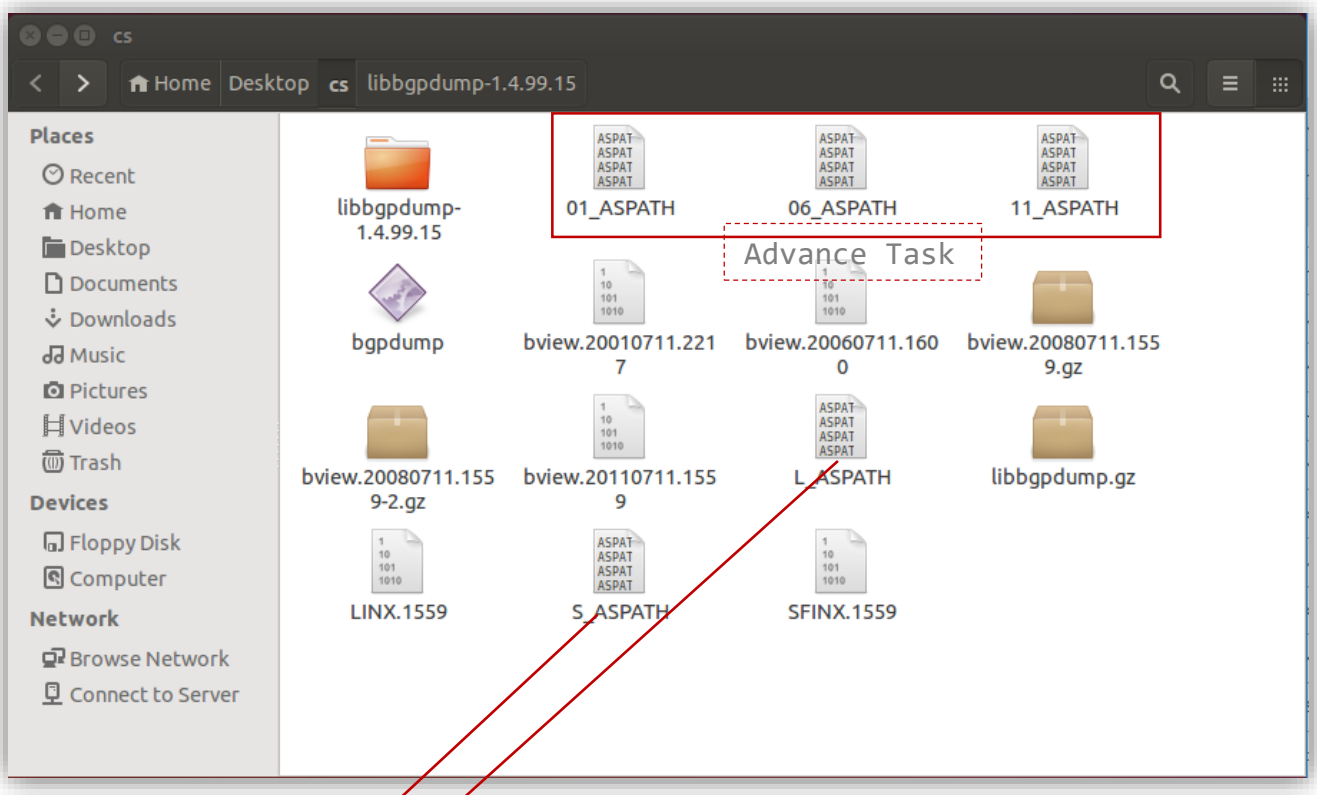
CS5229 REPORT

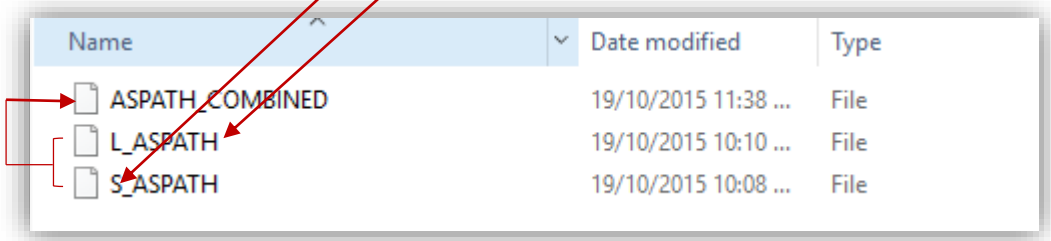
YANG MO A0091836X

1. Preparation

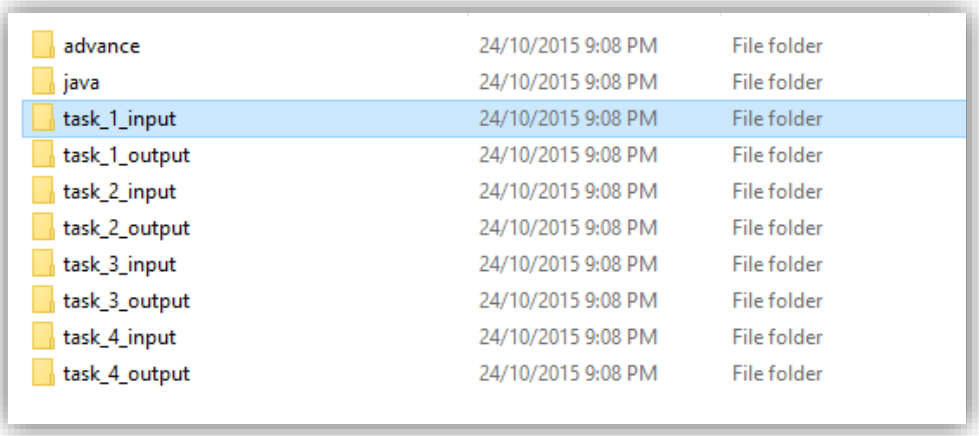
1. Unzip data set first at local Ubuntu virtual machine

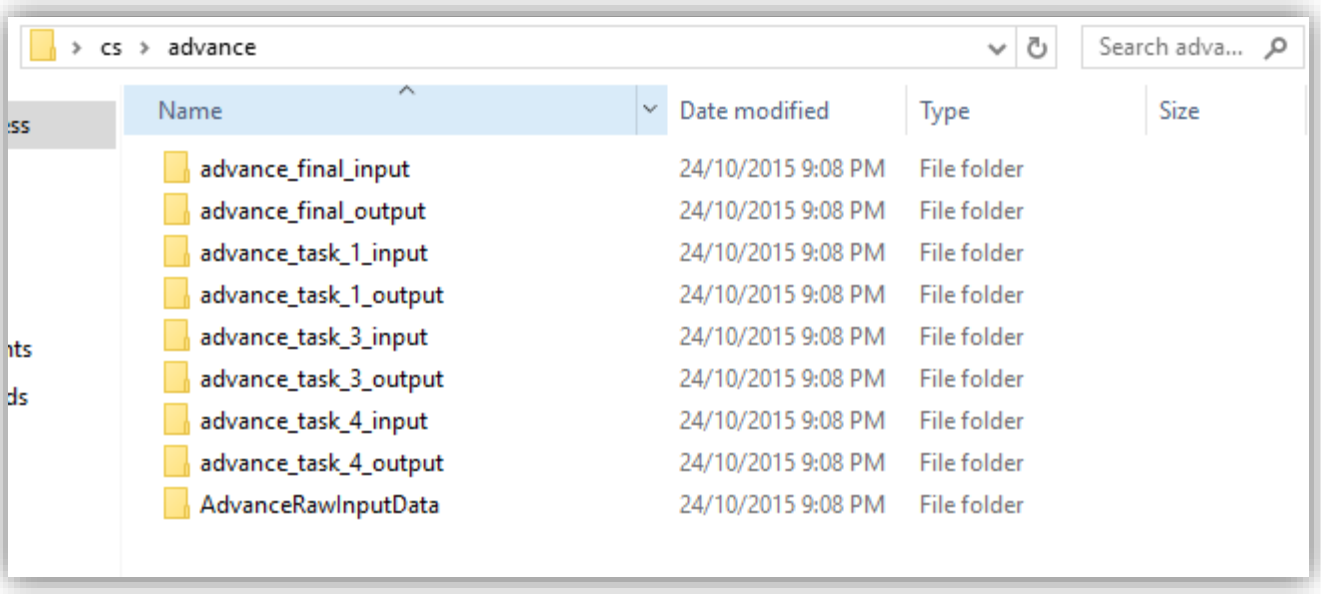


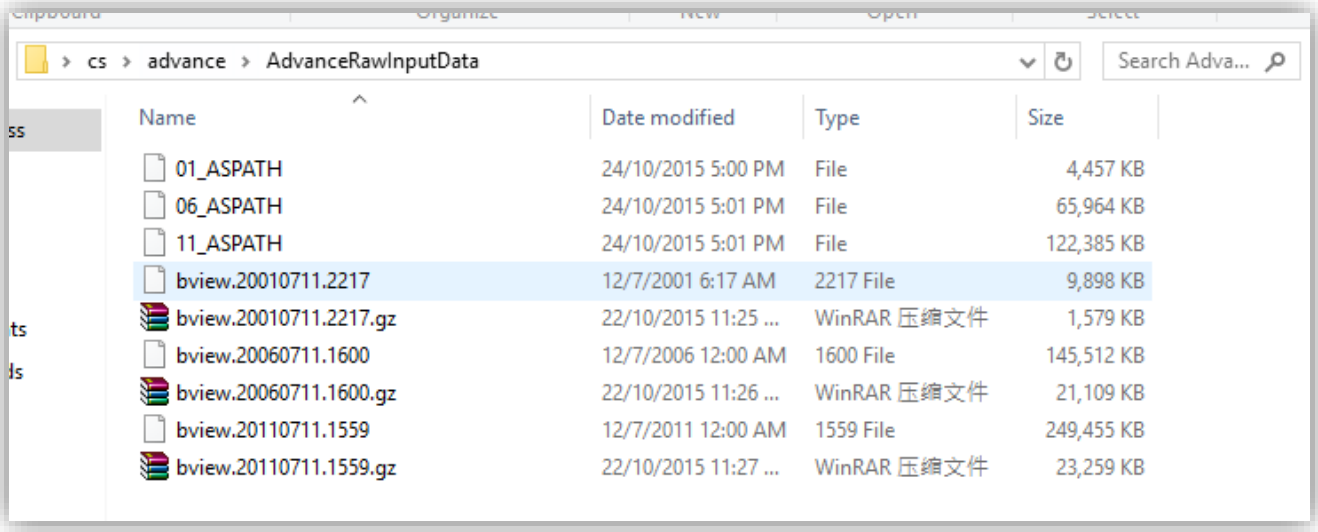
2. Combine two files for LINX and SFINX into one combined file for Elementary and Intermediate Tasks



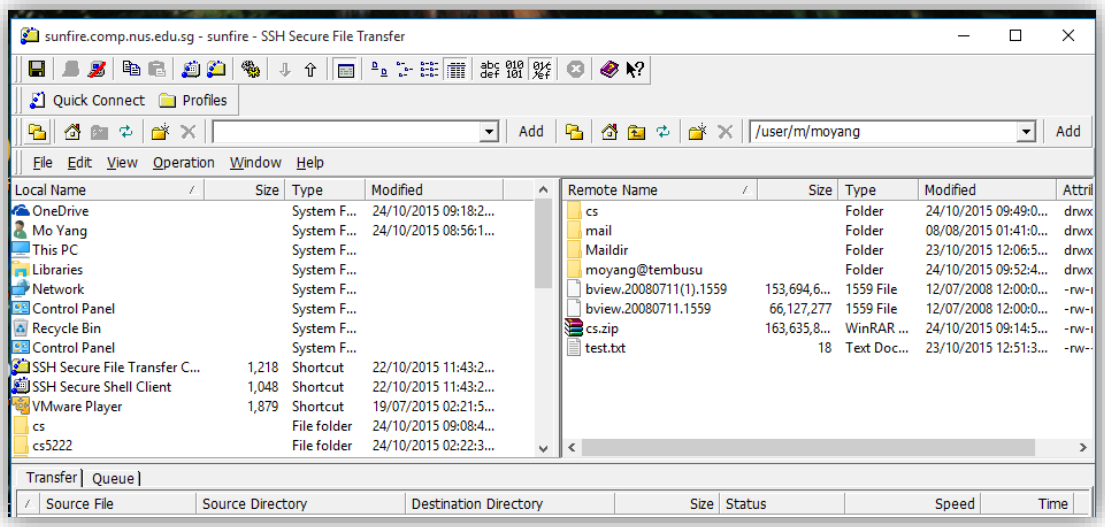
3. Create local folder first







4. Zip folder and upload to sunfire and unzip it



5. Copy unzipped cs folder to Tembusu Cluster and rename it to cs5229 using command:

```
scp -r cs moyang@tembusu:cs5229
```

6. Log into Tembusu :

```
ssh -X tembusu
```

```
Last login: Fri Oct 23 00:27:10 2015 from sunfire0.comp.nus.edu.sg
                                Important Message
Usage Policy concerning long jobs on Access Nodes in
  <https://docs.comp.nus.edu.sg/node/1818>

NO BACKUP for cluster user files. Homedir quota 200G.

List of compute cluster hardware in <https://docs.comp.nus.edu.sg/node/1814>

!!! SGE deprecated as of 1 May 2015 !!!

compg21, compg35, compg37 offline due to hardware faults

centos6.7 64bit with 9TB disk on compg0-19
centos6.7 64bit with 1TB disk on compg54-75, AMD processors with 32 cores

compg25-56 reserved from Aug 21 - Nov 25
compg10-19 reserved from Sept 25 - Oct 26

--------------------------------------------------------------------------------
moyang@access10:~$ ls
cs5229
moyang@access10:~$
```

7. Verify all content is present in tembusu

```
moyang@access10:~$ cd cs5229
moyang@access10:~/cs5229$ ls
advance  task_1_input   task_2_input   task_3_input   task_4_input
java     task_1_output  task_2_output  task_3_output  task_4_output
moyang@access10:~/cs5229$
```

2. Elementary Part --- Task 1

2.1 Run Task 1

Compile all the java files first

```
moyang@access10:~/cs5229/java$ ls
Advance.class  Task1.class  Task2$1.class  Task2.java  Task3.java  Task4.java
Advance.java   Task1.java   Task2.class    Task3.class  Task4.class
moyang@access10:~/cs5229/java$
```

Command Pattern : `java Task1 <inputfile> <outputfile>`

Real Command to execute task 1 :

```
java Task1 ~/cs5229/task_1_input/ASPATH_COMBINED ~/cs5229/task_1_output/ASPATH_TASK_1_OUT
```

Task 1 will take in the combined ASPATH file and run line by line to count distinct number of AS and AS paths. The result will be written in `ASPATH_TASK_1_OUT`

2.2 Result:

ASPATH_TASK_1_OUT

```
649351  6067 3356 7018 18903
649352  8468 1299 2828 11874
649353  5511 3356 12530 39399
649354  286 1299 20965 5408 9069
649355  20932 3356 7963 10340
649356  29636 2914 1239 19024 25782
649357  6762 6453 4755 24391
649358  6762 702 5486 20623
649359  8468 1299 2914 7385 19640
649360  6067 3356 7132 22410
649361  29636 3491 6846 41867 41631
649362  3257 9050 42497
649363  30126 2200 1273 5511 28683
649364  6067 6453 2828 40510
649365  6762 3356 9002 43975 39229
649366  5511 1239 32665
649367  13237 3320 209 19936
649368  30126 2200 3356 7018 40419
649369  13237 20485 31425
649370  6762 6453 4755 24397
649371  30126 2200 1273 4323 36664
649372  8419 19151 11013 10942
649373  6067 6453 1239 4775 17894
649374  30126 174 33363 14184
649375  286 22773 14849
649376  6762 1299 29209 38996
649377  5400 2856 34737
649378  8419 6461 701 27030
649379  5511 1239 32647
649380  34997 12566 174 10400
649381  39202 174 7473 7713 17974 24213 38505
649382  30844 3491 18042
649383  5511 1239 32652
649384  6939 19166 36874
649385  30126 2200 12956 18881 28599
649386  8419 3491 9299 17452
649387  5511 3356 20960 31304
649388  286 3257 3313 12498
649389  6067 3491 17557 17911
649390  34997 6939 4725 7514 2515
649391  Number of Ases: 28871
649392  Number of AS paths: 649390
```

To be able to verify the output file locally, the output file is copied back to sunfire and then downloaded to local windows machine. We then open it using notedpad ++.

The result is clear that:

Number of distinct Ases is 28871

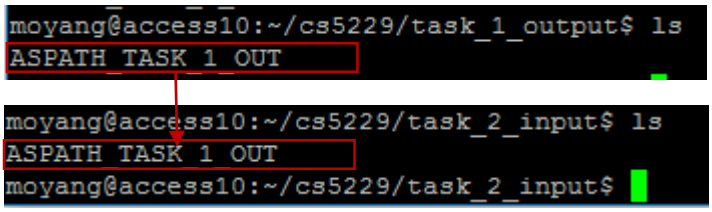
Number of distinct AS Paths is 649390

The ASPATH_TASK_1_OUT file does not contain any duplicate AS Path.

3. Elementary Part --- Task 2

3.1 Prepare input data

Since task 2 need to use the output from task 1. `ASPATH_TASK_1_OUT` is copied from `task_1_output` folder to `task_2_input` folder:



3.2 Run Task 2

Command Pattern : `java Task2 <inputfile> <outputfile>`

Real Command to execute task 2 :

```
java Task2 ~/cs5229/task_2_input/ASPATH_TASK_1_OUT ~/cs5229/task_2_output/ASPATH_TASK_2_OUT
```

Task 2 will first collect connection information for each AS node first. Then it will compute the degree for each AS node. Before it write to the file, it will sort based on the degree and list on Top 10 AS nodes.

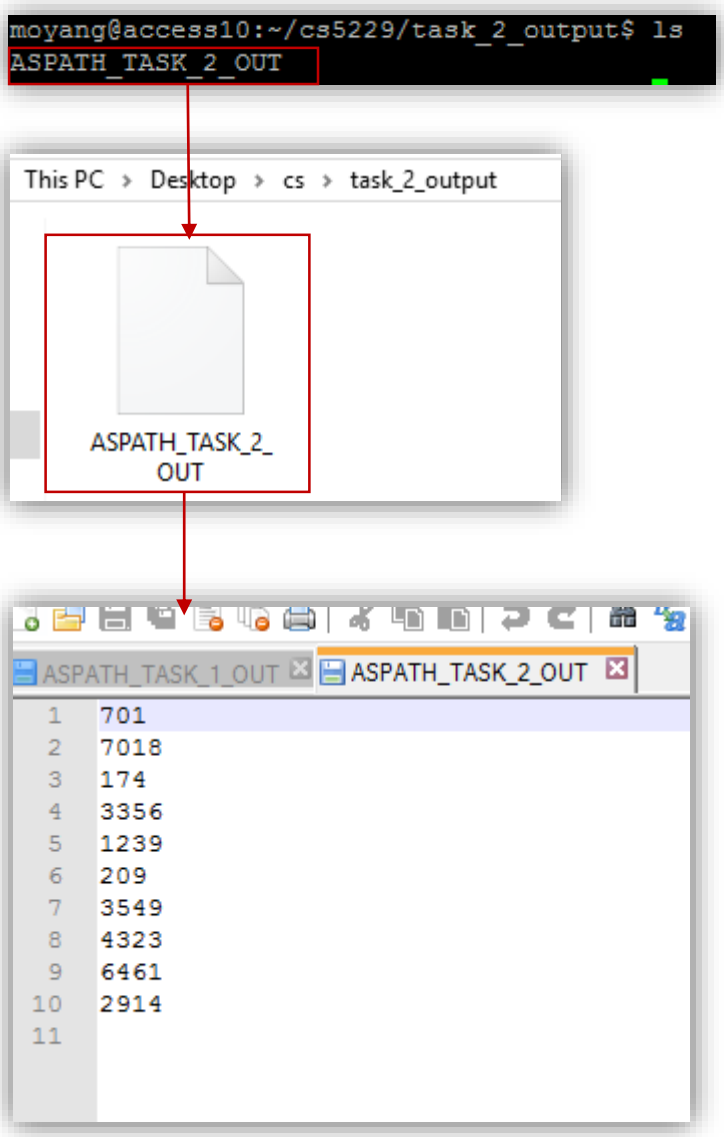
```
while(curLine != null) {
    //collect connection information for each AS node first
    processAndUpdateMap(curLine);
    printStatus();
    curLine = br.readLine();
}

//Convert the asSetMap to an sorted list of Node-Degree Entry
sortedEntryList = ConvertAndSortMap(asSetMap);
int i = 0;

for(Map.Entry<String, Integer> entry : sortedEntryList){
    bw.write(entry.getKey() + "\n");
    System.out.println(entry.getKey() + " ===== " + entry.getValue());
    i ++;
    if(i == 10){
        break;
    }
}
```

3.3 Result

To be able to verify the output file locally, the output file is copied back to sunfire and then downloaded to local windows machine. We then open it using notedpad ++.



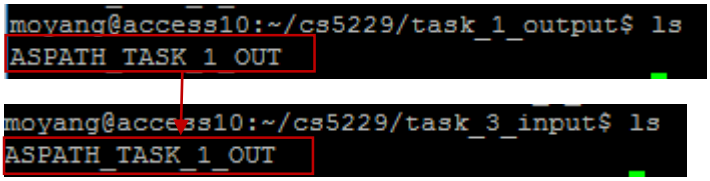
The results is clear that only top ten AS nodes are ouput in the output file.

And 701 is the AS of highest degree.

4. Intermediate Part --- Task 3

4.1 Prepare input data

Since task 3 need to use the output from task 1. `ASPATH_TASK_1_OUT` is copied from `task_1_output` folder to `task_3_input` folder:



4.2 Run Task 3

Command Pattern : `java Task3 <L> <R> <inputfile> <outputfile>`

Real Command to execute task 3 :

```
java Task3 1 60 ~/cs5229/task_3_input/ASPATH_TASK_1_OUT ~/cs5229/task_3_output/ASPATH_TASK_3_OUT
```

Here `L` is set to 1 and `R` is set to 60. These values are chosen due the original paper.

Following the algorithm given, the program contains five parts:

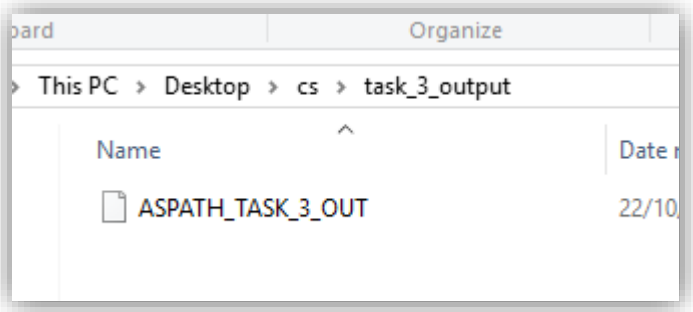
- Non p2p Phase 1
- Non p2p Phase 2
- Non p2p Phase 3
- p2p Phase 2
- p2p Phase 3

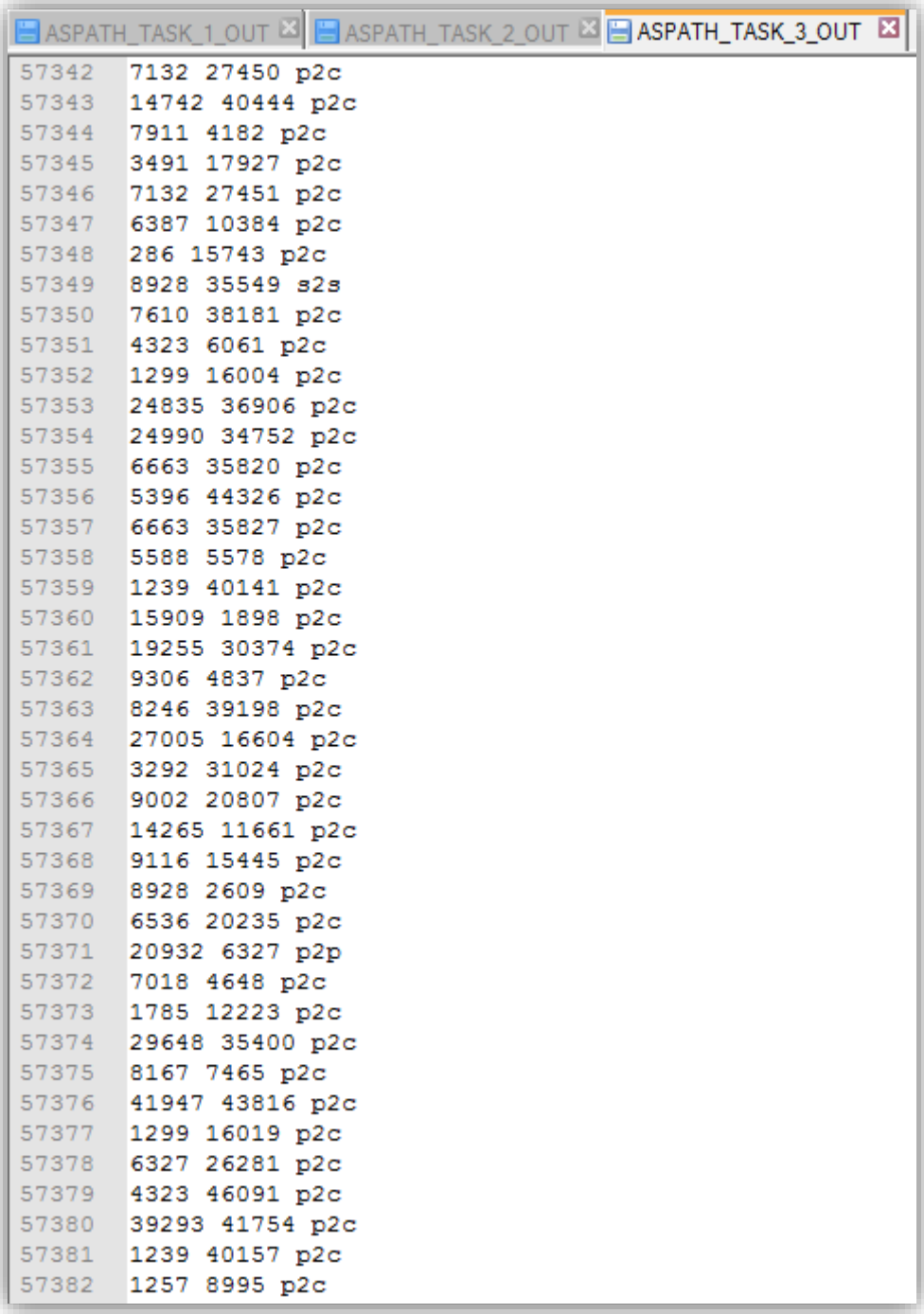
Where Non p2p phases annotate only s2s, c2p, p2c relationships.

P2p phases will annotate those node pairs who has p2p relationship.

4.3 Result:

To be able to verify the output file locally, the output file is copied back to sunfire and then downloaded to local windows machine. We then open it using notedpad ++.



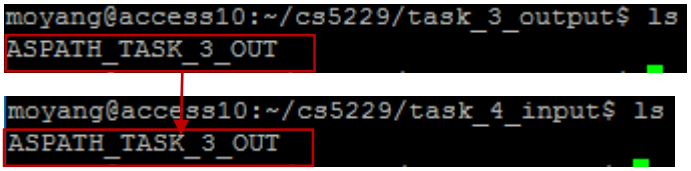


It is clear that there are totally 57382 node pairs which have been annotated.

5. Intermediate Part --- Task 4

5.1 Prepare input data

Since task 4 need to use the output from task 3. `ASPATH_TASK_3_OUT` is copied from `task_3_output` folder to `task_4_input` folder:



5.2 Run Task 4

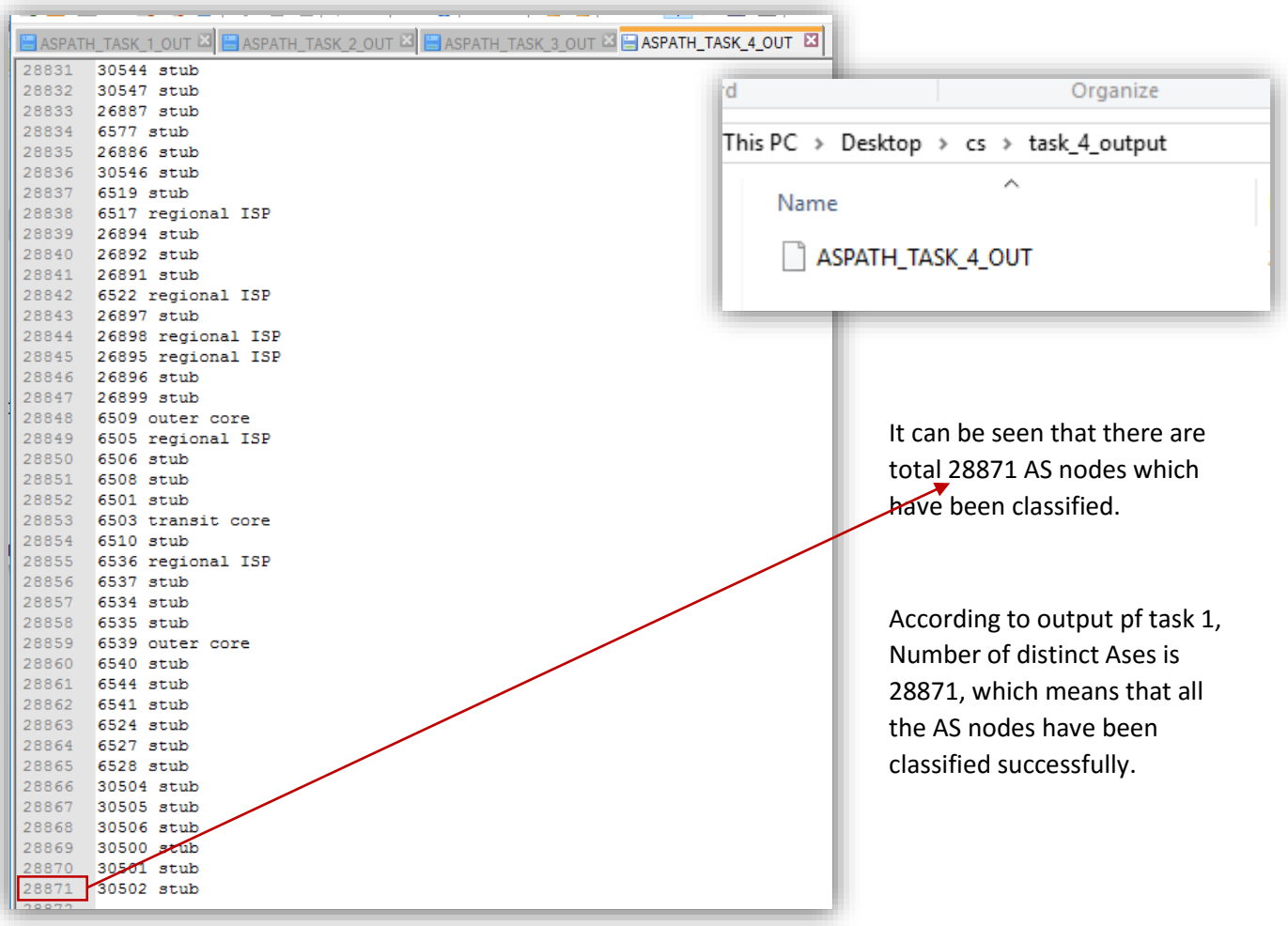
Command Pattern : `java Task4 <inputfile> <outputfile>`

Real Command to execute task 4 :

```
java Task4 ~/cs5229/task_4_input/ASPATH_TASK_3_OUT ~/cs5229/task_4_output/ASPATH_TASK_4_OUT
```

5.3 Result

To be able to verify the output file locally, the output file is copied back to sunfire and then downloaded to local windows machine. We then open it using notedpad ++.

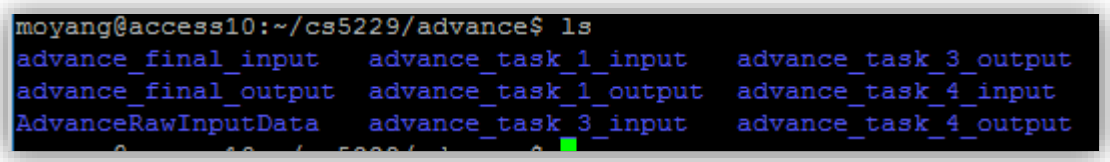


It can be seen that there are total 28871 AS nodes which have been classified.

According to output pf task 1, Number of distinct Ases is 28871, which means that all the AS nodes have been classified successfully.

6. Advance Part

6.1 Prepare input data



To get the edge count as well as distribution of degree, all the data for year 01, 06 and 11 need to go through task 1, task 3 and task 4 to get the output files from each task. These three kinds of output files will be copied into `advance_final_input` folder to enable the Advance task.

6.2 Run step:

1. Run task 1 for data from year 01, 06, 11
2. Run task 3 based on the results from step 1 for year 01, 06 , 11
3. Run task 4 based on the results from step 1 for year 01, 06, 11
4. Run advance task based on the results from step 1, 2 and 3

6.3 Run Command

Step 1:

```
java Task1 ~/cs5229/advance/advance_task_1_input/01_ASPATH ~/cs5229/advance/advance_task_1_output/01_TASK_1_OUT

java Task1 ~/cs5229/advance/advance_task_1_input/06_ASPATH ~/cs5229/advance/advance_task_1_output/06_TASK_1_OUT

java Task1 ~/cs5229/advance/advance_task_1_input/11_ASPATH ~/cs5229/advance/advance_task_1_output/11_TASK_1_OUT
```

Step 2:

```
java Task3 1 60 ~/cs5229/advance/advance_task_3_input/01_TASK_1_OUT ~/cs5229/advance/advance_task_3_output/01_TASK_3_OUT

java Task3 1 60 ~/cs5229/advance/advance_task_3_input/06_TASK_1_OUT ~/cs5229/advance/advance_task_3_output/06_TASK_3_OUT

java Task3 1 60 ~/cs5229/advance/advance_task_3_input/11_TASK_1_OUT ~/cs5229/advance/advance_task_3_output/11_TASK_3_OUT
```

Step 3:

```
java Task4 ~/cs5229/advance/advance_task_4_input/01_TASK_3_OUT
~/cs5229/advance/advance_task_4_output/01_TASK_4_OUT

java Task4 ~/cs5229/advance/advance_task_4_input/06_TASK_3_OUT
~/cs5229/advance/advance_task_4_output/06_TASK_4_OUT

java Task4 ~/cs5229/advance/advance_task_4_input/11_TASK_3_OUT
~/cs5229/advance/advance_task_4_output/11_TASK_4_OUT
```

Step 4 (final Step):

Command Pattern : `java Advance <asPathFile> <edgeFile> <classFile> <outputFile>`

Real Command to run advance task:

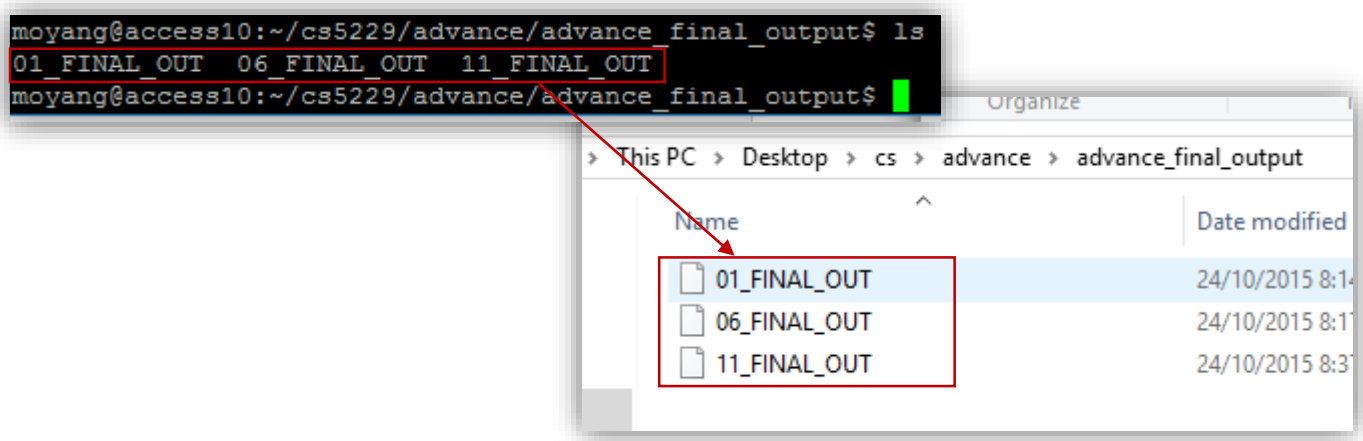
```
java Advance ~/cs5229/advance/advance_final_input/01_TASK_1_OUT
~/cs5229/advance/advance_final_input/01_TASK_3_OUT
~/cs5229/advance/advance_final_input/01_TASK_4_OUT
~/cs5229/advance/advance_final_output/01_FINAL_OUT

java Advance ~/cs5229/advance/advance_final_input/06_TASK_1_OUT
~/cs5229/advance/advance_final_input/06_TASK_3_OUT
~/cs5229/advance/advance_final_input/06_TASK_4_OUT
~/cs5229/advance/advance_final_output/06_FINAL_OUT

java Advance ~/cs5229/advance/advance_final_input/11_TASK_1_OUT
~/cs5229/advance/advance_final_input/11_TASK_3_OUT
~/cs5229/advance/advance_final_input/11_TASK_4_OUT
~/cs5229/advance/advance_final_output/11_FINAL_OUT
```

6.4 Result

To be able to verify the output file locally, the output file is copied back to sunfire and then downloaded to local windows machine. We then open it using notedpad ++.



6.4.1 Output file for Year 2001

ASPATH_TASK_1_OUT | ASPATH_TASK_2_OUT | ASPATH_TASK_3_OUT

```
Level,0,1,2,3,4
0,12,101,202,372,2934
1,129,67,468,419,3042
2,6,73,403,662,3515
3,2,2,6,342,2646

1,0.0,0.0,0.0,0.0,76.78459
2,6.060606,21.428572,13.221153,30.418945,97.03789
3,19.19192,21.428572,31.730768,57.1949,99.34059
4,22.222223,24.285715,43.269234,72.22222,99.80113
5,29.292929,41.428574,50.240387,80.327866,99.91627
6,36.363636,44.285713,55.528843,85.974495,99.979065
7,41.414143,47.142857,59.375,88.52459,99.979065
8,45.454548,48.57143,64.90385,90.80146,100.0
9,49.49495,50.0,68.269226,92.89618,100.0
10,52.525253,51.428574,71.15385,94.626595,100.0
20,66.66667,62.857143,87.74039,99.08926,100.0
30,81.818184,64.28571,92.06731,99.635704,100.0
40,89.898994,68.571434,94.47115,100.0,100.0
50,95.959595,78.57143,94.95192,100.0,100.0
60,95.959595,80.0,96.394226,100.0,100.0
70,95.959595,82.85715,96.875,100.0,100.0
80,96.969696,84.28571,97.83653,100.0,100.0
100,96.969696,84.28571,98.79808,100.0,100.0
200,97.9798,92.85714,99.27885,100.0,100.0
300,98.9899,94.28571,99.75961,100.0,100.0
400,98.9899,95.71428,99.75961,100.0,100.0
500,98.9899,97.14286,100.0,100.0,100.0
600,98.9899,97.14286,100.0,100.0,100.0
700,98.9899,98.57143,100.0,100.0,100.0
800,98.9899,98.57143,100.0,100.0,100.0
900,98.9899,100.0,100.0,100.0,100.0
1000,98.9899,100.0,100.0,100.0,100.0
2000,100.0,100.0,100.0,100.0,100.0
3000,100.0,100.0,100.0,100.0,100.0
4000,100.0,100.0,100.0,100.0,100.0
5000,100.0,100.0,100.0,100.0,100.0
6000,100.0,100.0,100.0,100.0,100.0
7000,100.0,100.0,100.0,100.0,100.0
8000,100.0,100.0,100.0,100.0,100.0
9000,100.0,100.0,100.0,100.0,100.0
10000,100.0,100.0,100.0,100.0,100.0
```

Inter-connectivity of ASes across levels.

Cumulative distribution of AS degree by level.

6.4.1 Output file for Year 2006

```
Level,0,1,2,3,4
0,1,57,80,218,2081
1,81,263,2675,1082,4630
2,11,317,2613,3246,17137
3,5,23,54,1210,8544
```

Inter-connectivity of ASes across levels.

```
1,0.0,0.0,0.0,0.0,45.314453
2,0.0,10.294118,2.141058,18.14394,90.39463
3,0.0,11.764706,4.911839,41.742424,97.30321
4,0.0,14.705883,8.816121,58.522724,99.0681
5,0.0,14.705883,14.105794,68.86364,99.635574
6,0.0,14.705883,19.269522,76.590904,99.79696
7,0.0,14.705883,25.566751,81.40151,99.88026
8,0.0,14.705883,31.3602,85.11364,99.921906
9,0.0,14.705883,37.153652,87.878784,99.94273
10,9.090909,14.705883,42.06549,90.26515,99.95835
20,36.363636,14.705883,69.39546,96.93182,100.0
30,54.545456,14.705883,80.85642,98.71212,100.0
40,63.636364,25.0,86.77582,99.431816,100.0
50,72.72727,47.058823,89.042816,99.65909,100.0
60,81.818184,50.0,90.42821,99.77273,100.0
70,81.818184,55.88235,91.8136,99.84849,100.0
80,90.909096,61.764706,93.324936,99.84849,100.0
100,90.909096,69.117645,95.08816,99.92424,100.0
200,90.909096,77.94118,97.607056,100.0,100.0
300,90.909096,86.7647,98.23678,100.0,100.0
400,90.909096,91.17647,98.740555,100.0,100.0
500,90.909096,95.588234,98.99245,100.0,100.0
600,90.909096,98.52941,99.24433,100.0,100.0
700,90.909096,98.52941,99.37028,100.0,100.0
800,90.909096,98.52941,99.49622,100.0,100.0
900,90.909096,98.52941,99.49622,100.0,100.0
1000,90.909096,98.52941,99.49622,100.0,100.0
2000,90.909096,100.0,100.0,100.0,100.0
3000,100.0,100.0,100.0,100.0,100.0
4000,100.0,100.0,100.0,100.0,100.0
5000,100.0,100.0,100.0,100.0,100.0
6000,100.0,100.0,100.0,100.0,100.0
7000,100.0,100.0,100.0,100.0,100.0
8000,100.0,100.0,100.0,100.0,100.0
9000,100.0,100.0,100.0,100.0,100.0
10000,100.0,100.0,100.0,100.0,100.0
```

Cumulative distribution of AS degree by level.

6.4.1 Output file for Year 2011

```
Level,0,1,2,3,4
0,0,58,279,585,2328
1,29,505,5226,2931,7585
2,12,333,2815,5726,28225
3,3,30,113,2453,16674
```

Inter-connectivity of ASes across levels.

```
1,0.0,0.0,0.0,0.0,48.769276
2,0.0,0.0,2.0952382,16.652683,89.2628
3,0.0,0.0,4.6666665,38.926174,96.40654
4,20.0,0.0,7.1428576,54.467285,98.41765
5,20.0,0.0,10.285714,64.555374,99.1826
6,40.0,0.0,14.761906,72.294464,99.55583
7,40.0,0.0,17.619047,77.60067,99.74399
8,40.0,0.0,20.952381,81.92114,99.84577
9,40.0,0.0,24.47619,85.57047,99.90746
10,40.0,0.0,30.285713,87.856544,99.95682
20,40.0,0.0,62.0,96.26678,99.98766
30,40.0,0.0,72.952385,98.11242,100.0
40,40.0,0.0,80.190475,98.993286,100.0
50,60.000004,0.0,85.333336,99.49664,100.0
60,60.000004,16.41791,87.42857,99.72735,100.0
70,60.000004,31.343285,89.71428,99.81124,100.0
80,60.000004,38.80597,90.66667,99.85319,100.0
100,60.000004,52.238804,92.571434,99.895134,100.0
200,80.0,76.11941,96.666664,100.0,100.0
300,80.0,80.597015,98.0,100.0,100.0
400,80.0,86.56716,98.57143,100.0,100.0
500,80.0,86.56716,98.952385,100.0,100.0
600,80.0,88.0597,99.2381,100.0,100.0
700,80.0,89.55224,99.2381,100.0,100.0
800,80.0,92.537315,99.333336,100.0,100.0
900,80.0,95.522385,99.42857,100.0,100.0
1000,80.0,97.01493,99.42857,100.0,100.0
2000,80.0,98.50746,99.90476,100.0,100.0
3000,80.0,100.0,100.0,100.0,100.0
4000,100.0,100.0,100.0,100.0,100.0
5000,100.0,100.0,100.0,100.0,100.0
6000,100.0,100.0,100.0,100.0,100.0
7000,100.0,100.0,100.0,100.0,100.0
8000,100.0,100.0,100.0,100.0,100.0
9000,100.0,100.0,100.0,100.0,100.0
10000,100.0,100.0,100.0,100.0,100.0
```

Cumulative distribution of AS degree by level.

6.5 Analysis of Results:

6.5.1 Inter-connectivity tables for year 2001, 2006 and 2011

Level	0	1	2	3	4
0	12	101	202	372	2934
1	129	67	468	419	3042
2	6	73	403	662	3515
3	2	2	6	342	2646

2001

Level	0	1	2	3	4
0	1	57	80	218	2081
1	81	263	2675	1082	4630
2	11	317	2613	3246	17137
3	5	23	54	1210	8544

2006

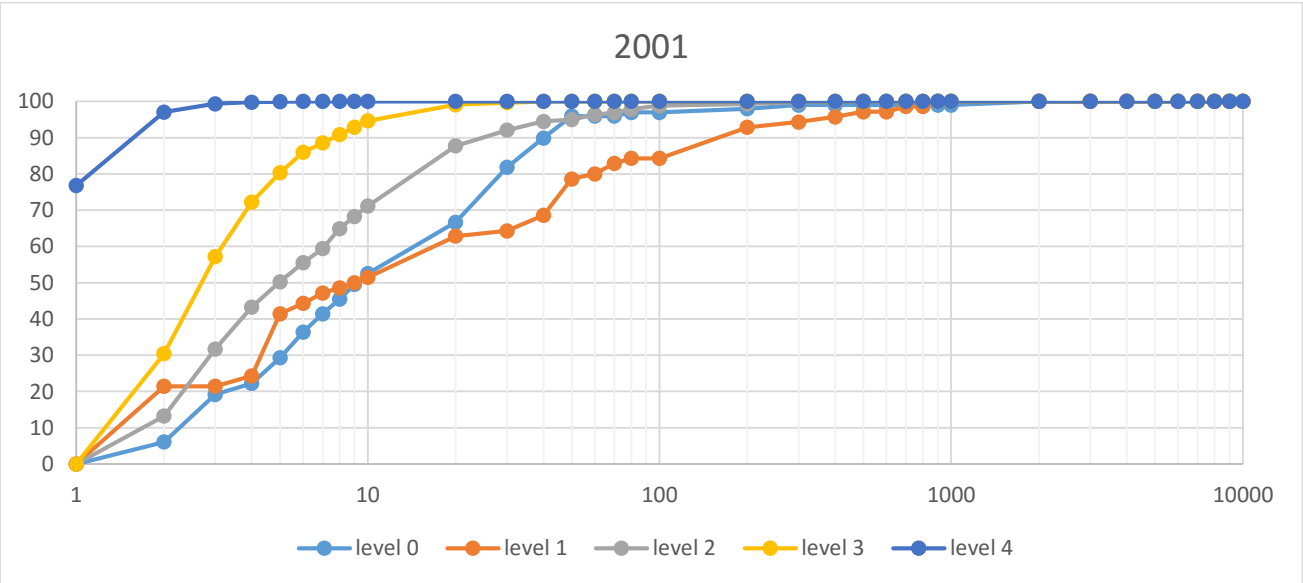
Level	0	1	2	3	4
0	0	58	279	585	2328
1	29	505	5226	2931	7585
2	12	333	2815	5726	28225
3	3	30	113	2453	16674

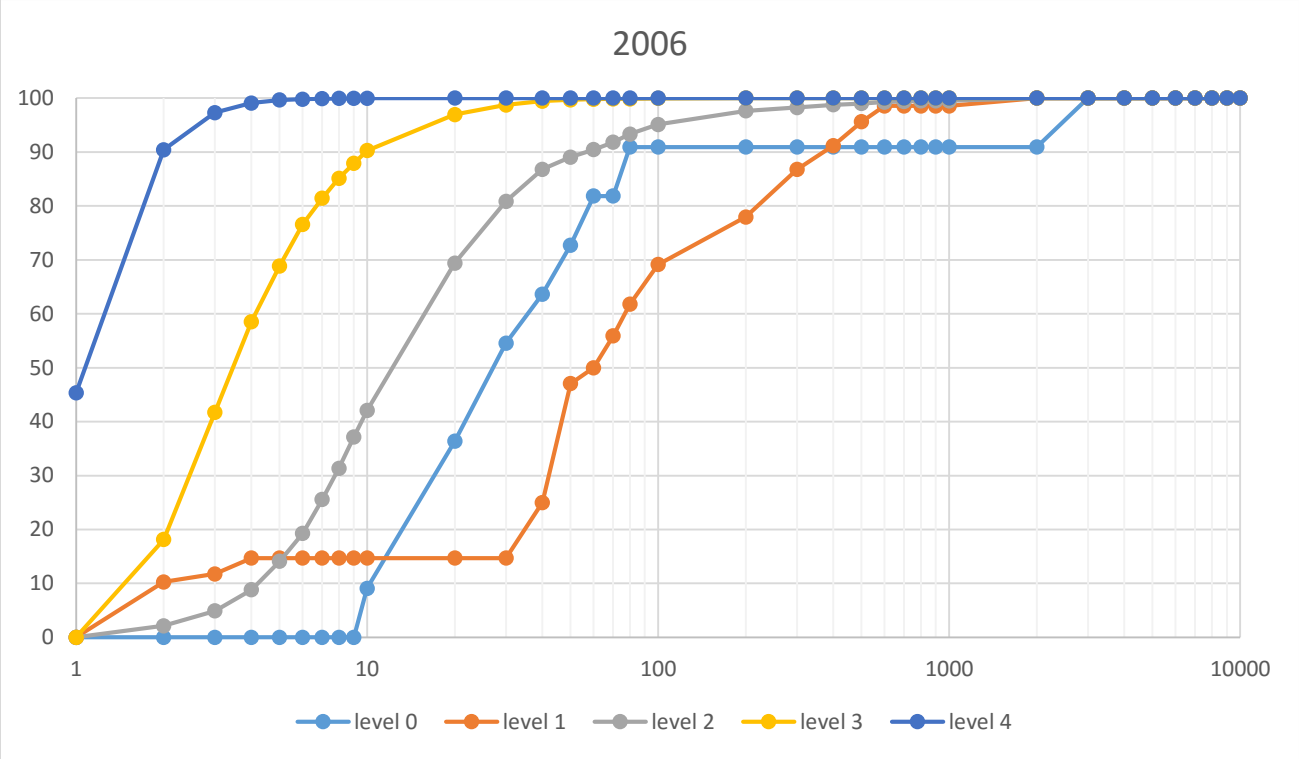
2011

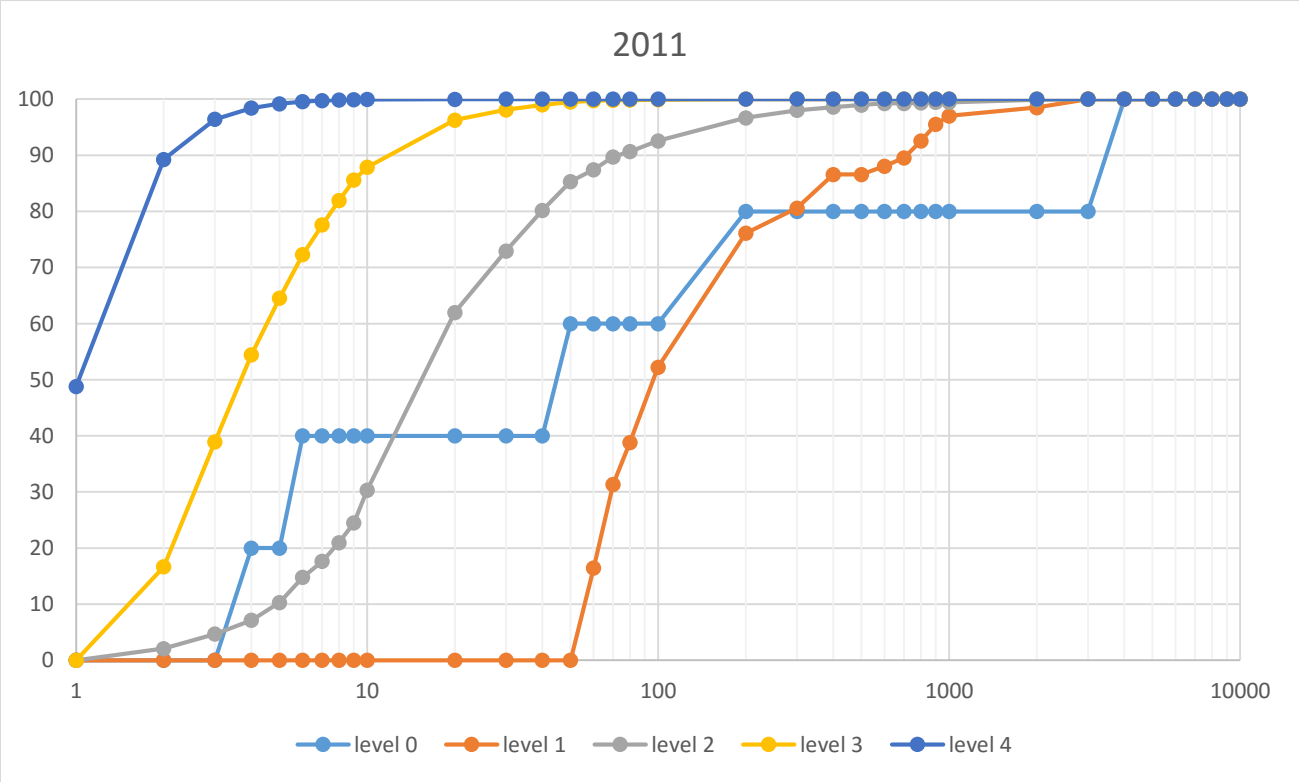
It can be seen that from 2001 to 2011

- Inter connectivity between level 0 AS is decreasing, which means dense cores are less connected with each other than before.
- Level 3 to level 4 connections have increased hugely. This means that regional ISP are more closely connected to normal users.

6.5.1 Cumulative distribution tables for year 2001, 2006 and 2011







From above, we can see that:

- Level 1 transit cores distribution are shifting from low degree to high degree
- Level 4 Customer distribution have also seen an huge shift from low degree to high degree, which could mean that there are more and more c2p relations from level 4 to upper levels.

We can predict that level 4 and level 1 distribution will keep shift to higher degree while others may see a much slower change.