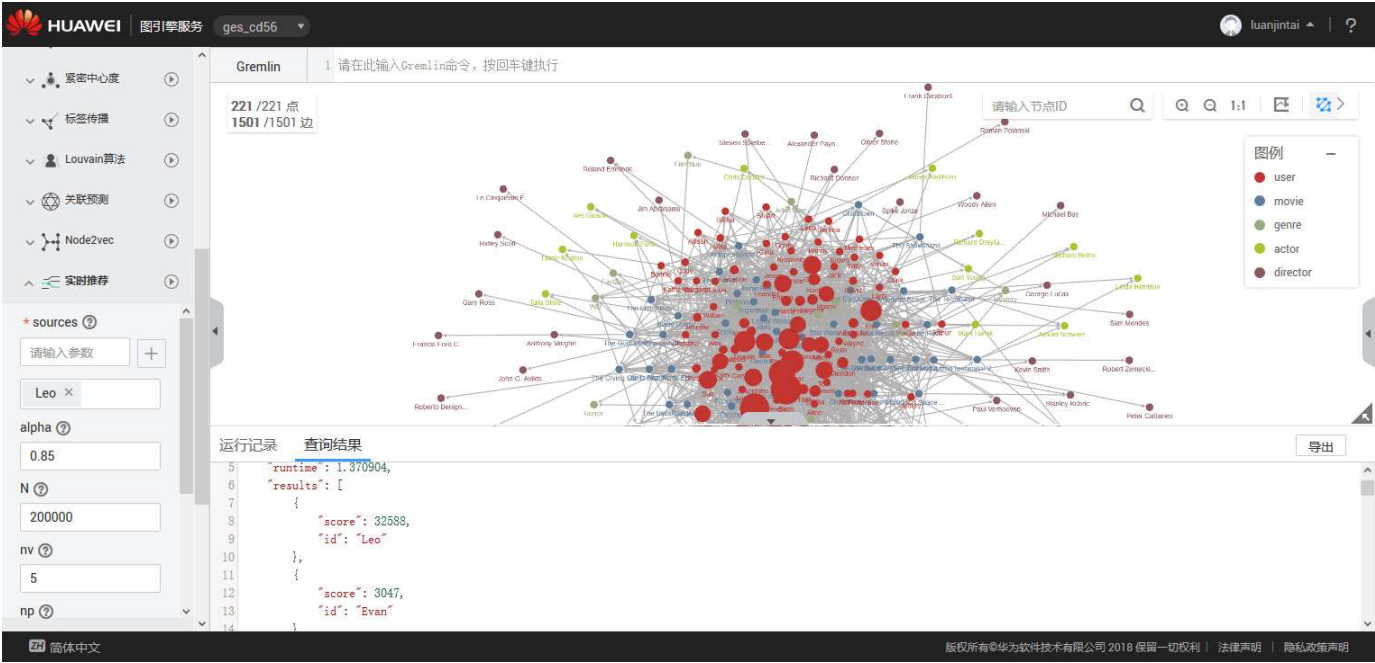
N = 10000:



N = 100000:



N = 200000:



结论

随着游走步数N的增加，算法运行时长增加，节点score增加，可以猜测，对于固定边和点的图，当N达到一定步数时，即得到全局最优解，合理的设置游走步数N可以节约计算资源。

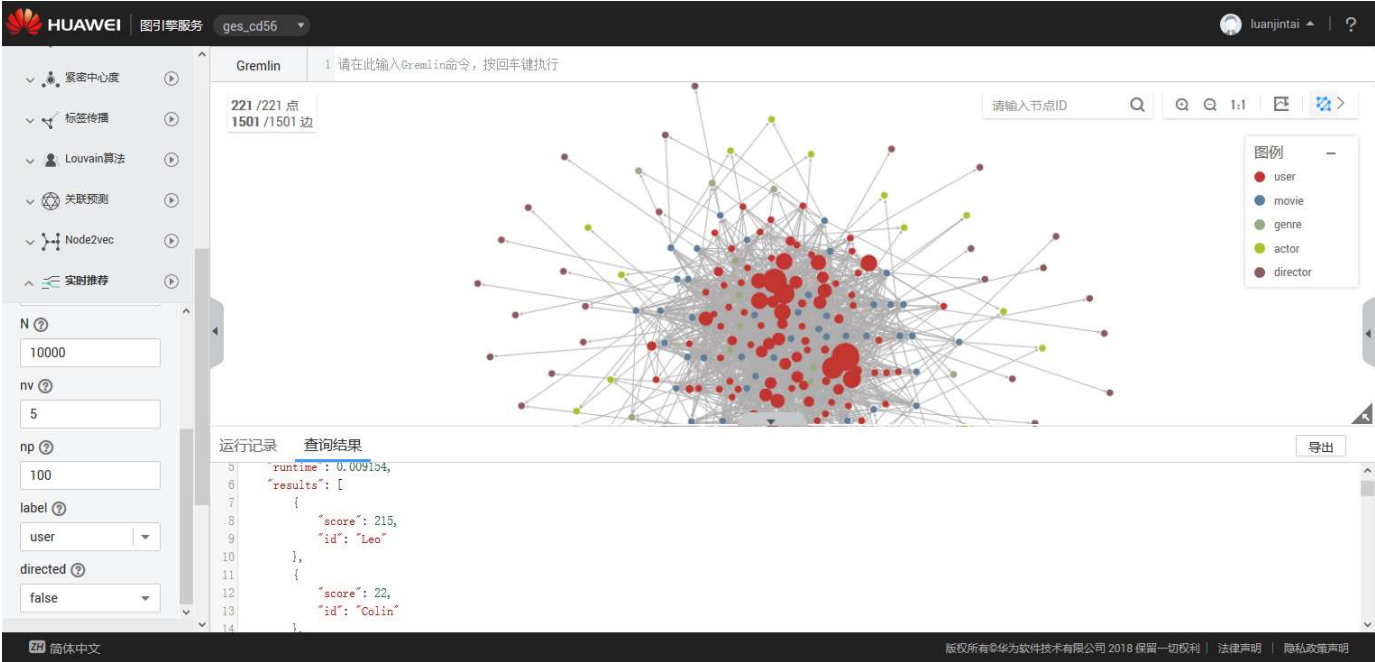
nv np

随机游走提前结束参数，当候选节点被访问次数达到nv或者候选节点个数达到np时，候选节点加入推荐。

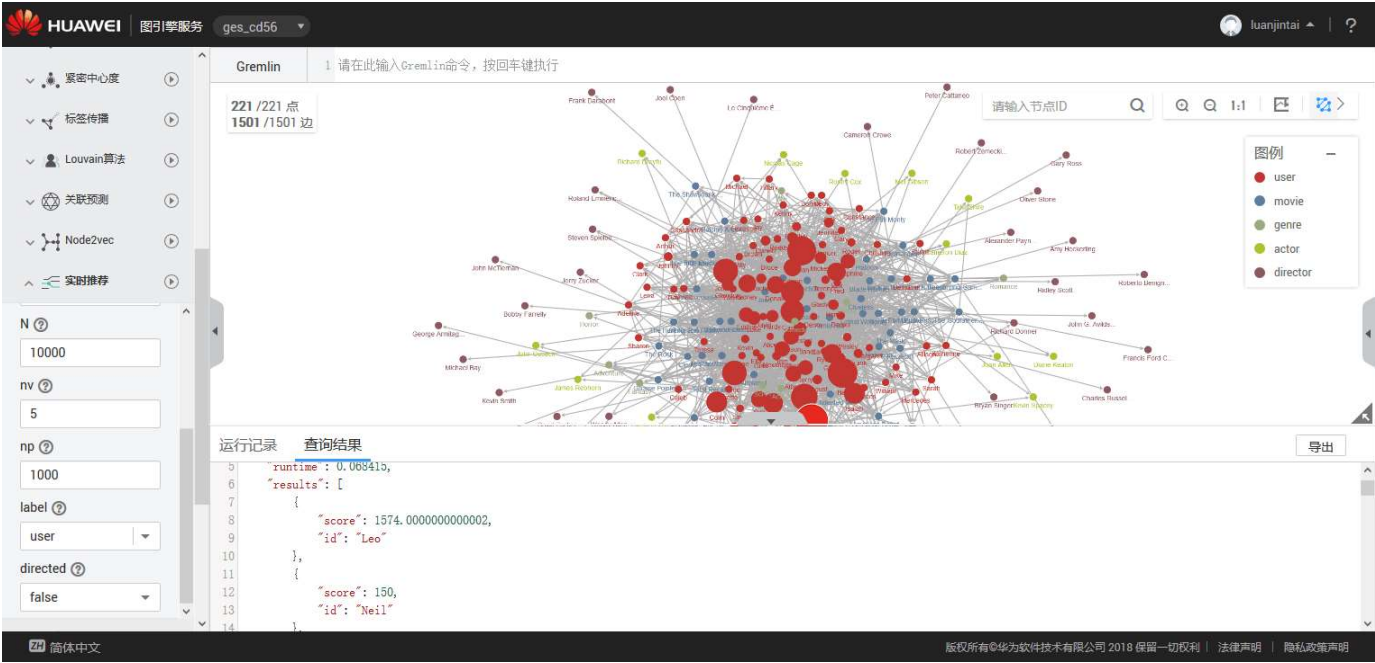
实验

nv - np 5 - 100:

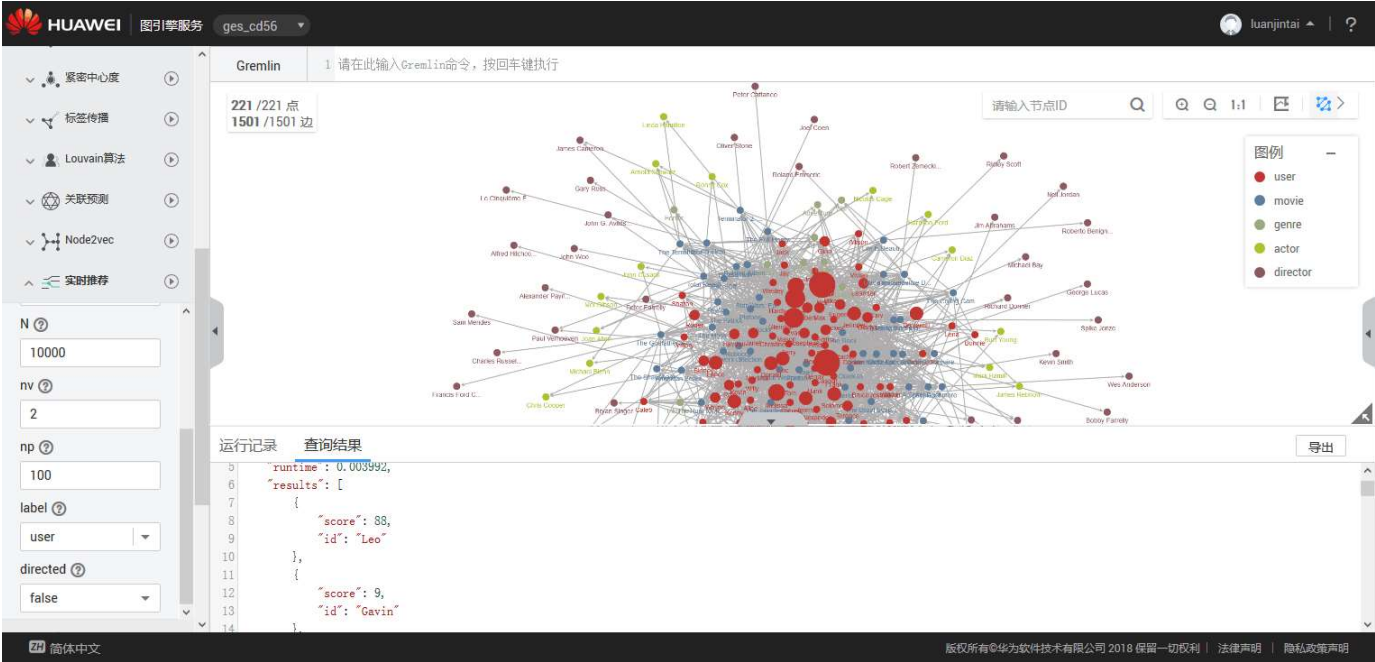




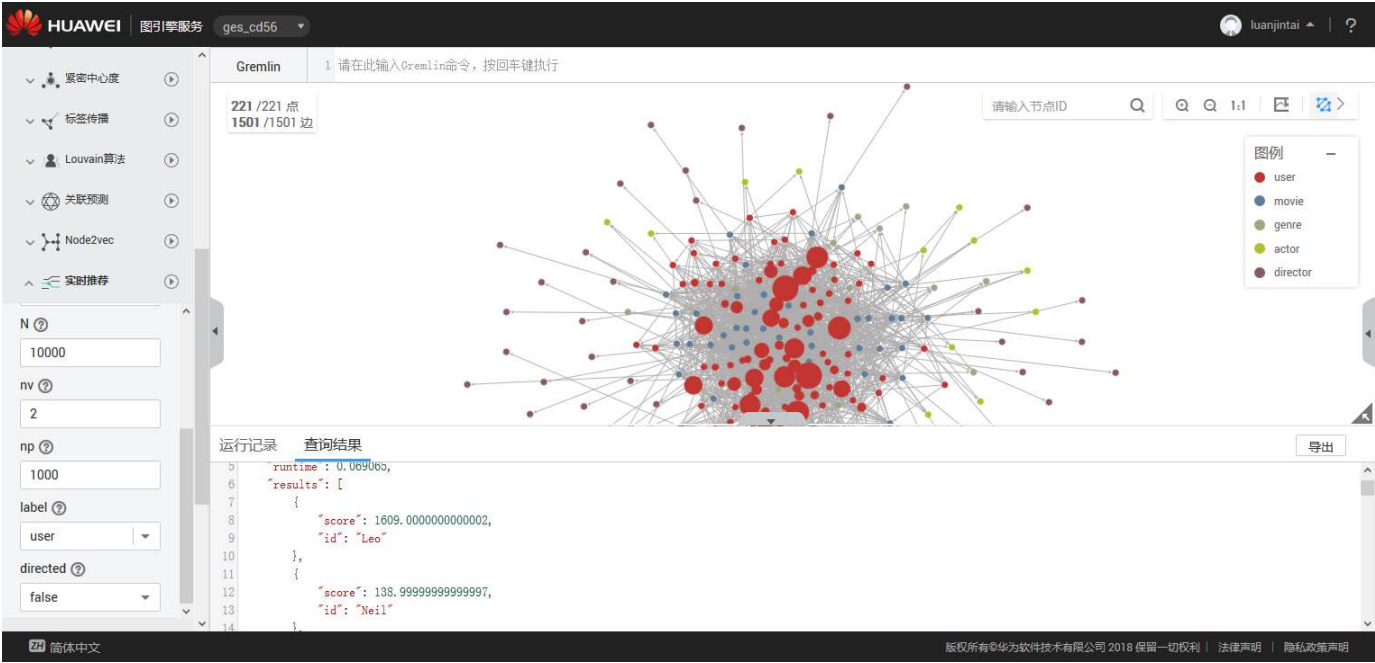
nv - np 5 - 1000:



nv - np 2 - 100:



nv - np 2 - 1000:



## 结论

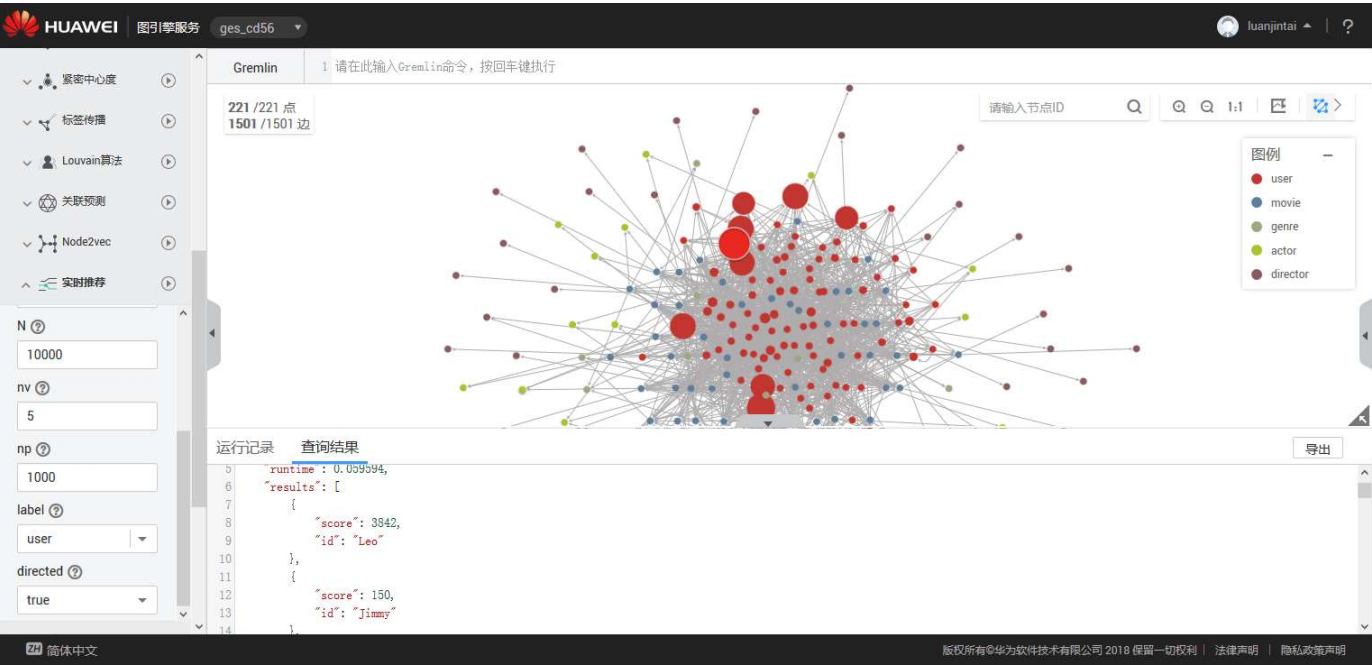
- 随着nv的增加, 运行时间增加, 节点的得分score增加, 可见当访问次数达到nv, 节点将加入推荐, 这将提前结束游走算法。
- 随着np的增加, 运行时间增加, 节点的得分score增加, 同样可见当访问次数达到nv, 节点将加入推荐, 这将提前结束游走算法。
- 类似神经网络中的dropout, 选择合适的nv和np可以减少计算量, 更快的得到需要的结果。

## directed

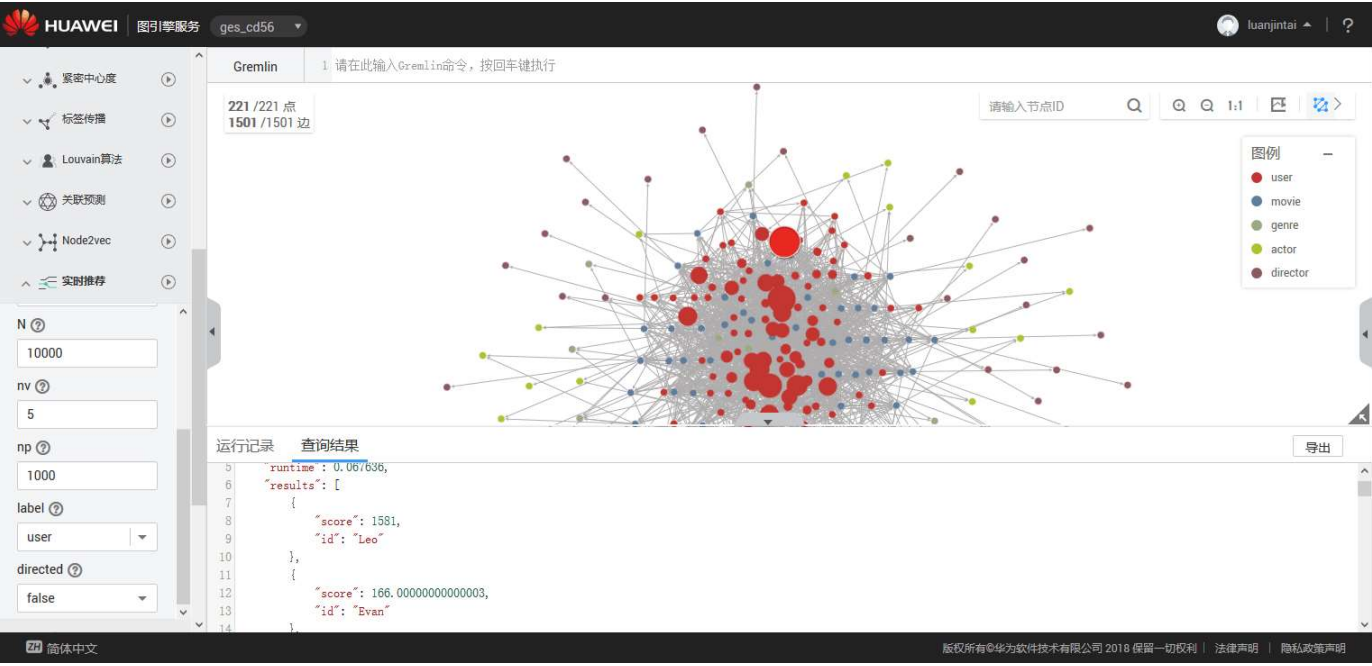
是否考虑图的边的方向。

## 实验

True:



False:



结论

考虑边的方向后，实际上限制了随机游走算法，计算复杂度降低，计算时间减少，