**Project 2 Testing**

Tested the correctness of the graph ADT implementation by jUnit testing individually each method and any other methods that was needed. The method, printAdjacencyList() was created to view the correctness of graph representation. Lines have been commented out such as in iterators test, next() and remove() when not possible will throw the right error matching the impossible call.

Graph is analyzed correctly with accurate information results. |A| gives the correct count as it would be twice the number of edges for undirected graph as the edge will have two arcs for both direction. The Graph ADT is intended for directed graphs as it reads in arcs but it is able to toggle to undirected. Graph results is expected with the strongly connected components as seen with the smallest components almost having no arc to any vertex. Adding more arcs makes it possible for the total number of strongly connected components count to decrease.

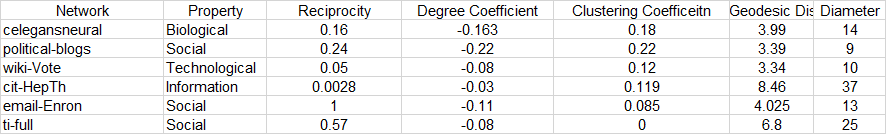
jUnit 33/33 Runs, 0 Errors, 0 Failures

**Project 3 Testing**

The algorithms for the metrics were tested by comparing the graph results with the statistics provided. The graph is read in and working correctly by having the same results with the statistics for the project 2. In general, the network metrics matches with the statistics provided except for some minor differences.   
 The degree correlation results are not the same as the statistics but are within the range. The degree correlation differences is seen more in the regular graphs such as celegansneural.vna has a 7.0992% difference from expected whereas the random has a 2.0421% difference from expected . The differences with the degree correlation is highly likely caused by rounding error from the computation involving four variables in the equation.   
 The graph is implemented that allows self-loop if there is an arc that points to itself as the destination (node) is added to the source (linked list). The results in comparison with the statistics show that the density for celegansneural and political-blogs ignore loops. The political-blogs is the only graph with self-loop arcs as seen in the reciprocity that it allows self-loop.

If there is differences then the discrepancies are under the graph result that occurs.

**Conclusion**



This project helps with the understanding of different types of networks and how they are modeled by researching the network metrics and comparing the graph results. The values from the metrics can be analyzed to determine the graph characteristics such as the reciprocity value. Social networks tend to have higher reciprocity as shown in the results. It may be possible to classify a graph with a very small percent reciprocity, almost complete disassortative mixing, and large mean geodesic distance to be an information graph as seen with the information graph cit-HepTh.  
 The differences between random graphs and regular graphs are mainly in the distributions of the arcs as seen with in the random graphs, the minimum and max degree for both ‘in’ and ‘out’ are relatively the same. Most of the metrics in the random graph are smaller than the metrics in the regular graphs except for the random and regular graph for wiki-Vote.vna.

**Reciprocity**: Values are higher in social networks as seen with political-blogs, email-enron, and ti-full. There is 100% reciprocity for undirected graph email-enron because the arcs are both ways. Popularity disparities is seen with the information graph cit-HepTh and technological graph wiki-Vote as the reciprocity is less than 6%.

**Degree Correlation**: The correlation has values from -1 to 1. A value of 1 would mean that the network has perfect assortative mixing. Value of 0 means that the network is non-assortative. Value of -1 would mean the network is completely disassortative. As shown in the results, all the regular graphs have disassortative mixing but the graphs with the less disassortative mixing is celegansneural, political-blogs, and email-enron. These three graphs show the "love of same" as there is collaboration or similar links in the neural network. The graph with the most disassortative value is the information graph, cit-HepTh which may be the result of the papers being in the same research field of high-energy physics so there is no general assortative mixing.

**Clustering Coefficient**: The graph with the highest percent for clustering coefficient is political-blogs which states that two references from a user tend to reference each other in collaboration. The opposite is with the graph ti-full as there is no clustering in the graph of actors, chat rooms, threaded discussions, and files shared.

**Mean Geodesic Distance (a.k.a. average path length**): The “small world effect” is shown in political-blogs as similar to the degree correlation that there is collaboration to what they are referencing. The two graphs wiki-Vote and ti-full have the largest distance and diameter which shows the graph having a large world.

Graph Results  
 Regular  
 Random

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\celegansneural.vna>

-----------------------------------------------------------------------------

|V| = 297

|A| = 2345

Density: 0.026674

Degree distribution: minimum average maximum

inDegree: 0 7.8956 134

outDegree: 0 7.8956 39

Reciprocity: 0.168017

Degree Correlation: -0.163199

Clustering Coefficient: 0.180711

Directed Clustering Coefficient Cycle: 19.023202

Directed Clustering Coefficient Middle: 0.136264

Directed Clustering Coefficient In: 0.230843

Directed Clustering Coefficient Out: 0.259937

Mean Geodesic Distance: 3.991884

Diameter: 14

The diameter for celegansneural.vna is not the same as the diameter in the statistics but the result matches with the diameter in Pajek results.

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-297-2345.vna>

-----------------------------------------------------------------------------

|V| = 297

|A| = 2345

Density: 0.026674

Degree distribution: minimum average maximum

inDegree: 2 7.8956 18

outDegree: 2 7.8956 19

Reciprocity: 0.028998

Degree Correlation: -0.054104

Clustering Coefficient: 0.051443

Clustering Coefficient: 0.025458

Directed Clustering Coefficient Cycle: 40.754967

Directed Clustering Coefficient Middle: 0.023270

Directed Clustering Coefficient In: 0.063965

Directed Clustering Coefficient Out: 0.051456

Mean Geodesic Distance: 2.976772

Diameter: 5

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\political-blogs.vna>

-----------------------------------------------------------------------------

|V| = 1490

|A| = 19025

Density: 0.008575

Degree distribution: minimum average maximum

inDegree: 0 12.7685 337

outDegree: 0 12.7685 256

Reciprocity: 0.242681

Degree Correlation: -0.221175

Clustering Coefficient: 0.226020

Directed Clustering Coefficient Cycle: 9.655261

Directed Clustering Coefficient Middle: 0.193499

Directed Clustering Coefficient In: 0.179540

Directed Clustering Coefficient Out: 0.309575

Mean Geodesic Distance: 3.390184

Diameter: 9

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-1490-19025.vna>

-----------------------------------------------------------------------------

|V| = 1490

|A| = 19025

Density: 0.008575

Degree distribution: minimum average maximum

inDegree: 2 12.7685 25

outDegree: 2 12.7685 30

Reciprocity: 0.007359

Degree Correlation: 0.010224

Clustering Coefficient: 0.016814

Directed Clustering Coefficient Cycle: 120.210188

Directed Clustering Coefficient Middle: 0.007846

Directed Clustering Coefficient In: 0.015968

Directed Clustering Coefficient Out: 0.016856

Mean Geodesic Distance: 3.136125

Diameter: 5

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\wiki-Vote.vna>

-----------------------------------------------------------------------------

|V| = 7115

|A| = 103689

Density: 0.002049

Degree distribution: minimum average maximum

inDegree: 0 14.5733 457

outDegree: 0 14.5733 893

Reciprocity: 0.056457

Degree Correlation: -0.083052

Clustering Coefficient: 0.125479

Mean Geodesic Distance: 3.341011

Diameter: 10

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-7115-103689.vna>

-----------------------------------------------------------------------------

|V| = 7115

|A| = 103689

Density: 0.002049

Degree distribution: minimum average maximum

inDegree: 4 14.5733 30

outDegree: 1 14.5733 36

Reciprocity: 0.001755

Degree Correlation: 0.001535

Clustering Coefficient: 0.003952

Mean Geodesic Distance: 3.612734

Diameter: 6

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\cit-HepTh.vna>

-----------------------------------------------------------------------------

|V| = 27770

|A| = 352807

Density: 0.000458

Degree distribution: minimum average maximum

inDegree: 0 12.7046 2414

outDegree: 0 12.7046 562

Reciprocity: 0.002849

Degree Correlation: -0.030305

Clustering Coefficient: 0.119612

Mean Geodesic Distance: 8.460137

Diameter: 37

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-27770-352807.vna>

-----------------------------------------------------------------------------

|V| = 27770

|A| = 352807

Density: 0.000458

Degree distribution: minimum average maximum

inDegree: 2 12.7046 30

outDegree: 2 12.7046 28

Reciprocity: 0.000465

Degree Correlation: -0.002448

Clustering Coefficient: 0.000953

Mean Geodesic Distance: 4.310337

Diameter: 7

The random graph for cit-HepTh.vna differs from the statistics in the average degree and network clustering coefficient as they are half of the given results. It is possible that the results for the two given are incorrect as the regular and random graph have the same numbers for vertices, arcs, density, and average degree. The results from the previous graphs for network clustering coefficient matches with the statistics provided so the algorithm is working correctly.

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\email-Enron.vna>

-----------------------------------------------------------------------------

|V| = 36692

|A| = 367662

Density: 0.000273

Degree distribution: minimum average maximum

inDegree: 1 10.0202 1383

outDegree: 1 10.0202 1383

Reciprocity: 1.000000

Degree Correlation: -0.110764

Clustering Coefficient: 0.085311

Mean Geodesic Distance: 4.025143

Diameter: 13

Random graph of email-Enron.vna as read in as directed  
-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-36692-183831.vna>

-----------------------------------------------------------------------------

|V| = 36692

|A| = 183831

Density: 0.000137

Degree distribution: minimum average maximum

inDegree: 0 5.0101 16

outDegree: 0 5.0101 16

Reciprocity: 0.000141

Degree Correlation: -0.001669

Clustering Coefficient: 0.000298

Mean Geodesic Distance: 6.701647

Diameter: 13

Random graph of email-Enron.vna as read in as undirected  
-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\g-36692-183831.vna>

-----------------------------------------------------------------------------

|V| = 36692

|E| = 183818

Density: 0.000273

Degree distribution: minimum average maximum

inDegree: 0 10.0195 25

outDegree: 0 10.0195 25

Reciprocity: 1.000000

Degree Correlation: -0.001669

Clustering Coefficient: 0.000298

Mean Geodesic Distance: 4.812111

Diameter: 8

The random graph for email-Enron.vna differs from the statistics in the average degree and network clustering coefficient. Email-Enron is undirected but the random graph shows that it is a directed graph. The average degree is half of the result in the statistics but the clustering coefficient is twice what is given. The random graph has half of the arcs in the regular graph which results in half the average degree. The clustering coefficient is twice the number in the statistics. Reading in the random graph as undirected solves the problem with the number of arcs and the average degree. The clustering coefficient doesn’t change regardless of reading it in as directed or undirected as no new arcs closes the triangle so the number is the statistics for clustering coefficient must be wrong.

-----------------------------------------------------------------------------

Graph <c:\\users\david\desktop\graphtests\ti-full.vna>

-----------------------------------------------------------------------------

|V| = 38804

|A| = 197082

Density: 0.000131

Degree distribution: minimum average maximum

inDegree: 0 5.0789 889

outDegree: 0 5.0789 944

Reciprocity: 0.570341

Degree Correlation: -0.083348

Clustering Coefficient: 0.000000

Mean Geodesic Distance: 6.816900

Diameter: 25