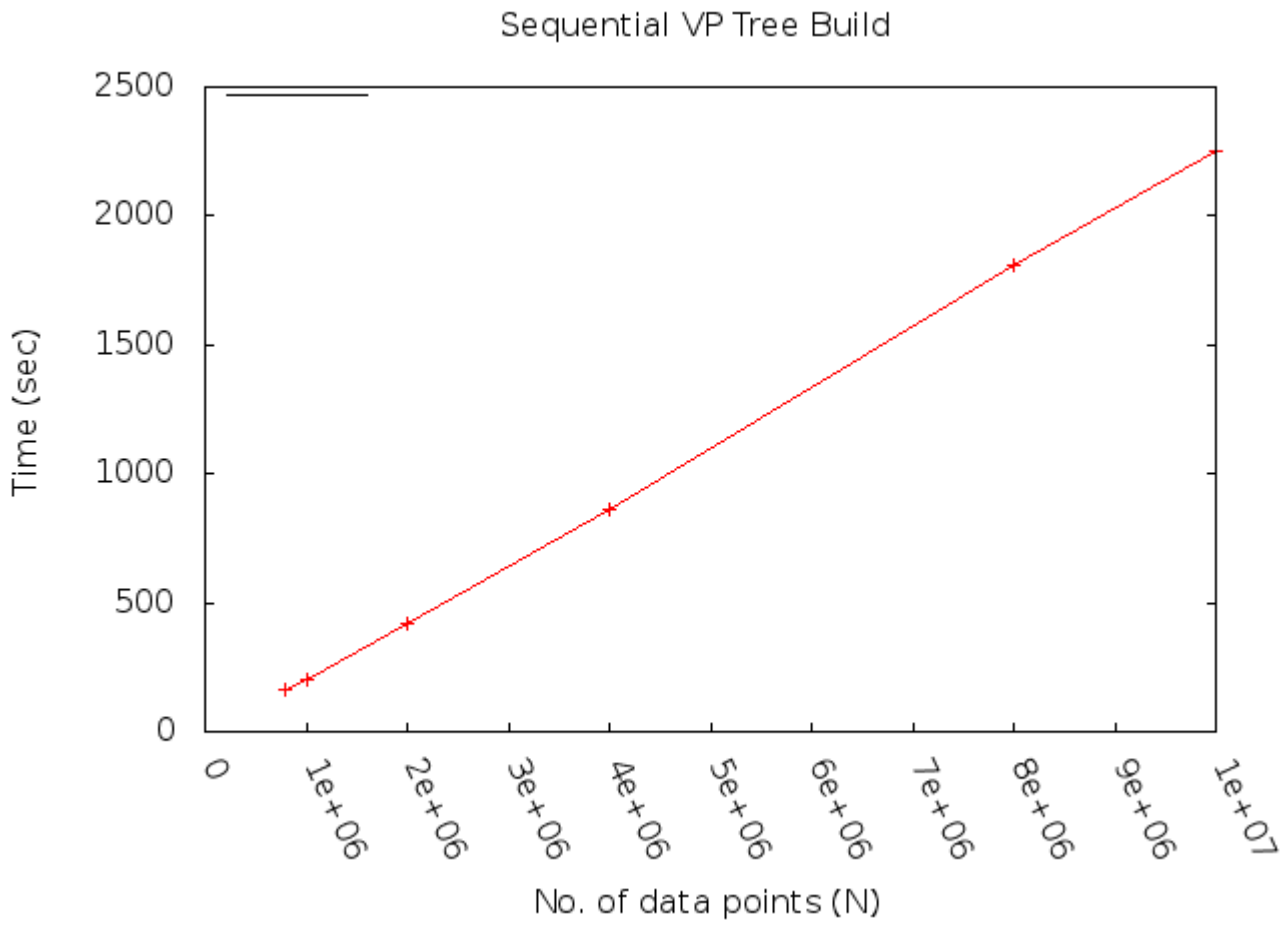
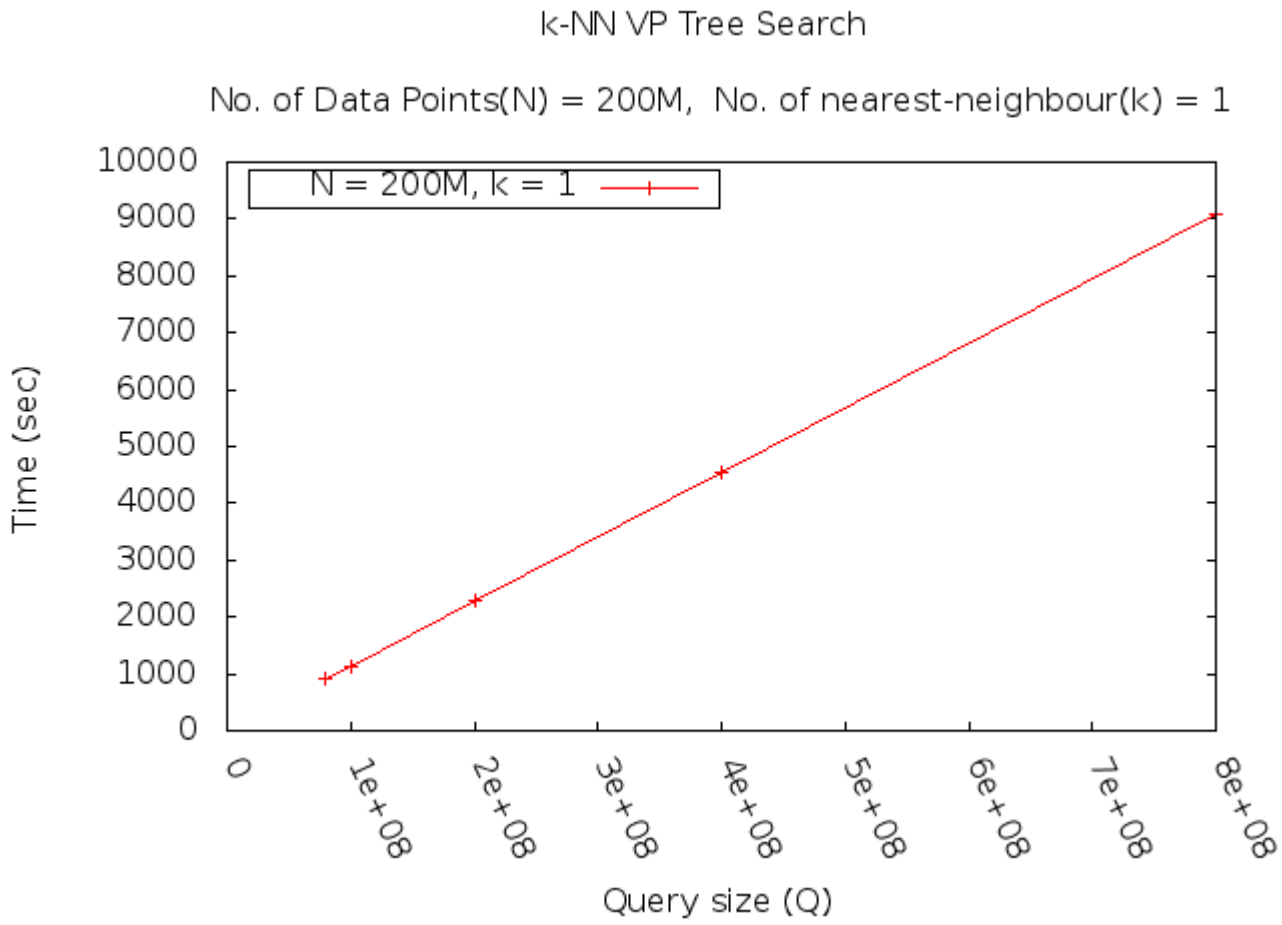
I. **Sequential VP Tree Build Process**



Graph 1: VP Tree Build Process

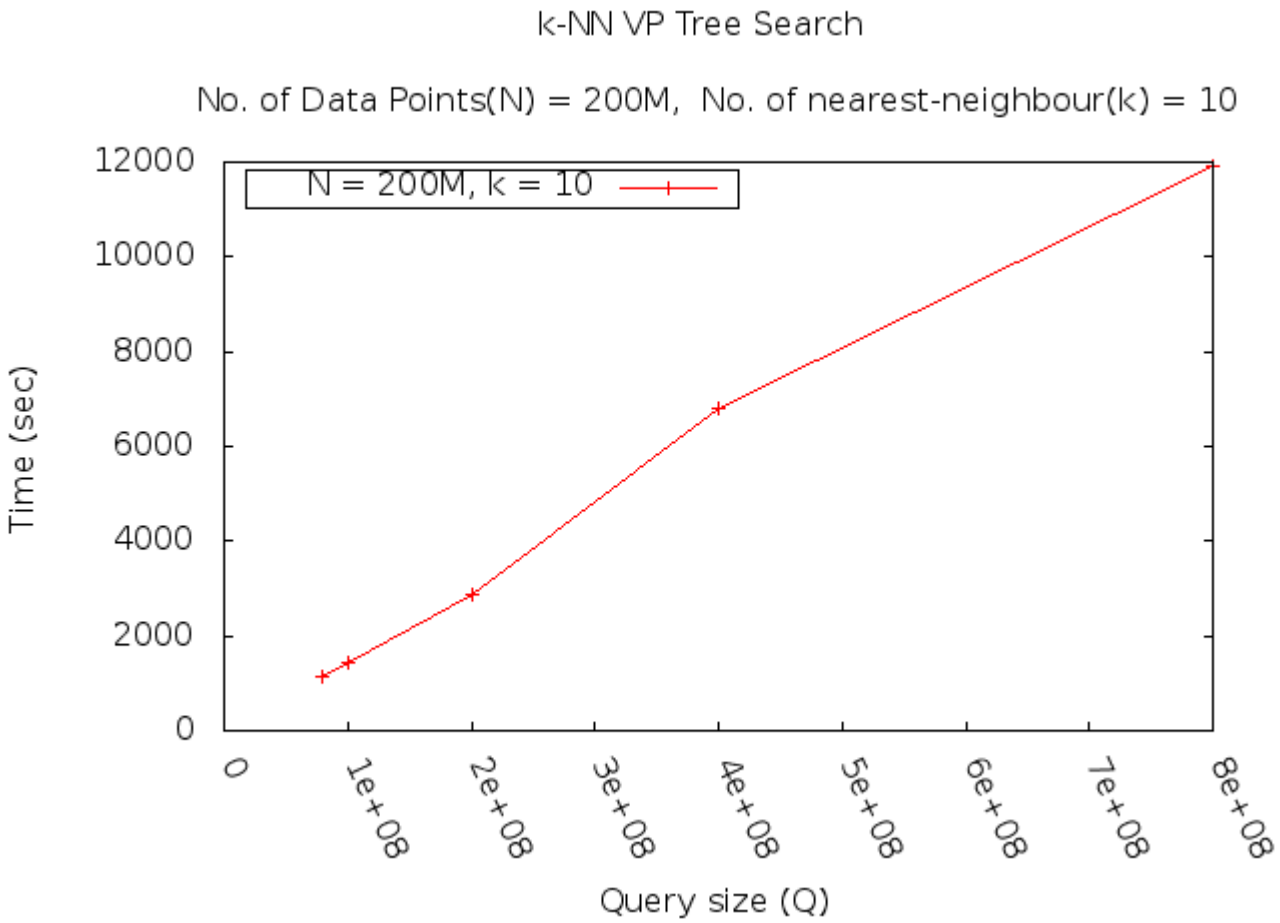
The Graph 1 shows time taken for the sequential build process for Vantage Point Tree on CGM7.

II. **Sequential Nearest Neighbour Search Queries**



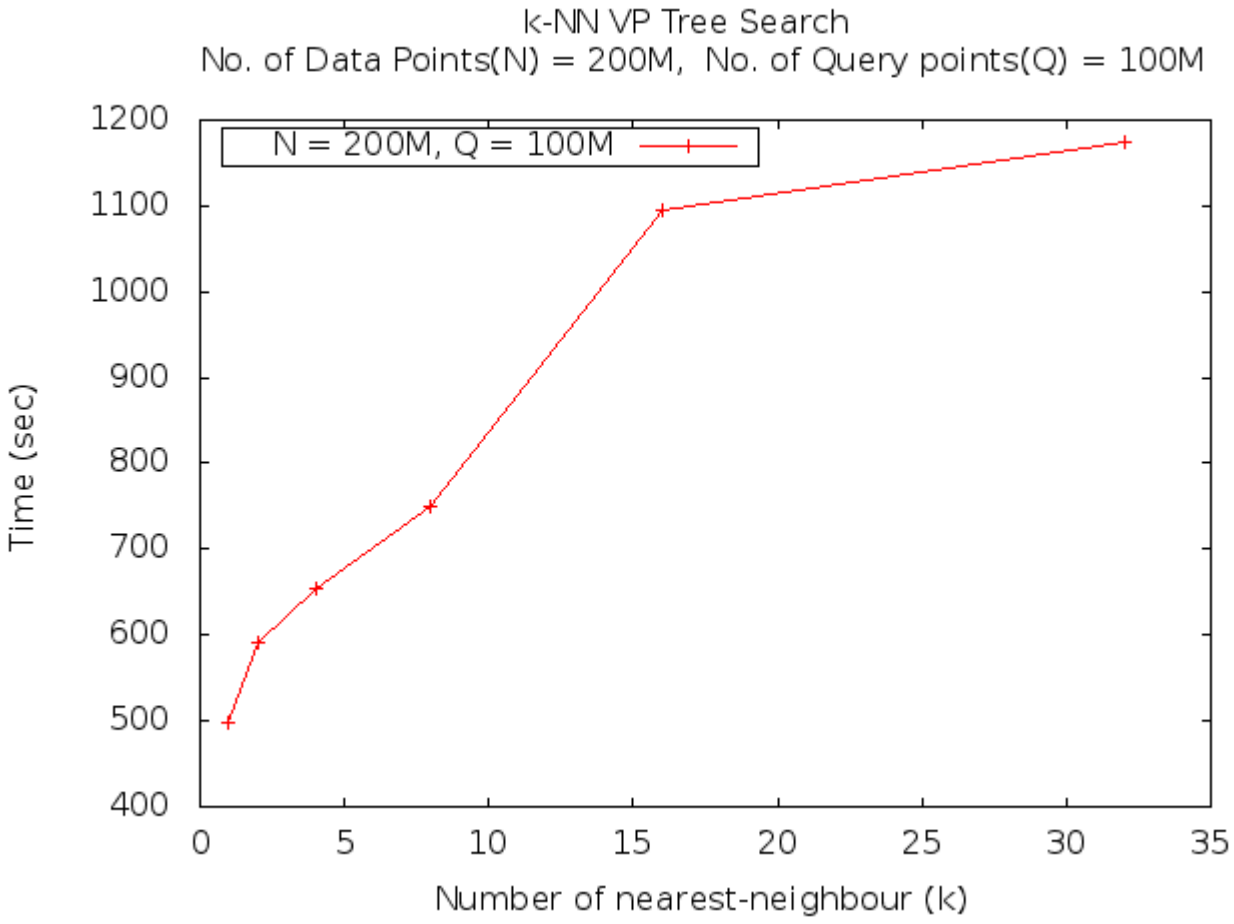
Graph 2

Graph 2 shows Varying querying size vs time taken to query “1” nearest neighbour on a tree size of 200M points. Data points following uniform distribution.



Graph 3: k-Nearest Neighbour Search, k = 10

Graph 2 shows Varying querying size vs time taken to query “1” nearest neighbour on a tree size of 200M points. Data points following uniform distribution.

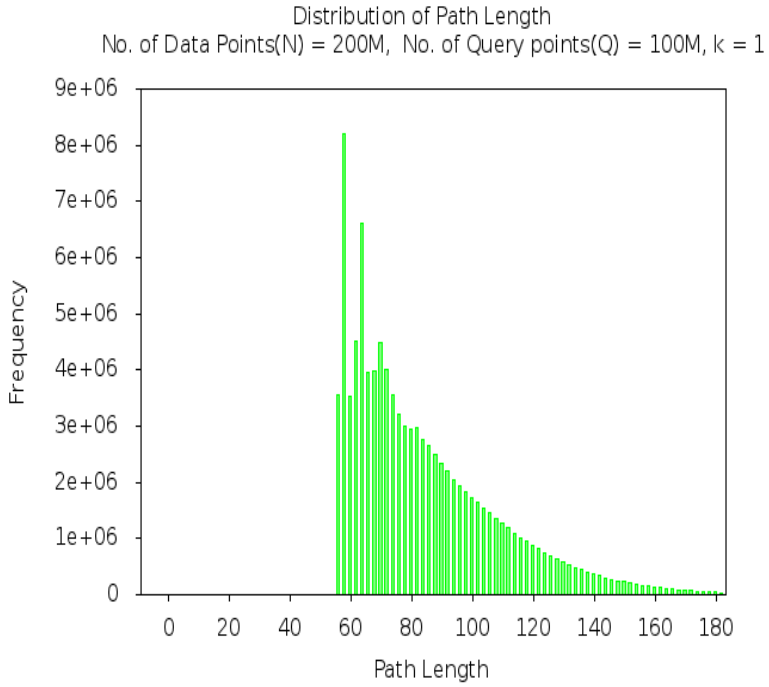


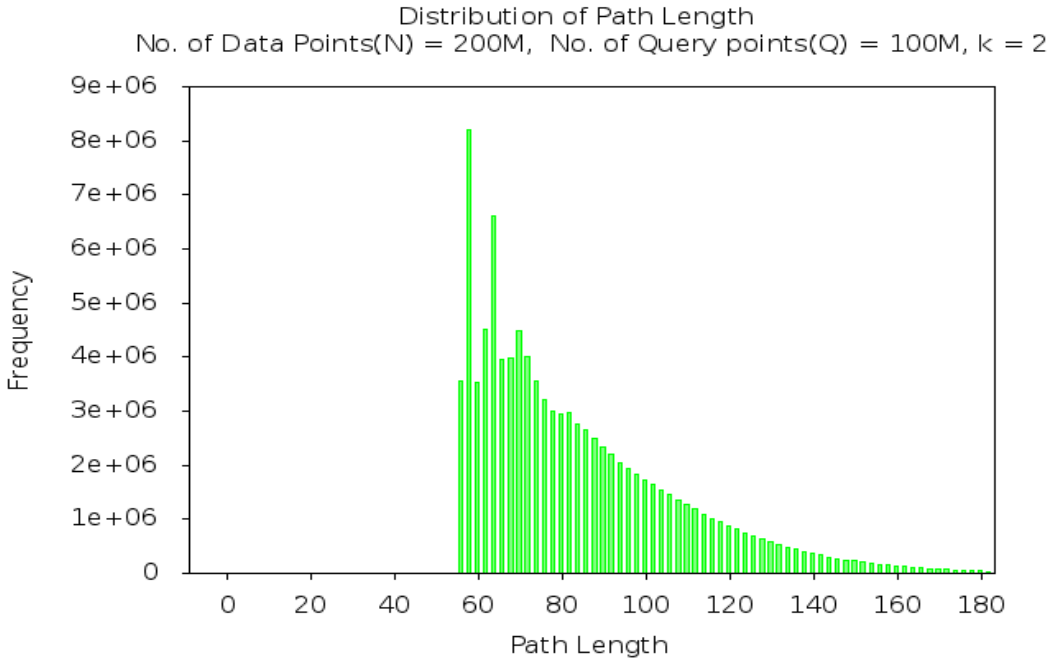
Graph 4

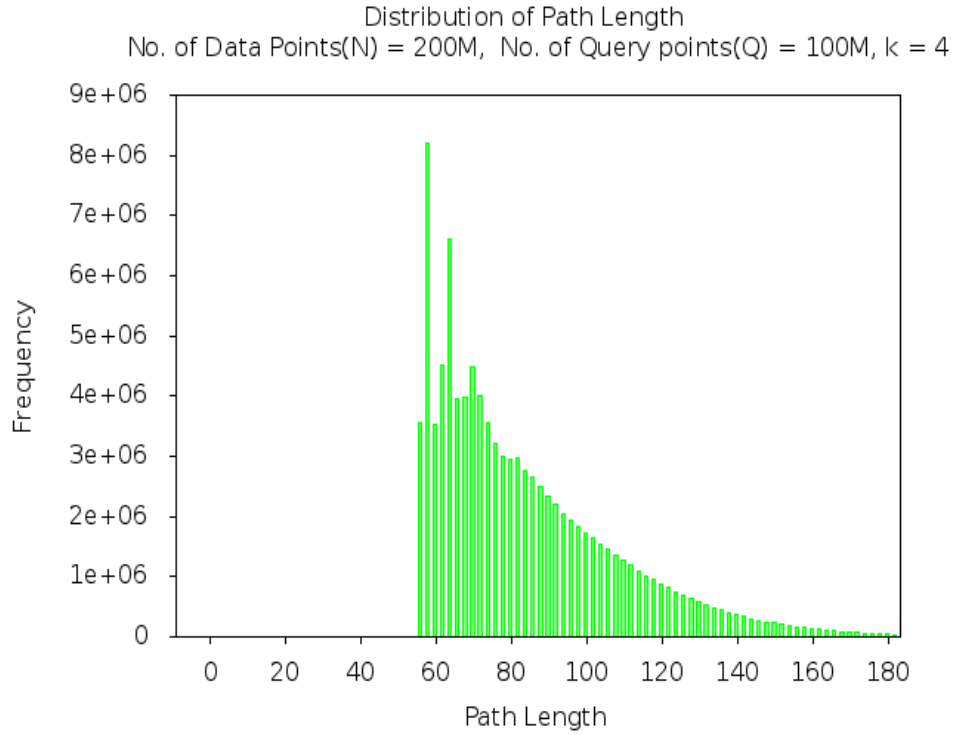
Graph 4 shows the time taken in querying k-NN on a VP Tree with 200M points and Query Set(Q) size being 100M points. K k varies from 1 , 2, 4, 8,16, 32

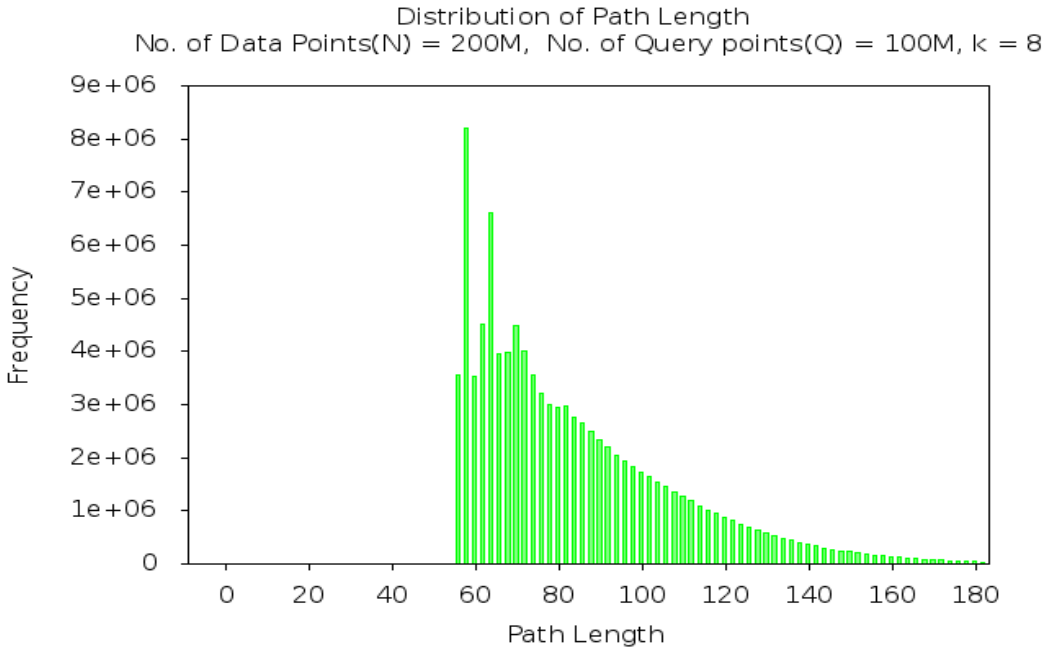
The following set of graphs show the path length(x-axis) for a particular value of k vs frequency

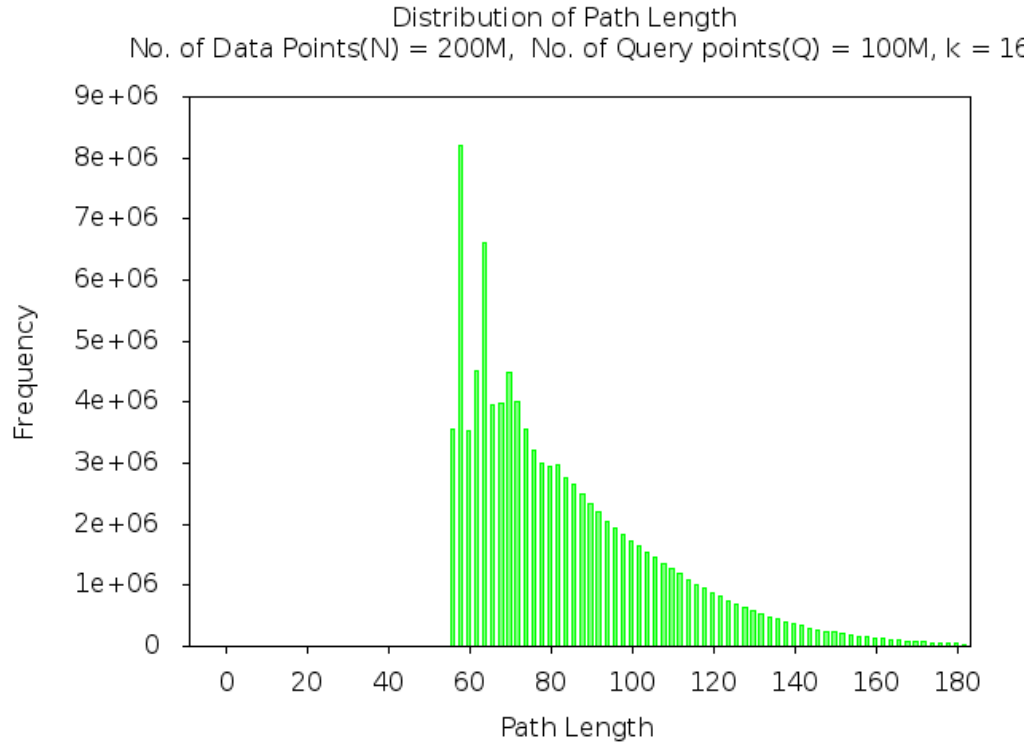
k = 1 , 2, 4, 8, 16, 32

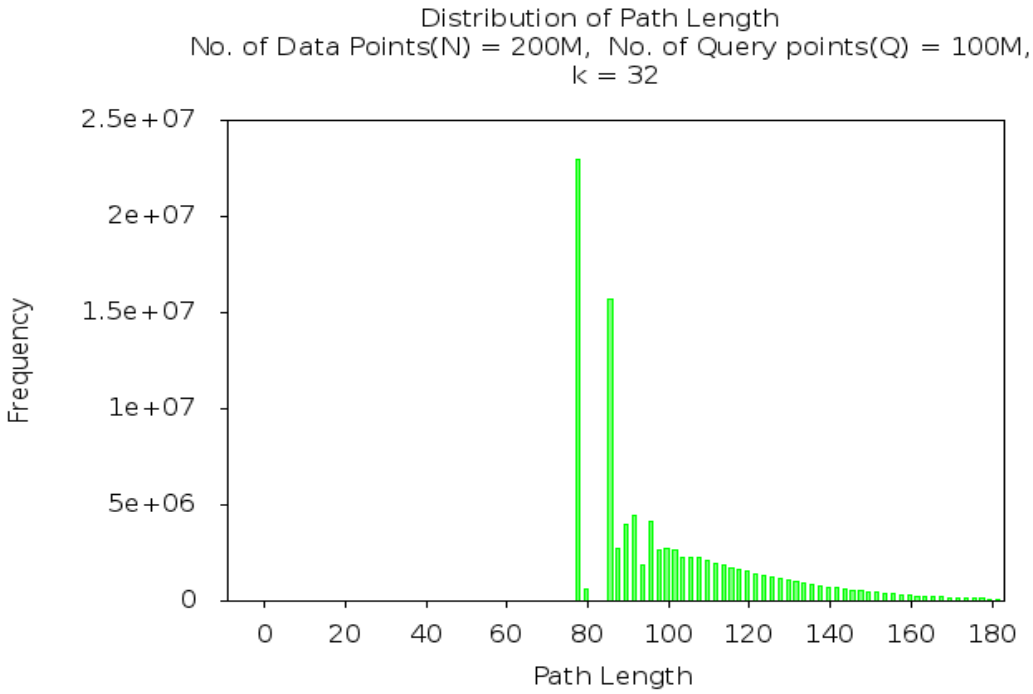




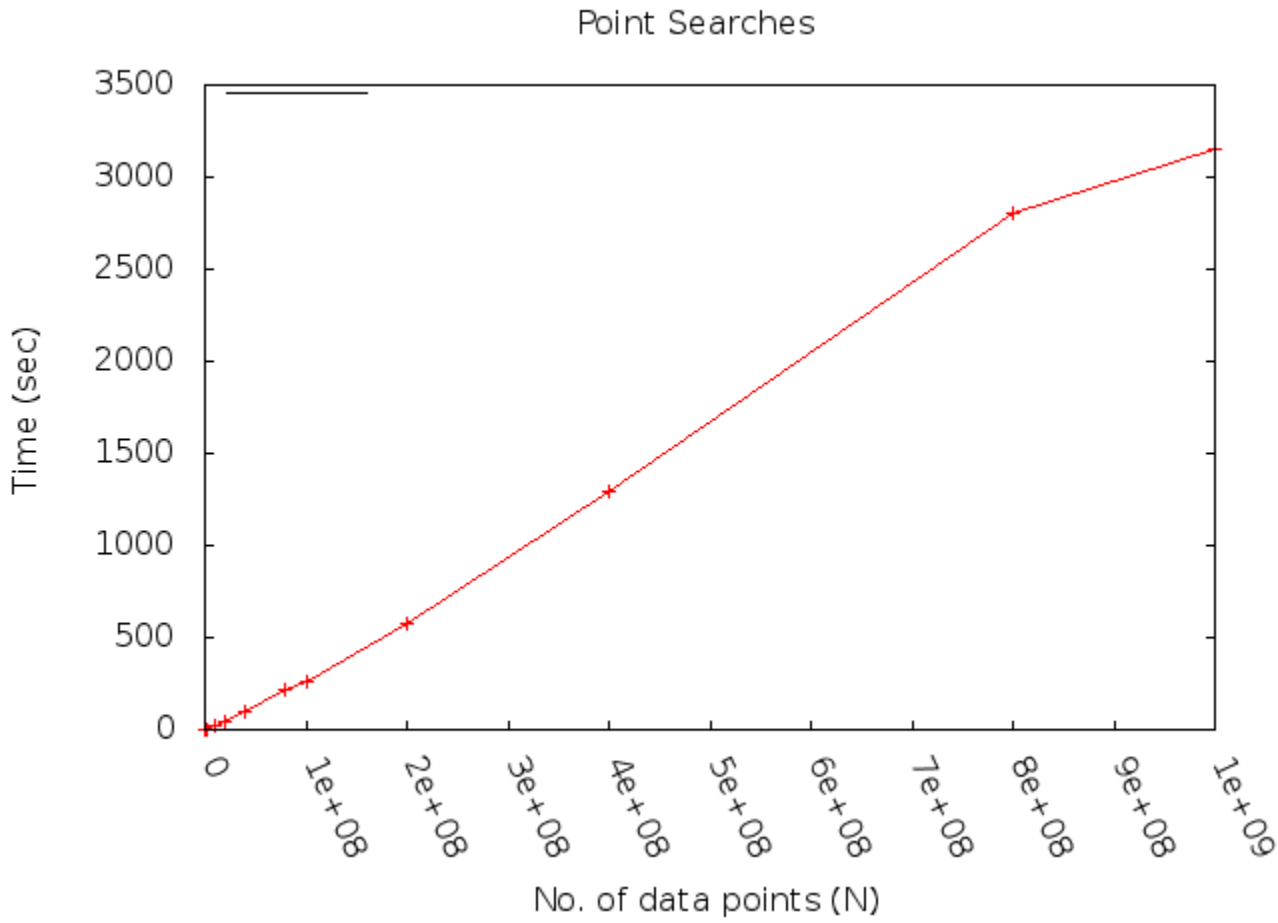

Distribution of Path Length
No. of Data Points(N) = 200M, No. of Query points(Q) = 100M, k = 4
Frequency
9e+06
8e+06
7e+06
6e+06
5e+06
4e+06
3e+06
2e+06
1e+06
0
0
20
40
60
80
100
120
140
160
180
Path Length


Distribution of Path Length
No. of Data Points(N) = 200M, No. of Query points(Q) = 100M, k = 8
Frequency
9e+06
8e+06
7e+06
6e+06
5e+06
4e+06
3e+06
2e+06
1e+06
0
0
20
40
60
80
100
120
140
160
180
Path Length


Distribution of Path Length
No. of Data Points(N) = 200M, No. of Query points(Q) = 100M, k = 16
Frequency
9e+06
8e+06
7e+06
6e+06
5e+06
4e+06
3e+06
2e+06
1e+06
0
0
20
40
60
80
100
120
140
160
180
Path Length

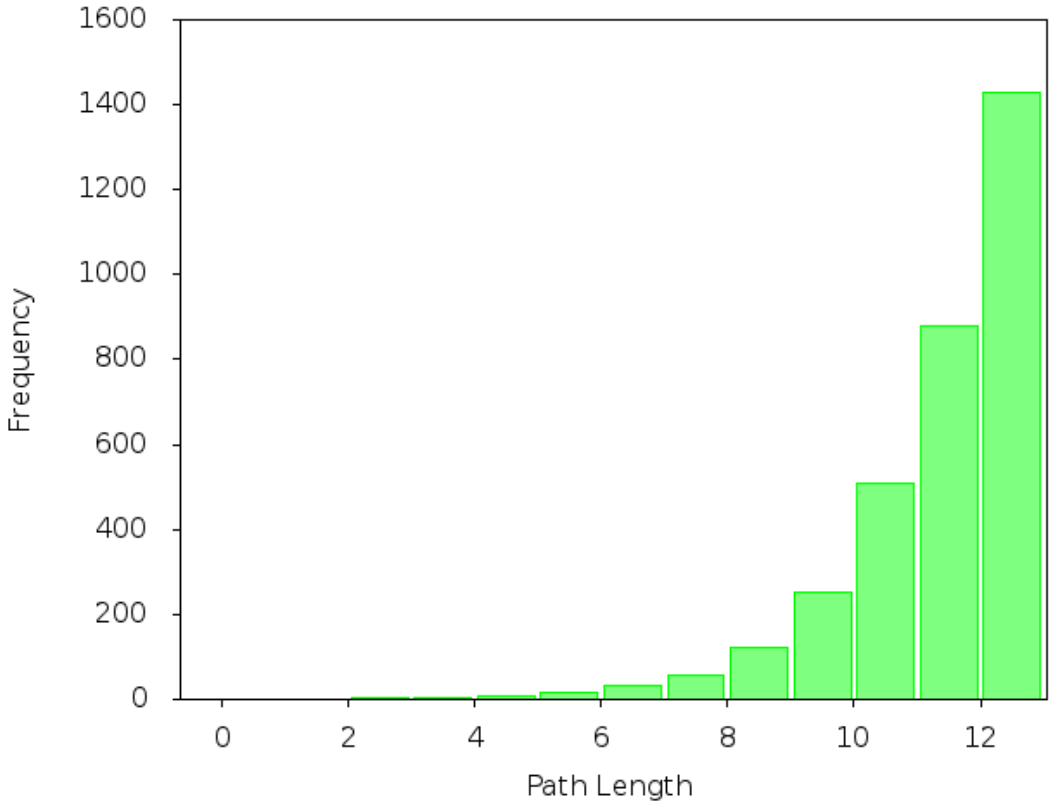

Distribution of Path Length
No. of Data Points(N) = 200M, No. of Query points(Q) = 100M,
k = 32
Frequency
2.5e+07
2e+07
1.5e+07
1e+07
5e+06
0
0
20
40
60
80
100
120
140
160
180
Path Length

III. Point Searches in VP Tree (exact query match)

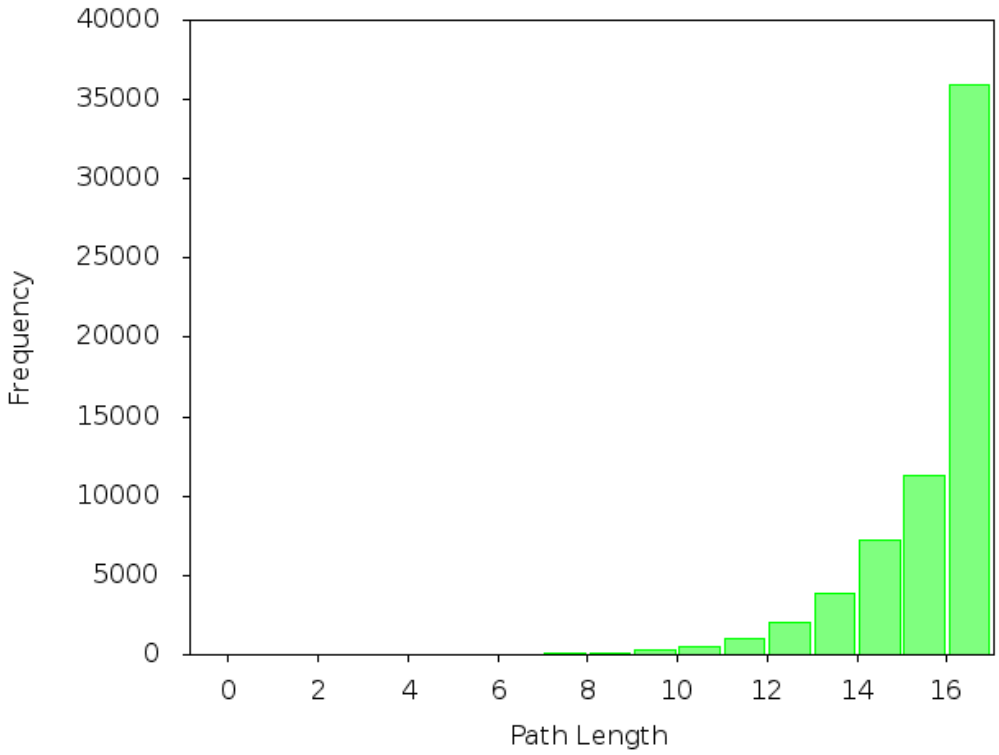


Following histogram depict the distribution of path length for data size varying from 10000 to 100M.

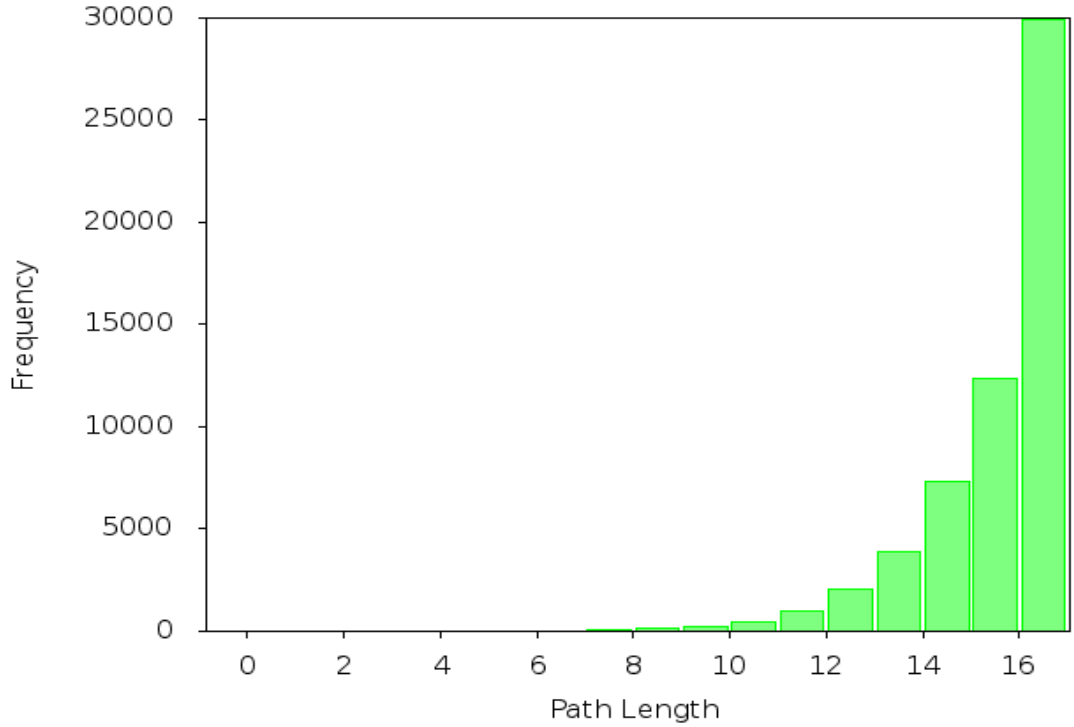
1. **10k**



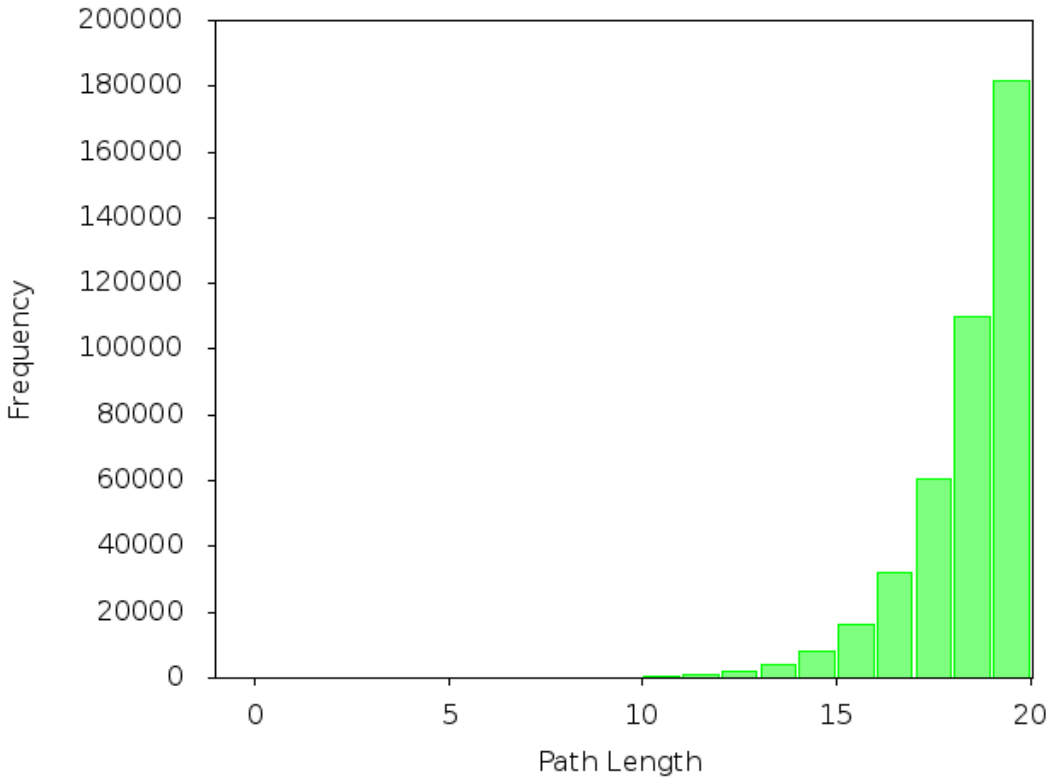
2. **80k**



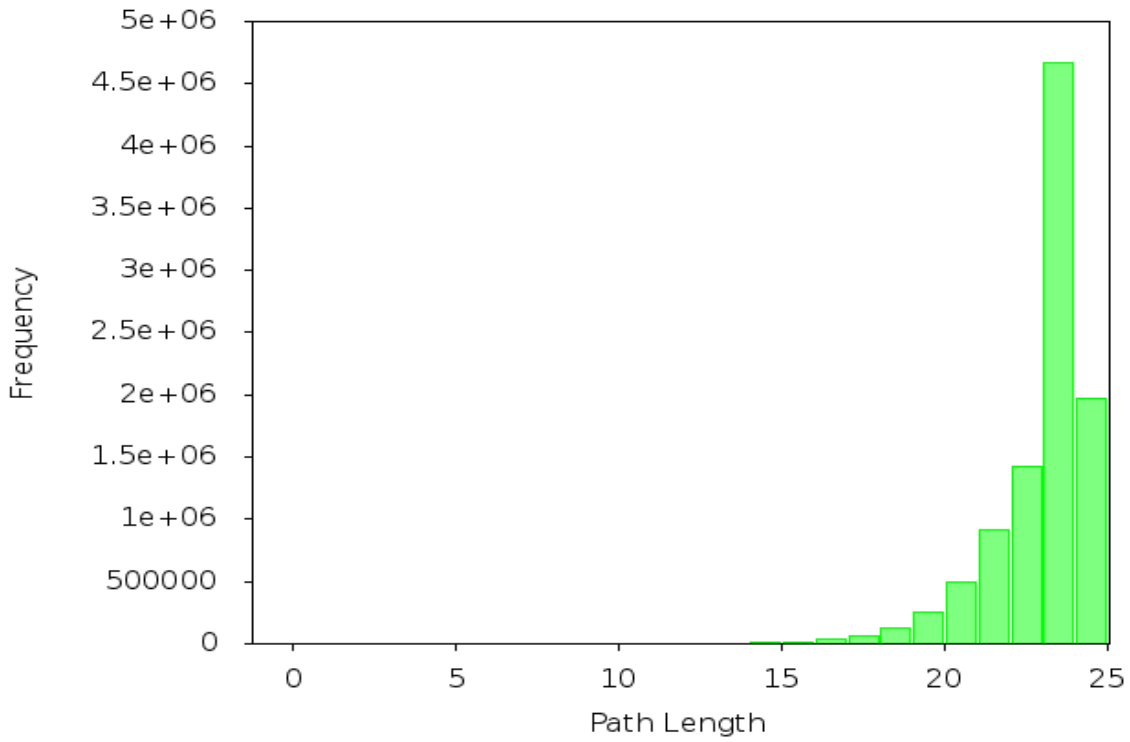
3. **100k**



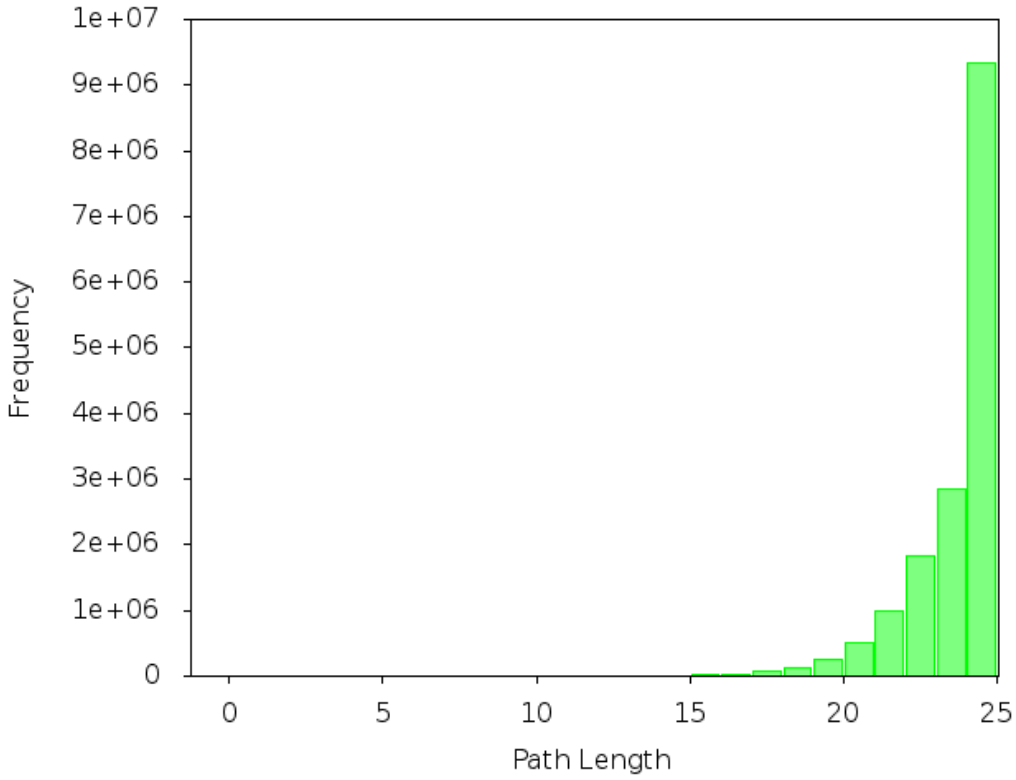
4. **1M**


200000
180000
160000
140000
120000
100000
80000
60000
40000
20000
0
Frequency
0
5
10
15
20
Path Length

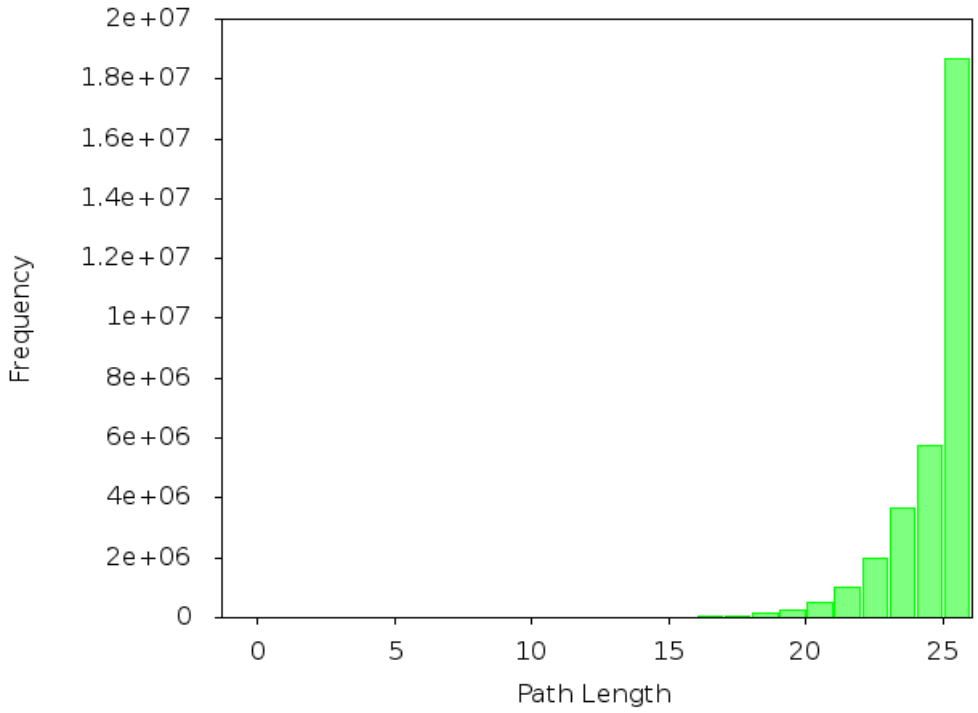
5. 10M


5e+06
4.5e+06
4e+06
3.5e+06
3e+06
2.5e+06
2e+06
1.5e+06
1e+06
500000
0
Frequency
0
5
10
15
20
25
Path Length

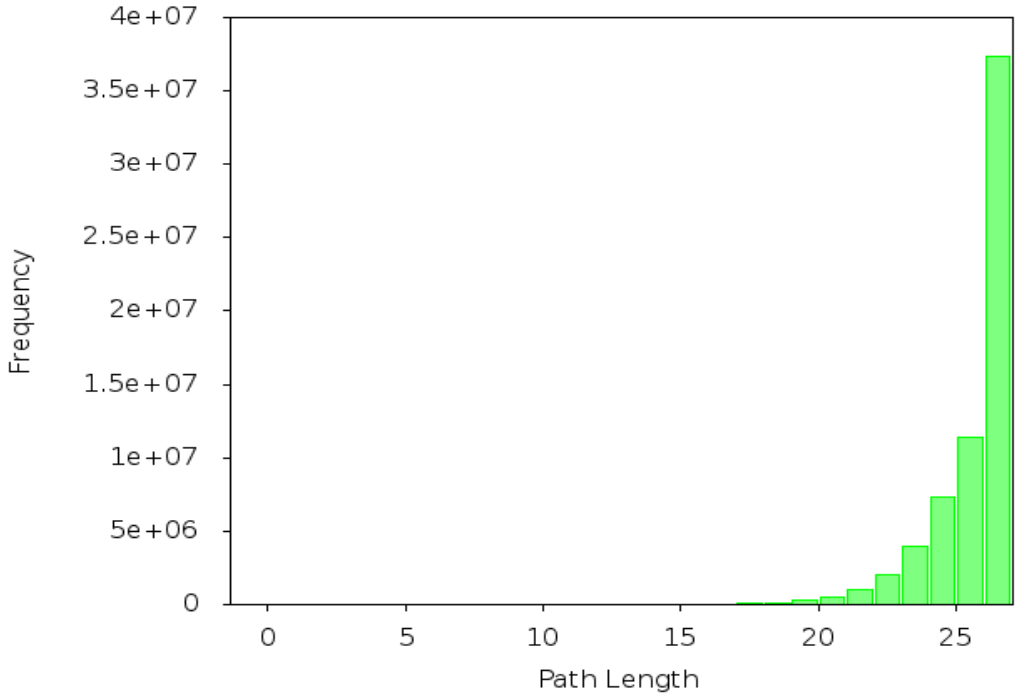
6. 20M


1e+07
9e+06
8e+06
7e+06
6e+06
5e+06
4e+06
3e+06
2e+06
1e+06
0
Frequency
0
5
10
15
20
25
Path Length

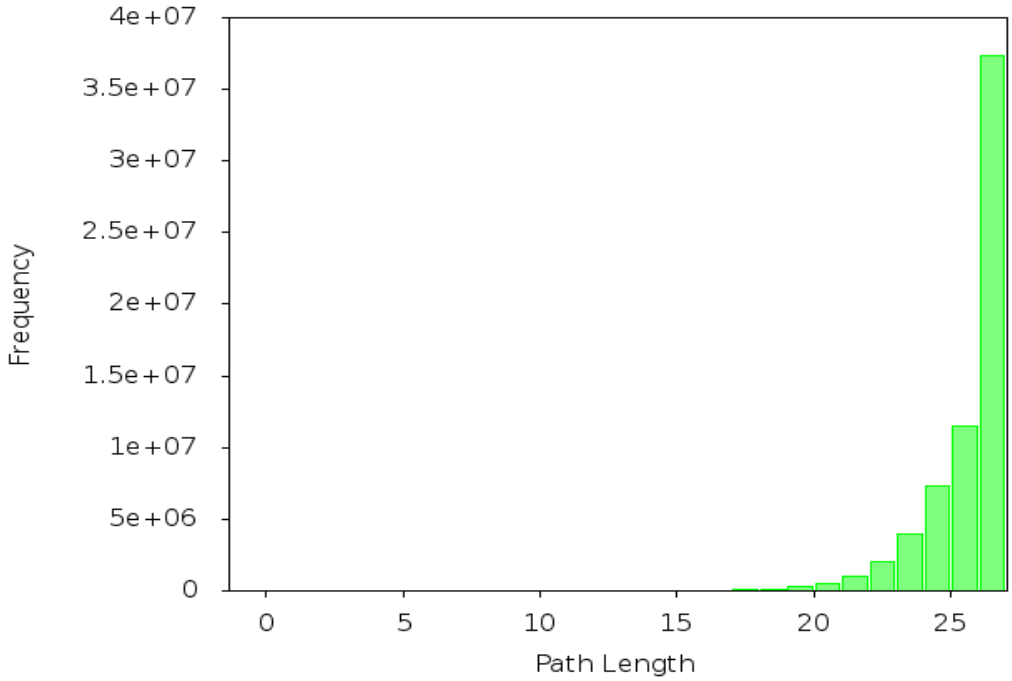
7. 40M


2e+07
1.8e+07
1.6e+07
1.4e+07
1.2e+07
1e+07
8e+06
6e+06
4e+06
2e+06
0
Frequency
0
5
10
15
20
25
Path Length

8. 80M


4e+07
3.5e+07
3e+07
2.5e+07
2e+07
1.5e+07
1e+07
5e+06
0
Frequency
0
5
10
15
20
25
Path Length

9. 100M


4e+07
3.5e+07
3e+07
2.5e+07
2e+07
1.5e+07
1e+07
5e+06
0
Frequency
0
5
10
15
20
25
Path Length