Program Structures and Algorithms

Spring 2023 (Section 3)

Assignment 4 – WQUPC

Name: Abhinav Choudhary

NUID: 002780326

Task

- To implement height-weighted quick union with path compression by modifying find(), mergeComponents(), and doPathCompression() methods of the UF_HWQUPC class.
- Create a new class called "UnionFind"
  - This class will have 2 methods main() and count().
  - The main() method's task is to get user input regarding number of sites or run a predefined set of number of sites, pass those values to count() method and print the result.
  - The count() method is responsible for creating an object of type UF_HWQUPC, generate random numbers between 0 and n-1 and loop through n until all sites are connected. It counts the number of connections and returns the number of connections as output.
- To determine a relationship between number of sites (n) and number of connections (m). This relationship is determined by plotting a graph between n and m.

Relationship Conclusion

By running the quick union program multiple times with increasing values of n, the number of connections (m) comes closest to the value of $nlog_{10}n$.

Thus, the most likely relationship between number of sites (n) and number of connections (m) with weighted quick union and random values of n is:

$$m = nlog_{10}n$$

Evidence to support conclusion

After running the program multiple times (i.e. multiple runs) for increasing values of n, following details were gathered:

Number of Sites (n)	Connections (m)						$n*log_2n$	$n*log_{10}n$
	Run 1	Run 2	Run 3	Run 4	Run 5	Average		
500	1974	1507	1474	2655	1655	1853	4482.89	1349.48
1000	3498	4325	3891	3071	4456	3848.2	9965.78	3000
2000	10080	9054	11130	7952	6666	8976.4	21931.56	6602.05
4000	17221	19498	16256	14915	18645	17307	47863.13	14408.23
8000	33999	49055	57480	36233	38183	42990	103726.27	31224.71
16000	77575	67104	82416	96114	67451	78132	223452.54	67265.91

In this table, by comparing average of number of connections of all the runs, we get the closest to the values of $nlog_{10}n$. Linear values like that of n would be very small, especially at larger values of n and when compared with other graphs like n^2 or n^3 or something else, their values tend to be very large. Even when compared to $nlog_2n$, there is large disparity from the start.

Therefore, even though there could be outliers in the data, the relationship between number of sites (n) and number of connections (m) is most likely or most closest to that of $nlog_{10}n$.

Following are the screenshots from the UnionFind class:

```
Console ×   Debug Shell   Problems   Executables
<terminated> UnionFind [Java Application] C:\Program Files\Java\jdk-18.0.2.1\bin\javaw.exe  (09-Feb-2023, 10:38:46 pm – 10:38:47 pm) [pid: 15312]
Runs: 1
Sites: 500
Connections: 1974
Sites: 1000
Connections: 3498
Sites: 2000
Connections: 10080
Sites: 4000
Connections: 17221
Sites: 8000
Connections: 33999
Sites: 16000
Connections: 77575
Runs: 2
Sites: 500
Connections: 1507
Sites: 1000
Connections: 4325
Sites: 2000
Connections: 9054
Sites: 4000
Connections: 19498
Sites: 8000
Connections: 49055
Sites: 16000
Connections: 67104
```

```
Runs: 3
Sites: 500
Connections: 1474
Sites: 1000
Connections: 3891
Sites: 2000
Connections: 11130
Sites: 4000
Connections: 16256
Sites: 8000
Connections: 57480
Sites: 16000
Connections: 82416
Runs: 4
Sites: 500
Connections: 2655
Sites: 1000
Connections: 3071
Sites: 2000
Connections: 7952
Sites: 4000
Connections: 14915
Sites: 8000
Connections: 36233
Sites: 16000
Connections: 96114
```

```
Runs: 5
Sites: 500
Connections: 1655
Sites: 1000
Connections: 4456
Sites: 2000
Connections: 6666
Sites: 4000
Connections: 18645
Sites: 8000
Connections: 38183
Sites: 16000
Connections: 67451
```

Graphical Representation

Following is the graph plotted between number of sites (n) and number of connections (m), with number of sites (n) along the y axis and number of connections (m) along the x axis. Solid blue line represents the function $nlog_{10}n$, and Solid Red line represents average of number of connections (m) obtained from the excel sheet.

(Desmos Graphing Calculator was used to plot the points and create the graph: Desmos | Graphing Calculator)
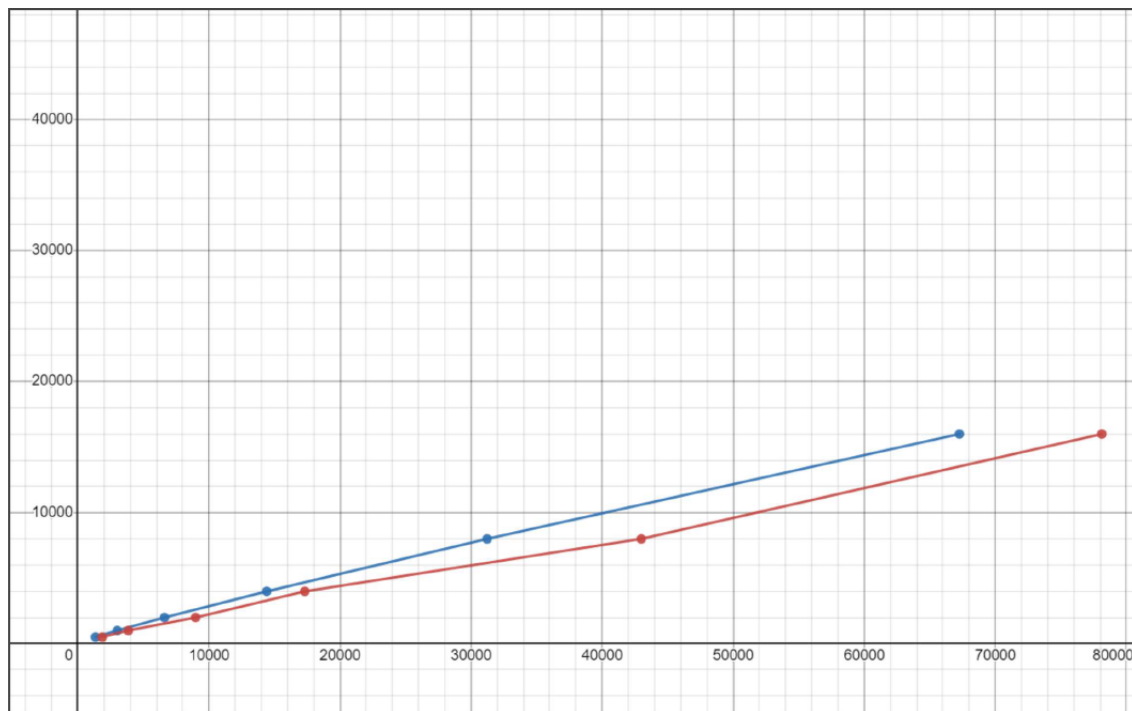
Graph created with Excel:
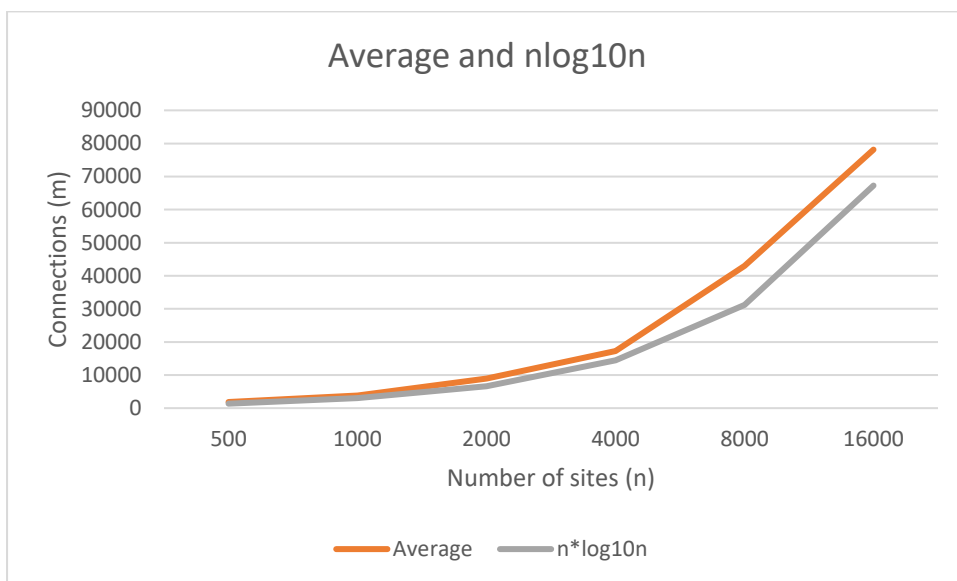
Even though the graph created with excel may appear different than traditional graph of $n\log_{10}n$ we can see that in both the graph created by excel and plotted using Desmos Graphing Calculator, the number of connections (m) is relatively closer to $n\log_{10}n$ .Thus supporting our argument.

Unit Test

Below is the screenshot of runs of different test files:

UF_HWQUPC_Test.java