1b. Compute the connected components.

The size of the 10 largest components are as follows:

1. 33696
2. 20
3. 16
4. 14
5. 13
6. 13
7. 13
8. 12
9. 12
10. 12

---

3. Implement the exact algorithm for closeness centrality

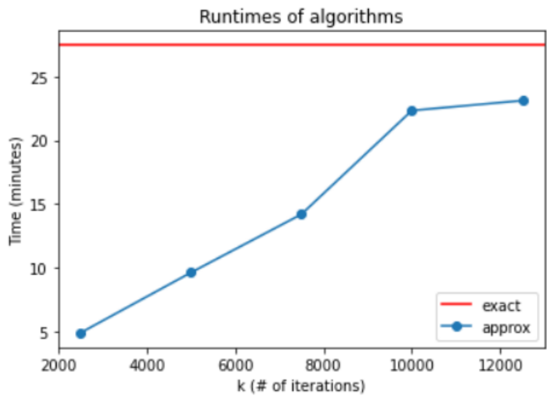
The indices of the 15 nodes with the highest closeness centrality in the largest connected component of the email-Enron graph and their closeness centrality scores are as follows:

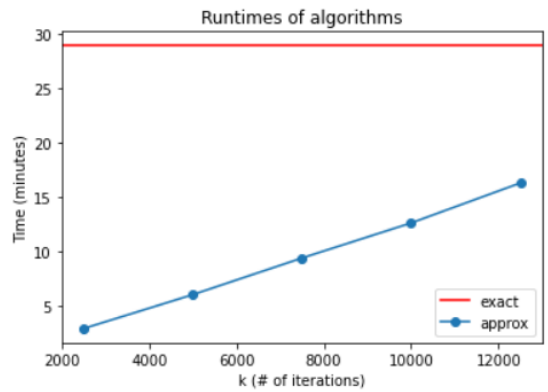
1. index: 136
   score: 0.3873700910512278
2. index: 76
   score: 0.3861183049526734
3. index: 46
   score: 0.3790810701347794
4. index: 140
   score: 0.37475531630928016
5. index: 370
   score: 0.374522052285257
6. index: 292
   score: 0.37433481830402277
7. index: 195
   score: 0.37398996625821346
8. index: 734
   score: 0.3739567610760898
9. index: 175
   score: 0.373790823571175
10. index: 416
    score: 0.3723410133156528
11. index: 1139
    score: 0.36917126829695857
12. index: 458
    score: 0.36829564209905014
13. index: 444
    score: 0.36808643121661333
14. index: 566
    score: 0.36778510303877054
15. index: 353
    score: 0.36742015331435984

The time needed to run the algorithm is 27.424920189380646 minutes. I used NetworkX's implementation of BFS, and the runtime ranged from 20-33 minutes on my Dell XPS 13.

---

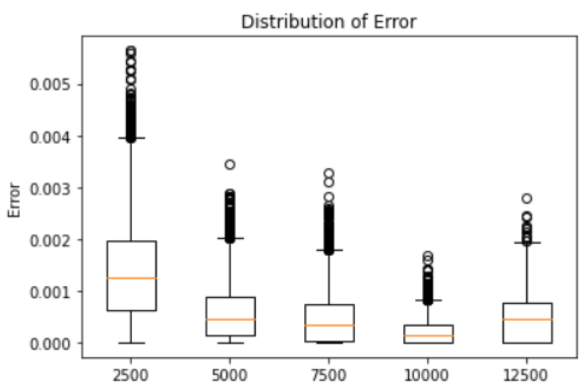
4. Implement the Eppstein-Wang approximation algorithm

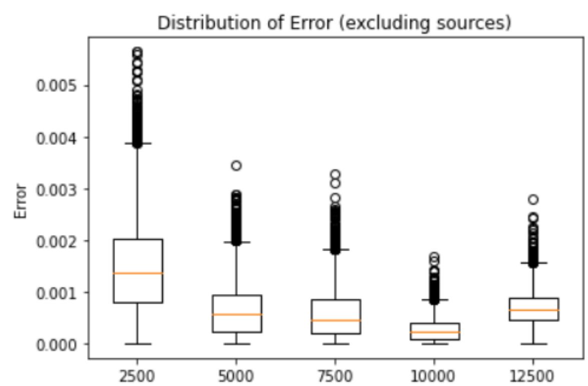





```
{2500: 4.862294209003449, 5000: 9.6032501856486, 7500: 14.183581137657166,
10000: 22.33936465581258, 12500: 23.120983095963798}
```

I ran the algorithms multiple times. The left plot above is what I will be using to analyze the distribution of errors. I noticed that in most runs, the Eppstein-Wang approximation algorithm was almost at least twice as fast, even with $k = 12500$.





As k increases, the errors generally decrease. However, at 12500 iterations, the distribution of error increases. We can see that when the sources are included, and k is large, then the distributions are skewed towards 0. This is most apparent for $k \geq 7500$ since the minimum and 1st quartile are about the same in the left plot above.