.

Social Network Analysis (SNA) Tool Comparison

Working Paper

(b)(6) DARC,

(b)(6) TID

COIC

JIEDDO

November 28, 2011

Table of Contents

Figures

Tables

Executive Summary

This is a working paper for an ongoing SNA Tool Comparison effort at the Counter-IED Operations/Intelligence Center's (COIC) Data Analysis Research and Collaboration (DARC) Cell. It contains the results of the first phase of this effort. A PowerPoint presentation summarizing this paper is also available. This paper will be edited and amended as additional results become available from subsequent phases.

The objective of this study was to compare and analyze four different Social Network Analysis (SNA) tools for the basic measures of Centrality (Degree, Closeness, Betweenness and Eigenvector), in order to set a baseline for further evaluation of the tools and their capabilities. The four tools compared were Analyst Notebook (ANB), Palantir, UCINet and ORA.

The results of this analysis are unclassified but for official use only (U//FOUO). The data and intermediate products (e.g. tool outputs, Excel sheets used for manual calculations etc.) will become available once they have been anonymized and unclassified.

Five data sets of varying sizes were used in this analysis, ranging from four agents to over two thousand agents. The data sets also had different densities, ranging from 0.6250 to 0.0006790. The smaller data sets consisting of four and sixteen agents allowed for manual calculation checks to be made on some of the Centrality measures.

For each data set in this Phase I, the data was binarized/dichotomized and used in symmetrized form to set a baseline for the results produced by the different tools.

All tools produced the same rankings by Degree Centrality for all five data sets, even though there were some differences in the centrality numbers produced by the tools.

For the smaller data sets, with four or sixteen or 209 agents, the minor 3rd and 4th decimal place differences in the various centralities did not produce any noticeable difference in the ranking of the agents for centralities other than Degree.

As stated above, for the 5th data set[1], the normalized values for the Degree Centrality were the same for UCINet and ORA, while ANB and Palantir were slightly different for the Top 20 agents. However, the ranks by Degree Centrality were the same for all four tools. There were bigger differences in Closeness and Betweenness centrality values for the 5th data set seen at the 2nd, 3rd and 4th decimal place. For Betweeness centrality, ANB and Palantir were on one side

[1] Network with 2049 agents and a density of 0.0006790.

with similar values and UCINet and ORA were on the other with similar values. For Closeness centrality, Palantir was on one side and the rest of the tools on the other[2] in the results that they produced. These differences among the tools are possibly due to different internal precision or rounding techniques or algorithmic differences between the tools that are manifested when working with larger and lower density networks. It is possible that this effect could be exaggerated further in even larger data sets that are on the order of 10K+ or 100K+ agents.

The differences seen in June 2011 between UCINet and ORA were probably a combination of default decimal settings in UCINet or default options settings in ORA which were hidden behind menus or the usage of an older ORA version (1.9.5.4.0). ORA 2.3.5, the version tested in this effort, now allows easier access to options for symmetrizing and dichotomizing data resulting in better transparency.

Working with non-symmetric and weighted (non-binarized) links data can be more involved, and while non-symmetric data was analyzed to some degree during this phase, the detailed results from that will be addressed in the follow-on phases. It is worth noting, however, that when the data was non-symmetric, the various tools produced different results in many instances. This was again likely due to different default settings in the tools, different precision or rounding techniques, or possibly algorithmic differences between the tools.

It is recommended that at this time either UCINet or ORA be used for SNA with the appropriate options settings. ANB and Palantir could be considered for SNA once the discrepancies for the largest data set have been resolved.

It is also recommended that at least two larger data sets on the order of 10K+ and 100K+ agents be analyzed for all the tools in order to investigate further the relationship between network size/densities and the number of decimal places needed for the results.

[2] Note that the 5th data set is very sparse and has many singletons and disconnected dyads and triads, making Closeness harder to calculate.

1.0 Introduction

In June 2011, an RFS requested a comparison between ANAT/ORA[3] and UCINet for IJC (ISAF Joint Command) Network Effects Cell (NEC) for use in network analysis. This RFS led to preliminary comparisons of specific centrality measures across UCINet, ORA and ANB. These preliminary comparisons revealed that while ORA and UCINet Centrality metrics seemed to mostly match, there were some differences. In addition, some ORA metrics also changed significantly depending on if the data was symmetrized or not. These preliminary findings also led to the decision to perform further investigation of UCINet (v6.35), ORA (v2.3.5[4]), ANB (v8.5.1) and Palantir (v3.4.1.4 with SNA Helper 2.1)[5] as these different tools are used to varying degrees by analysts at JIEDDO to perform SNA.

The users of the SNA tools are expected to be forward deployed analysts as well as analysts at the COIC. The analysts are expected to use the tools to investigate and analyze networks, applying traditional SNA analysis which includes, studying the structural properties of a network consisting of entities and relations, link mining which is data mining on networks seeking to be descriptive or predictive, and visual analysis which involves rendering data using different visual techniques [Loscalzo 2009]. In traditional SNA analysis the metrics usually studied are the various Centralities (Degree, Betweenness, Closeness, and Eigenvector) and other metrics such as K-Core, N-Clique, and N-Clan. In link mining the focus tends to be on prediction of existence or types of links, estimation of cardinality of links, entity reconciliation and group and sub-graph detection.

Besides documenting the comparisons across the tools, the objective of this study will also be to recommend a tool.

The rest of this document is divided into eight additional sections from 2.0 to 9.0. Section 2.0 covers Scope, 3.0 covers Limitations and 4.0 covers Assumptions. Section 5.0 starts off the Tool Evaluations by discussing Definitions and Formulae[6]. Section 6 introduces the different data

[3] ANAT stands for Advanced Network Analysis and Targeting. ORA stands for Organization Risk Analyzer.

[4] In August 2011 ORA 2.3.5 became available. While the initial analysis for this study used ORA 2.3.2, all the data sets were re-analyzed using ORA 2.3.5.

[5] Palantir has been included as it is often used in conjunction with ANB at the COIC and in theater.

[6] While the formulae for ORA are published in the tool's online help, the formulae were not available for the other tools at the time of publication for this report and are listed as TBD.

sets used in the analysis while Section 7.0 details the results from the different SNA Tools for the different data sets. Section 8.0 covers conclusions and Section 9.0 finishes up with the Next Steps.

2.0 Scope

The analysis will be divided into multiple phases, an initial Phase I and follow-on subsequent phases. This working paper covers Phase I.

In Phase I, the basic SNA Centrality measures are compared across the four tools in order to identify the similarities or differences observed across the tools. Where possible, explanations are provided for the differences.

The data consists of five data sets - one small data set of four agents which can be checked manually (e.g. using Excel); a second data set of sixteen agents building on the small data set; a third data set with the same actors as the second data set but with fewer links between the actors; a fourth larger data set consisting of 209 agents; and finally, a large data set made up of 2049 agents.

Subsequent phases of the study, when executed, are expected to be more comprehensive in the comparison of the tools, touching upon additional metrics such as K-Core, N-Cliques, N-Clans, Density etc.; different aspects of usability - from ease of input/output and import/export to availability and content of user manuals, technical support, and speed of performance (especially on larger data sets). In addition, the subsequent phases may cover additional SNA techniques that can be applied to covert networks. These could include additional Measures/Concepts such as those taught in ANAT, e.g. Structural Balance (0, 1 or -1 relationships), Clusterability, Transitivity, Diffusion, and Subgroup Analysis (CONCOR, Newman).

Based on communications with II MEF plan cell in AF RC (SW), it is understood and recognized that working with large network data sets is hard. Often there is too much data and understanding it in order to determine what is relevant in the data in real computational time becomes a challenge. To assist with this data challenge, the future phases of this effort may also include an investigation into how the SNA tools can be used to help with the analysis and understanding of large sets of data, e.g. in finding networked communities or local structures in the global structure (hairball). For the subsequent phases, an attempt may also be made to construct data sets that mimic real-world difficulties such as missing or incorrect data. This would help determine how the tools might perform in real-world situations.

3.0 Limitations –This study and analysis was constrained by having to run the latest versions of the various tools on different networks. The latest version of ANB (v8.5.1) and Palantir (v3.4.1.4) are approved to run on SLAN. However the latest versions of UCINet (v6.35) and ORA (v2.3.5) had to be approved for this study and were only allowed to run in the SDEV enclave. Since transfer of files between the two networks involves paperwork and scans by Information Assurance (IA), it slowed the movement of data, results and reports between the two networks.

4.0 Assumptions – The following assumptions were made for this study:

1. The versions of the tools evaluated are close to what will be used downrange for SNA.

2. The tools will be run downrange in an SLAN or SDEV like environment.

3. The larger data sets chosen are representative of the data sets that will be used downrange with the SNA tools.

4. The data being analyzed is symmetrized and binarized/dichotomized.

5.0 Tool Evaluation

Phase I

As mentioned above, in this Phase I of the analysis, the focus is on the basic centrality measures of Degree, Closeness, Betweenness, and Eigenvector. We first document the theoretic definitions of these measures from the academic literature. We follow that up with the theoretic and where available, tool specific formulae for the measures, noting when possible any departures from the theoretic formulae.

5.1 Metric Definitions

The approach in this study was to identify the theoretical formulae for these measures as provided in the SNA literature such as [Wasserman and Faust 1994] or [Bonacich 1972 & 1987] along with the formulae used by each of the tools, as defined in their respective user/training manuals or documentation from the tool vendor. Note that the various Centrality measures (Degree, Closeness, Betweenness and Eigenvector) used here are non-directional individual (actor) node level measures rather than directional or network level measures.

(U//FOUO) Table 1: SNA Measure Definitions for Phase I

SNA Measure	Theoretical Definition[7]	Considerations
Degree Centrality	A measure of the degree of an actor (node), defined as the number of lines incident on it	This measure depends on the network size *g* so it should be standardized before use in comparisons between networks. Related degree measures are Indegree and Outdegree which imply direction. Indegree is a measure of receptivity or popularity and Outdegree is a measure of expansiveness. Degree Centrality as defined in Phase I is on an undirected graph. This means that Indegree link from actor A to actor B will be treated the same as an Outdegree link from actor B to actor A.
Closeness Centrality	A measure of how close, in terms of links, an actor is to all the other actors in the set of actors	An actor is central if it can quickly interact with all others
Betweenness Centrality	A measure of how many times an individual is involved in the communication between two actors	Betweenness indices can be quite different measures of actor centrality than degree- and closeness-based indices
Eigenvector[8] Centrality	A measure of an actor's summed connections to others	

5.2 Metric Formulae

For a given graph/network with p nodes,

Let $\mathcal{N}$ be the node set, and $\mathcal{N}_s$ a subset on that node set,

Let the adjacency or sociomatrix be denoted by X,

Let the entries in the sociomatrix, x_{ij}, record which pairs of nodes are adjacent,

Let $d(n_i)$ be the degree of a node n_i,

Let $d(n_i, n_j)$ be the number of lines in the geodesic linking actors *i* and *j* (a geodesic is the shortest path between two nodes),

Let g_{jk} be the number of geodesics linking two actors *j* and *k*,

Let $g_{jk}(n_i)$ be the number of geodesics linking the two actors that contain actor *i*,

\

[7] From [Wasserman & Faust 1994], unless noted otherwise

[8] From [Bonacich 1972 and 1987]

Let *L* be the number of links present in the network

~~(U//FOUO)~~ Table 2: SNA Measure Theoretical Formulae

SNA Measure	Theoretical Formula[9]	Standardized Theoretical Formula[10]
Degree Centrality $C_D(n_i)$	$C_D(n_i) = d(n_i) = x_{i+} = \sum_j x_{ij} = \sum_j x_{ji}$	$C'_D(n_i) = \frac{d(n_i)}{g-1}$
Closeness Centrality $C_C(n_i)$	$C_C(n_i) = \left[\sum_{j=1}^{g} d(n_i, n_j)\right]^{-1}$	$C'_C(n_i) = (g-1)C_C(n_i)$
Betweenness Centrality $C_B(n_i)$	$C_B(n_i) = \sum_{j<k} g_{jk}(n_i)/g_{jk}$	$C'_B = C_B(n_i)/[(g-1)(g-2)/2]$
Eigenvector Centrality[11] $C_E(n_i)$	$C_E(n_i) = \frac{1}{\lambda}\sum_j x_{ij} C_E(n_j)$	N/A

~~(U//FOUO)~~ Table 3: ANB 8.5.1 Formulae

SNA Measure	ANB 8.5.1[12]	Comments
Degree Centrality	TBD	TBD
Closeness Centrality	TBD	TBD
Betweenness Centrality	TBD	TBD
Eigenvector Centrality	TBD	TBD

[9] From [Wasserman & Faust 1994]

[10] Results $\in [0, 1]$

[11] Eigenvector Centrality formula is based on [Bonacich 1972 and 1987]

[12] ANB has recently changed the underlying algorithms for the centrality measures that it computes to align with those used by UCINet (personal communication with ANB Technical Support).

(U//FOUO) Table 4: Palantir Formulae

SNA Measure	Palantir 3.4.1.4	Comments
Degree Centrality	$C_D(n_i) = d(n_i)$	TBD
Closeness Centrality	TBD	TBD
Betweenness Centrality	TBD	TBD
Eigenvector [13]Centrality	TBD	TBD

(U//FOUO) Table 5: UCINet Formulae

SNA Measure	UCINet 6.35[14]	Comments
Degree Centrality	TBD	TBD
Closeness Centrality	TBD	TBD
Betweenness Centrality	TBD	TBD
Eigenvector Centrality	TBD	TBD

(U//FOUO) Table 6: ORA Formulae

SNA Measure	ORA 2.3.5[15]	Comments
Degree Centrality [Wasserman and Faust, 1994 (pg 199)]	For a given matrix X of a square network with *n* entities, Total Degree Centrality for entity $i = \frac{1}{2(n-1)} \sum_{i=1}^{n} \sum_{j=1, i \neq j}^{n} X(i,j)$	The Total Degree Centrality here is a sum of the row and column degrees for a node. This sum is standardized via a division by *2(n-1)*. Note that the assumption here is of a symmetrized network

13 The current version of Palantir does not provide Eigenvector Centralities.

14 The current UCINet version at the COIC is 6.217

15 The current ORA version on the SLAN at the COIC is 1.9.5.4. These definitions are taken from the ORA Help file.

SNA Measure	ORA 2.3.5[15]	Comments
Closeness Centrality [Freeman 1979]	For a given graph G = (*V*, *E*) of a square matrix, fix $v \in V$ $dist = \sum_{i \in V} d_G(v,i)$ $d_G(v,i) = \lvert V \rvert$ if *i* is unreachable from *v*. Then Closeness Centrality of entity $v = (\lvert V \rvert - 1)/dist$	Once the notation is standardized to match that in Table 2 above this is the same as the Standardized Theoretical Formula in Table 2
Betweenness Centrality [Freeman 1979]	For graph G, Let n=|V|, and fix an entity $v \in V$ For $(u, w) \in VxV$, let $\sigma_G(u,w)$ be the number of geodesics in G from *u* to *w*. If $(u, w) \in E$, then set $\sigma_G(u,w) = 1$ let $S = \{(u,w) \in VxV \mid d_G(u,w) = d_G(u,v) + d_G(v,w)\}$ $between = \sum_{(u,w) \in S} (\sigma_G(u,v) \cdot \sigma_G(v,w))/\sigma_G(u,w)$ Then Betweenness Centrality of the entity $v = \frac{between}{((n-1)(n-2)/2)}$	Once the notation is standardized to match that in Table 2 above this is the same as the Standardized Theoretical Formula in Table 2
Eigenvector Centrality [Bonacich 1972]	Calculates the eigenvector of the largest positive eigenvalue of the adjacency matrix representation of a square network. A Jacobi method is used to compute the eigenvalues and vectors	The formula is not provided in ORA Help

5.3 Subsequent Phases

In subsequent phases, other functionality of the SNA tools will be evaluated. More details are covered in the Next Steps section at this end of this document.

6.0 Data

There were five datasets used in Phase I. All datasets were adapted from ANB files created at the COIC, as ANB data would be the most likely source of network data for JIEDDO users. All the datasets are Agent x Agent (unimodal) symmetric (NxN) datasets for the sake of simplicity. In this initial Phase, all datasets are dichotomized (0 or 1) and used in symmetrized (reciprocal relations) mode. Symmetrical matrices were used for the smaller data sets of four to sixteen agents and edge lists were used as inputs for the larger data sets. Note that usually the network created in ANB is not symmetrized.

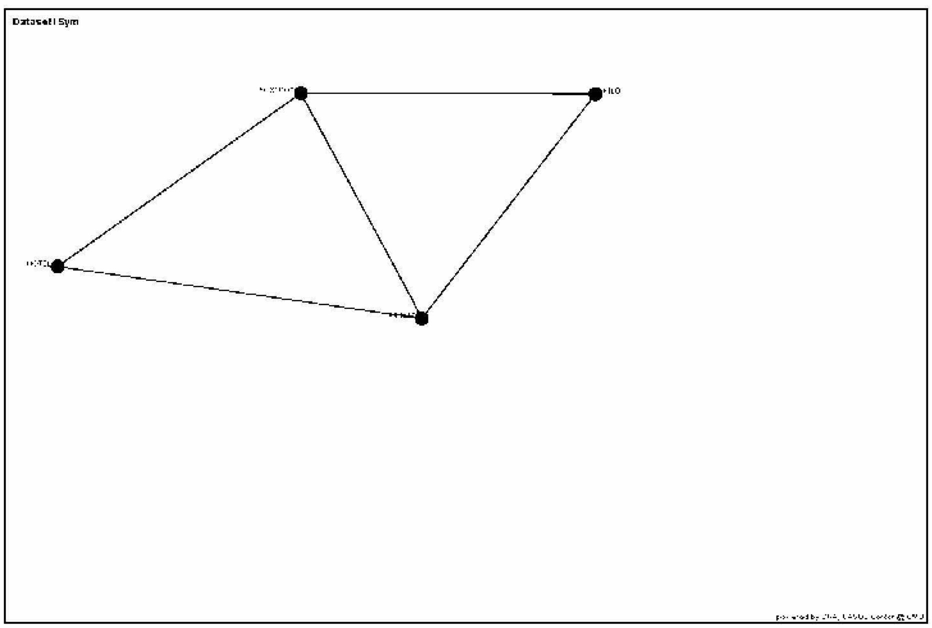
6.1 Data Set 1 (Four Agents)

The smallest dataset consists of four (4) agents. It was derived from an ANB data set made up of sixteen (16) agents that forms the basis of the second and third data sets discussed below. The following table shows the adjacency matrix representing Data Set 1 defines the links between the different agents. A zero (0) indicates no link and a one (1) indicates a link. Note that the data is symmetrized. All links that exist have a weight of one and are non-directional.

(U//FOUO) Table 7: Data Set 1 Adjacency Matrix- Symmetrized

Node ID	HOTEL	FOXTROT	KILO	JULIETT
HOTEL	0	1	0	1
FOXTROT	1	0	1	1
KILO	0	1	0	1
JULIETT	1	1	1	0

The following figure[16] shows a network representation of Data Set 1.



(U//FOUO) Figure 1: Network for Data Set 1

[16] ORA has been used here to draw the network. UCINet could also have been used.

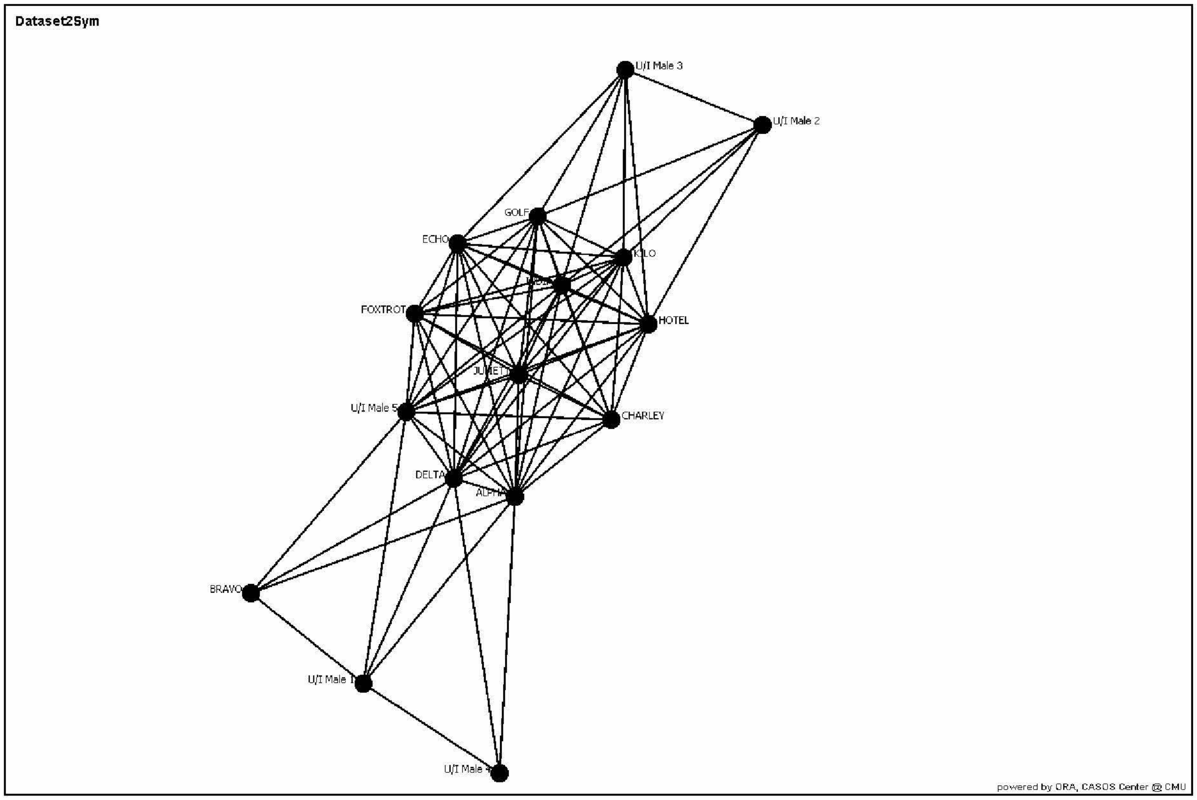
6.2 Data Set 2 (Sixteen Agents)

The second dataset consisted of sixteen (16) agents. The following adjacency matrix in (U//FOUO) Table 8 defines the links between the different agents. Once again, a zero (0) indicates no link and a one (1) indicates a link. Note that this data has also been symmetrized and all links that exist have a weight of one and are non-directional.

(U//FOUO) Table 8: Data Set 2 Adjacency Matrix- Symmetrized

	U/I Male 2	HOTEL	INDIA	U/I Male 5	CHARLEY	FOXTROT	KILO	JULIETT	ECHO	U/I Male 3	GOLF	DELTA	U/I Male 4	U/I Male 1	ALPHA	BRAVO
U/I Male 2	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0
HOTEL	1	0	1	1	1	1	1	1	1	1	1	1	0	0	1	0
INDIA	1	1	0	1	1	1	1	1	1	1	1	1	0	0	1	0
U/I Male 5	0	1	1	0	1	1	1	1	1	0	1	1	0	1	1	1
CHARLEY	0	1	1	1	0	1	1	1	1	0	1	1	0	0	1	0
FOXTROT	0	1	1	1	1	0	1	1	1	0	1	1	0	0	1	0
KILO	1	1	1	1	1	1	0	1	1	1	1	1	0	0	1	0
JULIETT	0	1	1	1	1	1	1	0	1	0	1	1	0	0	1	0
ECHO	0	1	1	1	1	1	1	1	0	1	1	1	0	0	1	0
U/I Male 3	1	1	1	0	0	0	1	0	1	0	1	0	0	0	0	0
GOLF	1	1	1	1	1	1	1	1	1	1	0	1	0	0	1	0
DELTA	0	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1
U/I Male 4	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
U/I Male 1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1
ALPHA	0	1	1	1	1	1	1	1	1	0	1	1	1	1	0	1
BRAVO	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0

The following (U//FOUO) Figure 2 shows a network representation of Data Set 2. As can be seen from the network diagram, this is a highly connected graph with a dense core.



(U//FOUO) Figure 2: Network for Data Set 2

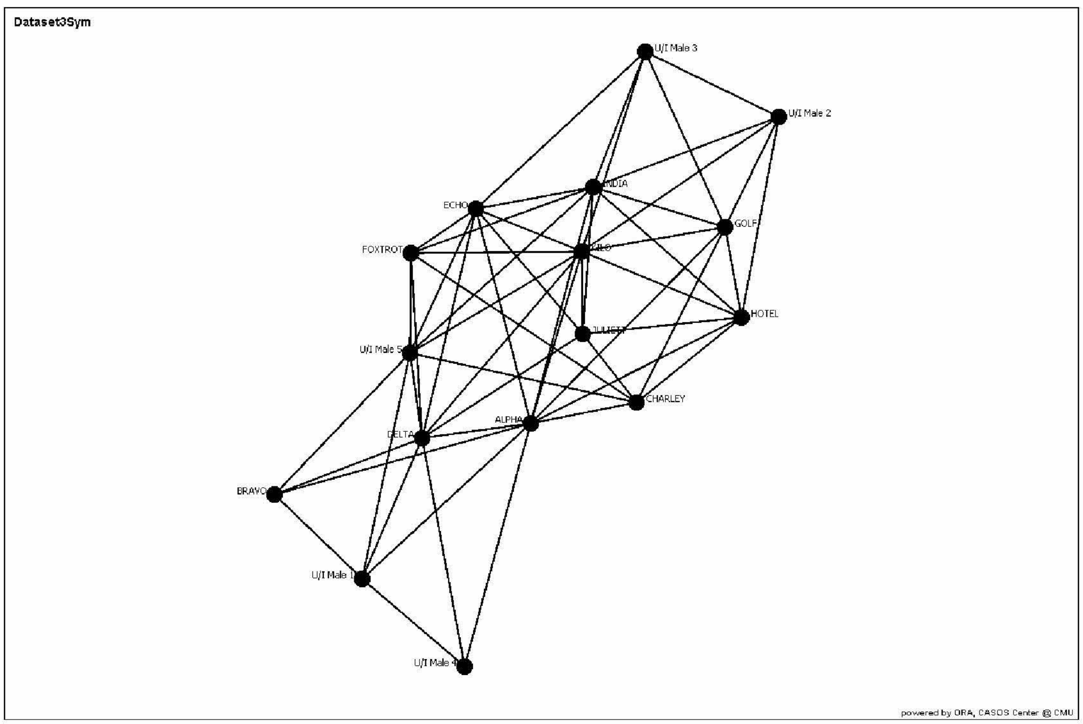
6.3 Data Set 3 (Similar to Data Set 2 but sparser)

The third dataset consisted of a variation on the second data set. To construct this data set, several links were removed from the second data set. The following figures show the network for this dataset for symmetrized data.

(U//FOUO) Table 9: Data Set 3 Adjacency Matrix- Symmetrized

	U/I Male 2	HOTEL	INDIA	U/I Male 5	CHARLEY	FOXTROT	KILO	JULIETT	ECHO	U/I Male 3	GOLF	DELTA	U/I Male 4	U/I Male 1	ALPHA	BRAVO
U/I Male 2	0	1	1	0	0	0	1	0	0	1	1	0	0	0	0	0
HOTEL	1	0	1	0	1	0	1	1	0	0	1	0	0	0	1	0
INDIA	1	1	0	1	0	1	0	1	1	1	1	0	0	0	1	0
U/I Male 5	0	0	1	0	1	1	1	0	1	0	0	1	0	1	0	1
CHARLEY	0	1	0	1	0	1	0	1	0	0	1	0	0	0	1	0
FOXTROT	0	0	1	1	1	0	1	0	1	0	0	1	0	0	0	0
KILO	1	1	0	1	0	1	0	1	1	1	1	1	0	0	1	0
JULIETT	0	1	1	0	1	0	1	0	1	0	0	1	0	0	0	0
ECHO	0	0	1	1	0	1	1	1	0	1	0	1	0	0	1	0
U/I Male 3	1	0	1	0	0	0	1	0	1	0	1	0	0	0	0	0
GOLF	1	1	1	0	1	0	1	0	0	1	0	0	0	0	1	0
DELTA	0	0	0	1	0	1	1	1	1	0	0	0	1	1	1	1
U/I Male 4	0	0	0	0	0	0	0	0	0	0	0	1	0	1	1	0
U/I Male 1	0	0	0	1	0	0	0	0	0	0	0	1	1	0	1	1
ALPHA	0	1	1	0	1	0	1	0	1	0	1	1	1	1	0	1
BRAVO	0	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0

The following (U//FOUO) Figure 3 shows a network representation of symmetrized Data Set 3.

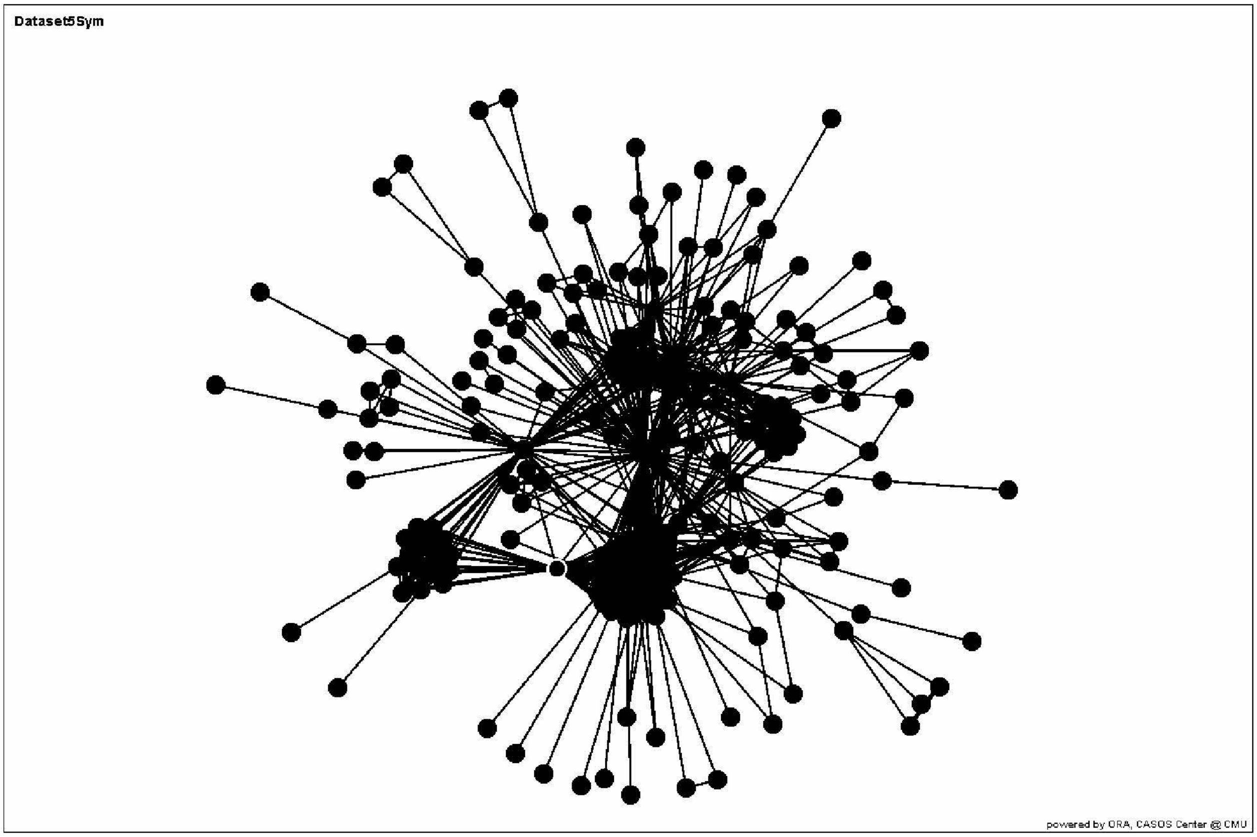


(U//FOUO) Figure 3: Network for Data Set 3 (Symmetrized)

6.4 Data Set 4

The fourth data set was the same one used in June for RFS 80857. At 209 agents, it was an order of magnitude larger than Data Sets 2 and 3. Data Set 4 had 3648 links once it was symmetrized.

The following (U//FOUO) Figure 4 shows the network for Data Set 4. Due to the large number of agents (209) in this data set, a matrix representation is not included here.



(U//FOUO) Figure 4: Network for Data Set 4 (Symmetrized) without node labels

6.5 Data Set 5

The fifth data set was an order of magnitude larger than the fourth data set and consisted of 2,049 agents. While this network had a larger set of nodes than the previous data sets, it was not as highly connected. In fact, many of the nodes in this network were singletons or independent dyads or triads. The network diagram for this data set is not shown here due to the large number of nodes.

6.6 Data Set Densities

The following table compares the number of agents and densities for the different data sets.

(U//FOUO) Table 10: Data Sets - Agents and Densities

	Number of Agents	Density
Data Set 1	4	0.6250000
Data Set 2	16	0.5859375
Data Set 3	16	0.4218750
Data Set 4	209	0.0832080
Data Set 5	2049	0.0006790

7.0 Tool Runs – In order to evaluate the various tools, the different data sets were run through the four tools and the results were compared between the tools. Where the data set was small enough, manual calculations were made for comparison purposes. The results of each run were normalized in order to enable fair comparisons between the manual calculations and the numbers reported by the tools. The normalization process divided each number in a column with the largest number in that column. This normalization was mostly done in Excel outside of the various tools.

7.1 Data Set 1

7.1.1 Manual Computation

For the smallest dataset, the following table shows the manually[17] computed Centrality measures for symmetrized data. The Standardized Theoretical Formulae from (U//FOUO) Table 2 above were used and the table has been sorted by Degree Centrality, highest to lowest.

(U//FOUO) Table 11: Manually computed Centrality values for Symmetrized Data Set 1

Node Num	Node ID	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
1	FOXTROT	1.0000	1.0000	0.3333	1.0000
2	JULIETT	1.0000	1.0000	0.3333	1.0000
3	HOTEL	0.6667	0.7500	0	0.7808
4	KILO	0.6667	0.7500	0	0.7808

[17] An Excel spreadsheet (DataSet1ManualCalcs.xlsx) was used to assist in the manual calculation.

(U//FOUO) Table 12: Manually computed Centrality values for Symmetrized Data Set 1 (Normalized))

Node Num	Node ID	Degree Centrality	Closeness Centrality	Betweenness Centrality	Eigenvector Centrality
1	FOXTROT	1.0000	1.0000	1.0000	1.0000
2	JULIETT	1.0000	1.0000	1.0000	1.0000
3	HOTEL	0.6667	0.7500	0	0.7808
4	KILO	0.6667	0.7500	0	0.7808

7.1.2 SNA Tool Computations

The following (U//FOUO) Table 13 shows the normalized Centrality values as computed and reported by the different SNA Tools for the symmetrized data version of Data Set 1.

For **ANB**, the **Social Network Analysis** functionality was used to compute these values. As can be seen from the table below, the centrality values computed by ANB match the manual calculations for symmetrized data.

The numbers produced by **Palantir** also match the manually computed numbers in

(U//FOUO) Table 12. Note that the version of Palantir used in this study does not provide Eigenvector centralities.

In **UCINet** the **Network | Centrality and Power | Multiple Measures** menu option was used for calculating the Degree, Betweenness and Eigenvector Centrality values. The **Network | Centrality and Power | Closeness** menu option was used for obtaining the Closeness centrality value. Once normalization is taken into account, the UCINet values in (U//FOUO) Table 13 are the same as the normalized manually computed values in
(U//FOUO) Table 12 above.

In ORA, the **Analysis | Generate Reports | Show me everything (All Measures)** option was used to generate the centrality measures. All the Centrality measures reported by ORA also match the manually computed ones in
(U//FOUO) Table 12.

(U//FOUO) Table 13: Centrality values for Symmetrized Data Set 1

		Degree Centrality				Closeness Centrality				Betweenness Centrality				Eigenvector Centrality			
Node Num	Node ID	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA
1	FOXTROT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	N/A	1.0000	1.0000
2	JULIETT	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	N/A	1.0000	1.0000
3	HOTEL	0.6667	0.6667	0.6667	0.6667	0.7500	0.7500	0.7500	0.7500	0.0000	0.0000	0.0000	0.0000	0.7808	N/A	0.7808	0.7808
4	KILO	0.6667	0.6667	0.6667	0.6667	0.7500	0.7500	0.7500	0.7500	0.0000	0.0000	0.0000	0.0000	0.7808	N/A	0.7808	0.7808

7.1.3 Discussion of Differences

The results of the runs for Data Set 1, once normalization and scaling is taken into account show that all Centrality values match exactly the manual calculations across the tools.

7.2 Data Set 2

Due to the larger size of the second data set, only the Degree Centrality was calculated manually, as shown in (U//FOUO) Table 14 below. While the Degree Centrality values produced by the different tools are compared against the manually computed values, the other centrality values are compared between the different tools.

7.2.1 Manual Computation

For the second dataset, the following table shows the manually[18] computed Degree Centrality measures for symmetrized data. Once again, the Standardized Theoretical Formulae from (U//FOUO) Table 2 above were used. The table below has been normalized and sorted by Degree Centrality, highest to lowest.

(U//FOUO) Table 14: Manually computed Degree Centrality values for Data Set 2 (Normalized)

Node Num	Node ID	Degree Centrality
1	ALPHA	1.0000
2	DELTA	1.0000
3	GOLF	0.9231
4	HOTEL	0.9231
5	INDIA	0.9231
6	KILO	0.9231
7	U/I Male 5	0.9231
8	ECHO	0.8462
9	CHARLEY	0.7692
10	FOXTROT	0.7692
11	JULIETT	0.7692
12	U/I Male 3	0.4615
13	U/I Male 1	0.3846
14	U/I Male 2	0.3846
15	BRAVO	0.3077
16	U/I Male 4	0.2308

[18] Once again, an Excel spreadsheet was used to assist in the manual calculation.

7.2.2 SNA Tool Computations

The following (U//FOUO) Table 15 shows the normalized centrality measures reported by the different SNA Tools for symmetrized Data Set 2.

The Degree centrality values produced by ANB match the manually computed ones in (U//FOUO) Table 14, except for one agent in the 4th decimal place.

The following table also shows the normalized centrality measures reported by Palantir for symmetrized Data Set 2. The Degree centrality values produced by Palantir match the manually computed ones. The Closeness values are the same as those produced by ANB as are the Betweenness values.

The normalized centrality measures reported by UCINet, when compared with the manually computed Degree Centrality in (U//FOUO) Table 14, show that the corresponding numbers reported by UCINet match quite closely, and any differences are in the 4th decimal place. The Closeness Centrality numbers match the ANB numbers exactly. The Betweenness and Eigenvector Centralities from UCINet match the ANB numbers closely except for in the 4th decimal place in some instances.

The normalized centrality measures reported by ORA for symmetrized Data Set 2 show that the Degree Centrality reported by ORA closely matches the manually computed values in (U//FOUO) Table 14. Any differences that exist are in the 4th decimal place. The Closeness Centrality computed by ORA is almost exactly the same as those produced by ANB, Palantir and UCINet, differing in some cases in the 4th decimal place. The Betweenness Centrality numbers in ORA are different from the ANB and Palantir numbers in the 4th decimal place for some agents but are the same as those from UCINet. The Eigenvector Centralities computed by ORA are the same as those computed by ANB, and close to the ones produced by UCINet, differing in the 4th decimal place.

~~(U//FOUO)~~ Table 15: Centrality values for Symmetrized Data Set 2

		Degree Centrality				Closeness Centrality				Betweenness Centrality				Eigenvector Centrality			
Node Num	Node ID	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA
1	ALPHA	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9935	N/A	0.9936	0.9935
2	DELTA	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9935	N/A	0.9936	0.9935
3	GOLF	0.9231	0.9231	0.9230	0.9231	0.9444	0.9444	0.9444	0.9444	0.3446	0.3446	0.3447	0.3447	1.0000	N/A	1.0000	1.0000
4	HOTEL	0.9231	0.9231	0.9230	0.9231	0.9444	0.9444	0.9444	0.9444	0.3446	0.3446	0.3447	0.3447	1.0000	N/A	1.0000	1.0000
5	INDIA	0.9231	0.9231	0.9230	0.9231	0.9444	0.9444	0.9444	0.9444	0.3446	0.3446	0.3447	0.3447	1.0000	N/A	1.0000	1.0000
6	KILO	0.9231	0.9231	0.9230	0.9231	0.9444	0.9444	0.9444	0.9444	0.3446	0.3446	0.3447	0.3447	1.0000	N/A	1.0000	1.0000
7	U/I Male 5	0.9231	0.9231	0.9230	0.9231	0.9444	0.9444	0.9444	0.9444	0.5405	0.5405	0.5404	0.5404	0.9747	N/A	0.9748	0.9747
8	ECHO	0.8461	0.8462	0.8461	0.8462	0.8947	0.8947	0.8947	0.8947	0.1459	0.1459	0.1455	0.1455	0.9635	N/A	0.9635	0.9635
9	CHARLEY	0.7692	0.7692	0.7692	0.7693	0.8500	0.8500	0.8500	0.8500	0.0000	0.0000	0.0000	0.0000	0.9198	N/A	0.9197	0.9198
10	FOXTROT	0.7692	0.7692	0.7692	0.7693	0.8500	0.8500	0.8500	0.8500	0.0000	0.0000	0.0000	0.0000	0.9198	N/A	0.9197	0.9198
11	JULIETT	0.7692	0.7692	0.7692	0.7693	0.8500	0.8500	0.8500	0.8500	0.0000	0.0000	0.0000	0.0000	0.9198	N/A	0.9197	0.9198
12	U/I Male 3	0.4615	0.4615	0.4615	0.4615	0.6296	0.6296	0.6296	0.6296	0.0162	0.0162	0.0162	0.0162	0.5075	N/A	0.5076	0.5075
13	U/I Male 1	0.3846	0.3846	0.3846	0.3846	0.6296	0.6296	0.6296	0.6296	0.0541	0.0541	0.0536	0.0536	0.3287	N/A	0.3288	0.3287
14	U/I Male 2	0.3846	0.3846	0.3846	0.3846	0.6071	0.6071	0.6296	0.6071	0.0000	0.0000	0.0000	0.0000	0.4246	N/A	0.4246	0.4246
15	BRAVO	0.3077	0.3077	0.3077	0.3077	0.6071	0.6071	0.6296	0.6071	0.0000	0.0000	0.0000	0.0000	0.3099	N/A	0.3100	0.3100
16	U/I Male 4	0.2308	0.2308	0.2308	0.2307	0.5862	0.5862	0.5862	0.5861	0.0000	0.0000	0.0000	0.0000	0.2181	N/A	0.2182	0.2181

7.2.3 Discussion of Differences

The values produced by the different tools are again very similar across the different centralities. The differences that appear in the 3rd and 4th decimal places do not cause any changes in ranking across the tools when sorted by Degree Centrality.

7.3 Data Set 3

As discussed above, Data Set 3 is a variation on Data Set 2, constructed by removing selected links from Data Set 2. The decision to construct and analyze Data Set 3 was made in order to see if the density of the network has any impact on the fidelity of the results produced by the tools.

7.3.1 Manual Computation

For the third dataset, the following table shows the manually computed Degree Centrality measures. Once again, the Standardized Theoretical Formulae from (U//FOUO) Table 2 above were used. The table below has been normalized and sorted by Degree Centrality, highest to lowest. Even though the Agents in this data set are the same as in Data Set 2, note that these values are different from those in Data Set 2 because for this data set several links were removed.

(U//FOUO) Table 16: Manually computed Degree Centrality values for Data Set 3 (Normalized)

Node Num	Node ID	Degree Centrality
1	ALPHA	1.0000
2	KILO	1.0000
3	DELTA	0.9000
4	INDIA	0.9000
5	ECHO	0.8000
6	U/I Male 5	0.8000
7	GOLF	0.7000
8	HOTEL	0.7000
9	CHARLEY	0.6000
10	FOXTROT	0.6000
11	JULIETT	0.6000
12	U/I Male 1	0.5000
13	U/I Male 2	0.5000
14	U/I Male 3	0.5000
15	BRAVO	0.4000
16	U/I Male 4	0.3000

7.3.2 SNA Tool Computations

The following (U//FOUO) Table 17 shows the centrality measures reported by the different SNA Tools for symmetrized Data Set 3 after they have been normalized.

As can be seen from the (U//FOUO) Table 17 below, the Degree centralities produced by ANB are the same as from the manual calculation.

The Degree centralities produced by Palantir are also the same as in the manual calculation. The Closeness and Betweenness values are the same as from ANB.

The normalized centrality measures reported by UCINet for symmetrized Data Set 3 show that the Degree Centrality values are quite close to the manually computed values, differing only in the 4th decimal place for a few agents. The Closeness values are the same as those from ANB and Palantir. The Betweenness and Eigenvector Centralities are again quite similar to those produced by ANB, differing mostly in the 4th decimal place.

From the table below, it can be seen that the Degree Centrality values reported by ORA are quite close to those computed manually, differing only in the 4th decimal place. The other centralities are quite similar to those produced by UCINet and ANB, differing only in the 4th decimal place, in a manner similar to those for Data Set 2.

(U//FOUO) Table 17: Centrality values for Symmetrized Data Set 3

		Degree Centrality				Closeness Centrality				Betweenness Centrality				Eigenvector Centrality			
Node Num	Node ID	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA
1	ALPHA	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9328	N/A	0.9327	0.9328
2	KILO	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.5900	0.5900	0.5901	0.5901	1.0000	N/A	1.0000	1.0000
3	DELTA	0.9000	0.9000	0.9000	0.8999	0.9524	0.9524	0.9524	0.9524	0.5546	0.5546	0.5546	0.5546	0.8352	N/A	0.8351	0.8351
4	INDIA	0.9000	0.9000	0.9000	0.8999	0.9524	0.9524	0.9524	0.9524	0.4180	0.4180	0.4178	0.4178	0.8858	N/A	0.8857	0.8858
5	ECHO	0.8000	0.8000	0.7999	0.7999	0.9091	0.9091	0.9091	0.9091	0.2130	0.2130	0.2128	0.2128	0.8680	N/A	0.8680	0.8680
6	U/I Male 5	0.8000	0.8000	0.7999	0.7999	0.9091	0.9091	0.9091	0.9091	0.3602	0.3602	0.3604	0.3604	0.7863	N/A	0.7862	0.7863
7	GOLF	0.7000	0.7000	0.7000	0.6999	0.8696	0.8696	0.8696	0.8696	0.1914	0.1914	0.1914	0.1914	0.7181	N/A	0.7181	0.7181
8	HOTEL	0.7000	0.7000	0.7000	0.6999	0.8696	0.8696	0.8696	0.8696	0.1303	0.1304	0.1301	0.1301	0.7335	N/A	0.7336	0.7335
9	CHARLEY	0.6000	0.6000	0.6000	0.6000	0.8333	0.8333	0.8333	0.8333	0.1337	0.1337	0.1334	0.1334	0.6195	N/A	0.6195	0.6195
10	FOXTROT	0.6000	0.6000	0.6000	0.6000	0.8333	0.8333	0.8333	0.8333	0.0614	0.0614	0.0614	0.0614	0.6831	N/A	0.6830	0.6831
11	JULIETT	0.6000	0.6000	0.6000	0.6000	0.8333	0.8333	0.8333	0.8333	0.0927	0.0927	0.0929	0.0929	0.6759	N/A	0.6759	0.6759
12	U/I Male 1	0.5000	0.5000	0.4999	0.5000	0.7407	0.7407	0.7407	0.7408	0.0524	0.0524	0.0524	0.0524	0.4469	N/A	0.4468	0.4469
13	U/I Male 2	0.5000	0.5000	0.4999	0.5000	0.7143	0.7143	0.7143	0.7143	0.0194	0.0194	0.0191	0.0191	0.5313	N/A	0.5312	0.5313
14	U/I Male 3	0.5000	0.5000	0.4999	0.5000	0.7143	0.7143	0.7143	0.7143	0.0372	0.0372	0.0372	0.0372	0.5475	N/A	0.5474	0.5475
15	BRAVO	0.4000	0.4000	0.4000	0.4000	0.7143	0.7143	0.7143	0.7143	0.0077	0.0077	0.0079	0.0079	0.4105	N/A	0.4103	0.4105
16	U/I Male 4	0.3000	0.3000	0.3000	0.2999	0.6897	0.6897	0.6897	0.6896	0.0000	0.0000	0.0000	0.0000	0.3029	N/A	0.3029	0.3029

7.3.6 Discussion of Differences

In general, the results line up well across the tools and do not result in changes in rank, just as for Data Set 2. Changes in density do not appear to impact the results materially, at least at the level of a smaller data set and the density levels seen in Data Set 2 and Data Set 3.

7.4 Data Set 4

7.4.1 SNA Tool Computations

The following (U//FOUO) Table 18 shows the normalized centrality measures, for the Top 20 Agents (selected on the basis of the highest Degree Centrality), reported by the different SNA Tools for symmetrized Data Set 4, which had a total of 209 agents.

The Top 20 agents are the same across the four tools. The order of the Top 20 agents, based on Degree Centrality, is also the same across the tools.

The Degree Centrality numbers produced by Palantir are the same as those produced by ANB, as are the Closeness and Betweenness values.

The Degree centrality values produced by UCINet are almost the same as those from ANB and Palantir, differing only in the 4th decimal place for two agents. The Closeness values are the same as ANB and Palantir. The Betweenness and Eigenvector values are similar to ANB, differing in the 4th decimal place.

The results reported by ORA are quite similar to the other tools. The names and order of agents is the same and the numbers are similar, differing only in the 4th decimal place.

7.4.2 Discussion of Differences

From the table below it can be seen that the results of the Top 20 agents for all the tools match. Any differences among the four tools are in the 4th decimal place.

~~(U//FOUO)~~ Table 18: Centrality values for Symmetrized Data Set 4

		Degree Centrality				Closeness Centrality				Betweenness Centrality				Eigenvector Centrality			
Node Num	Node ID	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA
1	a080	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	N/A	1.0000	1.0000
2	a170	0.5339	0.5339	0.5339	0.5340	0.7835	0.7835	0.7835	0.7835	0.1626	0.1626	0.1626	0.1626	0.9668	N/A	0.9667	0.9668
3	a023	0.5085	0.5085	0.5085	0.5086	0.7657	0.7657	0.7657	0.7657	0.0924	0.0924	0.0925	0.0925	0.9646	N/A	0.9648	0.9646
4	a113	0.4915	0.4915	0.4915	0.4916	0.7896	0.7896	0.7896	0.7897	0.0555	0.0555	0.0555	0.0555	0.9719	N/A	0.9718	0.9718
5	a059	0.4661	0.4661	0.4661	0.4662	0.7855	0.7855	0.7855	0.7856	0.0780	0.0780	0.0780	0.0780	0.9662	N/A	0.9662	0.9662
6	a043	0.4576	0.4576	0.4576	0.4577	0.7415	0.7415	0.7415	0.7414	0.0247	0.0247	0.0246	0.0246	0.9602	N/A	0.9601	0.9601
7	a123	0.4407	0.4407	0.4407	0.4407	0.7815	0.7815	0.7815	0.7815	0.3934	0.3934	0.3935	0.3935	0.1800	N/A	0.1802	0.1800
8	a128	0.4322	0.4322	0.4322	0.4322	0.7415	0.7415	0.7415	0.7414	0.0324	0.0324	0.0325	0.0325	0.9571	N/A	0.9573	0.9571
9	a112	0.4237	0.4237	0.4238	0.4237	0.7343	0.7343	0.7343	0.7343	0.0105	0.0105	0.0106	0.0106	0.9562	N/A	0.9564	0.9562
10	a117	0.4237	0.4237	0.4238	0.4237	0.7343	0.7343	0.7343	0.7343	0.0416	0.0416	0.0416	0.0416	0.9556	N/A	0.9554	0.9556
11	a003	0.4153	0.4153	0.4153	0.4153	0.7290	0.7290	0.7290	0.7290	0.0166	0.0166	0.0166	0.0166	0.9545	N/A	0.9545	0.9545
12	a070	0.4153	0.4153	0.4153	0.4153	0.7290	0.7290	0.7290	0.7290	0.0166	0.0166	0.0166	0.0166	0.9545	N/A	0.9545	0.9545
13	a108	0.4153	0.4153	0.4153	0.4153	0.7290	0.7290	0.7290	0.7290	0.0303	0.0303	0.0302	0.0302	0.9533	N/A	0.9535	0.9533
14	a110	0.4153	0.4153	0.4153	0.4153	0.7562	0.7562	0.7562	0.7562	0.0120	0.0120	0.0120	0.0120	0.9585	N/A	0.9587	0.9584
15	a007	0.4068	0.4068	0.4068	0.4068	0.7308	0.7308	0.7308	0.7308	0.0015	0.0015	0.0014	0.0014	0.9576	N/A	0.9578	0.9575
16	a008	0.4068	0.4068	0.4068	0.4068	0.7308	0.7308	0.7308	0.7308	0.0015	0.0015	0.0014	0.0014	0.9576	N/A	0.9578	0.9575
17	a011	0.4068	0.4068	0.4068	0.4068	0.7308	0.7308	0.7308	0.7308	0.0015	0.0015	0.0014	0.0014	0.9576	N/A	0.9578	0.9575
18	a014	0.4068	0.4068	0.4068	0.4068	0.7415	0.7415	0.7415	0.7414	0.0124	0.0124	0.0124	0.0124	0.9545	N/A	0.9545	0.9545
19	a028	0.4068	0.4068	0.4068	0.4068	0.7308	0.7308	0.7308	0.7308	0.0015	0.0015	0.0014	0.0014	0.9576	N/A	0.9578	0.9575
20	a048	0.4068	0.4068	0.4068	0.4068	0.7273	0.7273	0.7273	0.7273	0.0200	0.0200	0.0201	0.0201	0.9528	N/A	0.9531	0.9528

7.5 Data Set 5

7.5.1 SNA Tool Computations

The following (U//FOUO) Table 19 shows the normalized centrality measures, for the Top 20 agents (selected on the basis of the highest Degree Centrality), reported by the different SNA Tools for symmetrized Data Set 5.

The Top 20 agents are the same across the four tools. The order of the Top 20 names, based on Degree Centrality, is also the same across the tools.

As reported by Palantir, the Degree Centrality values are similar to those from ANB, differing in the 4th decimal place. The Closeness values produced by Palantir are quite different from those produced by ANB. The Betweenness values produced by Palantir are the same as ANB.

For UCINet, the Degree and Closeness Centralities are close to those computed by ANB, differing in the 3rd or 4th decimal place. The Betweenness centrality is quite different from ANB and Palantir. The Eigenvector centralities are very close to the ones from ANB, differing in the 4th decimal place in some cases.

The Degree Centrality for the Top 20 agents reported by ORA is close to the values for ANB and Palantir, differing in the 3rd or 4th decimal place and the same as those produced by UCINet. The Closeness values are similar to ANB and UCINet, with the differences in the 2nd decimal place where they exist. The Betweenness values are similar to those from ANB, sometimes differing in the 2nd or 3rd decimal place and the same as UCINet. The Eigenvector values are quite similar to the ones produced by ANB and the same as UCINet, differing in the 4th decimal place where the differences exist.

7.5.2 Discussion of Differences

As can be seen from the table below, the Top 20 list ranking for the agents is the same between ANB, Palantir, UCINet and ORA when sorted by Degree Centrality. The Closeness and Betweenness values show more variability across the tools with Palantir being the outlier.

~~(U//FOUO)~~ Table 19: Centrality values for Symmetrized Data Set 5

		Degree Centrality				Closeness Centrality				Betweenness Centrality				Eigenvector Centrality			
Node Num	Node ID	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA	ANB	Palantir	UCINet	ORA
1	a181	1.0000	1.0000	1.0000	1.0000	0.9806	0.2907	0.9749	1.0000	0.3602	0.3602	0.3500	0.3500	0.0000	N/A	0.0000	0.0001
2	a02A	0.9333	0.9286	0.9315	0.9315	0.9587	0.8759	0.9498	1.0000	0.0275	0.0275	0.0500	0.0500	0.0000	N/A	0.0000	0.0000
3	a1F7	0.8666	0.8571	0.8630	0.8630	1.0000	0.2111	1.0000	1.0000	0.8913	0.8913	0.9000	0.9000	1.0000	N/A	1.0000	1.0000
4	a47F	0.8666	0.8571	0.8630	0.8630	1.0000	0.1531	1.0000	1.0000	0.3488	0.3488	0.3500	0.3500	0.2050	N/A	0.2049	0.2050
5	a78B	0.7334	0.7143	0.7397	0.7397	0.9806	0.2823	0.9749	1.0000	0.3440	0.3440	0.3500	0.3500	0.0000	N/A	0.0000	0.0000
6	a436	0.6667	0.6429	0.6712	0.6712	1.0000	0.1207	1.0000	1.0000	0.4363	0.4363	0.4500	0.4500	0.0415	N/A	0.0416	0.0414
7	a95B	0.6667	0.6429	0.6712	0.6712	1.0000	0.2095	1.0000	1.0000	0.9209	0.9209	0.9500	0.9500	0.3645	N/A	0.3644	0.3645
8	a3E6	0.6000	0.5714	0.6027	0.6027	0.9684	0.1712	0.9614	1.0000	0.0669	0.0669	0.0500	0.0500	0.0000	N/A	0.0000	0.0000
9	a734	0.6000	0.5714	0.6027	0.6027	1.0000	0.2300	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.7698	N/A	0.7698	0.7698
10	a967	0.6000	0.5714	0.6027	0.6027	0.9854	0.2148	0.9807	1.0000	0.5406	0.5406	0.5500	0.5500	0.0000	N/A	0.0000	0.0000
11	a191	0.5333	0.5000	0.5342	0.5342	0.9587	0.4955	0.9498	1.0000	0.0257	0.0257	0.0500	0.0500	0.0000	N/A	0.0000	0.0000
12	a664	0.5333	0.5000	0.5342	0.5342	0.9684	0.3044	0.9614	1.0000	0.1318	0.1318	0.1500	0.1500	0.0000	N/A	0.0000	0.0000
13	a72F	0.5333	0.5000	0.5342	0.5342	1.0000	0.1558	1.0000	1.0000	0.4509	0.4509	0.4500	0.4500	0.3391	N/A	0.3391	0.3392
14	a9D0	0.5333	0.5000	0.5342	0.5342	0.9806	0.3087	0.9749	1.0000	0.3463	0.3463	0.3500	0.3500	0.0000	N/A	0.0000	0.0001
15	a080	0.4666	0.4286	0.4658	0.4658	0.9563	0.6276	0.9479	1.0000	0.0139	0.0139	0.0000	0.0000	0.0000	N/A	0.0000	0.0000
16	a1EB	0.4666	0.4286	0.4658	0.4658	1.0000	0.1186	1.0000	1.0000	0.1301	0.1301	0.1500	0.1500	0.0541	N/A	0.0540	0.0540
17	a545	0.4666	0.4286	0.4658	0.4658	0.9854	0.1454	0.9807	1.0000	0.2962	0.2962	0.3000	0.3000	0.0000	N/A	0.0000	0.0000
18	a741	0.4666	0.4286	0.4658	0.4658	1.0000	0.1498	1.0000	1.0000	0.2220	0.2220	0.2000	0.2000	0.3348	N/A	0.3347	0.3348
19	a927	0.4666	0.4286	0.4658	0.4658	1.0000	0.0518	1.0000	1.0000	0.2150	0.2149	0.2000	0.2000	0.0059	N/A	0.0060	0.0060
20	a9B3	0.4666	0.4286	0.4658	0.4658	0.9854	0.1842	0.9807	1.0000	0.3786	0.3786	0.4000	0.4000	0.0000	N/A	0.0000	0.0000

7.6 Discussion of Differences across the Data Sets

The density of the data when it gets below a certain level, has an impact on the results/rankings produced by the various tools, unless the calculations are done to a sufficient number of decimal places. In going from four to sixteen to 209 agents (densities varying from 0.625 to 0.0832080), all the SNA tools reviewed matched across the different centrality measures. However, differences started to emerge when the data set with 2049 agents and a density of 0.0006790 was used. When the number of decimal places for the output in UCINet and Palantir was increased from 3 to 4, it made the results closer to the ones from ORA and ANB, especially for UCINet.

7. Impact of density of data on Results

In order to attempt to establish a clearer relationship between the density of the data and the number of decimal places needed in the results, an experiment using a manual binary search was performed that involved running ORA and UCINet for different data sets. The data sets with the number of links, agents and densities are shown in the table below.

(U//FOUO) Table 20: Densities for different Data Sets

	Links	Agents	Density
Data Set A	713	1120	0.00113578
Data Set B	356	577	0.00213489
Data Set C	535	860	0.00144505
Data Set D	916	1412	0.00091822
Data Set E	624	989	0.00127463
Data Set F	668	1052	0.00120604
Data Set G	**690**	**1084**	**0.00117333**
Data Set H	702	1103	0.00115298

It was found that when the data sets had a density greater than 0.00117333, the results in ORA and UCINet matched (when UCINet was producing results to the 3rd decimal place). As a rule of thumb, four or more decimal places should be used in the SNA tools when the density of the network being analyzed falls below 0.00117333. Note that additional decimal places may be needed as the density of the network decreases further.

8.0 Conclusions

This phase looked at four different SNA tools and the results they produce for the basic Centralities across five different data sets using data sets that were symmetrized and dichotomized/binarized.

UCINet and ORA produce the same results, if the options/settings are correctly applied to treat the data as symmetrized and binarized[19].

Specifically, it was observed during the analysis that when the data sets became large and less dense, the smaller density of the network resulted in the centrality measures having smaller un-scaled values and thus smaller normalized values. This effect combined with metrics calculated to only the 3rd decimal place resulted in several agents receiving the same centrality values, making it difficult to differentiate between agents. This was the case for UCINet and Palantir for Data Set 5. When the settings for UCINet and Palantir were changed to provide results to the 4th decimal place, the numbers better matched ANB and ORA, although ANB Betweenness and Palantir Closeness and Betweenness numbers were still different enough from UCINet and ORA to produce rank changes if only Closeness or Betweenness were considered. Nevertheless, it is recommended that as a rule of thumb, four or more decimal places should be used in the SNA tools when the density of the network being analyzed falls below 0.00117333.

As can be seen from (U//FOUO) Table 13, (U//FOUO) Table 15, (U//FOUO) Table 17, (U//FOUO) Table 18, and (U//FOUO) Table 19, all four tools provide similar results with a slight bifurcation for the 5th data set where UCINet and ORA tend to be "more" similar to each other, and ANB and Palantir tend to produce "more" similar results.

Based on the above study, UCINet or ORA can be used for SNA, if the options/settings are correctly applied to treat the data as symmetrized and binarized. ANB and Palantir could be considered for SNA once the discrepancies seen for the Data Set 5 have been resolved.

As this study has only looked at dichotomized (binarized) and symmetrized data, the results do not apply to data that is not dichotomized (binarized) and symmetrized. At this time it is recommended that, if possible, any network data be dichotomized (binarized) and symmetrized before analysis.

[19] Some additional button clicks and check boxes/flags may have to be set

9.0 Next Steps

It is recommended that follow-on efforts should first focus on:

1. Learning the state of the art in other IC agencies and establishing contacts to form (or join, should one exist) a Community of Interest;

2. Deciding the extent of COIC's participation in the larger SNA community;

3. Understanding the properties of SNA measures with regard to the use of non-binarized and non-symmetric data - something academia tends to avoid due to the blooming complexity such data produces (along with the concomitant difficulty to evaluate such data) although such data is more representative of real-world networks;

4. Developing proper processes analysts can use to apply the tool, reasonably and consistently, in targeting/capture/kill decision-making when dealing with covert networks—the COIC "336" course being a humble beginning. As part of that look at reasons people interact in a network (homophily; expertise/economic; jointly assigned to a task) using the attributes and skills of the agents; the impact of missing data (structural holes in the network); the abilities of covert networks to heal etc.; and,

5. Capturing the knowledge and processes in deliverable lessons.

After the above steps have been undertaken, in the subsequent phases, additional functionality of the tools could be covered and compared. For example, the kind of support available, user friendliness, how the tools respond for networks with a large number (> 100,000 nodes) of nodes, the impact of weighted links etc? Which tool is better (helps with identification of "structural holes", "incomplete triads" etc) for networks which may be missing nodes or connections? Just as in Phase I, the data would need to cover multiple sets – smaller ones which can be checked manually (e.g. using Excel) and larger ones to stress the SNA tool. Which tool is best for analyzing multi-modal networks (ingesting the data, collapsing into different modes etc)? (Example: some analysts focus more on SIGINT networks, while others look at facilitation networks, those involving people, materials, places, events, 'groups', etc.) Different tools may be optimal for different methods of Intel analysis. Which tool is best (provides many different ways to view the data) for visualization? Ease of data input/output and import/export? These subsequent phases could also include a ranking and weighting of the criteria to allow for an objective final scoring at the end.

(U//FOUO) Table 21: Notional Evaluation Criteria and Weights for Phase II

Evaluation Criteria	Weight[20]
Correctness of Metrics	Wt1
Comprehensiveness of Metrics	Wt2
Ease of Data Input	Wt3
Ease of Data Output	Wt4
Visualization	Wt5
Ease of Use (GUI)	Wt6
Level of Training Required	Wt7
…	…

[20] All weights will add up to 1.0

10.0 References

1. Huisman, M. & A.J. van Duijn Marijtje (2005). Software for Social Network Analysis. In P.J. Carrington, J. Scott, and S. Wasserman (Eds.) Models and Methods in Social Network Analysis (pp 270 – 316). New York: Cambridge University Press, 2005.

2. LEANS LLC (2010). Advanced Network Analysis and Targeting Course: A Social Networks Approach to Targeting. Course Guide, August 2010.

3. Loscalzo, S. & Lei Yu (2008). Social Network Analysis: Tasks and Tools. In H. Liu, J. Salerno, and M.J. Young (Eds.) Social Computing, Behavioral Modeling, and Prediction, 2008.

Appendix

The following are the URLs for the different SNA tools analyzed in this study.

ANB
www.i2.com

Palantir
www.palantir.com

UCINet
www. analytictech.com/ucinet

ORA
www. casos.cs.cmu.edu/projects/ora/