第二周进展汇报

一、已完成

1. 对 FaSTest 项目进行修改，无需与真实值对比，修改后的项目命名为"FaSTest_backup"。

2. FaSTest_backup 在数据集 dblp、human、patents、wordnet 以及 youtube 上的运行结果如下：

Data Set	#Vertex #Edge	Total Time(ms)	Note
dblp	317080 1049866	2463445.85	执行命令"./Fastest -d dblp"后大概半分钟后终端有输出，sparse 图的运行时间显著低于 dense 图
human	4674 86282	11331936.00	采用分批次处理（**此方法的可行性后有验证**），sparse 图的运行时间显著低于 dense 图，具体情况见后。
patents	3774768 16518947	-	不管如何减少每次查询数量以及更换查询图，始终会被系统杀死
wordnet	76853 120399	787380.57	采用分批次处理，具体情况见后
youtube	1134890 2987624	-	不管如何减少每次查询数量以及更换查询图，始终会被系统杀死

3. 对于"获得子图基数准确值（即运行 DAF）"，目前只是时间问题，经过多次试验，将 timeout 设置为 600s 最为合适。DAF 在大部分的查询图上的运行时间在 300s 以上，若将 timeout 设置过低，会导致得到的值较少；对于运行时间超过 600s 的查询图，将 timeout 设置为 1200s 时，仅有一小部分（约 1/5）会运行得到结果，时间开销太大。故将 timeout 设置为 600s 较为合适。

4. 已将批量运行 DAF 的脚本修改为根据不同的命令行参数运行，这样方便在不同的数据集的情况下，动态调整 timeout，以期最优效果。

5. 已与葛秀京同学确认参加此次大创项目，目前已将讲述 FaSTest 算法的论文发送给她，大概本周内会进行进展同步工作，包括 FaSTest 算法的大概内容、代码，以及我目前的进展。

二、正在进行

1. 运行 DAF 以获得*_ans.txt 文件，目前相关脚本已经成熟，不会出现 bug。总共有 dblp、human、patents、wordnet 和 youtube 五个数据集，每个数据集的查询图文件有 1800 个，每个查询图文件在 DAF 上的运行时间按最坏时间(600ms)计算,总计 1500 h, 后面会与葛秀京同学进行分工合

作，若每天 24h 运行，也大概需要 30 天。当然，这是最坏情况，可以通过减小 timeout 、使用多台电脑等方式进行减少时间。

2. 最近在学习"正则表达式"相关知识，学习路线是从图书馆借一本小册子，里面有对正则表达式的大致讲解，在实际运用中可结合小册子和网络进行学习并运用。在实际操作中，会涉及到很多与文件相关的操作，以及对终端输出结果的处理，学习"正则表达式"是十分必要的。

三、将要进行

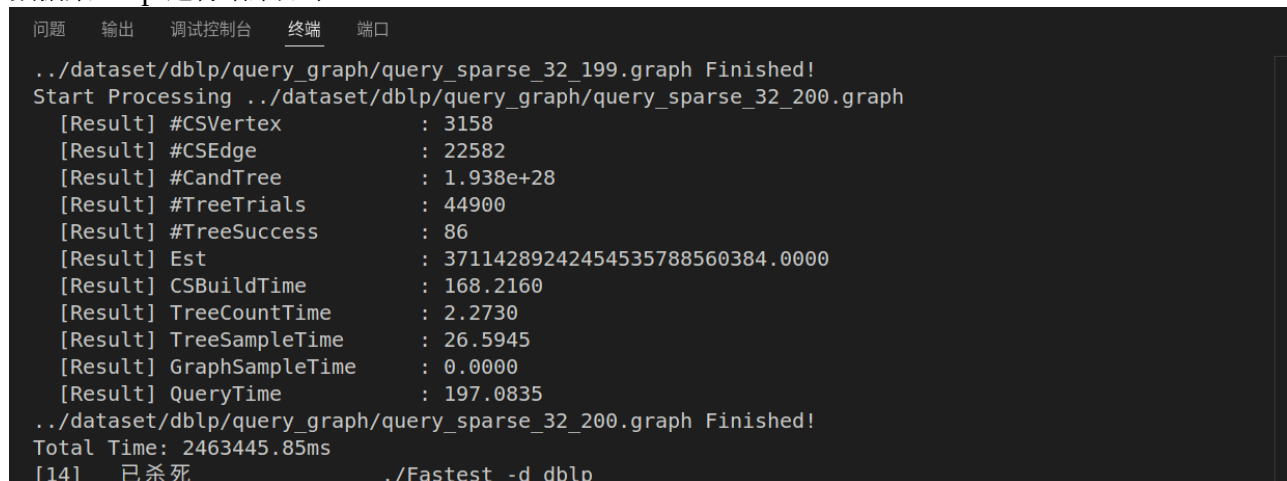1. 预计在本周日（3.9）会和葛秀京同学进行商讨后面分工的具体细节，会制定相关的文档格式、代码风格、数据记录、修改代码时的注意事项等等，做到一致性。

四、疑惑

1. 对于数据集 patents、youtube 未运行成功的这部分问题，该如何解决？之前我的想法是设置 timeout ，跳过运行不成功的部分，但查看日志后发现不是运行时间长的问题，而是系统内存超过 50 GB,被系统杀死。

2. 在大创开展的过程中，王老师您对于文档格式、代码风格、数据记录、修改代码时的注意事项等等有无任何建议？目前我的想法是规范文档的字体、型号等，代码风格主要涉及变量命名、注释、对代码进行文档式编写等。

———

数据集 dblp 运行结果如下：

```
问题    输出    调试控制台    终端    端口
../dataset/dblp/query_graph/query_sparse_32_199.graph Finished!
Start Processing ../dataset/dblp/query_graph/query_sparse_32_200.graph
  [Result] #CSVertex        : 3158
  [Result] #CSEdge          : 22582
  [Result] #CandTree        : 1.938e+28
  [Result] #TreeTrials      : 44900
  [Result] #TreeSuccess     : 86
  [Result] Est              : 3711428924245435788560384.0000
  [Result] CSBuildTime      : 168.2160
  [Result] TreeCountTime    : 2.2730
  [Result] TreeSampleTime   : 26.5945
  [Result] GraphSampleTime  : 0.0000
  [Result] QueryTime        : 197.0835
../dataset/dblp/query_graph/query_sparse_32_200.graph Finished!
Total Time: 2463445.85ms
[14]    已杀死              ./Fastest -d dblp
```

运行 human 数据集时，被系统杀死

查看 `/var/log/syslog` 日志：

Mar 2 13:54:23 programFaSTest kernel: [319.251398] Out of memory: Killed process 2908 (Fastest) total-vm:66164908kB, anon-rss:50854464kB, file-rss:256kB, shmem-rss:0kB, UID:1000 pgtables:129440kB oom_score_adj:0

Mar 2 13:54:23 programFaSTest systemd[1]: user@1000.service: A process of this unit has been killed by the OOM killer.

Mar 2 13:54:23 programFaSTest systemd[1004]: snap.code.code-4d64cb4c-38df-4fcd-9be8-8135369550dd.scope: A process of this unit has been killed by the OOM killer.

可以看出程序因占用过多内存而被 OOM Killer 杀死。

下面对数据集 human 的查询图分批次进行处理

query_dense_4_1.graph – query_dense_4_100.graph:

Total Time: 16839.85ms

query_dense_4_101.graph – query_dense_4_200.graph:

```
问题    输出    调试控制台    终端    端口

../dataset/human/query_graph/query_dense_4_199.graph Finished!
Start Processing ../dataset/human/query_graph/query_dense_4_200.graph
Constructing Candidate Space: 12 12
   [Result] #CSVertex          : 701
   [Result] #CSEdge            : 126564
   [Result] #CandTree          : 5.078e+07
   [Result] #TreeTrials        : 1000
   [Result] #TreeSuccess       : 752
   [Result] Est                : 38189388.2720
   [Result] CSBuildTime        : 531.5454
   [Result] TreeCountTime      : 1.0987
   [Result] TreeSampleTime     : 0.4845
   [Result] GraphSampleTime    : 0.0000
   [Result] QueryTime          : 533.1286
../dataset/human/query_graph/query_dense_4_200.graph Finished!
 Total Time: 18528.43ms
```

Total Time: 18528.43ms

query_dense_8_1.graph – query_dense_8_200.graph:

```
问题    输出    调试控制台    终端    端口

../dataset/human/query_graph/query_dense_8_199.graph Finished!
Start Processing ../dataset/human/query_graph/query_dense_8_200.graph
Constructing Candidate Space: 12 56
   [Result] #CSVertex          : 498
   [Result] #CSEdge            : 194140
   [Result] #CandTree          : 6.933e+11
   [Result] #TreeTrials        : 1000
   [Result] #TreeSuccess       : 739
   [Result] Est                : 512356448290.1090
   [Result] CSBuildTime        : 12622.2348
   [Result] TreeCountTime      : 0.0478
   [Result] TreeSampleTime     : 0.7299
   [Result] GraphSampleTime    : 0.0000
   [Result] QueryTime          : 12623.0125
../dataset/human/query_graph/query_dense_8_200.graph Finished!
 Total Time: 316559.73ms
```

Total Time: 316559.73ms

query_dense_12_1.graph – query_dense_12_200.graph:

```
问题    输出    调试控制台    终端    端口

../dataset/human/query_graph/query_dense_12_199.graph Finished!
Start Processing ../dataset/human/query_graph/query_dense_12_200.graph
Constructing Candidate Space: 12 132
   [Result] #CSVertex          : 298
   [Result] #CSEdge            : 73626
   [Result] #CandTree          : 9.424e+14
   [Result] #TreeTrials        : 1000
   [Result] #TreeSuccess       : 188
   [Result] Est                : 177176023434362.8750
   [Result] CSBuildTime        : 916.8800
   [Result] TreeCountTime      : 0.0737
   [Result] TreeSampleTime     : 0.5261
   [Result] GraphSampleTime    : 0.0000
   [Result] QueryTime          : 917.4799
../dataset/human/query_graph/query_dense_12_200.graph Finished!
 Total Time: 1070061.74ms
```

Total Time: 1070061.74ms

query_dense_16_1.graph – query_dense_16_200.graph:

```
问题    输出    调试控制台    终端    端口
 ../dataset/human/query_graph/query_dense_16_199.graph Finished!
 Start Processing ../dataset/human/query_graph/query_dense_16_200.graph
 Constructing Candidate Space: 16 240
    [Result] #CSVertex          : 1568
    [Result] #CSEdge            : 468822
    [Result] #CandTree          : 1.708e+24
    [Result] #TreeTrials        : 2100
    [Result] #TreeSuccess       : 86
    [Result] Est                : 69963457499844348739584.0000
    [Result] CSBuildTime        : 7510.5922
    [Result] TreeCountTime      : 1.3530
    [Result] TreeSampleTime     : 1.9586
    [Result] GraphSampleTime    : 0.0000
    [Result] QueryTime          : 7513.9038
 ../dataset/human/query_graph/query_dense_16_200.graph Finished!
 Total Time: 4356820.74ms
```

Total Time: 4356820.74ms

query_dense_20_1.graph – query_dense_20_100.graph:

```
问题    输出    调试控制台    终端    端口                                                          + ∨
 ../dataset/human/query_graph/query_dense_20_99.graph Finished!
 Start Processing ../dataset/human/query_graph/query_dense_20_100.graph
 Constructing Candidate Space: 20 380
    [Result] #CSVertex          : 938
    [Result] #CSEdge            : 298826
    [Result] #CandTree          : 1.21e+29
    [Result] #TreeTrials        : 50000
    [Result] #TreeSuccess       : 1
    [Result] Est                : 30546888053023915733155584.0000
    [Result] CSBuildTime        : 5211.5701
    [Result] TreeCountTime      : 1.3964
    [Result] TreeSampleTime     : 18.9214
    [Result] GraphSampleTime    : 11090.8780
    [Result] QueryTime          : 16322.7659
 ../dataset/human/query_graph/query_dense_20_100.graph Finished!
 Total Time: 2959895.40ms
 myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 2959895.40ms

query_dense_20_101.graph – query_dense_20_200.graph:

```
问题    输出    调试控制台    终端    端口                                                          + ∨
 ../dataset/human/query_graph/query_dense_20_199.graph Finished!
 Start Processing ../dataset/human/query_graph/query_dense_20_200.graph
 Constructing Candidate Space: 20 344
    [Result] #CSVertex          : 685
    [Result] #CSEdge            : 117188
    [Result] #CandTree          : 7.163e+19
    [Result] #TreeTrials        : 1000
    [Result] #TreeSuccess       : 503
    [Result] Est                : 36030721405570170880.0000
    [Result] CSBuildTime        : 17121.9619
    [Result] TreeCountTime      : 0.1895
    [Result] TreeSampleTime     : 1.6587
    [Result] GraphSampleTime    : 0.0000
    [Result] QueryTime          : 17123.8101
 ../dataset/human/query_graph/query_dense_20_200.graph Finished!
 Total Time: 2437417.83ms
 myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 2437417.83ms

query_sparse_8_1.graph – query_sparse_8_200.graph:

```
Start Processing ../dataset/human/query_graph/query_sparse_8_200.graph
Constructing Candidate Space: 12 22
  [Result] #CSVertex        : 844
  [Result] #CSEdge          : 109554
  [Result] #CandTree        : 9.555e+11
  [Result] #TreeTrials      : 1000
  [Result] #TreeSuccess     : 812
  [Result] Est              : 775895363348.3081
  [Result] CSBuildTime      : 1153.6314
  [Result] TreeCountTime    : 0.4661
  [Result] TreeSampleTime   : 0.6221
  [Result] GraphSampleTime  : 0.0000
  [Result] QueryTime        : 1154.7196
../dataset/human/query_graph/query_sparse_8_200.graph Finished!
Total Time: 27525.87ms
[1]+  已杀死                ./Fastest -d human
```

Total Time: 27525.87ms

query_sparse_12_1.graph – query_sparse_12_200.graph:

```
问题   输出   调试控制台   终端   端口                                          + ∨
 ../dataset/human/query_graph/query_sparse_12_199.graph Finished!
 Start Processing ../dataset/human/query_graph/query_sparse_12_200.graph
 Constructing Candidate Space: 12 32
   [Result] #CSVertex        : 439
   [Result] #CSEdge          : 21744
   [Result] #CandTree        : 1.617e+15
   [Result] #TreeTrials      : 1000
   [Result] #TreeSuccess     : 341
   [Result] Est              : 551300215699673.4375
   [Result] CSBuildTime      : 33.1333
   [Result] TreeCountTime    : 0.1512
   [Result] TreeSampleTime   : 0.4780
   [Result] GraphSampleTime  : 0.0000
   [Result] QueryTime        : 33.7625
 ../dataset/human/query_graph/query_sparse_12_200.graph Finished!
 Total Time: 26735.99ms
○ myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 26735.99ms

query_sparse_16_1.graph – query_sparse_16_200.graph:

```
问题   输出   调试控制台   终端   端口                                          + ∨
 ../dataset/human/query_graph/query_sparse_16_199.graph Finished!
 Start Processing ../dataset/human/query_graph/query_sparse_16_200.graph
 Constructing Candidate Space: 16 44
   [Result] #CSVertex        : 1714
   [Result] #CSEdge          : 278058
   [Result] #CandTree        : 2.229e+24
   [Result] #TreeTrials      : 1000
   [Result] #TreeSuccess     : 173
   [Result] Est              : 3856883026047200951009208.0000
   [Result] CSBuildTime      : 628.1015
   [Result] TreeCountTime    : 1.1119
   [Result] TreeSampleTime   : 0.9470
   [Result] GraphSampleTime  : 0.0000
   [Result] QueryTime        : 630.1603
 ../dataset/human/query_graph/query_sparse_16_200.graph Finished!
 Total Time: 50372.67ms
○ myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 50372.67ms

query_sparse_20_1.graph – query_sparse_20_200.graph:

```
问题    输出    调试控制台    终端    端口                                                          +

 ../dataset/human/query_graph/query_sparse_20_199.graph Finished!
 Start Processing ../dataset/human/query_graph/query_sparse_20_200.graph
 Constructing Candidate Space: 20 56
   [Result] #CSVertex          : 1423
   [Result] #CSEdge            : 81566
   [Result] #CandTree          : 2.58e+24
   [Result] #TreeTrials        : 1000
   [Result] #TreeSuccess       : 243
   [Result] Est                : 6269232385609567976816664.0000
   [Result] CSBuildTime        : 694.6420
   [Result] TreeCountTime      : 0.6087
   [Result] TreeSampleTime     : 0.8838
   [Result] GraphSampleTime    : 0.0000
   [Result] QueryTime          : 696.1345
 ../dataset/human/query_graph/query_sparse_20_200.graph Finished!
 Total Time: 51213.80ms
 myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 51213.80ms

Query Range	Time (ms)
query_dense_4_1.graph – query_dense_4_100.graph	16839.85
query_dense_4_101.graph – query_dense_4_200.graph	18528.43
query_dense_8_1.graph – query_dense_8_200.graph	316559.73
query_dense_12_1.graph – query_dense_12_200.graph	1070061.74
query_dense_16_1.graph – query_dense_16_200.graph	4356820.74
query_dense_20_1.graph – query_dense_20_100.graph	2959895.40
query_dense_20_101.graph – query_dense_20_200.graph	2437417.83
query_sparse_8_1.graph – query_sparse_8_200.graph	27525.87
query_sparse_12_1.graph – query_sparse_12_200.graph	26735.99
query_sparse_16_1.graph – query_sparse_16_200.graph	50372.67
query_sparse_20_1.graph – query_sparse_20_200.graph	51213.80
Total	11331936.00

为确保最终 Total 的正确性，即各 Query Range 的运行时间之间具有可加性，挑选
query_dense_4_1.graph – query_dense_4_100.graph,将其分为 query_dense_4_1.graph –
query_dense_4_25.graph 和 query_dense_4_26.graph – query_dense_4_100.graph 进行验证：

query_dense_4_1.graph – query_dense_4_25.graph:

```
问题    输出    调试控制台    终端    端口

 ../dataset/human/query_graph/query_dense_4_24.graph Finished!
Start Processing ../dataset/human/query_graph/query_dense_4_25.graph
Constructing Candidate Space: 12 12
    [Result] #CSVertex         : 888
    [Result] #CSEdge           : 73536
    [Result] #CandTree         : 1.703e+07
    [Result] #TreeTrials       : 1000
    [Result] #TreeSuccess      : 869
    [Result] Est               : 14797690.0280
    [Result] CSBuildTime       : 275.9922
    [Result] TreeCountTime     : 0.3282
    [Result] TreeSampleTime    : 0.2934
    [Result] GraphSampleTime   : 0.0000
    [Result] QueryTime         : 276.6137
 ../dataset/human/query_graph/query_dense_4_25.graph Finished!
 Total Time: 3458.36ms
```

Total Time: 3458.36ms

query_dense_4_26.graph – query_dense_4_100.graph:

```
问题    输出    调试控制台    终端    端口

 ../dataset/human/query_graph/query_dense_4_99.graph Finished!
Start Processing ../dataset/human/query_graph/query_dense_4_100.graph
Constructing Candidate Space: 12 12
    [Result] #CSVertex         : 448
    [Result] #CSEdge           : 60816
    [Result] #CandTree         : 1.509e+07
    [Result] #TreeTrials       : 1000
    [Result] #TreeSuccess      : 764
    [Result] Est               : 11528318.4080
    [Result] CSBuildTime       : 64.4351
    [Result] TreeCountTime     : 0.3291
    [Result] TreeSampleTime    : 0.5713
    [Result] GraphSampleTime   : 0.0000
    [Result] QueryTime         : 65.3355
 ../dataset/human/query_graph/query_dense_4_100.graph Finished!
 Total Time: 13479.25ms
```

Total Time: 13479.25ms

3458.36 + 13479.25 = 16937.61, 当 Query Range 为 query_dense_4_1.graph –
query_dense_4_100.graph 时，运行时间为 16839.85，误差为 0.58%,这可能与系统有关，可以忽略不
计。由此可知，将查询图分批次处理的办法是可行的。

对于数据集 patents,不管如何调整查询图数量的大小，运行后终端均输出 "被杀死"，原因是运行时所需内存已超出 50 GB，被 OOM 杀死：

```
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$ ./Fastest -d patents
 Reading from: ../dataset/patents/patents_ans.txt
 已杀死
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$ ./Fastest -d patents
 Reading from: ../dataset/patents/patents_ans.txt
 已杀死
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$ ./Fastest -d patents
 Reading from: ../dataset/patents/patents_ans.txt
 已杀死
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

对于 wordnet 数据集，如果也将所有查询图文件一次性读入，也存在被系统杀死的问题，但 wordnet 数据集较小，因此只用分两批处理即可：

query_dense_4_1.graph – query_dense_dense_20_200.graph:

```
问题    输出    调试控制台    终端    端口

 ../dataset/wordnet/query_graph/query_dense_20_199.graph Finished!
 Start Processing ../dataset/wordnet/query_graph/query_dense_20_200.graph
 Constructing Candidate Space: 20 76
   [Result] #CSVertex         : 67943
   [Result] #CSEdge           : 780296
   [Result] #CandTree         : 3.623e+25
   [Result] #TreeTrials       : 50000
   [Result] #TreeSuccess      : 0
   [Result] Est               : 6845954.9870
   [Result] CSBuildTime       : 617.0599
   [Result] TreeCountTime     : 14.2304
   [Result] TreeSampleTime    : 15.8009
   [Result] GraphSampleTime   : 80.4848
   [Result] QueryTime         : 727.5761
 ../dataset/wordnet/query_graph/query_dense_20_200.graph Finished!
 Total Time: 293863.46ms
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 293863.46ms

query_sparse_8_1.graph – query_sparse_20_200.graph:

```
问题    输出    调试控制台    终端    端口

 ../dataset/wordnet/query_graph/query_sparse_20_199.graph Finished!
 Start Processing ../dataset/wordnet/query_graph/query_sparse_20_200.graph
 Constructing Candidate Space: 20 56
   [Result] #CSVertex         : 216471
   [Result] #CSEdge           : 1490920
   [Result] #CandTree         : 1.793e+27
   [Result] #TreeTrials       : 50000
   [Result] #TreeSuccess      : 0
   [Result] Est               : 102556108531291.7812
   [Result] CSBuildTime       : 638.0355
   [Result] TreeCountTime     : 31.9182
   [Result] TreeSampleTime    : 21.5718
   [Result] GraphSampleTime   : 167.8883
   [Result] QueryTime         : 859.4139
 ../dataset/wordnet/query_graph/query_sparse_20_200.graph Finished!
 Total Time: 493517.11ms
myran@programFaSTest:~/FaSTest_backup/FaSTest/build$
```

Total Time: 493517.11ms

Query Range	Time (ms)
query_dense_4_1.graph – query_dense_dense_20_200.graph	293863.46
query_sparse_8_1.graph – query_sparse_20_200.graph	493517.11
Total	787380.57

运行 youtube 数据集时，出现了与 patents 数据集相同的问题：