Assignment 4

Web Science CS595

Name: Amara Naas

Answer to question 1

In the Python program (q1_2.py)[1] will extract all of the links from a selected 100[2] pages to other pages. For each URI it will create a text file of all of the outbound links from that page to other URIs including the main URI and save it in the files directory[1].

Answer to question 2

The same Python program (q1_2.py)[1] will create a single GraphViz "dot" file called gefiDoc.gv[1] of the resulting graph.

Answer to question 3

In this question I used Gephi and gefiDoc.gv[1] to visualize the graph as shown in figure 1. Also I find out that the most in-link was twitter.com as shown in figure 2 and the must out-link was finance.detik.com as shown in figure 3.

From this figures we can observe that many links have a very high out degree and very few have high in degree.

Also figure 4 shows graph of the HITS, figure 5 shows PageRank, figure 6 shows avg degree, figure 7 shows network diameter, and figure 8 shows connected components.

[1]File uploaded to github

[2]Saved in res.txt

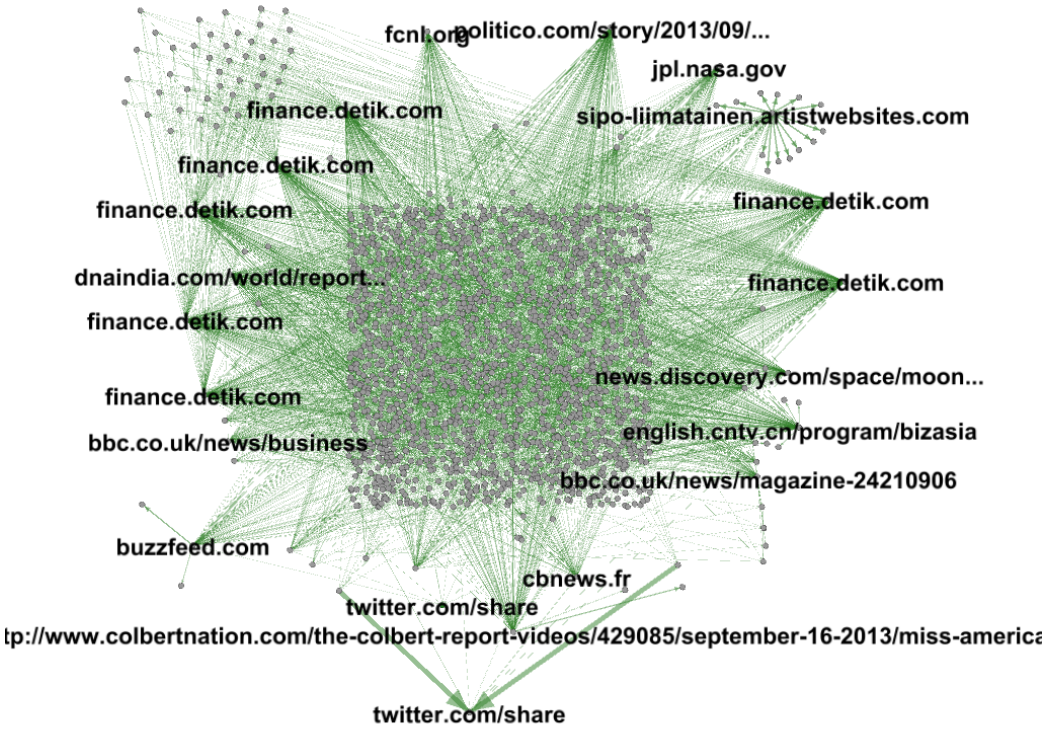


Figure 1: Gephi Graph of "dot" file

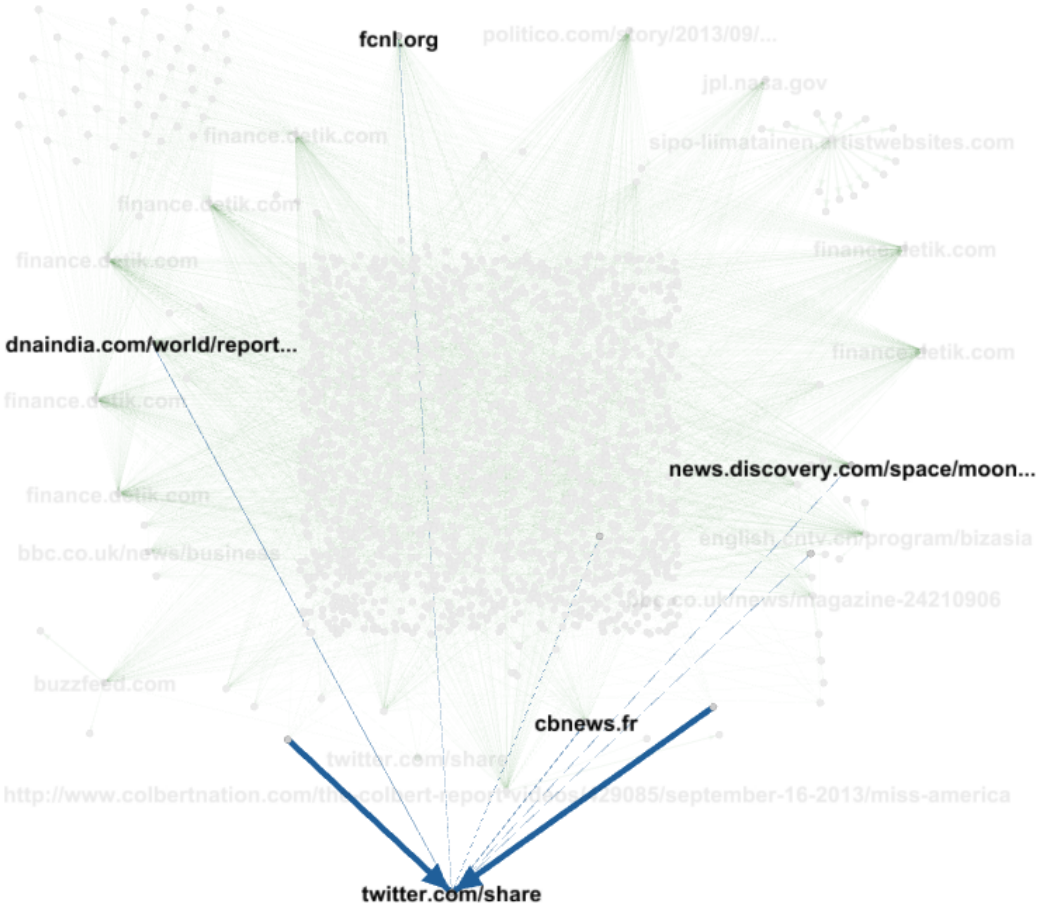
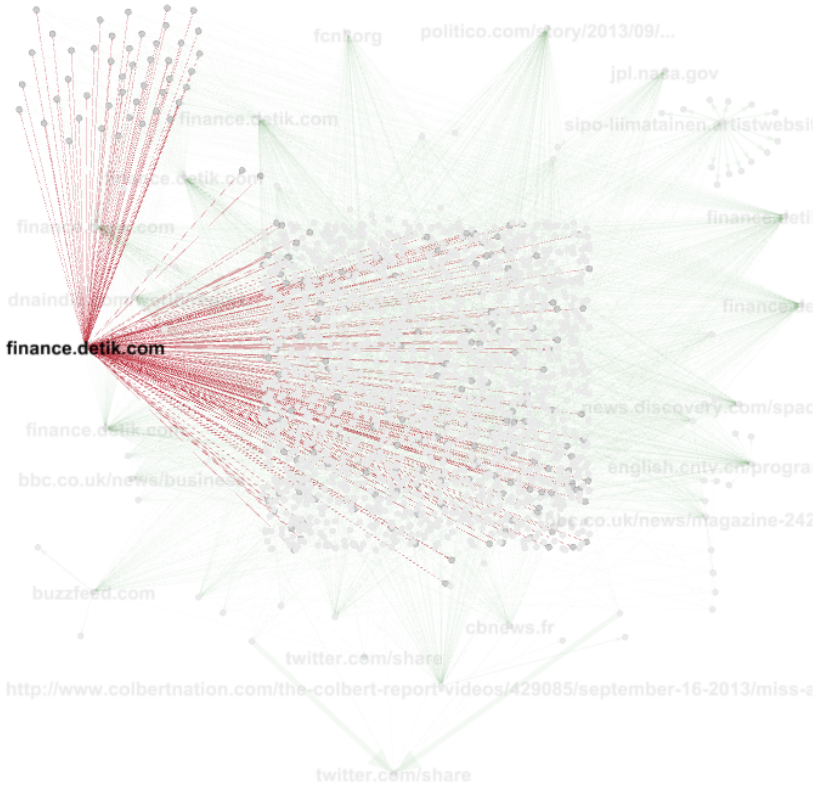
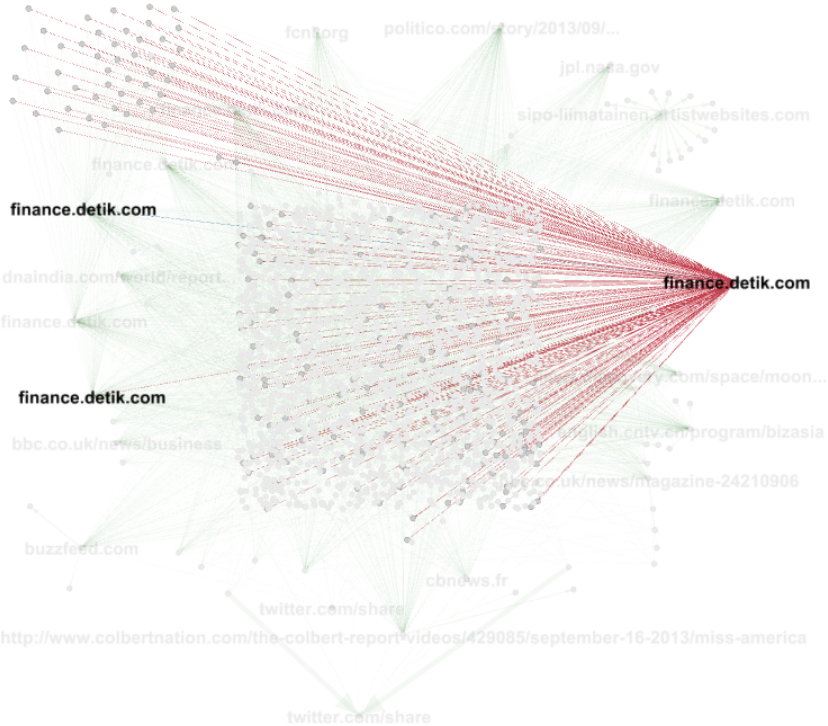


Figure 2: twitter.com in-degree

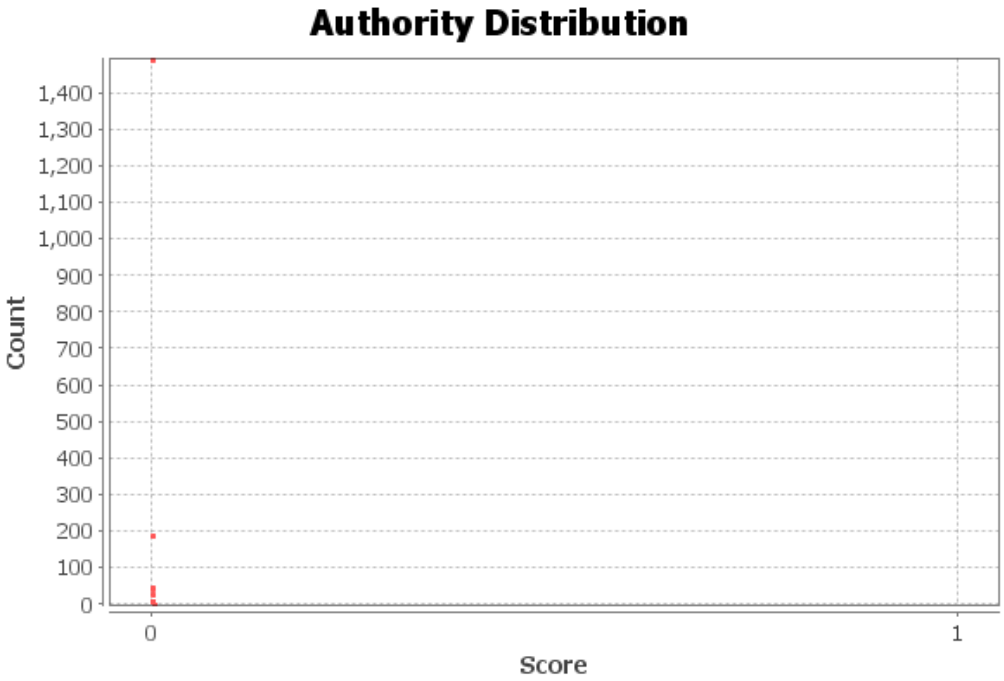


(a) No link to other finance.detik.com

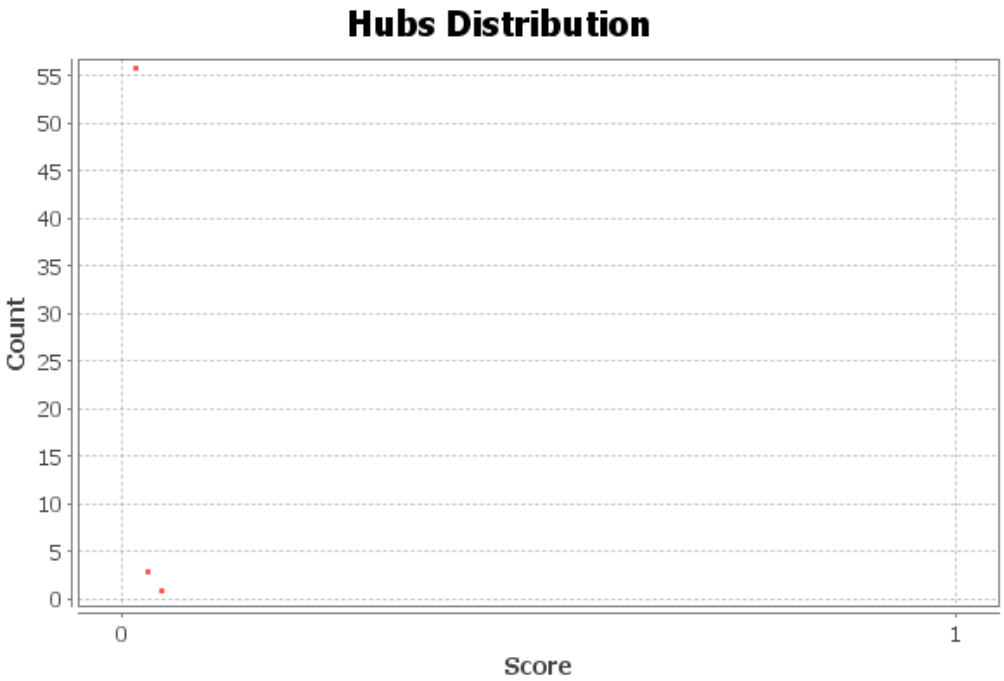


(b) some links to other finance.detik.com

Figure 3: finance.detik.com out-degree



(a) Authorities Distribution



(b) Hubs Distribution

Figure 4: HITS

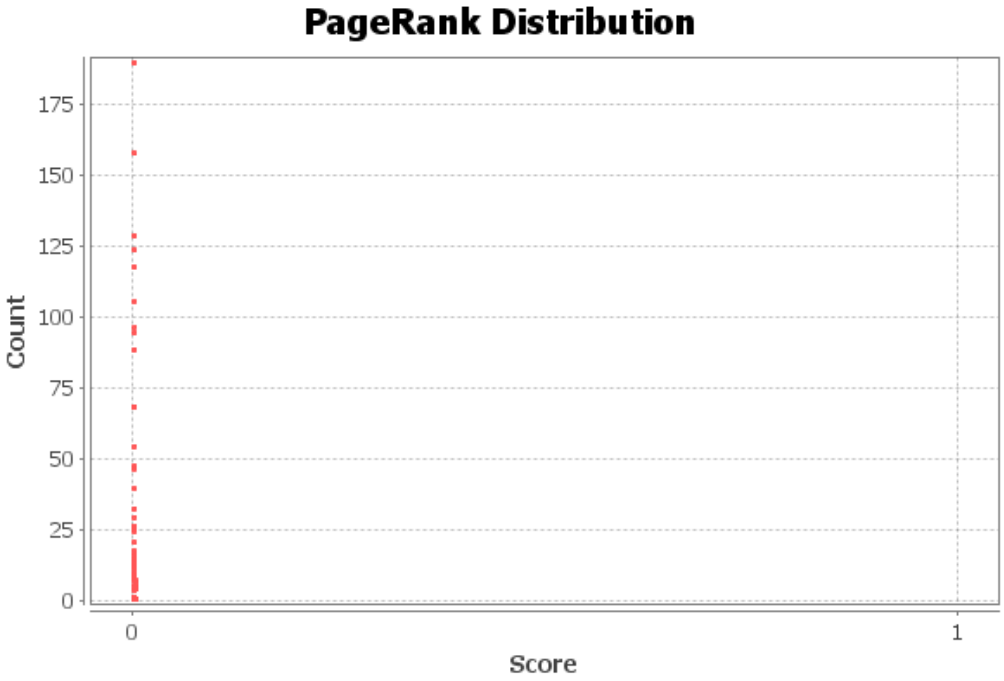
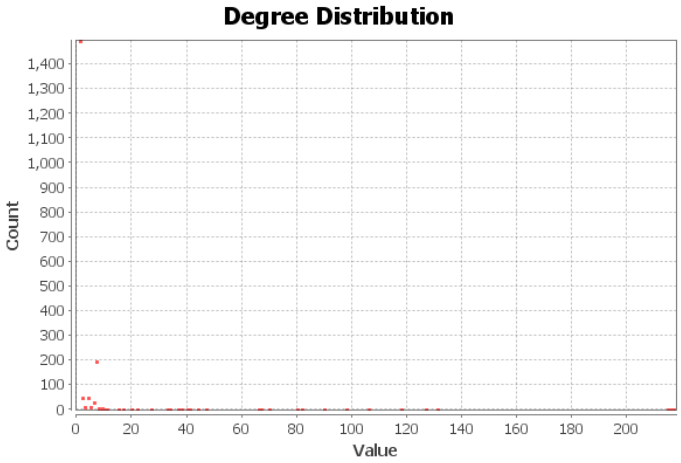
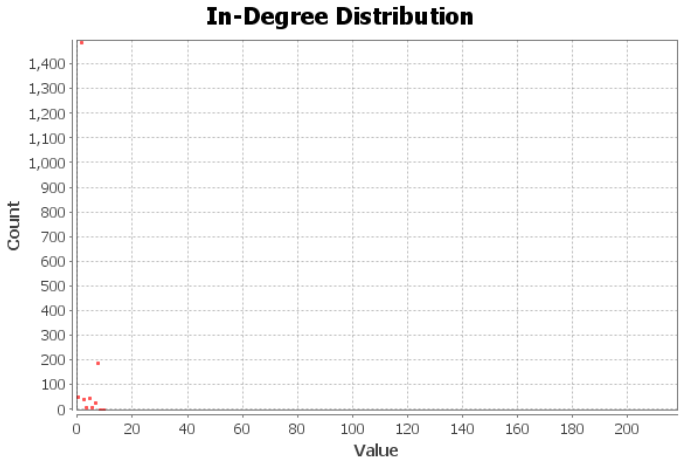


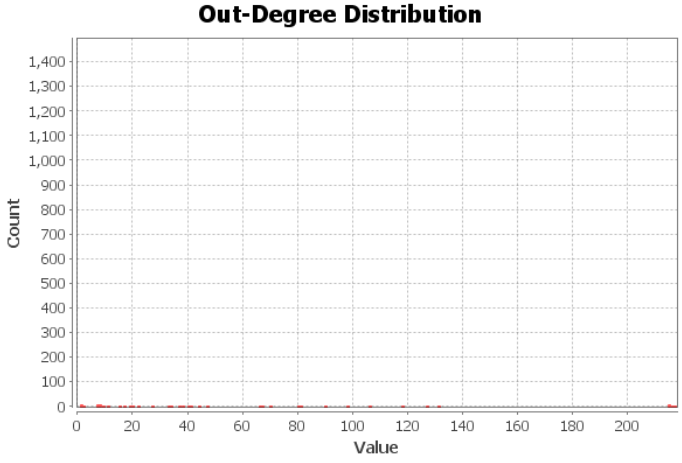
Figure 5: PageRank



(a) Degree Distribution

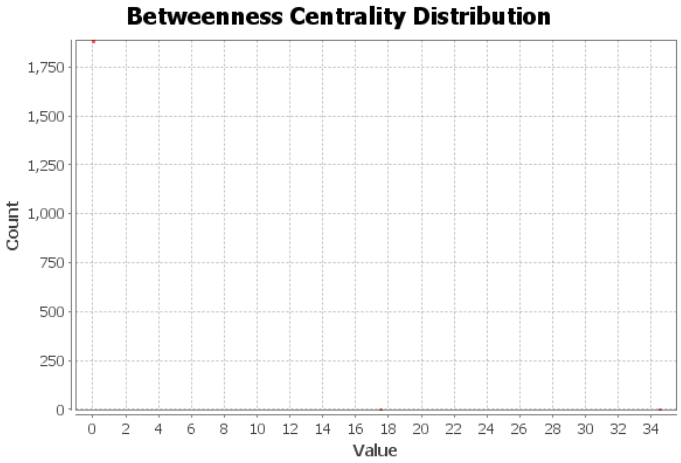


(b) In-degree Distribution

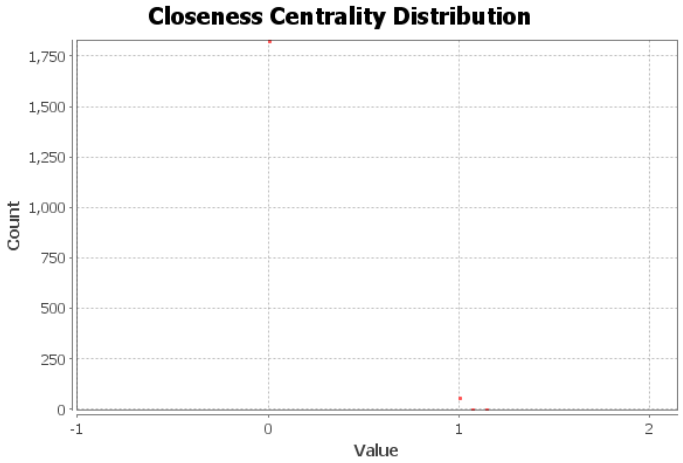


(c) Out-degree Distribution

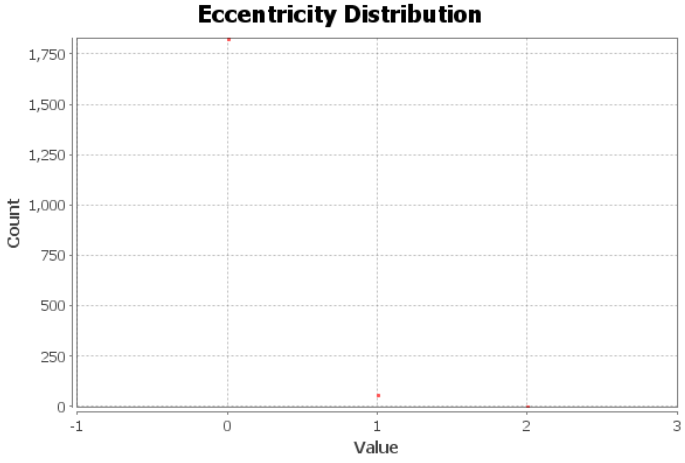
Figure 6: HITS



(a) Betweenness Centrality Distribution



(b) Closeness Centrality Distribution



(c) Eccentricity Distribution

Figure 7: Network Diameter

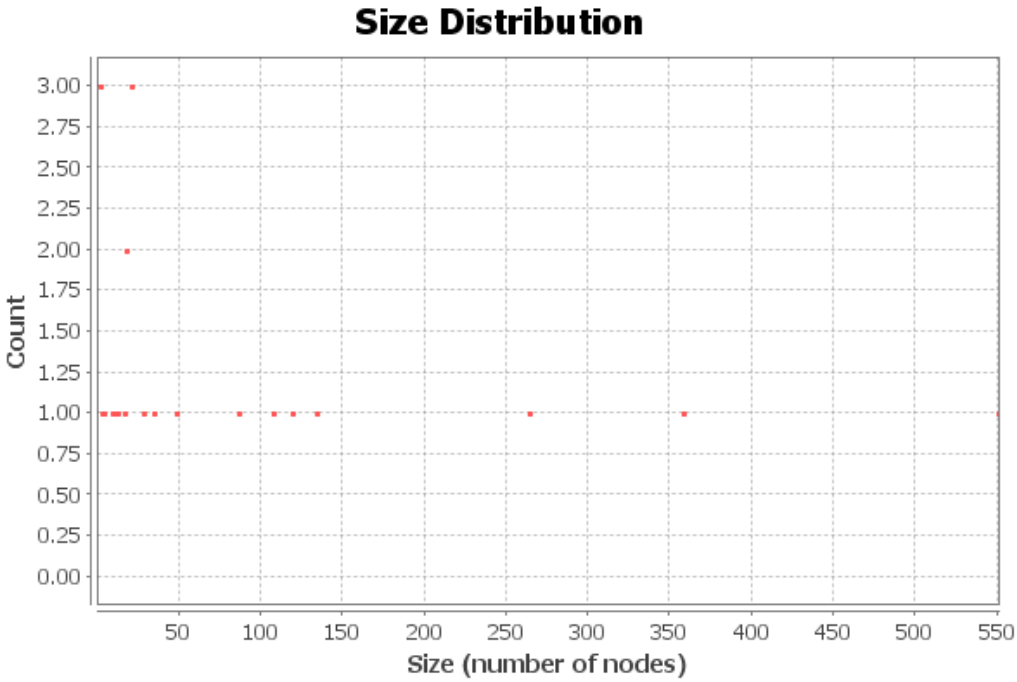


Figure 8: Connected Components