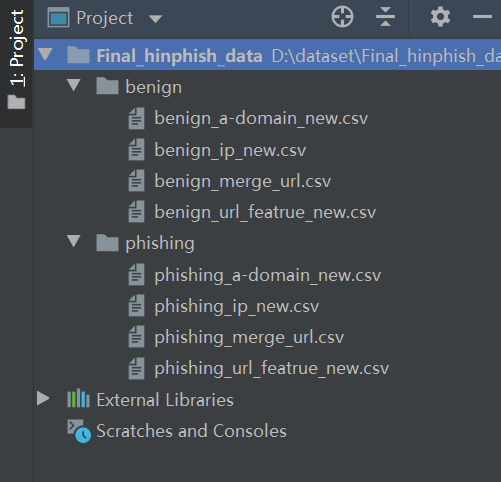
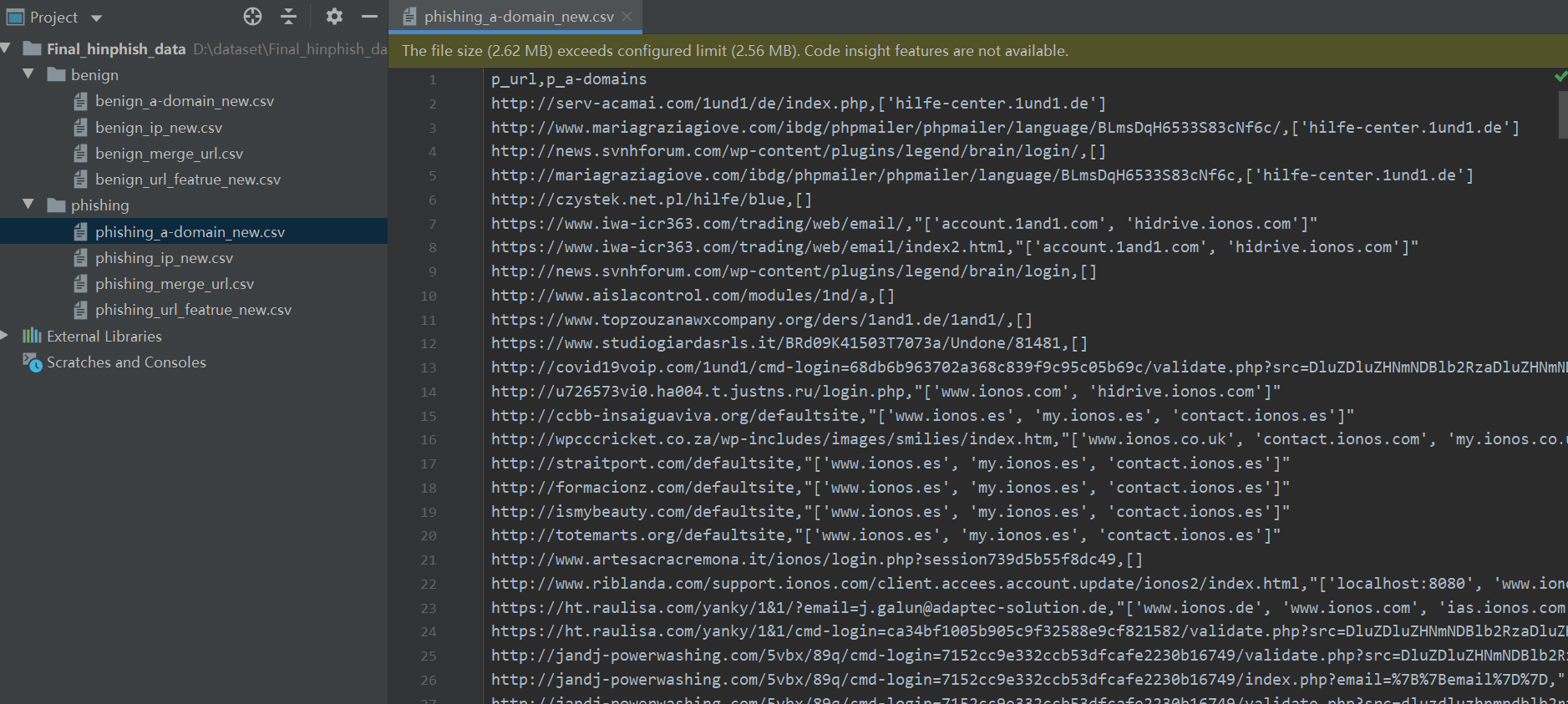
**Final\_hinphish\_data**的说明：

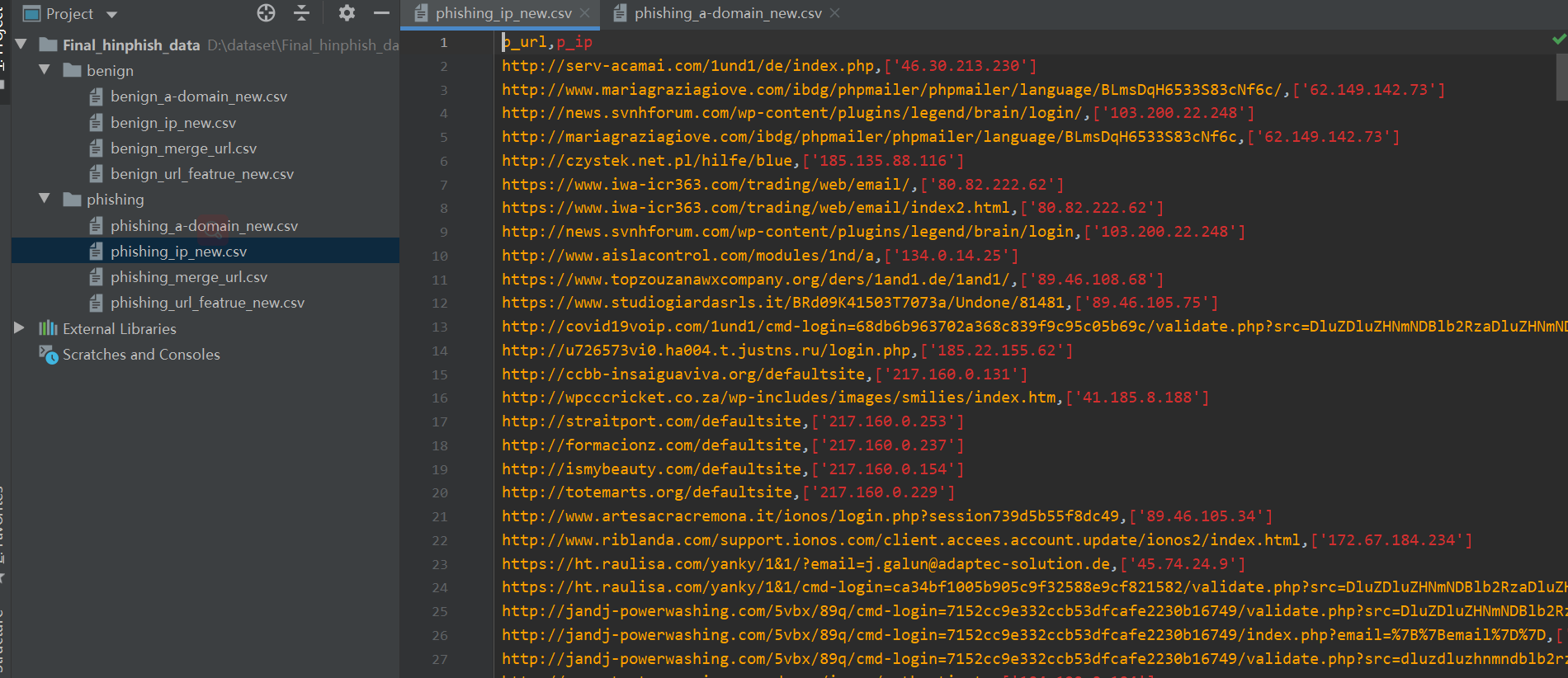


以phishing为例：

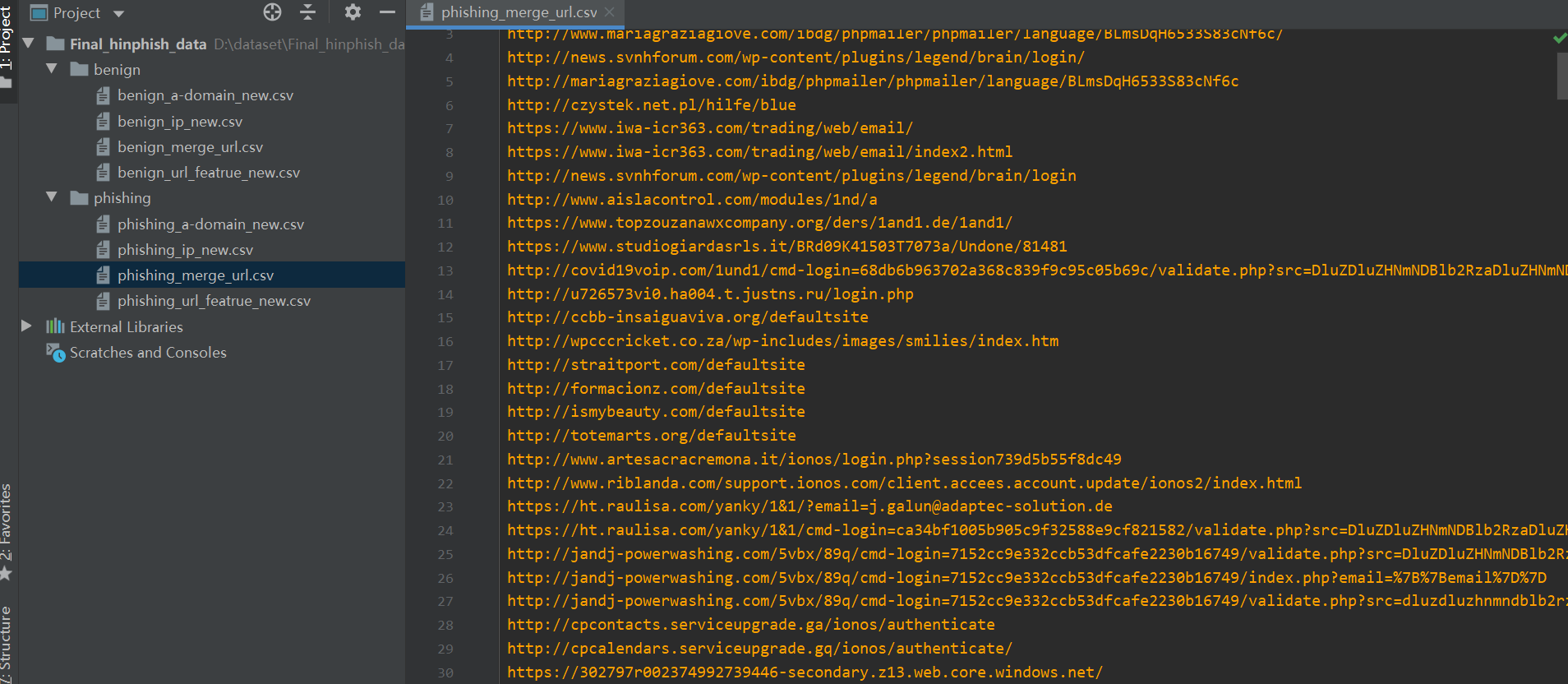
1. “**phisng\_a-domain\_new.csv**”这个文件描述的是一条url对应的外部链接a-domain的关系，这个表主要是构造“**url-alink-url**”这个邻接矩阵用的，这个**邻接矩阵的意思**，这里**举个例子**: url1连alink1，url2连alink1、alink3，url3连alink3，邻接矩阵中就要体现url1和url2相连，url2和url3相连。这个csv的内容如下：p\_url是url，p\_a-domain是对应连接的alink。



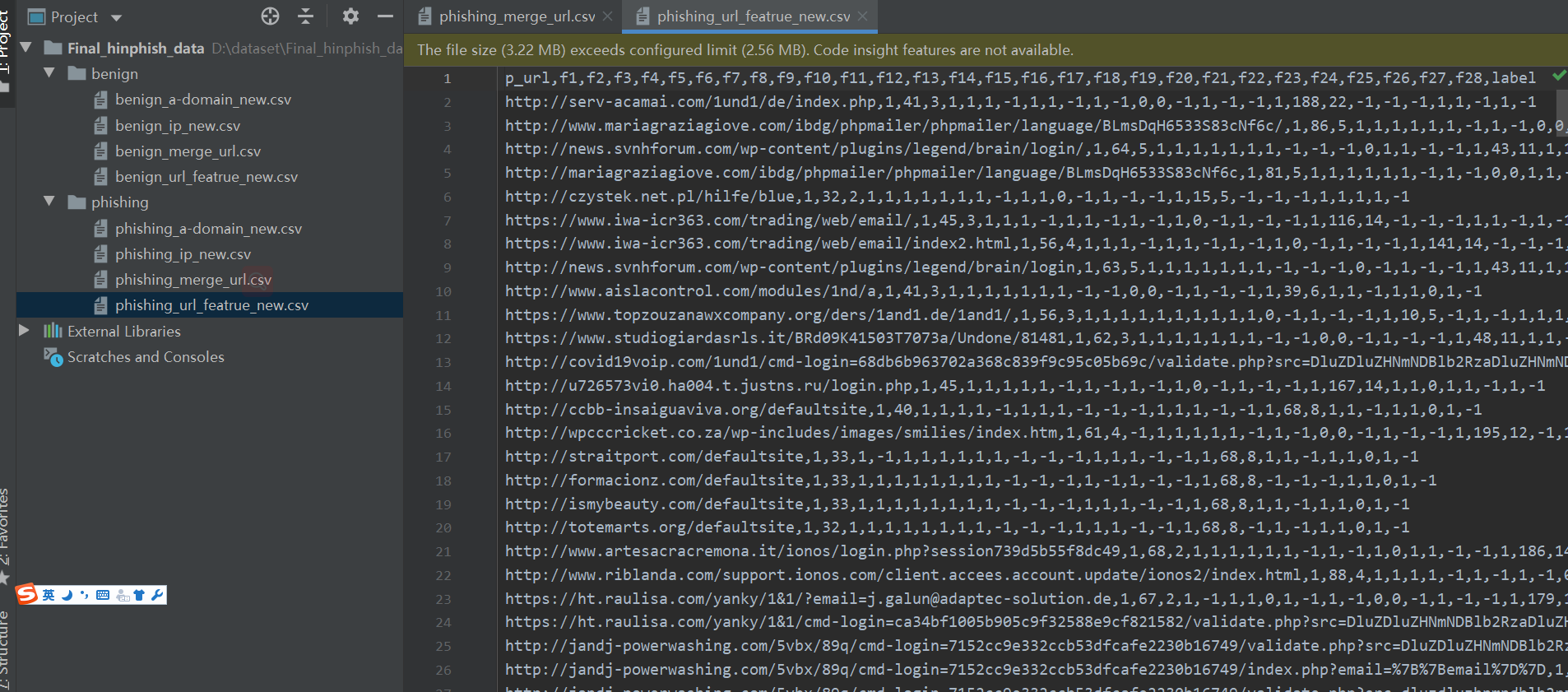
1. “**phishing\_ip\_new.csv**”这个文件描述的是一条url对应的ip，这个表主要是构造“**url-ip-url**”这个邻接矩阵用的，这个**邻接矩阵的意思**，这里**举个例子**：url1连ip1，url2连ip1，领接矩阵里就要体现url1和url2相连。csv的内容如下：p\_url是待分类的url，p\_ip是对应的ip。



1. “**phishing\_merge\_url.csv**”这个文件描述的是phishing类别下用到的所有url。



1. “**phishing\_url\_featrue\_new.csv**”这个文件描述的是phishing url对应的特征（28维）和label。



Han这个模型主要就是得出待分类的节点嵌入，然后再经过一个全连接做分类。Han就是一个双层的注意力机制，先是节点级别的，再是语义级别的。所以老师，模型输入的label应该是待分类url的label，这里做二分类，label应该是两维的（0、1），然后输入的featrue是url包括phishing和benign所有url的featrue就好了，两条元路径对于刚刚说的两个邻接矩阵，应该就可以了。