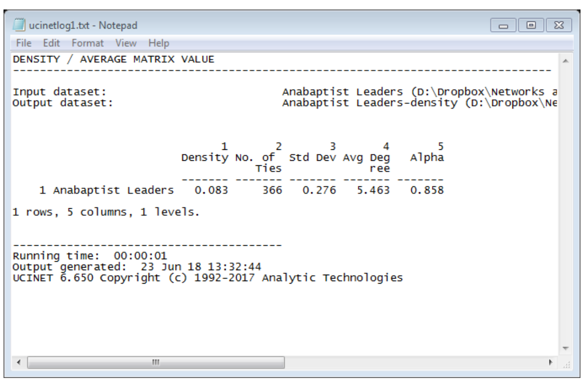
SNA Basics #5
Topography of Social Networks

In this lab, we consider methods for examining network topography. We begin with network size, and then explore several that measure network cohesiveness: density, average degree, clustering coefficient, fragmentation, the E-I index, average distance, and diameter. Size is the number of actors in a network. Density is the number of actual ties divided by the total possible ties. Unfortunately, it is inversely related to network size (i.e., the larger the network, the lower the density) because the number of possible lines increases exponentially as actors join the network, which is why SNA researchers sometimes turn to other measures, such as average degree and the clustering coefficient. Network fragmentation is the proportion of all pairs of actors that are not tied (directly or indirectly) with one another. The E-I index measures the ratio of external to internal ties (good for determining how isolated a network is). Average distance is the average length of all the shortest paths (i.e., geodesics) between all actors in a network, and diameter is a network's longest geodesic. Finally, we consider several metrics that capture how hierarchical a network is: centralization, variance, and standard deviation. This lab shows how to estimate all of the measures in UCINET and Pajek. Once again we will use the Anabaptist Leader network data, which are included in the zipped file, `SNA Basics #5(Data).zip`.

Part I – Network Topography in UCINET

[UCINET] Network>Cohesion >Density >Density Overall

1. Let's begin by calculating network density. Select the *Network>Cohesion> Density>(new)Density Overall* command. This callus up a dialog box (not shown). Select the `Anabaptist Leaders.##h` file. Click OK. The resulting output (Figure 1) gives you density of the network. As you can see, it equals



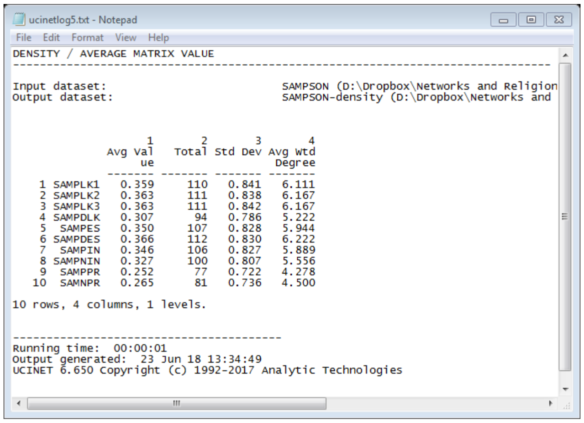
```
DENSITY / AVERAGE MATRIX VALUE
------------------------------------------------------------------------------

Input dataset:                          Anabaptist Leaders (D:\Dropbox\Networks a
Output dataset:                         Anabaptist Leaders-density (D:\Dropbox\Ne



                               1       2       3       4       5
                         Density No. of  Std Dev Avg Deg   Alpha
                                    Ties             ree
                         ------- ------- ------- ------- -------
    1 Anabaptist Leaders   0.083     366   0.276   5.463   0.858

1 rows, 5 columns, 1 levels.


----------------------------------------
Running time:  00:00:01
Output generated:  23 Jun 18 13:32:44
UCINET 6.650 Copyright (c) 1992-2017 Analytic Technologies
```

Figure 1: Output of Overall Density Command (Anabaptist Leadership Network)

Network>Cohesion >Density >Density Overall

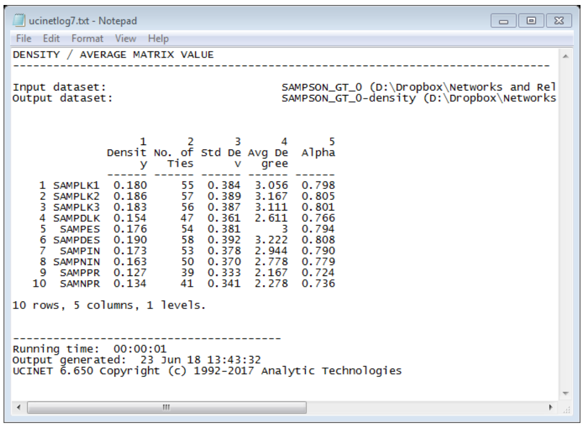
Transform>Dichotomize

2. If we issued this command with a stacked network (e.g., the Sampson novice data that we used in the previous lab), it gives us scores for all of the networks in the file (Figure 2, next page). Notice, however, that what UCINET reports is a bit different from that above. Because the Sampson data are weighted (the novices ranked their fellow novices from 3 to 1), UCINET calculates average value and weighted average degree. If we dichotomize (binarized) the network, and then issue the command, UCINET gives us an output like that in Figure 3 (next page).



```
DENSITY / AVERAGE MATRIX VALUE
------------------------------------------------------------------------------

Input dataset:                          SAMPSON (D:\Dropbox\Networks and Religion
Output dataset:                         SAMPSON-density (D:\Dropbox\Networks and



                    1       2       3       4
              Avg Val   Total Std Dev Avg Wtd
                   ue                  Degree
              ------- ------- ------- -------
    1 SAMPLK1   0.359     110   0.841   6.111
    2 SAMPLK2   0.363     111   0.838   6.167
    3 SAMPLK3   0.363     111   0.842   6.167
    4 SAMPDLK   0.307      94   0.786   5.222
    5  SAMPES   0.350     107   0.828   5.944
    6 SAMPDES   0.366     112   0.830   6.222
    7  SAMPIN   0.346     106   0.827   5.889
    8 SAMPNIN   0.327     100   0.807   5.556
    9  SAMPPR   0.252      77   0.722   4.278
   10  SAMNPR   0.265      81   0.736   4.500

10 rows, 4 columns, 1 levels.


----------------------------------------
Running time:  00:00:01
Output generated:  23 Jun 18 13:34:49
UCINET 6.650 Copyright (c) 1992-2017 Analytic Technologies
```

Figure 2: Output of Overall Density Command (Sampson's Novices – Weighted)



```
DENSITY / AVERAGE MATRIX VALUE
------------------------------------------------------------------------------

Input dataset:                          SAMPSON_GT_0 (D:\Dropbox\Networks and Rel
Output dataset:                         SAMPSON_GT_0-density (D:\Dropbox\Networks



                   1      2      3      4      5
              Densit No. of Std De Avg De  Alpha
                   y   Ties      v   gree
              ------ ------ ------ ------ ------
    1 SAMPLK1  0.180     55  0.384  3.056  0.798
    2 SAMPLK2  0.186     57  0.389  3.167  0.805
    3 SAMPLK3  0.183     56  0.387  3.111  0.801
    4 SAMPDLK  0.154     47  0.361  2.611  0.766
    5  SAMPES  0.176     54  0.381      3  0.794
    6 SAMPDES  0.190     58  0.392  3.222  0.808
    7  SAMPIN  0.173     53  0.378  2.944  0.790
    8 SAMPNIN  0.163     50  0.370  2.778  0.779
    9  SAMPPR  0.127     39  0.333  2.167  0.724
   10  SAMNPR  0.134     41  0.341  2.278  0.736

10 rows, 5 columns, 1 levels.


----------------------------------------
Running time:  00:00:01
Output generated:  23 Jun 18 13:43:32
UCINET 6.650 Copyright (c) 1992-2017 Analytic Technologies
```

Figure 3: Output of Overall Density Command (Sampson's Novices – Binarized)

3. Network density is the ratio of actual ties divided by all possible ties. However, density scores tend to decrease as social networks get larger because the number of possible ties increases exponentially, whereas the number of ties that each actor can maintain is limited. Consequently, we can only use it to compare networks of the same size. An alternative is average degree centrality, which is not sensitive to network size and thus can be used to compare different sized networks. If you look again at Figures 1 and 3, you will note that UCINET's density output also provides the average degree of each network (column 4).

4. Another measure used to capture the overall level of cohesion is the clustering coefficient. It is estimated by first identifying each actor's ego network (i.e., each actor's ties to other actors – "alters" – and the ties between them), then calculating the density of each ego network (without including ego's ties), and then averaging the ego-network density scores for all of the actors in the network. It is also known as ego-network density. To get this score, select the *Network>Cohesion>Clustering Coefficient* command. In the resulting dialog box (not shown) select the "`Anabaptist Leaders.##h`" network and click OK.

Network>Cohesion >Clustering Coefficient

Network >Centrality and Power >Fragmentation

Data>Unpack

5. Fragmentation and cohesiveness are additional measures for estimating a network's internal connectivity. They are additive inverses of each and range from 0.0 to 1.0 where a network composed entirely of isolates will have a cohesiveness score of 0.0 (and a fragmentation score of 1.0) and a network where every actor is connected (either directly or indirectly) to every other actor will have a cohesive score of 1.0 (and a fragmentation score of 0.0). You estimate these using UCINET's *Network>Centrality and Power>Fragmentation* command. Unfortunately, it only works on single networks, so if you have a stacked network, you need to unpack it with the *Data>Unpack* command. UCINET reports two sets of scores: (1) the standard fragmentation score, which is the proportion of all pairs of actors that are not connected with one another, and (2) a distance-weighted score that weights this score by the path distance between all pairs. Another helpful feature is that it reports a series of scores that indicate (1) what the network fragmentation would be if a particular actor is removed ("Frag"); (2) what the distance weighted network fragmentation will be if the actor is removed ("DwFrag"); (3) the change in network fragmentation when a particular node is removed ("FragDiff"); (4) the change in distance-weighted network fragmentation when node is removed ("DwFragDiff"); (5) the percent of fragment change ("PctFragChg"); and (6) the percent of distance-weighted fragment change ("PctFragChg").

Network >Cohesion>E-I Index

Data>Display

6. The E-I index compares the ratio of internal to external ties. It ranges from -1.00 to 1.00 where a score of -1.00 indicates that the network has only internal ties and a score of 1.00 indicates that it only has external ties. It's calculated using the *Network>Cohesion>E-I Index* command and requires both a network and an attribute file. The latter is needed to indicate which groups actors belong to. Here we will use the attribute that distinguishes between those Anabaptists who Melchiorites and those who were not. You can view this attribute file with the *Data>Display* command; the "Melchiorite" attribute is located in column 6 (see Figure 4). Note that the "L" speed button is a way to select the attribute that you hope to use.

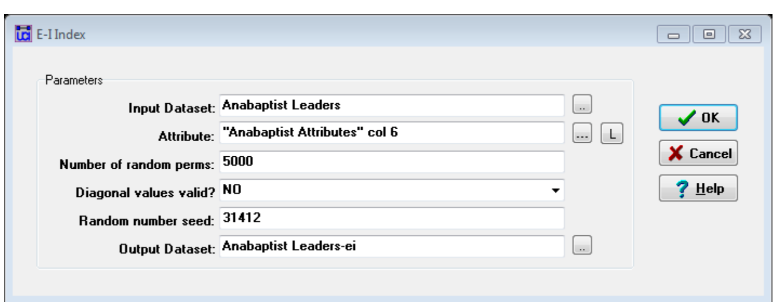
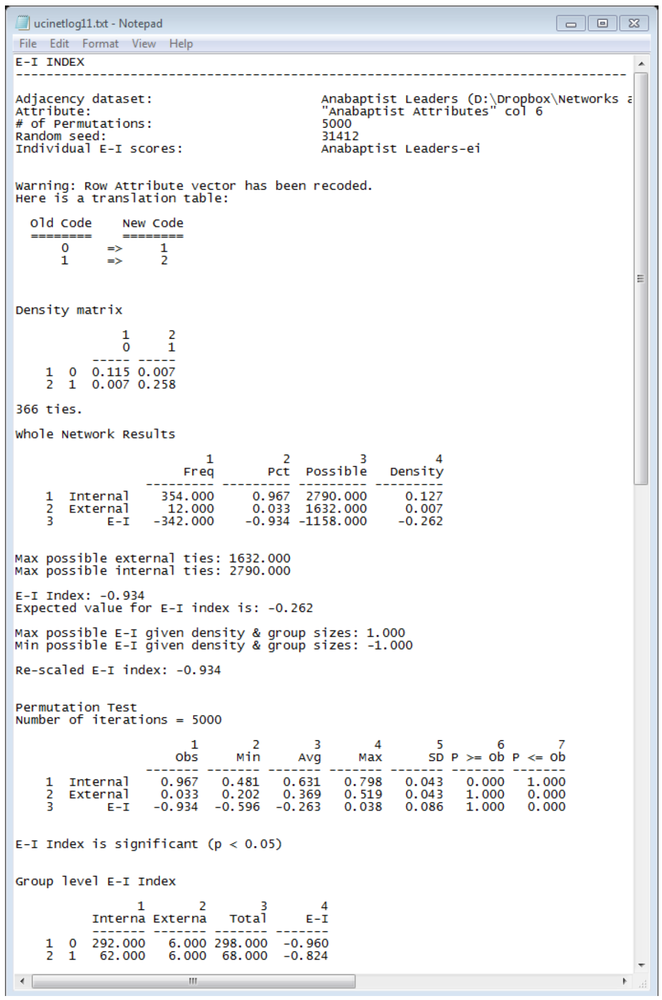


Figure 4: UCINET's E-I Index Dialog Box

7. After clicking OK, UCINET's output will be similar to that shown in Figure 5. Note that it is quite extensive. The top of the report provides overall information on both networks; to get the E-I index of the two subgroups, you have to scroll down the report until you find the "Group Level E-I Index." There you can see that the E-I index for the Melchiorites (group 2) equals -0.824, which tells us that the Melchiorite leaders had very few ties to non-Melchiorites. The score for the other group is somewhat meaningless because it really isn't a distinct group; it includes Anabaptists, Lutherans, and Reformed Protestants. All they share in common is that they aren't Melchiorites.



```
ucinetlog11.txt - Notepad
File  Edit  Format  View  Help
E-I INDEX
------------------------------------------------------------------------

Adjacency dataset:                     Anabaptist Leaders (D:\Dropbox\Networks a
Attribute:                             "Anabaptist Attributes" col 6
# of Permutations:                     5000
Random seed:                           31412
Individual E-I scores:                 Anabaptist Leaders-ei


Warning: Row Attribute vector has been recoded.
Here is a translation table:

  Old Code     New Code
  ========     ========
      0     =>      1
      1     =>      2



Density matrix

              1     2
              0     1
          ----- -----
   1  0   0.115 0.007
   2  1   0.007 0.258

366 ties.

Whole Network Results

                         1         2         3         4
                      Freq       Pct  Possible   Density
                 --------- --------- --------- ---------
   1  Internal     354.000     0.967  2790.000     0.127
   2  External      12.000     0.033  1632.000     0.007
   3       E-I    -342.000    -0.934 -1158.000    -0.262


Max possible external ties: 1632.000
Max possible internal ties: 2790.000

E-I Index: -0.934
Expected value for E-I index is: -0.262

Max possible E-I given density & group sizes: 1.000
Min possible E-I given density & group sizes: -1.000

Re-scaled E-I index: -0.934


Permutation Test
Number of iterations = 5000

                       1       2       3       4       5       6       7
                     Obs     Min     Avg     Max      SD P >= Ob P <= Ob
                 ------- ------- ------- ------- ------- ------- -------
   1  Internal     0.967   0.481   0.631   0.798   0.043   0.000   1.000
   2  External     0.033   0.202   0.369   0.519   0.043   1.000   0.000
   3       E-I    -0.934  -0.596  -0.263   0.038   0.086   1.000   0.000

E-I Index is significant (p < 0.05)


Group level E-I Index

               1       2       3       4
         Interna Externa   Total     E-I
         ------- ------- ------- -------
   1  0  292.000   6.000 298.000  -0.960
   2  1   62.000   6.000  68.000  -0.824
```

Figure 5: UCINET's E-I Index Output

Network>Cohesion >Geodesic Distance (legacy)

8. Average distance and diameter can be calculated in UCINET using the *Network> Cohesion>Geodesic Distance (old)* command. Like the fragmentation command, this command only works on single networks. The output (Figure 6) first indicates the average (path) distance among reachable pairs (since not all actors can reach one another). This is followed by the distance-weighted measures of fragmentation and cohesiveness (rather than the standard measures of

fragmentation and cohesiveness), which in turn are followed by the number of geodesics at various distance levels (from shorter to longer). The largest geodesic level equals the network's diameter. ***Just looking at the output in Figure 6, what is the leadership network's average distance and diameter?*** Now compare it to the output in Figure 8 to make sure that, like Indiana Jones, you chose wisely.

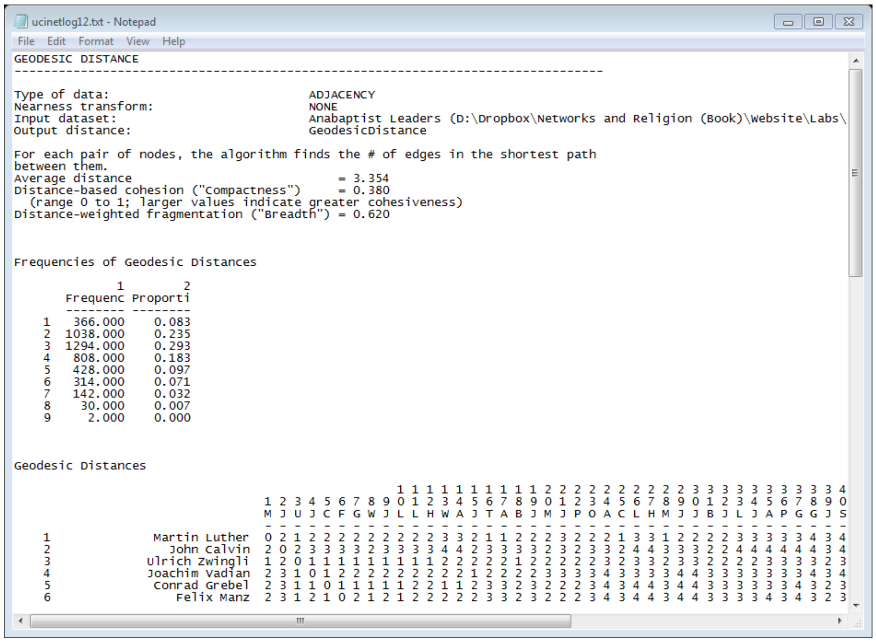


Figure 6: UCINET's Old Geodesic Output

Network >Multiple cohesion measures

9. Finally, UCINET recently added a function that calculates multiple cohesion measures: *Network>Multiple cohesion measures*, which calls up a dialog box similar to Figure 7. It calculates 23 different measures, including many of the ones discussed above (except size, clustering coefficient, and the E-I index) plus degree centralization. Thus, it can be quite useful (see Figure 8).

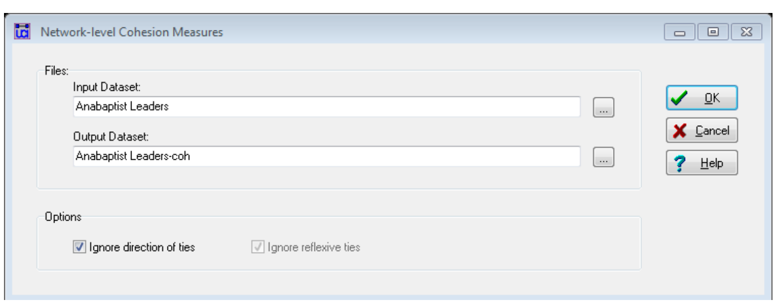
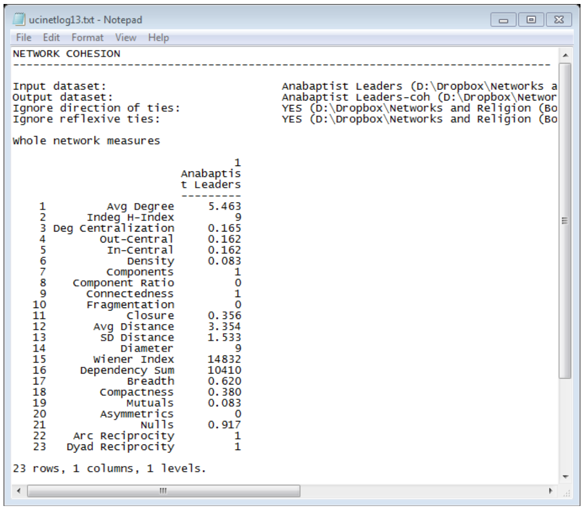


Figure 7: UCINET's Multiple Cohesion Measures Dialog Box




```
NETWORK COHESION
--------------------------------------------------------------------------------

Input dataset:                          Anabaptist Leaders (D:\Dropbox\Networks a
Output dataset:                         Anabaptist Leaders-coh (D:\Dropbox\Networ
Ignore direction of ties:               YES (D:\Dropbox\Networks and Religion (Bo
Ignore reflexive ties:                  YES (D:\Dropbox\Networks and Religion (Bo

Whole network measures

                                  1
                          Anabaptis
                          t Leaders
                          ---------
    1         Avg Degree      5.463
    2      Indeg H-Index          9
    3 Deg Centralization      0.165
    4        Out-Central      0.162
    5         In-Central      0.162
    6            Density      0.083
    7         Components          1
    8    Component Ratio          0
    9       Connectedness         1
   10      Fragmentation          0
   11            Closure      0.356
   12       Avg Distance      3.354
   13        SD Distance      1.533
   14           Diameter          9
   15       Wiener Index      14832
   16     Dependency Sum      10410
   17            Breadth      0.620
   18        Compactness      0.380
   19            Mutuals      0.083
   20        Asymmetrics          0
   21              Nulls      0.917
   22    Arc Reciprocity          1
   23   Dyad Reciprocity          1

23 rows, 1 columns, 1 levels.
```

Figure 8: UCINET's Multiple Cohesion Measures Output

10. Network centralization, variance, and standard deviation are measures that can capture the hierarchical dimension of a network's topography. They help researchers determine how hierarchical (or non-hierarchical) a network. Centralization uses the variation in actor centrality (as compared to the highest centrality score) within the network to measure the level of centralization. More variation yields higher network centralization scores, while less yields lower scores. In general, the larger a centralization index is, the more likely it is that a single actor is very central while the other actors are not. Thus, the index can be seen as measuring how unequal the distribution of individual actor scores are. Because we can calculate centralization using different measures of centrality (e.g., degree, betweenness, closeness, and eigenvector), we need to interpret the results in light of the type of centrality used.

11. Variance and standard deviation are similar to centralization. They differ from centralization in that rather comparing individual scores to the highest centrality score, they compare individual scores to the average centrality score. Because standard deviation is the square root of the variance, it is probably preferable to variance because it returns it returns to the original unit of analysis.[1] If you look at Figure 9 (next page), you will note that the network centralization score appears just below the descriptive statistics, whereas the variance and standard deviation are "buried" in the descriptive statistics themselves. Centralization scores range from 0.00 – 1.00 (or 0 – 100%) when analyzing dichotomized data. If you are analyzing valued data, centralization scores will sometimes be larger than 1.00; thus, it's generally a good idea to dichotomize your data before estimating network centralization.

---

[1] Variance is calculated by taking the sum of the *squared* differences between the mean and each score. Thus, by taking its squared root, we return to the original unit of analysis.

Network >Centrality and Power >_Degree (legacy), Closeness (old), Freeman Betweenness, Eigenvector centrality

12. The various centralization scores are calculated with UCINET's *Network> Centrality and Power>_Degree (legacy), Closeness (old), Freeman Betweenness, Eigenvector centrality* commands. If you use the degree command, a dialog box will appear (not shown) in which you should tell UCINET to treat the data as symmetric because you are working with undirected data. The default should be "Yes" but for some reason its "No." Click OK, and after UCINET generates the output file, scroll down past the individual centrality scores for each network, and you will find the centralization, standard deviation, and variance scores (circled in Figure 9). You will have to scroll to the bottom of each network to get the centralization scores.

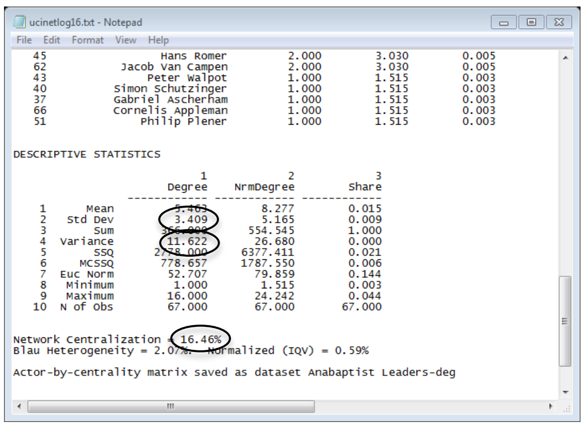


Figure 9: UCINET's Centralization Scores

Part II – Network Topography in Pajek

[Pajek] File>Pajek Project File >Read

Network>Info>General

Network>Create New Network>Transform >Arcs→Edges>All

1. Open the "`Anabaptist Leaders.paj`" project file using the *File>Pajek Project File>Read* command (or the F1 key) – ***be sure to read in the project file as a project file, not a network file***. Before calculating density and average degree, we need to first check if the network is symmetric. We do this by clicking on the network information button (circled in Figure 10) or by using the *Network>Info>General* command. This will bring up a dialog box, asking you whether you want the highest or lowest ranked ties; here, we don't care, so accept Pajek's default and click okay. This will generate a report (Figure 11) where we can see that the network only consists of edges, which means that the network is indeed symmetric. If the network wasn't symmetric, we'd first use the *Network> Create New Network>Transform>Arcs→Edges>All* command, which brings up a dialog box asking whether we want to create a new network. Typically, we select yes so that we don't overwrite any existing networks. Next, Pajek will ask us whether we want to remove multiple lines. Usually, that's a good idea, so we can select the "Single" option (#5).

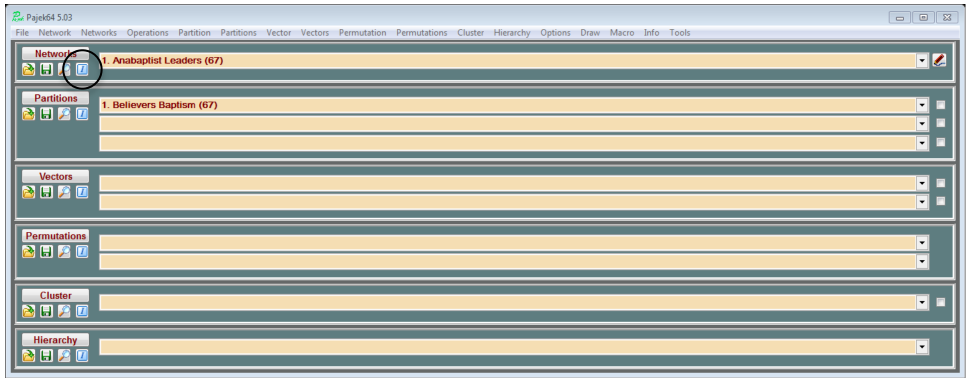


Figure 10: Pajek Main Screen

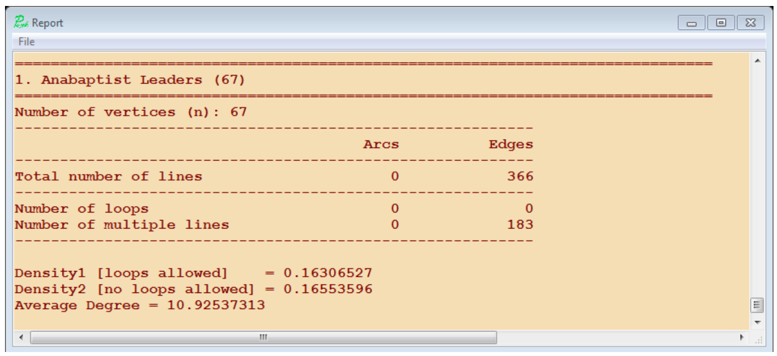


Figure 11: Pajek Network General Info Report

Network>Create New Network>Transform >Remove>Multiple Lines >Single Line

2. Looking at the network general information report in Figure 11, you can see that the network included multiple lines. Because this is a symmetric network, what this means is that Pajek is interpreting the edges as reciprocal and thus counting them twice (note that the number of multiple lines is exactly half of the total number of edges). Thus, we need to first simplify the network (get rid of the multiple lines) using the *Network>Create New Network>Transform>Remove> Multiple Lines>Single Line* command. Pajek will ask if you want to create a new network; click "Yes," and Pajek will generate a new network, probably called, "Simplification (Single Line) of N1 (67)."

Network>Info>General

3. With this network appearing in the top network dropdown box, once again use the *Network>Info>General* command, accept Pajek's defaults, and click "OK." This will generate a new report that looks quite similar to the previous report, except that the density and average degree scores are exactly half of what they are in Figure 11. See Figure 12, next page. Typically, we use the density score calculated without loops (i.e., not including the diagonal). You'll note that scores shown in Figure 12 are the same as those we found earlier (Figures 1 and 8).

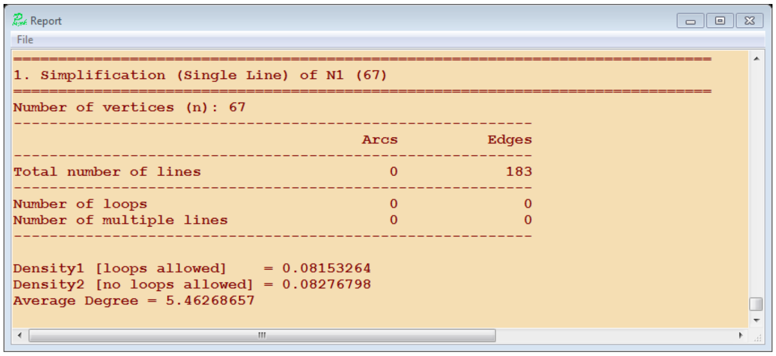


Figure 12: Pajek Network General Info Report

*Network >Create Vector >Distribution of Distances**

4. Let's check to see if Pajek gives us the same scores in terms of network size, average distance, and diameter as UCNET. Make sure that the network appears in the first network drop-down box. The number in parentheses indicates the size of the network (67). To obtain a network's average distance and diameter in Pajek, select the *Network>Create New Vector>Distribution of Distances** command. This will generate a report (Figure 13) that indicates the number of unreachable pairs (i.e., pairs of actors that are neither directly nor indirectly tied to one another), the average distance between reachable pairs (3.35414), and the names of the actors most distant from one another along with the distance/diameter (9). Both of these agree with what we found earlier with UCINET.

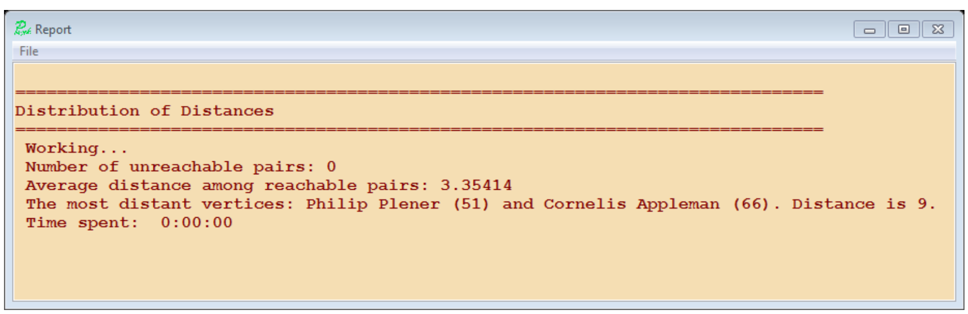


Figure 13: Pajek's Distribution of Distances Report

Network>Create Vector >Centrality>Degree>All

5. To calculate centralization, we use the same command for calculating degree centrality: *Network>Create Vector>Centrality>Degree>All*. However, we need do this with the network we created without multiple lines. Otherwise, Pajek will give an error, saying that it can't compute centralization because it the network has multiple lines. The command generates a report that gives network centralization scores (Figure 14), which is the same as UCINET's. If we want centralization based on closeness, betweenness, or eigenvector centrality we use the following commands: *Network>Create Vector>Centrality>Closeness>All*; *Network>Create Vector>Centrality>Betweenness*; and *Network>Create Vector>Centrality>Hubs-Authorities*. Note that for eigenvector centrality we use "Hubs-Authorities;" that is because with undirected networks, hubs and

authorities are the same as eigenvector centrality. We'll discuss this more in the lab that considers centrality measures.

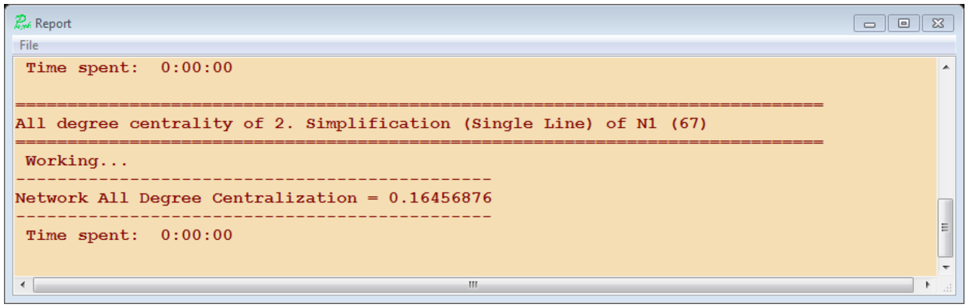


Figure 14: Pajek's Degree Centralization Report

Network>Create Vector >Clustering Coefficients >CC1

6. Finally, to calculate the clustering coefficient, use Pajek's *Network>Create Vector>Clustering Coefficients>CC1* command. This creates one partition and two vectors as well as generates a report that indicates the Watts-Strogatz clustering coefficient score (Figure 15).

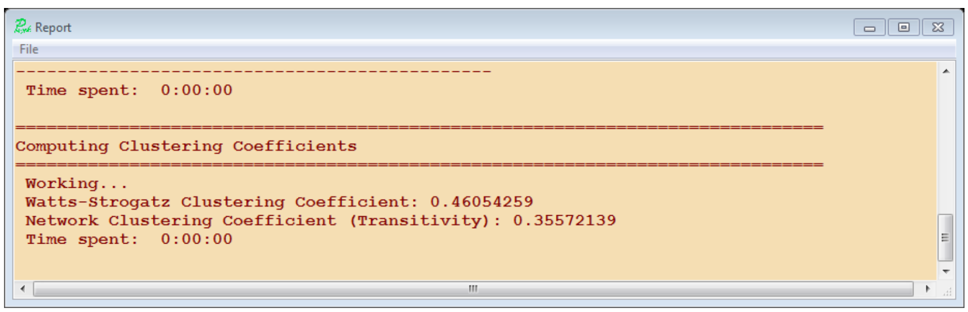


Figure 15: Pajek's Degree Clustering Coefficient Report

File>Pajek Project File >Save

7. In order to save all of your work, save it as a Pajek Project File: *File>Pajek Project File>Save* (or F2). As noted previously, project files are great for storing all of your work in a single file.