配置文件说明：

在配置文件中，含有多个变量，均采用$Attribute name$=$value$ 这样的形式；等号前后有空格之类的都无所谓，但是变量名不能错。

说明如下：

*=============== Sample =================*

*datafile=../data/biddata.full.csv*

*ROLLINGWINDOW = 200*

*STEPSIZE = 50*

*CLUSTER\_SIMILAR\_CUTOFF =0.98*

*POLICY\_CHOICE\_OPTION = 0*

*BINARY\_DISTANCE\_CUTOFF =0.3*

*RANKINGOPTION =RANKING*

*RESULT\_DIR =../output/output0.3\_0.98*

*CALCULATE\_SIMILARITY=1*

*CALCULATE\_CLIQUES=1*

*DISTANCE\_DIR=../output*

*========= Sample Ends =================*

datafile:

是数据文件的路径，如果有必要也可以更换数据文件的路径，但是要保持更换的数据文件和biddata.full.csv 一致；

ROLLINGWINDOW：

是在ROLLING的过程中，窗口的大小；

STEPSIZE：

在ROLLING的过程中，窗口移动的步长；举例，如果ROLLINGWINDOW=10，STEPSIZE=1，那么就是分别计算 1~10，2~11,3~12；

CLUSTER\_SIMILAR\_CUTOFF:

在寻找Clique的过程中，主要依据的距离的下限；如果某一条连接两个bidder的边权小于该值，就认为这两个bidder这件不存在关系；

BINARY\_DISTANCE\_CUTOFF:

在寻找Clique的过程中，参考的Participation 距离的cutoff，如果某一条连接两个bidder 在participation 算出来的距离的边权小于该值，就认为这两个bidder这件不存在关系；

补充说明：

1. 计算clique是依据距离计算的，这个距离可能是通过Participation得到的，也可能是通过Price得到的。如果这个距离是通过Price得到的，那么CLUSTER\_SIMILAR\_CUTOFF针对的是基于Price的距离，BINARY\_DISTANCE\_CUTOFF针对的是与Price相对应的Participation的值；

反过来，如果输入的距离文件是Participation得到的，那么CLUSTER\_SIMILAR\_CUTOFF 针对的就是基于Participation的距离，BINARY\_DISTANCE\_CUTOFF这个时候是失效的；

2. 程序怎么来判断一个distance\_file 是基于participation的，还是基于price的呢？就是通过文件名；所以不建议修改输出的文件名；

POLICY\_CHOICE\_OPTION:

选择是0或者1，如果选择0，在加载数据的时候会忽略掉policy flag为1的数据；

RESULT\_DIR：

对于一个config文件，基于这些参数算出来的所有文件都会被放到同一个路径下，result\_dir 就是指定这个路径；在编写config文件的时候，可以通过对不同的config文件编写不同的RESULT\_DIR，把输出放到不同的路径里面，批量运行；

CALCULATE\_SIMILARITY：

=1 的时候计算SIMILARITY，=0的时候不计算；

CALCULATE\_CLIQUES：

=1 的时候计算CLIQUES，=0的时候不算；

DISTANCE\_DIR：

这里需要说明，我们的计算分为2步，1: Similarity 2: Cliques；其中2的结果是依赖于1给出的一系列Distance File的。因此一般计算的时候，会同时执行1和2；如果执行1，DISTANCE\_DIR可以随便指定一个路径，计算出来的Distance 文件不会放在这个路径下；

如果不执行1，只执行2，那么这个时候可以指定一个输入的DISTANCE\_DIR的路径；

一般来说，现在看来计算1 的步骤最多只要几分钟的时间，推荐既执行1，又执行2；