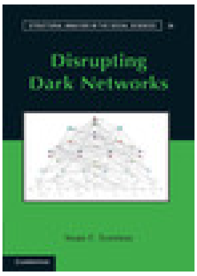


Disrupting Dark Networks

Sean F. Everton

Book DOI: http://dx.doi.org/10.1017/CBO9781139136877

Online ISBN: 9781139136877

Hardback ISBN: 9781107022591

Paperback ISBN: 9781107606685

Chapter

5 - Network Topography pp. 135-169

Chapter DOI: http://dx.doi.org/10.1017/CBO9781139136877.009

Cambridge University Press

5

Network Topography[1]

5.1 Introduction

As noted in the second chapter, although social network analysis appears to have wide appeal as a methodological tool for targeting members of dark networks (see, e.g., Gjelten 2010), it can be applied far more broadly than it has been thus far. Furthermore, strategy should drive the choice of metrics rather than the other way around. Unfortunately, just the opposite often occurs. The tail (i.e., the choice of metrics) is often found wagging the dog (i.e., the strategic choices) rather than the other way around. Indeed, the most common application of SNA to the study of dark networks has focused on targeting central actors within the network for elimination or capture (aka the "whack a mole" strategy). Although this focus is intuitively appealing and can provide short-term results, it may be misplaced and, in fact, make tracking, disrupting, and destabilizing terrorist networks more difficult. As Brafman and Beckstrom (2006) have noted, targeting key players in decentralized organizations seldom shuts them down. Instead, it only drives them to become more decentralized, making them harder to target. In terms of terrorist networks, such a strategy may in fact exacerbate what Sageman (2008) refers to as the "leaderless jihad," by which he means the numerous independent and local groups that have branded themselves with the Al Qaeda name and are attempting to emulate bin Laden and his followers by conceiving and executing terrorist operations from the bottom up.[2]

This chapter argues that analysts need to first explore a terrorist network's overall topography (i.e., its level of density, centralization, degree

[1] Portions of this chapter have been adapted from Sean Everton. 2012. "Network Topography, Key Players and Terrorist Networks." *Connections* 31(2):1–8. Used with permission.

[2] For critiques of Sageman's notion of leaderless jihad, see Hoffman (2008) and Tucker (2010).

of fragmentation, etc.) before identifying subgroups and estimating centrality, brokerage, and other types of metrics. This is not to say analysts have completely neglected the topographical dimensions of terrorist networks. There have been exceptions. Pedahzur and Perliger (2006), for example, noted that terrorist networks with a large number of cliques (see Chapter 6) appear to be more effective than those with few, and the most recent U.S. Army and Marine counterinsurgency manual (U.S. Army 2007) argues that network density is positively associated with network efficiency and, as such, should guide tactics. Perhaps the best-known example is Sageman's (2004) initial study of what he calls the global Salafi jihad (GSJ) in which he found that it exhibits the characteristics of a scale-free network. This led him to argue that the United States should focus its efforts on taking out hubs (i.e., nodes that have many connections) rather than randomly stopping terrorists at borders. "[The latter] may stop terrorists from coming here, but will leave the network undisturbed. However . . . if the hubs are destroyed, the system breaks down into isolated nodes or subgroups. The jihad will be incapable of mounting sophisticated large scale operations like the 9/11 attacks and be reduced to small attacks by singletons" (Sageman 2003). The simultaneous removal of 10 to 15 percent of a terrorist network's hubs is easier said than done, and subsequent research has found that hubs are often quickly replaced by other highly central and/or structurally equivalent actors (Pedahzur and Perliger 2006; Tsvetovat and Carley 2005), but it does not change the fact that Sageman's approach illustrates how the exploration of a network's overall topography can inform strategic decision making.

In this chapter we examine a series of metrics that capture various aspects of network topography. We begin by exploring a few basic metrics: network size, average distance, diameter, and network fragmentation. We then turn to two interrelated but analytically distinct dimensions of network topography: what I call the (1) provincial-cosmopolitan and (2) hierarchical-heterarchical (or decentralized) dimensions. With regards to the former, we consider "light network" research that suggest that networks that are too provincial (e.g., dense, high levels of clustering, an overabundance of strong ties) or too cosmopolitan (e.g., sparse, low levels of clustering, an overabundance of weak ties) tend to perform more poorly than do networks that maintain a balance between the two. With regards to the latter, we explore a series of studies that suggest that a similar dynamic is at work in terms of network hierarchy: Networks that are too hierarchical (e.g., centralized, high levels of variance) or too heterarchical (e.g., decentralized, low levels of variance) tend to underperform those that lie between the two extremes. We then consider the implications of these findings if similar dynamics hold true for dark networks. If so, then while targeting central actors may be an appropriate strategy in some instances, it may not be in others, all of which suggests that analysts need to take into account a network's overall topography before crafting strategies for its disruption. Finally, we turn to the techniques to calculating these metrics in UCINET, Pajek, and ORA.

5.2 Some Basic Topographical Metrics

In this section we consider a few basic network metrics – network size, average distance, diameter, and network fragmentation. Network size refers to the number of actors in a network. Average distance refers to the average length of all the shortest paths (i.e., geodesics) between all actors in a network and may be indicative of the speed that information spreads through a network. In other words, information should diffuse faster through networks with lower average distance than those with higher average distance. This could have implications for the success and failure of deception campaigns that seek to spread disinformation through criminal or terrorist networks. Obviously, one would suspect that such a campaign is more likely to be successful in dark networks with lower average distances; however, because networks constantly change (i.e., they are dynamic), their average distance will almost certainly vary over time, suggesting that deception campaigns may be more attractive options at particular points of time than at others. For now, we need to put this aside until we take up the topic of longitudinal network in Chapter 10.

Network diameter refers to a network's longest geodesic and could indicate how dispersed a network is. As we will see, because decentralized networks are better suited for solving nonroutine, complex, and/or rapidly changing problems or challenges because of their adaptability, dark networks are probably more likely to be decentralized than they are hierarchical. A network's diameter could possibly be used as a supplementary measure to the centralization measures we will discuss in more detail. That is, networks with large diameters may be more decentralized than those with small diameters. Because, however, network diameter is, in part, a function of network size (all else being equal, the diameters of larger networks – i.e., networks with more actors – are longer than those of smaller networks), diameter should be used carefully when comparing networks of different size. It may be more useful when examining the same network over time, but even then the network size will almost certainly vary as well.

Finally, network fragmentation, as its name implies, measures the degree to which a network is fragmented. The standard fragmentation measure is equal to the proportion of all pairs of actors that cannot either directly or indirectly reach one another. UCINET, however, calculates both the standard measure as well as a weighted one that takes into account the (path) distance between actors.[3] It also calculates measures of cohesion (or compactness), which are simply one minus the respective fragmentation scores. Network fragmentation could prove useful in the crafting of strategies. For instance, if analysts were interested in determining which scenarios would fragment a network more, one could take measure fragmentation before and after the various (hypothetical) scenarios. In fact, UCINET reports a series of scores for each actor in the network that indicates the degree of network fragmentation, the degree of distance-weighted network fragmentation, the change in network fragmentation, the change in distance-weighted network fragmentation, the percent of change in fragmentation, and the percent of change in distance-weighted fragmentation if a particular actor is removed from the network.

5.3 The Provincial-Cosmopolitan Dimension

Previously we saw how Mark Granovetter (1973, 1974) discovered that when it came to finding jobs, people were far more likely to use personal contacts than other means. Moreover, of those who found their jobs through personal contacts, most of those contacts were weak (i.e., acquaintances) rather than strong ties (i.e., close friends). This was because not only do we tend to have more weak ties than strong ties (because weak ties demand less of our time), but also because our weak ties are more likely to form the crucial bridges that tie together densely knit clusters of people (see Figure 5.1). In fact, if it were not for these weak ties, Granovetter argues, these clusters would not be connected at all. Thus, whatever is to be spread (e.g., information, influence, and other types of resources) will reach a greater number of people when it passes through weak ties rather than strong ones. Because of this, actors with few weak ties are more likely to "confined to the provincial news and views of their close friends" (Granovetter 1983:202).

Granovetter does not believe that strong ties are of no use, however. He notes that although weak ties provide individuals with access to information and resources beyond those available in their immediate social circles, strong ties have greater motivation to be sources of support in times of uncertainty. Others have noted this as well (see, e.g., Krackhardt 1992; Stark 2007). "There is a mountain of research showing that people with strong ties are happier and even healthier, because in such networks members provide one another with strong emotional and material support in times of grief or trouble and someone with whom to share life's joys and

[3] Distance-weighted fragmentation is one less the average reciprocal distance between all pairs of nodes.

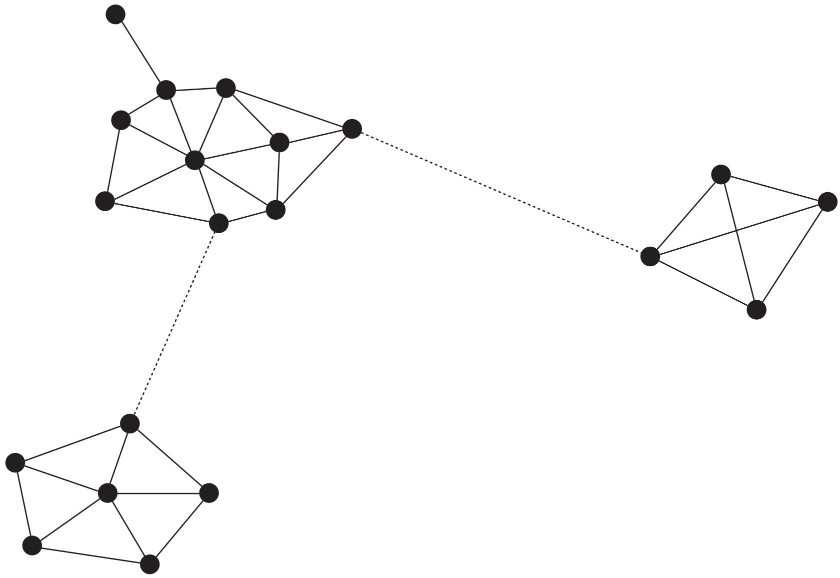


Figure 5.1. Strong and Weak Ties

triumphs" (Stark 2007:37). This suggests that people's networks differ in terms of their mix of weak and strong ties. Individuals' networks range from local or provincial ones, consisting primarily of strong, redundant ties and very few weak ties, to worldly or cosmopolitan ones, consisting of numerous weak ties and very few strong ties (Stark 2007:37–38). It also suggests that peoples' networks should ideally consist of a mix of weak and strong ties. Their networks should be neither too provincial nor too cosmopolitan but rather should land somewhere between the two extremes, not necessarily at the arithmetic mean, but rather at a "golden mean" of sorts (Aristotle 1998:36–43).

Pescosolido and Georgianna's (1989) study of suicide illustrates this dynamic. It found that the density of actors' social networks has a curvilinear (or inverted U) relationship to suicide. Individuals whose social networks are very sparse (i.e., cosmopolitan) or very dense (i.e., provincial) are far more likely to commit suicide than are those whose networks lie between the two extremes. Why? People located in sparse social networks often lack the social and emotional ties that provide them with the support they need during times of crisis. They also typically lack ties to others who might otherwise prevent them from engaging in self-destructive (i.e., deviant) behavior (Finke and Stark 2005; Granovetter 2005). On the other hand, individuals embedded in dense networks are often cut off from people outside of their immediate social group, which increases the likelihood that they will lack the ties to others who would otherwise prevent them from taking the final, fatal step.

An ideal mix of weak and strong ties appears to provide benefits not only at the individual level but also at the organizational level. In his study of the New York apparel industry, Brian Uzzi (1996) found that a mix of weak and strong ties proved beneficial to the long-term survival of apparel firms.[4] The firms he studied tended to divide their market interactions into two types: "market" or "arms-length" relationships (i.e., weak ties) and "special" or "close" relationships (i.e., strong ties), which Uzzi refers to as "embedded" ties. He found that, although market relationships were more common than embedded ties, they tended to be less important. Embedded ties were especially important in situations where fine-grained information had to be passed to the other party, and when certain types of joint problem solving were on the agenda (Uzzi 1996:677). According to Uzzi, embeddedness increases economic effectiveness along a number of dimensions crucial to competitiveness in the global economy: organizational learning, risk-sharing, and speed-to-market. He also found, however, that firms that are too embedded suffer because they lack access to information from distant parts of the social structure, rendering them vulnerable to rapidly changing situations. This led him to argue that firms should seek to maintain a balance of embedded and market ties. In support of this, he found that the topography of interfirm networks (i.e., in terms of embedded and market ties) varied and that a U-shaped association exists between the degree of embeddedness and the probability of firm failure (Uzzi 1996:675–676). Firms with high levels of embedded ties (i.e., provincial networks) or high levels of market ties (i.e., cosmopolitan networks) were much more likely to fail than were those that maintained a balance between the two.

Interestingly, Uzzi and Spiro (2005) found that an inverted U relationship also existed between the extent to which the networks of creative teams that produced Broadway musicals from 1945 to 1989 exhibited "small-worldness" (measured by what they called "small world *Q*") and the probability that a musical would be a critical and financial success. Building on the research of Stanley Milgram (Milgram 1967; Travers and Milgram 1969) and Duncan Watts and Steven Strogatz (1998), Uzzi and Spiro argue that this relationship existed because up to a point, connectivity and cohesion facilitated the flow of diverse and innovative material across the network. Moreover, connectivity and cohesion made risk-taking among the teams more likely because they were embedded in networks of trust:

> As the level of *Q* increases, separate clusters become more interlinked and linked by persons who know each other. The processes distribute creative material among teams and help to build

[4] Uzzi does not use the weak and strong tie terminology in the article.

> a cohesive social organization within teams that support risky collaboration around good ideas. (Uzzi and Spiro 2005:464)

However, as connectivity and cohesion increase, homogenization and imitation set in and returns become negative:

> Increased structural connectivity reduces some of the creative distinctiveness of clusters, which can homogenize the pool of creative material. At the same time, problems of excessive cohesion can creep in. The ideas most likely to flow can be conventional rather than fresh ideas because of the common information effect and because newcomers find it harder to land "slots" on productions. (Uzzi and Spiro 2005:464)

In other words, connectivity and cohesion initially increase a network's overall creativity by encouraging human innovation, but beyond a certain point, they can stifle it.

Although it may be difficult to conceive of criminal and terrorist networks as varying in their ability to encourage innovative thinking and creative risk-taking, these studies should give us pause. They suggest that in order to be successful, dark networks can be neither too provincial nor too cosmopolitan. What constitutes a particular dark network's optimum balance will likely vary depending on the environment in which it operates (e.g., the IRA can operate more openly in Ireland than Al Qaeda can in the United States). However, because the survival of dark networks depends largely on their recruiting members whom they can trust (Berman 2009; Tilly 2004, 2005), they tend to recruit through strong (rather than weak) ties, and networks formed primarily by strong ties become increasingly dense as ties form between previously unlinked actors (Granovetter 1973; Holland and Leinhardt 1971; Rapoport 1953a, b; Rapoport and Horvath 1961). Thus, we should expect that dark networks will be denser than light networks, all else being equal.

5.4 The Hierarchical-Heterarchical Dimension

Another well-developed body of research has explored how the degree to which an organization is hierarchically structured impacts its performance (see, e.g., Nohria and Eccles 1992; Podolny and Page 1998; Powell 1985, 1990; Powell and Smith-Doerr 1994). This literature typically identifies two ideal types of organizational form: networks and hierarchies. The former are seen as decentralized, informal, and/or organic, while the latter are seen as centralized, formal, and/or bureaucratic (see, e.g., Burns and Stalker 1961; Powell 1990; Ronfeldt and Arquilla 2001). As we have noted, however, although this distinction can be useful (see, e.g., Arquilla and Ronfeldt 2001; Castells 1996; Podolny and Page 1998; Powell and Smith-Doerr 1994; Ronfeldt and Arquilla 2001) within the world of SNA, all organizations are seen as networks, regardless of whether they are hierarchical or decentralized. Thus, it is better to think of these two ideal types as poles on either end of a continuum, running from highly decentralized forms on one end to highly centralized ones on the other.

More importantly, at least for our purposes here, research suggests that the hierarchical-heterarchical dimension impacts network performance much like the provincial-cosmopolitan one: That is, an optimal level of centralization or hierarchy appears to exist. For example, Rodney Stark (1987, 1996b), in his analysis of why some new religious movements succeed, identified centralized authority as an important factor. Nevertheless, he notes that too much centralization can be a bad thing and successful religious movements, such as the Mormon (Latter-day Saints; LDS) Church, balance centralized authority structures with decentralized ones:

> But it would be wrong to stress only the hierarchical nature of LDS authority and its authoritarian aspects, for the Latter-day Saints display an amazing degree of amateur participation at all levels of their formal structure. Moreover, this highly authoritarian body also displays extraordinary levels of participatory democracy – to a considerable extent the rank-and-file Saints are the church. A central aspect of this is that among the Latter-day Saints to be a priest is an unpaid, part-time role that all committed males are expected to fulfill. (Stark 2005:125)

Like the provincial-cosmopolitan dimension, the optimal level almost certainly varies depending on environmental context. As David Tucker (2008) notes:

> The most important issue is how well an organization's structure is adapted to its environment, which includes what its enemies are doing, given what the organization wants to achieve and the resources available to it. No one organizational structure is always inherently superior to another. Some are better for some things, some for others. These principles apply to al Qaeda as well as the governmental network (the federal, state, and local governments) in the United States. (Tucker 2008:2)

Because decentralized networks are better suited for solving nonroutine, complex, and/or rapidly changing problems or challenges because of their adaptability (Saxenian 1994, 1996), and because available evidence suggests that highly centralized networks are vulnerable to the removal of central, well-connected nodes (Albert, Jeong, and Barabási 2000; Bakker, Raab, and Milward 2011; Barabási 2002; Barabási and Bonabeau 2003), it is likely that, all else being equal, successful dark networks tend fall on the decentralized end of the continuum (Arquilla and Ronfeldt 2001; Ronfeldt and Arquilla 2001). Even here, though, dark networks that are too decentralized may find it difficult to mobilize resources, leading them to underperform, once again suggesting that analysts need to take into account this dimension of a network when crafting strategies to disrupt it.

5.5 Estimating Network Topographical Metrics

We now turn to the techniques for estimating the topographical metrics discussed previously. We will illustrate them using the four Noordin networks outlined in the previous chapter: namely, the trust, operational, communications, and business and finance networks.[5] We begin with UCINET before turning to Pajek and ORA.

Network Topography in UCINET

There are various ways to discover a network's size in UCINET. One way is by using UCINET's Matrix Browser, which is accessed with the *Data>Browse* command, which brings up a dialog box similar (but not identical) to Figure 5.2. Using the browser's *File>Open* command, open the stacked trust network file – `Trust Network (Stacked).##h` – each network visible, and on the right panel are statistics that indicate the size of the network. Here, we can see that this stacked network includes four networks (Mats), each of which includes seventy-nine actors, as indicated by the number of rows and columns. Of course, if we had opened a two-mode network in the browser, the number of rows and columns would most likely differ but would indicate the number of the two types of actors in the network.

[UCINET] Data>Browse

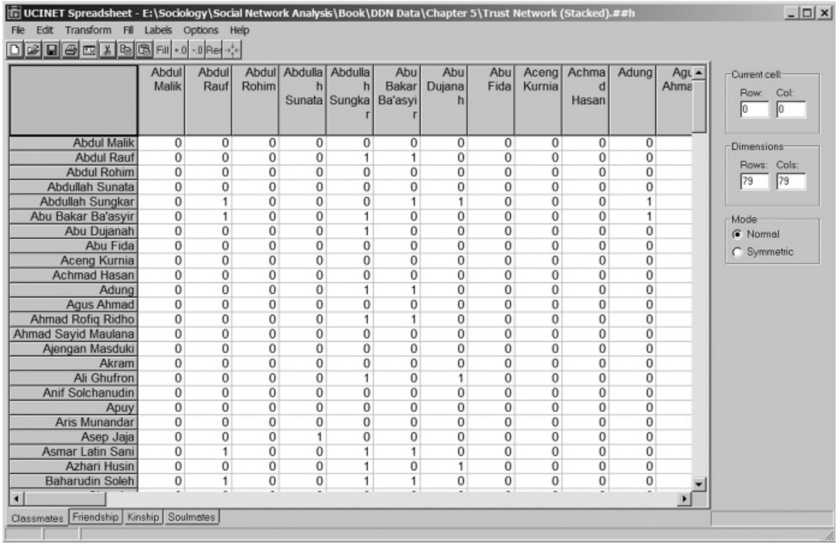
[Matrix Browser] File>Open

Average distance and diameter are calculated in UCINET using the *Network> Cohesion>Geodesic Distance (old)* command. Unfortunately, this command only works with individual rather than stacked networks, so in order to examine individual networks embedded in a stacked trust network, we need to first unpack them using UCINET's *Data>Unpack* command. Once this is done, use the *Geodesic Distance (old)* command, select the friendship network, accept UCINET's defaults, and click "OK." The output (Figure 5.3) first indicates the average (path) distance among reachable pairs (since not all actors can reach one another). This is followed by (path) distance-weighted measures of fragmentation and cohesiveness (remember, cohesion is simply 1.0 minus the fragmentation

Network >Cohesion >Geodesic Distance (old)

Data>Unpack

[5] Here, we will use the one-mode individual network derived from the business and finance affiliation network.

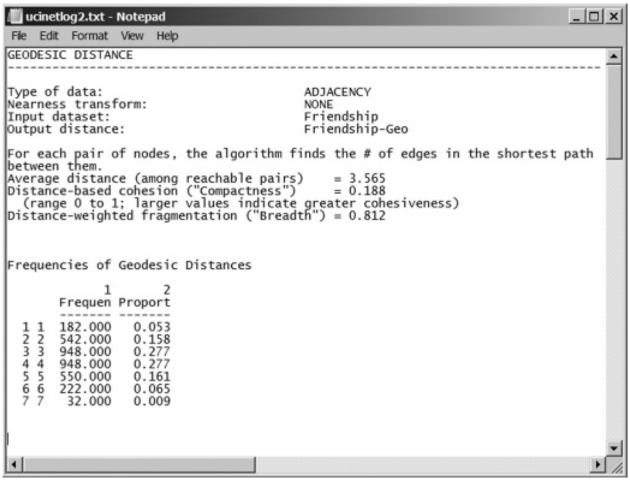


UCINET Spreadsheet - E:\Sociology\Social Network Analysis\Book\DDN Data\Chapter 5\Trust Network (Stacked).##h

	Abdul Malik	Abdul Rauf	Abdul Rohim	Abdullah Sunata	Abdullah Sungkar	Abu Bakar Ba'asyir	Abu Dujanah	Abu Fida	Aceng Kurnia	Achmad Hasan	Adung	Agu Ahma
Abdul Malik	0	0	0	0	0	0	0	0	0	0	0	
Abdul Rauf	0	0	0	0	1	1	0	0	0	0	0	
Abdul Rohim	0	0	0	0	0	0	0	0	0	0	0	
Abdullah Sunata	0	0	0	0	0	0	0	0	0	0	0	
Abdullah Sungkar	0	1	0	0	0	1	1	0	0	0	1	
Abu Bakar Ba'asyir	0	1	0	0	1	0	0	0	0	0	1	
Abu Dujanah	0	0	0	0	1	0	0	0	0	0	0	
Abu Fida	0	0	0	0	0	0	0	0	0	0	0	
Aceng Kurnia	0	0	0	0	0	0	0	0	0	0	0	
Achmad Hasan	0	0	0	0	0	0	0	0	0	0	0	
Adung	0	0	0	0	1	1	0	0	0	0	0	
Agus Ahmad	0	0	0	0	0	0	0	0	0	0	0	
Ahmad Rofiq Ridho	0	0	0	0	1	1	0	0	0	0	0	
Ahmad Sayid Maulana	0	0	0	0	0	0	0	0	0	0	0	
Ajengan Masduki	0	0	0	0	0	0	0	0	0	0	0	
Akram	0	0	0	0	0	0	0	0	0	0	0	
Ali Ghufron	0	0	0	0	1	0	1	0	0	0	0	
Anif Solchanudin	0	0	0	0	0	0	0	0	0	0	0	
Apuy	0	0	0	0	0	0	0	0	0	0	0	
Aris Munandar	0	0	0	0	0	0	0	0	0	0	0	
Asep Jaja	0	0	0	1	0	0	0	0	0	0	0	
Asmar Latin Sani	0	1	0	0	1	1	0	0	0	0	0	
Azhari Husin	0	0	0	0	1	0	1	0	0	0	0	
Baharudin Soleh	0	1	0	0	1	1	0	0	0	0	0	

Figure 5.2. UCINET's Matrix Browser

score). These measures are, in turn, followed by the number of geodesics at various distance levels (from shorter to longer). The largest level equals the network's diameter. Below these metrics (not shown in Figure 5.3) is geodesic distance matrix, which indicates the path distance between all pairs of actors. If a geodesic matrix is all you are interested in obtaining, you can use the *Network>Cohesion>Geodesic Distance* command,



```
GEODESIC DISTANCE
--------------------------------------------------------------------------------

Type of data:                          ADJACENCY
Nearness transform:                    NONE
Input dataset:                         Friendship
Output distance:                       Friendship-Geo

For each pair of nodes, the algorithm finds the # of edges in the shortest path
between them.
Average distance (among reachable pairs)    = 3.565
Distance-based cohesion ("Compactness")     = 0.188
  (range 0 to 1; larger values indicate greater cohesiveness)
Distance-weighted fragmentation ("Breadth") = 0.812


Frequencies of Geodesic Distances

            1       2
       Frequen Proport
       ------- -------
  1 1  182.000   0.053
  2 2  542.000   0.158
  3 3  948.000   0.277
  4 4  948.000   0.277
  5 5  550.000   0.161
  6 6  222.000   0.065
  7 7   32.000   0.009
```

Figure 5.3. UCINET's Old Geodesic Distance Output

Table 5.1. *Diameter, average distance, and fragmentation of Noordin's network*

Relation	Diameter	Average distance	Fragmentation	Distance-weighted fragmentation
Trust Network				
Classmates	7.000	2.213	0.784	0.877
Friendship	7.000	3.565	0.444	0.812
Kinship	2.000	1.158	0.994	0.994
Soulmates	2.000	1.313	0.995	0.996
Aggregated	7.000	2.834	0.216	0.664
Operational Network				
Logistics	4.000	1.913	0.970	0.981
Meetings	4.000	2.068	0.895	0.941
Operations	3.000	1.667	0.759	0.837
Training	4.000	1.870	0.863	0.912
Aggregated	5.000	2.109	0.261	0.595
Communications	6.000	2.703	0.123	0.623
Business & Finance	2.000	1.167	0.994	0.995
Combined Network	4.000	1.990	0.025	0.443

which can handle stacked networks but does not provide information on a network's average distance, fragmentation, or cohesion.

Network >Centrality and Power >Fragmentation

You can also estimate fragmentation scores using UCINET's *Network>Centrality and Power>Fragmentation* command. This command also reports an unweighted score, which is the same score reported by ORA. As noted previously, a helpful feature of this command is that it reports a series of scores (not shown) for each actor that indicates whether that actor is removed from the network: (1) what the network fragmentation will be ("Frag"); (2) what the distance-weighted network fragmentation will be ("DwFrag"); (3) what the change in fragmentation will be ("FragDiff"); (4) what the change in distance-weighted fragmentation will be ("DwFragDiff"); (5) what the percent change in fragmentation will be ("PctFragChg"); and (6) what the percent change in distance-weighted fragmentation will be ("PctFragChg").

Table 5.1 summarizes the diameter, average distance, fragmentation, and distance-weighted fragmentation for each of the four Noordin networks, as well as the aggregated trust, operational, and combined[6] networks. The size of each network is not listed because it is the same across all seventy-nine networks and, for our purposes here, uninteresting. Note the considerable variation across the networks. Their diameter ranges from as low as 2.0 to 7.0, and average distance ranges from 1.158 to

[6] The combined network is the aggregation of the trust, operational, communications, and business and finance networks.

3.565. It is helpful to recall that the latter score is the average distance between actors that are connected to one another (either directly or indirectly), so it is not unusual in highly fragmented networks (i.e., where there is a high proportion of pairs of actors that cannot directly or indirectly reach one another) for the average path distance to be quite small. The kinship and soulmates networks are examples of this. Both are highly fragmented, and the average distance (and diameter) of both are among the lowest of the networks. Note also how the distance-weighted fragmentation measure differs from the traditional fragmentation measure. When average path distance is not taken into account, the trust network is less fragmented than the operational network. However, when the average distance between connected actors is taken into account, the operational network is less fragmented. The trust network also appears to be more spread out than the operational network because both its diameter and average distance are larger than those of the operational network. Of course, although these and other comparisons can be interesting and provide insights into a network, it is by looking at them longitudinally that we can gain a sense of how they are changing over time and reacting to exogenous and endogenous factors. We will postpone taking up this task until Chapter 10.

Provincial-Cosmopolitan Dimension, Part I: Network Density and Average Degree. Two measures that help analysts tap into the provincial-cosmopolitan dimension of networks are network density and average degree. Network density (d) is formally defined as the total number of ties divided by the total possible number of ties:

$$d = \frac{L}{\frac{n(n-1)}{2}} \tag{5.1}$$

where L refers to the actual number of ties (or lines) in a network and n to the number of actors (or nodes) in the network. Because each actor can potentially be connected to all other actors in the network (except to himself), in an undirected network (i.e., where ties are reciprocal), the total possible number of ties equals $\frac{n(n-1)}{2}$. Calculating density with a directed graph is similar except that with a directed graph you do not have to divide the denominator by two:

$$d = \frac{L}{n(n-1)} \tag{5.2}$$

In practical terms what both of these equations mean is that network density scores range from 0.0 to 1.0. In networks with a density of 0.0 (or 0.0 percent), no ties exist between actors, whereas in networks with a density of 1.0 (or 100.0 percent), all possible ties exist between actors.

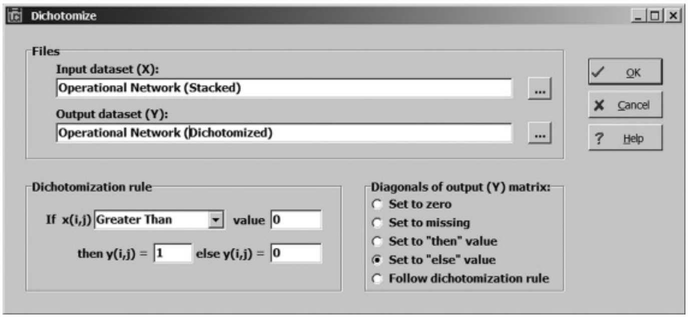


Figure 5.4. UCINET's Dichotomize Dialog Box

Unfortunately, the formal measure of density is inversely related to network size (i.e., all else being equal the larger the network, the lower the density) because the number of possible lines increases exponentially as actors are added to the network, while the number of ties that each actor can maintain tends to be limited. That is why social network analysts often turn to average degree, which is simply the average number of ties that each actor in the network has, in order to measure how "dense" a network is (de Nooy et al. 2005; Scott 2000).[7]

[UCINET] Network >Cohesion >Density (new)Density Overall

In UCINET you calculate network density using the *(new)Density Overall* command found under the *Network>Cohesion>Density* submenu. Before doing this, however, first dichotomize (binarize) the stacked operational network because some of the networks included therein are valued networks (e.g., the tie between some pairs of individuals are greater than one because they participated in more than one operation together, or attended two or more training events together), and typically you do not want to take into account tie values when calculating density. To dichotomize networks in UCINET, we use the *Transform>Dichotomize* command, which calls up a dialog box (Figure 5.4), which asks you to indicate your input dataset. The example uses the `Operational Network (Stacked)` file. UCINET provides a few "dichotomization rule" options that allow you to indicate what you want your cutoff value for dichotomizing the network to be. Here, we have chosen to accept UCINET's default because we want every tie that is currently greater than 0 to equal 1. UCINET also provides a default output file name, but here we have chosen our own: `Operational Network (Dichotomized)`.

Transform >Dichotomize

UCINET's overall density command calls up a dialog box (see Figure 5.5) where you indicate your input network (`Operational Network (Dichotomized)`). Click "OK."

The resulting output log (Figure 5.6) lists both the density and average degree for each network in the stacked network. As you can see, the

7 See Chapter 7 for the equation used for calculating degree centrality.

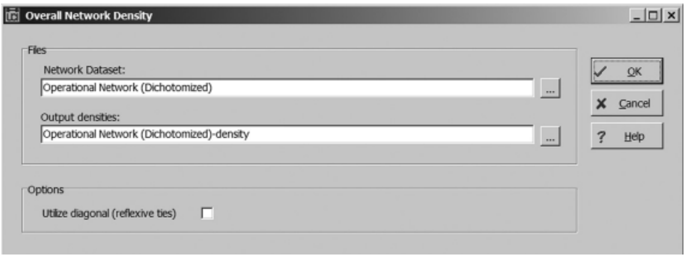


Figure 5.5. UCINET's Density Dialog Box

operations network is the densest (and has the highest average degree) of the four networks.

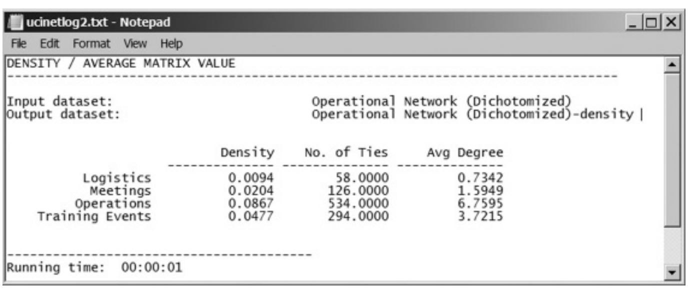
Transform >Matrix Operations >Within datasets >Aggregations

Transform >Dichotomize

Network >Cohesion >Density >(new) Density Overall

We may also be interested in aggregating the stacked trust (`Trust Network (Stacked.##h)`) and operational (`Operational Network (Stacked.##h)`) networks and calculating their density and average degree. To do this we use UCINET's *Transform>Matrix Operations>Within datasets>Aggregations* command, which we discussed at length in Chapter 4. After aggregating the networks, we then need to dichotomize the network (because some cells have values of greater than 1) using UCINET's *Transform>Dichotomize* command. After the aggregated networks are dichotomized, then we calculate the density of the overall network just as we did before, using the *Network>Cohesion>Density>(new) Density Overall* command.

Provincial-Cosmopolitan Dimension, Part II: Clustering Coefficient and Small World Q. Two other measures that are sometimes used to capture the provincial-cosmopolitan dimension are the clustering coefficient and the small world quotient (*SWQ*). The former is estimated directly, although as we will see, there are different methods for doing so. The latter is typically calculated by first estimating both the clustering coefficient (*CC*) and average path distance (*AP*). These are then normalized by calculating the ratio of each to the respective clustering coefficient



```
ucinetlog2.txt - Notepad
File  Edit  Format  View  Help
DENSITY / AVERAGE MATRIX VALUE
--------------------------------------------------------------------------------

Input dataset:                        Operational Network (Dichotomized)
Output dataset:                       Operational Network (Dichotomized)-density


                          Density     No. of Ties     Avg Degree
                       ------------ -------------- -------------
         Logistics        0.0094       58.0000        0.7342
          Meetings        0.0204      126.0000        1.5949
        Operations        0.0867      534.0000        6.7595
   Training Events        0.0477      294.0000        3.7215

----------------------------------------
Running time:  00:00:01
```

Figure 5.6. UCINET's Density and Average Degree Output

and average path distance of a random network of the same size and density:

$$CC_{Ratio} = \frac{CC_{Actual}}{CC_{Random}} \tag{5.3}$$

$$AP_{Ratio} = \frac{AP_{Actual}}{AP_{Random}} \tag{5.4}$$

The small world quotient is simply the ratio of the CC_{Ratio} to the AP_{Ratio}:

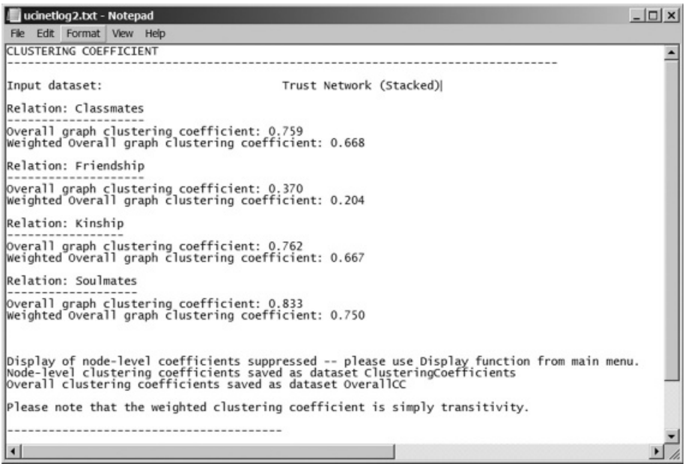
$$SWQ = \frac{CC_{Ratio}}{AP_{Ratio}} \tag{5.5}$$

Because the *CC* and density of a random network are approximately equal to one another and the latter is, by definition, equal to the density of the original small world network, dividing the original network's *CC* by its density is typically sufficient for calculating the CC_{Ratio}. Moreover, in a later analysis of the Broadway musical data, Uzzi (2008) found that the AP_{Ratio} almost always equaled one, and a subsequent analysis of simulated and real-world networks discovered that Uzzi's discovery applied to a larger range of networks than Broadway musicals (Everton 2012; Everton and Lieberman 2009). What this means is that a network's small world quotient can be estimated by simply dividing the original network's *CC* by its density.

The clustering coefficient is estimated by first taking the ego network of each actor (i.e., each actor's ties to other actors – aka an actor's "alters" – and the ties between them), then calculating the density of each ego network (but not including ego or ego's ties in the calculation – i.e., only the ties between ego's alters are used) and taking the average of these scores. Luckily, UCINET (and Pajek and ORA) has automated all of this for us. To obtain the clustering coefficient, select UCINET's *Network>Cohesion>Clustering Coefficient* command. In the resulting dialog box (not shown), select the dichotomized networks you wish to analyze as your input dataset, and click "OK." The resulting output (Figure 5.7 – here we have estimated the clustering coefficients for the trust network (`Trust Network (Stacked.##h)`) provides the clustering coefficient scores for each network in a file. And as we just noted, to estimate a network's small world quotient, we simply divide the clustering coefficient by network density. Note that UCINET provides two scores: an overall graph clustering coefficient and a weighted overall graph clustering coefficient. It is the former metric that interests us here.

Network >Cohesion >Clustering Coefficient

Table 5.2 summarizes the density, average degree, clustering coefficient, and small world quotient for each of the Noordin networks. Notice the variation in scores across the various networks. In terms of density and



```
ucinetlog2.txt - Notepad
File Edit Format View Help
CLUSTERING COEFFICIENT
--------------------------------------------------------------------------------

Input dataset:                          Trust Network (Stacked)

Relation: Classmates
--------------------
Overall graph clustering coefficient: 0.759
Weighted Overall graph clustering coefficient: 0.668

Relation: Friendship
--------------------
Overall graph clustering coefficient: 0.370
Weighted Overall graph clustering coefficient: 0.204

Relation: Kinship
-----------------
Overall graph clustering coefficient: 0.762
Weighted Overall graph clustering coefficient: 0.667

Relation: Soulmates
-------------------
Overall graph clustering coefficient: 0.833
Weighted Overall graph clustering coefficient: 0.750


Display of node-level coefficients suppressed -- please use Display function from main menu.
Node-level clustering coefficients saved as dataset ClusteringCoefficients
Overall clustering coefficients saved as dataset OverallCC

Please note that the weighted clustering coefficient is simply transitivity.

----------------------------------------
```

Figure 5.7. UCINET's Clustering Coefficient Output

average degree, the operational network is the densest of the four networks and the business and finance network is the sparsest.[8] Within the trust and operational networks, certain types of relations appear to play a more central role. For example, within the trust network, the classmate network is by far the densest, with the friendship network coming a somewhat distant second; by contrast, the kinship and religious networks are relatively sparse and contribute few ties at all to the network as a whole. This is somewhat in contrast to Sageman's (2004) finding that the global Salafi jihad has primarily been formed through friendship and kinship ties. Moreover, the prominence of classmate ties suggests that analysts may want to consider exploring potential vulnerabilities in Noordin's school network (where a tie between two schools indicates that at least one individual is affiliated with both). For example, they might want to develop a kinetic strategy that shuts down schools from which Noordin has recruited a number of followers; in fact, that is what the Indonesian authorities did with the Luqmanul Hakeim School. Or, they might consider a nonkinetic strategy of building alternative schools nearby, schools that promote moderate forms of Islam and instruct students in subjects other than the memorization of the Qur'an (Roberts and Everton 2011).

The clustering coefficient and small world quotient scores highlight one of the peculiarities of how the overall clustering coefficient is calculated. As one might guess, it equals the mean of the sum of the actors' clustering coefficient scores. However, UCINET, Pajek, and ORA only calculate the clustering coefficients for actors who have two or more alters, and when calculating the average, they divide the sum by the number of actors

[8] Because the networks are the same size, comparing their density levels is permissible.

Table 5.2. *Density, average degree, clustering coefficient, and small world Q*

Relation	Density	Average degree	Clustering coefficient WS/alternate	Small world quotient WS/alternate
Trust Network				
Classmates	0.057	4.430	0.759/0.308	13.32/5.40
Friendship	0.030	2.304	0.370/0.173	12.33/5.77
Kinship	0.005	0.405	0.762/0.068	152.40/13.60
Soulmates	0.004	0.278	0.833/0.074	208.25/18.50
Aggregated	0.084	6.557	0.528/0.381	6.29/4.54
Operational Network				
Logistics	0.009	0.734	0.749/0.104	83.22/11.56
Meetings	0.020	1.595	0.768/0.223	38.40/11.15
Operations	0.087	6.760	0.923/0.444	10.61/5.10
Training	0.048	3.722	0.833/0.401	17.35/8.35
Aggregated	0.142	11.063	0.791/0.661	5.57/4.66
Communications	0.065	5.063	0.562/0.441	8.65/6.79
Business & Finance	0.005	0.380	0.938/0.095	187.60/19.00
Combined Network	0.202	15.772	0.692/0.648	3.43/3.21

who have clustering coefficient scores (not by the number of actors in the network). This is the well-known Watts-Strogatz (WS) clustering coefficient (Watts and Strogatz 1998), but in sparse networks with a few well-connected actors, it can mislead one into thinking that a network is more clustered than it actually is. Interestingly, until recently, Pajek calculated an alternative clustering coefficient that divided the sum of the clustering coefficient scores by the number of actors in the network. Both scores are reported in Table 5.2. The Watts-Strogatz scores appear to the left of the slash mark; the alternative scores appear to the right. When comparing these scores to the density and average degree scores, the alternative scores appear more reasonable. Note, for example, that in terms of network density and average degree, the classmate and friendship networks are more provincial than the kinship and soulmate networks, but in terms of the WS clustering coefficient, the latter appear more provincial. That is not the case with the alternative scores, however, suggesting they more accurately reflect the overall clustering of the network. Note, however, that when working with a well-connected network (e.g., the combined network), the difference between the two scores is minimal.

The small world quotient scores calculated using the alternative clustering coefficient scores also appear to be more reasonable. There is less variation because the alternative clustering coefficients scores do not overstate a sparse network's clustering level. Still, the small world quotients for the kinship, soulmate, and business and finance networks are larger than the scores of the other networks (unlike their respective density, average degree, and clustering coefficient scores), which at first glance may seem incorrect. However, because the small world quotient is the ratio of the clustering coefficient to density, networks that are extremely sparse but have small and highly dense clusters should score quite high. They are also the type of network that will probably find it difficult to mobilize effectively, and as we saw, that is exactly what Uzzi and Spiro (2005) discovered. Networks that score too high (and too low) in terms of the small world quotient are less likely to succeed than are those whose scores land somewhere in between.

Hierarchical-Heterarchical Dimension: Centralization and Variance. Network centralization, variance, and standard deviation are measures that help researchers determine how hierarchical (or nonhierarchical) a network is. Centralization uses the variation in actor centrality within the network to measure the level of centralization. More variation yields higher network centralization scores, while less variation yields lower scores. Formally, centralization equals

$$C = \frac{\sum [Cmax - C(n_i)]}{\max \sum [Cmax - C(n_i)]} \tag{5.6}$$

where *Cmax* equals the largest centrality score for all actors and $C(n_i)$ is the centrality score for actor n_i, and $\max \sum [Cmax - C(n_i)]$ is the theoretical maximum possible sum of differences in actor centrality. In other words, network centralization is the ratio of the actual sum of differences in actor centrality over the theoretical maximum, yielding (like density) a score somewhere between 0.0 and 1.0. In general, the larger a centralization index is, the more likely it is that a single actor is very central, whereas the other actors are not (Wasserman and Faust 1994:176); thus, they can be seen as measuring how unequal the distribution of individual actor values are. Finally, because network centralization scores are based on the type of centrality estimated (e.g., degree, betweenness, closeness, and eigenvector), we need to interpret them in light of the centrality metric estimated. For example, degree centrality counts the number of ties each individual actor has; thus, we would expect that a centralization metric based on it would capture the extent to which one or a handful of actors possess a lot of ties, whereas other actors in the network do not. By contrast, a centralization measure based on betweenness centrality, which measures the extent to which actors lie between other actors in the network, could be interpreted as indicating the degree to which a handful of a network's actors are in a position of brokerage. Put differently, the higher the score, the more likely it is that only a few actors score high in terms of betweenness centrality.

An alternative measure to centralization that has been recommended by Coleman (1964), Hoivik and Gleditsch (1975), and Snijders (1981) is the variance (*V*) of actor centrality scores found in a network (see also Wasserman and Faust 1994:177, 180–181), although the standard deviation (*SD*) is preferable, as it brings us back to the original unit of measure (Hamilton 1996:72–73).

$$V = \frac{\left[\sum_{i=1}^{n} (C(n_i) - \overline{C})^z\right]}{n} \tag{5.7}$$

$$SD = \sqrt{\frac{\left[\sum_{i=1}^{n} (C(n_i) - \overline{C})^z\right]}{n}} \tag{5.8}$$

Those familiar with standard statistics will recognize these equations as standard measures of variance and standard deviation. The variance (5.7) equals the sum of the squared differences between each actor's centrality score ($C(n_i)$) and the average centrality score ($\overline{C}$), whereas the standard deviation (see equation 5.8) is simply the square root of the variance.[9] Comparing equations 5.6 and 5.7, one can see that they differ in that, whereas centralization (5.6) uses the difference between the network's largest centrality score and each actor's centrality score to estimate variance in actor centrality, the traditional variance (5.7) and standard deviation (5.8) measures use the difference between the network's mean centrality score and each actor's centrality score. Thus, although both attempt to capture the level of variance in centrality scores, they look to different baseline measures (i.e., largest centrality score vs. average centrality score) for doing this.[10]

To calculate these measures in UCINET, we use its various centrality commands. In other words, to calculate degree centralization, we use the *Network>Centrality and Power>Degree* command; for betweenness centralization, *Network>Centrality and Power>Freeman Betweenness*; for closeness, *Network> Centrality and Power>Closeness*, and for eigenvector, *Network>Centrality and Power>Eigenvector*. Within each of these commands, there are a few options that should be noted. For instance, when you issue the degree centrality command, the dialog box (Figure 5.8) that is called up asks whether the network data are symmetric.

Network >Centrality and Power >Degree Freeman Betweenness, Closeness, Eigenvector

[9] Because the sum of the differences always equals zero, we first square the differences. Squaring the differences eliminates negative values and gives us a measure that can be used to measure variation. However, because the variance is measured in unnatural squared units (e.g., centrality^2), taking the square root of the variance (i.e., the standard deviation) returns us to a more understandable unit of measure (e.g., centrality).

[10] One last set of measures worth noting are Krackhardt's (1994) graph theoretical measures of hierarchy. These can be quite informative, but they are intended for directed data, and to date, most SNA data collected on dark networks have been undirected, which is why we do not consider Krackhardt's measures at any length in this book.

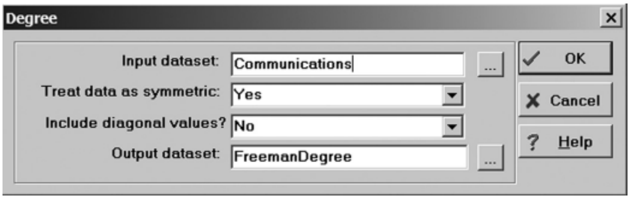


Figure 5.8. UCINET's Degree Centrality Dialog Box

Although UCINET's default is "Yes," it sometimes defaults to "No," so if you are analyzing symmetric data, you will want to make sure that this option is selected correctly (in Chapter 7 we will see how we use this option to analyze asymmetric data).

Network >Centrality and Power >Freeman Betweenness >Node Betweenness

In terms of betweenness centrality, we generally want to select the *Node Betweenness* menu option. This is the most common betweenness centrality measure and the one we will use here. At the resulting dialog box (not shown), accept UCINET's defaults and click "OK." With eigenvector centrality, you will typically accept UCINET's default options *except* make sure you check the box that "forces the majority of scores to be positive" (Figure 5.9).

As we have noted, closeness centrality measures how close, in terms of path distance, each actor is, on average, to every other actor in the network. Unfortunately, the standard closeness measure, which we will discuss in more detail in Chapter 7, cannot be calculated when some actors or clusters of actors are disconnected from others because the path distance between some actors is infinite.

One way to work around this problem is to extract the network's largest component (see Chapter 6), calculate closeness for the actors that are a part of it, and assign scores of "0.00" to those that that are not. An alternative approach, the one that is used here (but currently can only be calculated in UCINET), is to sum (and average) the reciprocal distance between all actors; this approach works with disconnected networks because the reciprocal of infinity is typically regarded as equaling

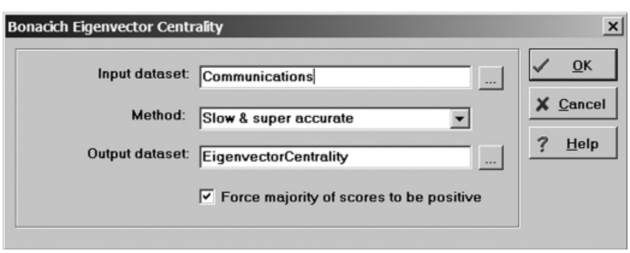


Figure 5.9. UCINET's Eigenvector Centrality Dialog Box

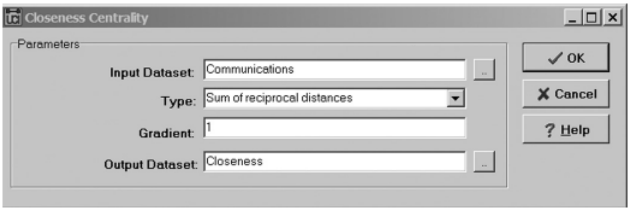


Figure 5.10. UCINET's Closeness Centrality Dialog Box

zero (Borgatti 2006).[11] To use this measure of closeness, after issuing the closeness centrality command, select the "Sum of reciprocal distances" option in the resulting dialog box (Figure 5.10) and click "OK."

Figure 5.11 presents a portion of the output from using UCINET's degree centrality command with the communication network (`Communications.##h`). At the bottom it displays the descriptive statistics for the network; just above these are some of the individual actor centrality scores. The descriptive statistics are what interest us here. We will consider actor centrality scores in Chapter 7. As you can see from Figure 5.11, the descriptive statistics section of the output provides users with a lot of information, more than most analysts will need. Moreover, most of the metrics are presented in terms of their raw, normalized, and share scores.[12] Two metrics that we considered previously, network size (N of Obs) and average (mean) degree, are also included in this report.[13] More pertinent to the current discussion is that this report includes the variance, standard deviation, and centralization scores.[14] A number of other measures not discussed in this book are also included in the report; readers who are interested in learning more about these should see UCINET's Help function, as well as consult Hanneman and Riddle (2005).

Figure 5.12 presents the output generated by UCINET's betweenness centrality command. As you can see, although the output is not identical to the degree centrality output, it is similar. This is true of the output generated by the other centrality commands as well. Looking at both figures, you will note that the network centralization score appears just below the descriptive statistics, whereas the average degree, variance, and standard deviation scores are "buried" in the descriptive statistics themselves. When you analyze dichotomized data, centralization scores

[11] It is this measure of closeness that is used for calculating distance-weighted fragmentation.

[12] Normalized scores have been adjusted for network size, whereas individual share scores have been adjusted so that they sum to 1.00.

[13] Close observers may have noted that normalized average degree equals network density.

[14] Note, however, that the centralization score appears separately at the bottom of the output. This is because it is the same whether the raw, normalized, or share scores are used to calculate it.

```
ucinetlog1.txt - Notepad
File  Edit  Format  View  Help
 79              Zulkarnaen      2.000      2.564      0.005
 49                  Marwan      2.000      2.564      0.005
 62             Rosihin Noor     2.000      2.564      0.005
 42           Iqbal Huseini      2.000      2.564      0.005
 63             Said Sungkar     2.000      2.564      0.005
 56            Mus'ab Sahidi     1.000      1.282      0.002
 40             Imam Samudra     1.000      1.282      0.002
 35                    Harun     1.000      1.282      0.002
 57      Muzayin Abdul Wahab     1.000      1.282      0.002
  9             Aceng Kurnia     1.000      1.282      0.002
  1              Abdul Malik     1.000      1.282      0.002
 30           Engkos Kosasih     1.000      1.282      0.002
 51            Mohamed Ihsan     1.000      1.282      0.002
 15           Ajengan Masduki    1.000      1.282      0.002
 14       Ahmad Sayid Maulana    1.000      1.282      0.002
 54                  Muchtar     1.000      1.282      0.002
 76               Umar Wayan     1.000      1.282      0.002
 36             Hence Malewa     0.000      0.000      0.000
 65                   Salman     0.000      0.000      0.000
 58               Nasir Abas     0.000      0.000      0.000
  2              Abdul Rauf      0.000      0.000      0.000
 41                    Iqbal     0.000      0.000      0.000

DESCRIPTIVE STATISTICS

                           1            2            3
                      Degree    NrmDegree        Share
                ------------ ------------ ------------
  1      Mean          5.063        6.491        0.013
  2   Std Dev          5.738        7.356        0.014
  3       Sum        400.000      512.820        1.000
  4  Variance         32.920       54.109        0.000
  5       SSQ       4626.000     7603.550        0.029
  6     MCSSQ       2600.684     4274.628        0.016
  7  Euc Norm         68.015       87.198        0.170
  8   Minimum          0.000        0.000        0.000
  9   Maximum         41.000       52.564        0.102
 10  N of Obs         79.000       79.000       79.000

Network Centralization = 47.27%
Blau Heterogeneity = 2.89%.  Normalized (IQV) = 1.65%

Actor-by-centrality matrix saved as dataset FreemanDegree

----------------------------------------
Running time:  00:00:01
```

Figure 5.11. UCINET's Degree Centrality Output

```
ucinetlog1.txt - Notepad
File  Edit  Format  View  Help
 21                Asep Jaja       0.000       0.000
 42            Iqbal Huseini       0.000       0.000
 62             Rosihin Noor       0.000       0.000
 35                    Harun       0.000       0.000
 26                  Cholily       0.000       0.000
 65                   Salman       0.000       0.000
 66                  Saptono       0.000       0.000
 67        Sardona Siliwangi       0.000       0.000
 49                   Marwan       0.000       0.000
 69           Subur Sugiarto       0.000       0.000
 51            Mohamed Ihsan       0.000       0.000
 71                    Tohir       0.000       0.000
 14      Ahmad Sayid Maulana       0.000       0.000
 54                  Muchtar       0.000       0.000
 74                     Umar       0.000       0.000
 36             Hence Malewa       0.000       0.000
 76               Umar Wayan       0.000       0.000
 58               Nasir Abas       0.000       0.000
  1              Abdul Malik       0.000       0.000
 40             Imam Samudra       0.000       0.000

DESCRIPTIVE STATISTICS FOR EACH MEASURE

                           1            2
                 Betweenness nBetweenness
                ------------ ------------
  1      Mean         58.241        1.939
  2   Std Dev        173.771        5.787
  3       Sum       4601.000      153.213
  4  Variance      30196.260       33.484
  5       SSQ    2653469.000     2942.411
  6     MCSSQ    2385504.500     2645.267
  7  Euc Norm       1628.947       54.244
  8   Minimum          0.000        0.000
  9   Maximum       1452.547       48.370
 10  N of Obs         79.000       79.000

Network Centralization Index = 47.03%

Output actor-by-centrality measure matrix saved as dataset FreemanBetweenness

----------------------------------------
Running time:  00:00:01
```

Figure 5.12. UCINET's Betweenness Centrality Output

Table 5.3. *Centralization scores*

Relation	Degree centralization	Closeness centralization	Betweenness centralization	Eigenvector centralization
Trust Network				
Classmates	29.69%	55.75%	5.77%	43.33%
Friendship	18.02%	46.52%	24.84%	75.25%
Kinship	3.41%	6.68%	0.07%	94.83%
Soulmates	4.90%	9.56%	0.10%	4.37%
Aggregated	28.21%	48.96%	14.73%	39.68%
Operational Network				
Logistics	12.19%	25.83%	1.18%	6.50%
Meetings	24.21%	46.88%	5.58%	72.82%
Operations	39.78%	64.89%	9.24%	34.39%
Training	17.47%	39.63%	3.66%	42.87%
Aggregated	42.01%	58.10%	15.67%	31.60%
Communications	47.27%	69.96%	47.03%	65.21%
Business & Finance	4.76%	9.36%	0.10%	80.19%
Combined Network	51.60%	58.96%	20.25%	29.43%

will range from 0 to 1 (or 0–100 percent), but if you analyze valued data, they will sometimes be larger than 1 (or 100 percent). As a general rule, you will want to dichotomize your data before estimating network centralization (the closeness and betweenness commands automatically do this if valued data are used).

Table 5.3 presents the centralization scores for each of the various networks, whereas Table 5.4 presents the corresponding standard deviation scores. The scores are somewhat consistent across types of centrality and between the two measures of centralization (i.e., centralization and standard deviation). Networks that are highly centralized in terms of degree centrality are often highly centralized in terms of closeness, betweenness, and eigenvector centrality. Likewise, networks that are identified as more centralized in terms of centralization tend to be similarly identified in terms of the standard deviation. There are, of course, a few exceptions. For instance, in terms of betweenness centrality, the communication network is far more centralized relative to the other networks than it is in terms of the other centrality measures. And, although in terms of degree, closeness, and betweenness centralization the kinship and business and finance networks are not in the least bit centralized, in terms of eigenvector centrality, they are. Or again, according to the degree and closeness centralization metrics, the communication network is more centralized than the operational network, but according to the corresponding standard deviation scores, the opposite is true. These are not the only exceptions, of course, but they illustrate that different metrics based on different

Table 5.4. *Standard deviation*

Relation	Degree standard deviation	Closeness standard deviation	Betweenness standard deviation	Eigenvector standard deviation
Trust Network				
Classmates	6.358	10.688	26.476	0.093
Friendship	2.587	9.203	121.902	0.092
Kinship	0.685	0.750	0.249	0.110
Soulmates	0.841	0.976	0.401	0.109
Aggregated	6.612	11.458	103.574	0.088
Operational Network				
Logistics	1.874	3.144	5.352	0.105
Meetings	3.188	6.696	20.415	0.098
Operations	8.315	13.084	33.945	0.092
Training	4.818	8.072	15.090	0.097
Aggregated	8.728	13.881	84.815	0.080
Communications	5.738	9.963	173.771	0.079
Business & Finance	0.959	1.039	0.335	0.109
Combined Network	10.260	8.581	84.884	0.064

centrality measures will sometimes provide different answers as to how centralized a particular network is.

Caution should be in order when interpreting centralization measures (both centralization and standard deviation). Networks that differ substantially in terms of centralization may look quite similar. Take, for instance, Noordin's trust and communication networks (Figure 5.13). They are almost mirror images of one another, but as we can see in Table 5.3, the communications network is far more centralized than the trust network, whereas according to Table 5.4, the trust network is more centralized. This illustrates the importance of using network metrics and visualization in conjunction with one another.

Network Topography in Pajek

Let's begin by checking to see if Pajek gives us the same scores in terms of network size, average distance, and diameter. Dichotomized versions of the four sets of networks are included in the project file, `Noordin's Network (Dichotomized)`.paj, having been exported from UCINET using the techniques discussed in Chapter 4. Read this file into Pajek using the *File>Pajek Project File>Read* command. Before making any calculations, we need to convert all arcs to edges (because Pajek reads ties in networks exported from UCINET as arcs rather than edges), remove any loops (i.e., ties along the diagonal) from the networks, and then extract

[Pajek-Main Screen] File >Pajek Project File >Read

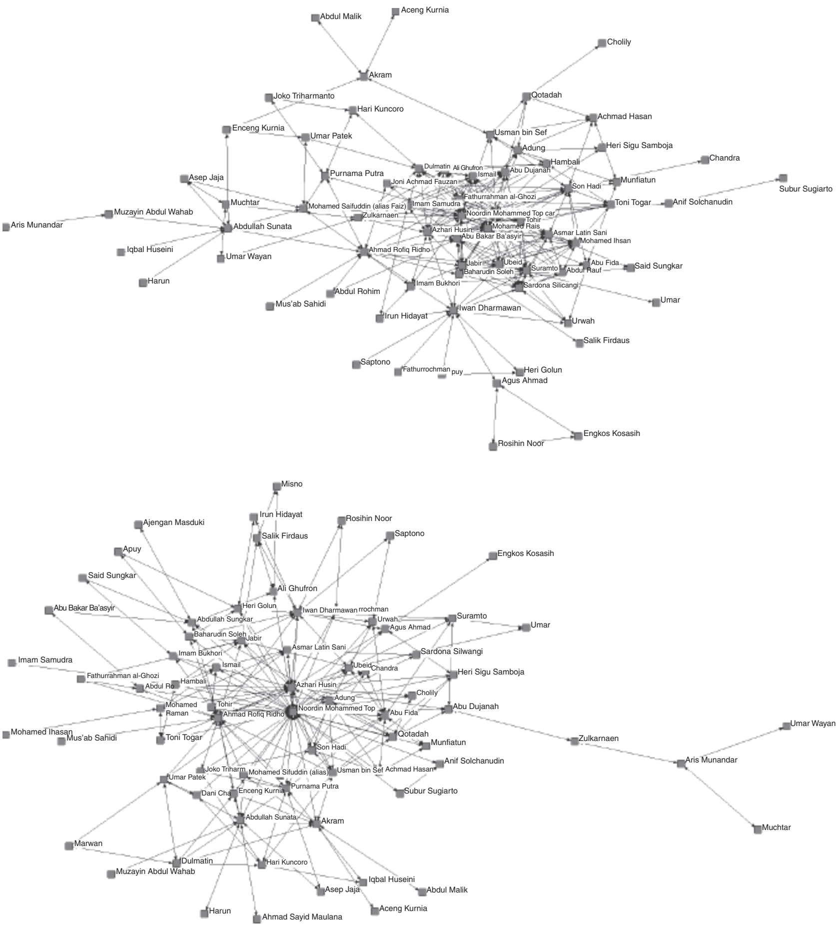


Figure 5.13. Noordin Trust and Communication Networks (NetDraw)

the individual relations from all of the multirelational (i.e., stacked) networks (i.e., trust and operational). The first of these steps is accomplished using Pajek's *Net>Transform>Arcs→Edges> All* command for *each of the four networks*. This brings up a dialog box asking whether you want to create a new network. Select "Yes." Pajek will then ask whether you want to remove multiple lines. Most of the time you will want to choose the "Single" option (#5). Pajek will assign the new networks with names such as "Symmetrizing (Single) of N1," which means that the new network was symmetrized (i.e., arcs-to-edges) using the "Single" option from the first network (N1) listed in the Network drop-down menu. You can rename them with the *File>Network>Change Label* command.

Net>Transform >Arcs→Edges >All

File>Network >Change Label

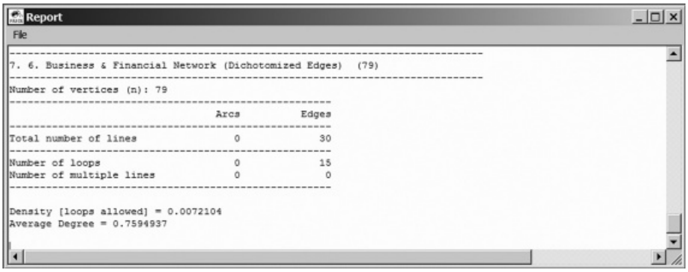


Figure 5.14. Pajek's Info>Network>General Report

Info>Network >General

Net>Transform >Remove Loops

Net>Transform >Multiple Relations >Extract Relation(s)

Next, we need to remove any loops that are in the networks so that Pajek does not take them into account when calculating density and average degree. Not all networks have loops, however. To determine whether they do, we use the *Info>Network>General* command for each network (note that the network for which we are seeking information needs to be highlighted in the first or top Network drop-down menu). Accept Pajek's defaults in the dialog box and click "OK." This calls up Pajek's report window with output similar to Figure 5.14, which indicates that the business and finance network has fifteen loops. To remove these, we select the *Net>Transform>Remove Loops* command. Once again, you will be asked to create a new network; select "Yes" and give the new network a name. Finally, for the (symmetrized) trust and operational networks, we need to extract each relation. To do this, we use Pajek's *Net>Transform>Multiple Relations>Extract Relation(s)* command. In the resulting dialog box tell Pajek that you want to extract relations 1 through 4 (because there are four relations in both the trust and operational networks).

Net>Paths between 2 vertices >Distribution of Distances >From All Vertices

The size of each network is indicated by the number in parentheses next to the network's name. As you can see in Figure 5.15, the trust network (as expected) contains seventy-nine actors. To obtain the average distance and diameter of a network in Pajek, we use the *Net>Paths*

Figure 5.15. Pajek's Main Screen

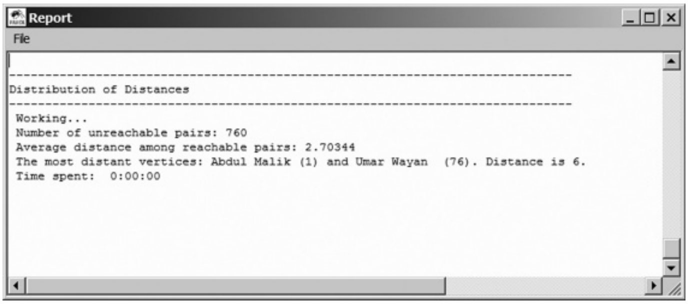


Figure 5.16. Pajek's Distribution of Distances Report

between 2 vertices>Distribution of Distances>From All Vertices command. The output contained in Pajek's report window provides the number of unreachable pairs, the average distance, and the most distant vertices (i.e., the network's diameter).

Figure 5.16 presents the output generated by Pajek for the (symmetrized) communications network. As you can see it indicates that the diameter of the network is 6.0 and the average distance between reachable pairs is 2.703. These scores agree with those listed in Table 5.1. Note that Pajek reports the names of the actors who are most distant as well as the number of unreachable pairs.

Unfortunately, Pajek does not currently provide a fragmentation score. However, you can calculate it by dividing the number of unreachable pairs (760) by the total number of possible pairs in the network, which is equal to network size multiplied by network size minus one, that is, $79^*(79-1) = 6{,}162$, and $760/6{,}162 = .123$, which agrees to the standard fragmentation score listed in Table 5.1.

Network Density and Average Degree. Whereas estimating density and average degree is very simple with Pajek when working with individual networks, it is somewhat more complex when working with multirelational (i.e., stacked) networks. This is why when working with the latter, most of the time you will probably want to estimate them in UCINET. Nevertheless, it is helpful to see how it is done in Pajek so you do not need to switch back and forth between programs unnecessarily. To obtain the density and average degree of the network appearing in the top or first Network drop-down menu, select the *Info>Network>General* command. This generates the same report (Figure 5.17) that we saw previously, at the bottom of which you will find the network's density and average degree. *Info>Network >General*

Unlike UCINET, in Pajek you can only estimate network density and average degree one network at a time. Here, we have requested information on the business and finance network. Note that the report indicates

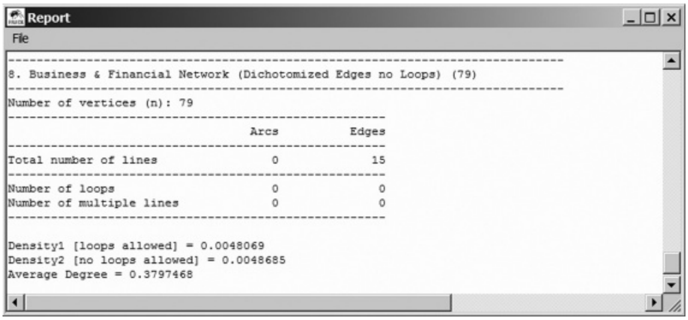


Figure 5.17. Pajek's Info>Network>General Report

network density with and without loops (i.e., including and not including the cells along the diagonal into the calculation). Looking at Figure 5.17, we can see that the density of the communications network is .005, while the average degree is .380, which agrees with UCINET's calculations (see Table 5.1).

Net>Vector >Clustering Coefficient >CC1

Clustering Coefficient and Small World Q. The clustering coefficient is calculated with the *Net>Vector>Clustering Coefficient>CC1* command. This creates one partition, two vectors, and a report (Figure 5.18) containing the Watts-Strogatz clustering coefficient (compare to Table 5.2 communication network score).

Net>Partitions >Degree Net>Vector >Centrality >Closeness

Centralization. Like UCINET, Pajek generates centralization scores when it calculates centrality. Degree centralization is obtained with the *Net>Partitions>Degree>All, Input, Output* command; closeness centralization with the *Net>Vector>Centrality>Closeness>All, Input,*

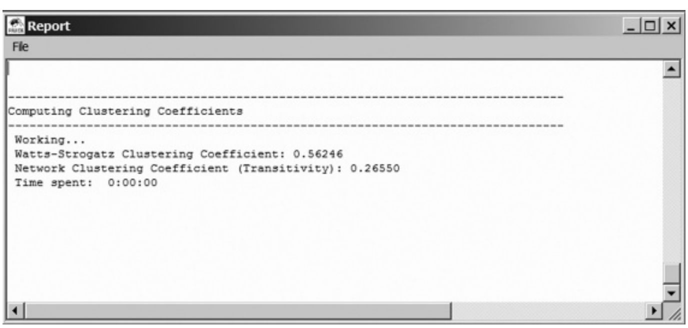


Figure 5.18. Pajek's Clustering Coefficient Report (Communication Network)

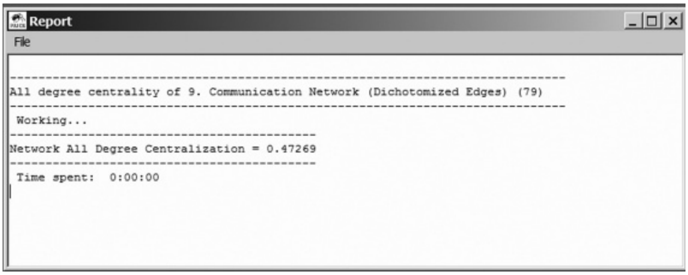


Figure 5.19. Pajek's Degree Centralization Report

Output command;[15] betweenness centralization with the *Net>Vector>Centrality>Betweenness* command; and eigenvector centralization with the *Net>Vector>Important Vertices>1-Mode: Hubs-Authorities* command. Figure 5.19 presents the report generated using Pajek's degree centrality command with the communication network. Note that it agrees with the score calculated by UCINET (Table 5.3). Unfortunately, Pajek does not calculate standard deviation. As we will see in the following section, ORA does.

Net>Vector >Centrality >Betweenness

Net>Vector >Important Vertices >1-Mode: Hubs-Authorities

Network Topography in ORA

To see how to calculate topographical metrics using ORA, we need to first load the four sets of networks (`Trust Network.xml`, `Operational Network.xml`, `Communications.xml`, and `Business and Finance Ties.xml`) using ORA's *File>Open Meta-Network* command. ORA automatically provides users with a count of the number of actors and the density of networks that are loaded into the program's Meta-Network Manager. Simply select a particular network and the number of actors (i.e., the count) and density appear in the "Statistics" portion of the Network Information/Editor panel. For example, in Figure 5.20 we selected the classmates subnetwork. We have also "deselected" the "Allow self-loops" box (this is similar to examining the "no loops allowed" density calculation in Pajek). Looking at the Network Information/Editor panel (admittedly quite small in the screen shot), we can see that the size of the network is seventy-nine and the density is .0568 (both are circled in Figure 5.20), both of which agree with our previous calculations.

[ORA] File>Open Meta-Network

Of course, we are typically interested in learning more than just the count and density of a particular network. To acquire the topographical characteristics discussed in this chapter we can use ORA's Standard

[15] Pajek calculates the traditional measure of closeness centrality, so it will not estimate closeness centralization with a disconnected network.

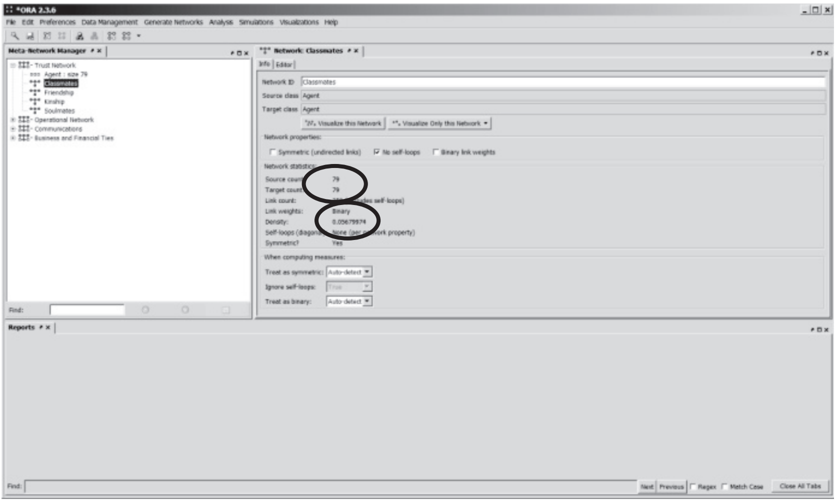


Figure 5.20. ORA's Main Screen with Network Statistics

Analysis >Generate Reports >Locate Key Entities >Standard Network Analysis

Network Analysis report, which is accessed with the *Analysis>Generate Reports>Locate Key Entities>Standard Network Analysis* command. At the first dialog box (not shown), make sure that the meta-network you are analyzing is selected (e.g., the trust network) and click "Next." This brings up a second dialog box (Figure 5.21) where you want to select all of the subnetworks you intend to analyze. Typically, you will want to select all of the subnetworks as we have here (i.e., classmates, friendship, kinship, and soulmates) although there may be occasions when you will only want to analyze a subset. After making your selection, click "Next."

At the next dialog box (not shown) accept ORA's defaults and click "Next." The next dialog box (not shown) will ask what type of output you want. For now, select "Text" (we will explore the other types of output further on), indicate where (i.e., in what folder) you want the report to be stored, and give the report a name (e.g., trust network). Click "Finish" and in relatively short order, a separate report for each network will appear under the Reports panel (each under its own tab). Click on one of the tabs, scroll down, and you will find ORA's calculations of *Network Level* measures, including count, density, characteristic path length (i.e., average distance), clustering coefficient, network levels (i.e., diameter), fragmentation, degree centralization, betweenness centralization, and closeness centralization. Scroll down a bit further, and you will discover a series of *Node Level* measures, including the average (and standard deviation) degree, eigenvector, closeness, and betweenness centrality – both normalized (scaled) and raw (unscaled). The unscaled scores should agree with what we calculated previously. Note that ORA

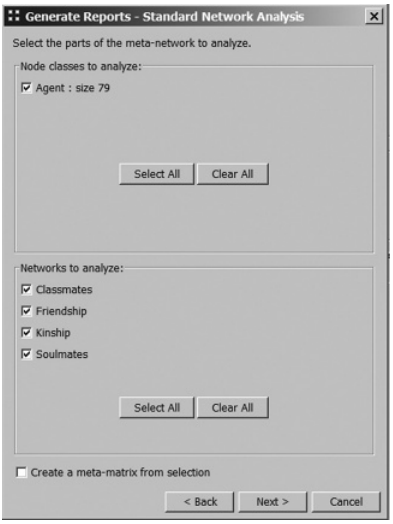


Figure 5.21. ORA's Standard Network Analysis Report Dialog Box

calculates total degree, indegree, and outdegree centrality. Although we will postpone our discussion of the difference between total indegree and outdegree until the Chapter 7, note that average total degree centrality is twice that of average degree in UCINET and Pajek, while average indegree and outdegree centrality equals average degree in UCINET and Pajek.

In other words, in this single report, ORA has given us all of the topographical metrics that we have been exploring in this chapter. This illustrates one of ORA's greatest strengths. Rather than issuing a different command for each set of metrics, ORA provides them all in a single report. You can even create your own custom reports with ORA's *Preferences>Other* command. That said, you need to be careful when using ORA's standard reports. They sometimes include metrics that are inapplicable for the network under analysis. For example, ORA Standard Network Analysis reports closeness centralization scores for disconnected networks, which it should not, and Krackhardt's measure of hierarchy for symmetric networks when the algorithm is only designed for asymmetric networks. This illustrates why it is important for you to have a working understanding of SNA. Otherwise, you could end up using an inappropriate metric when crafting strategies for the disruption of a particular dark network.

Preferences >Other

A Cautionary Note

This chapter's discussion has essentially treated the various dimensions as conceptually distinct from one another, but the formal measures of these dimensions are anything but independent from one another. For example, sensitive to network size but also affected by the degree of network centralization. "Because density is based on how many ties are present in the network, one or two individuals having a disproportionately high number of ties to others in the network might raise the density score" (Prell 2011:168). Thus, the two measures may often want to be used in conjunction:

> Using degree centralization alongside density is somewhat similar to what statisticians do when they make use of a mean and standard deviation. The mean is a measure of central tendency and the standard deviation is a measure of spread or variance. Similarly, centralization measures the extent to which ties hover around one actor, and density measures the extent to which all the ties are actually present. (Prell 2011:170)

To this we could also add that it would make sense to compare density with the variance and standard deviation scores discussed in previous sections.

As Noah Friedkin (1981) has noted, density is also sensitive to the number of cohesive groups within a network. It can be "a misleading indicator of cohesion in cases where the network in question has many cohesive subgroups . . . low densities in large networks may reflect more structural cohesion than higher densities in smaller networks, as such large networks have fewer cohesive subgroups, and hence, less amount of 'fragmentation'" (Prell 2011:171). Thus, in addition to considering density, average degree, the clustering coefficient, and the small world quotient as measures of network cohesiveness, we would be well advised to examine fragmentation as well. The broader point, of course, is that many network measures are interdependent with one another, and that they should not be examined in isolation but rather together.

5.6 Summary and Conclusion

In this chapter we explored how to use UCINET, Pajek, and ORA to calculate several different sets of network topography metrics: size, diameter, average distance, fragmentation, density, and centralization. UCINET is able to estimate all the metrics covered in this chapter, while Pajek and ORA are able to estimate most. ORA's report-based approach probably makes it the easiest of the programs to use, but as we saw, it will report metrics even when they do not apply to the network being analyzed. As mentioned in Chapter 3, Pajek's strength lies in its ability to calculate these (and other) metrics with very large networks. This ability did not come into play in this chapter because the Noordin network is relatively small, but not all dark networks are small (e.g., the FARC).

We also briefly considered a series of studies of "light" networks that suggest that effective networks can be neither too dense nor too sparse, neither too centralized nor too decentralized. Instead, they must land somewhere on a continuum between the two sets of extremes. If these same dynamics are at work in dark networks, then analysts need to take into account network topography in the crafting of strategies to disrupt them. For example, imagine a scenario where the dark network we are seeking to disrupt is located toward the dense or provincial side of the cosmopolitan-provincial continuum. Moreover, let's assume that the network has crossed the tipping point and has become too dense and is no longer as effective as it used to be. If, in such a situation, we targeted a central actor for capture or elimination, that could have the unintended effect of making the network less dense and (consequently) more effective. Instead, we may want to adopt a strategy (or an array of strategies) that pushes the network in the opposite direction (e.g., peeling off peripheral members through rehabilitation and reintegration programs), causing it to become even more dense and hopefully less effective. Of course, the crafting of strategies is seldom so simple. We often need to take into consideration second-order effects. As we saw in the first chapter, for example, research has found that extremely dense networks that are effectively cut off from the wider society, tend to become more extreme (Sunstein 2003, 2009), and, in fact, this is what appears to have happened with the Hamburg cell whose members participated in the 9/11 attacks. Sageman (2004) found that they had met at a local mosque, moved into the same apartment where they underwent a long period of intense social interaction, progressively adopted the beliefs and faith of their most extreme members, and then eventually joined the global salafi jihad. Thus, if we do attempt to force a network to become more dense, then we need to be prepared (at least in the short term) for the effects that such a strategy could have.

With regards to Noordin's network, we have drawn a few tentative conclusions and suggested possible strategies for disruption. For example, we saw that the classmates network is relatively dense (Table 5.2) and somewhat centralized (Tables 5.3 and 5.4). This suggests that strategists may want to target the schools that contributed most to the forming of these ties. Figure 5.22 presents the school network, which was derived from the two-mode school network (see Appendix 1). Here, a tie between two schools indicates that at least one individual is associated with both schools. For example, if a person attended one school and teaches at another, then a tie has been drawn between the two schools.

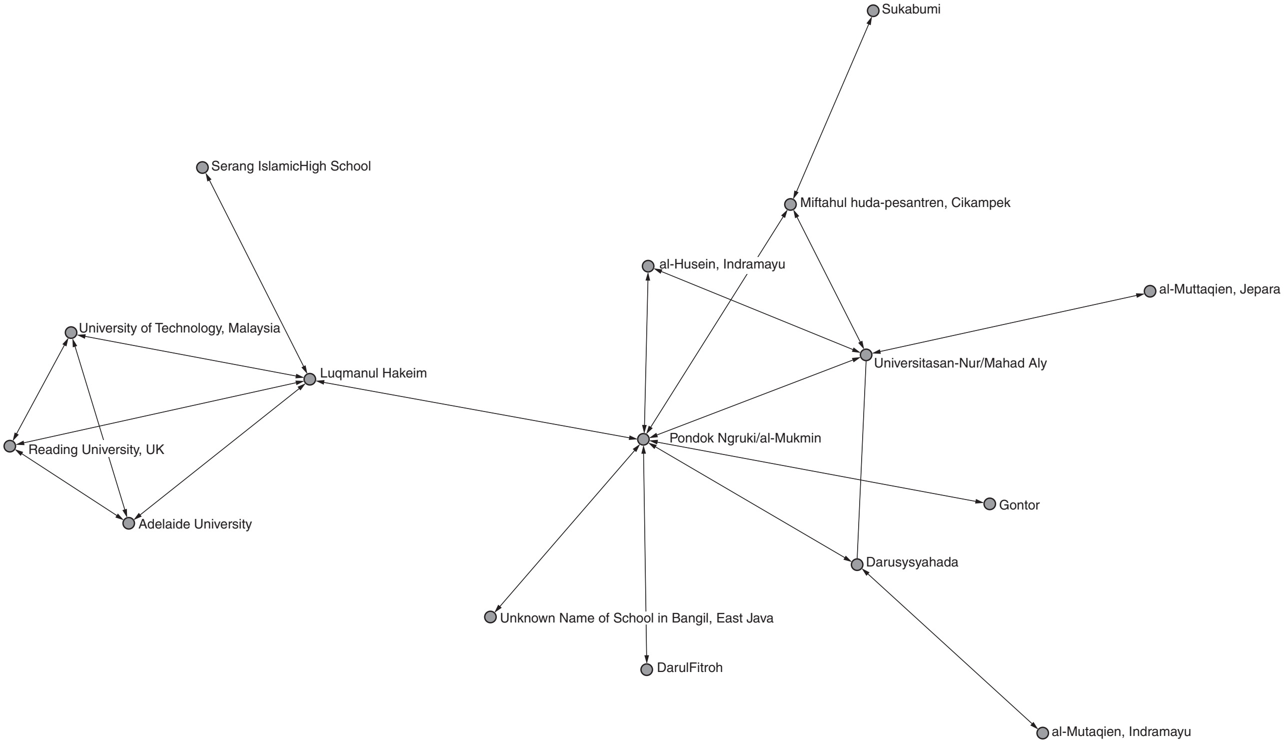


Figure 5.22. Noordin School Network (Isolates Removed)

Although it will not be until Chapter 7 that we examine various measures of centrality, clearly certain schools are more central in Noordin's network than are others: namely, the Universitas an-Nur, Pondok Ngruki, and Luqmanul Hakeim schools. As noted previously, a kinetic strategy might seek to close down one of these schools, which is exactly what the Indonesian authorities did to the Luqmanul Hakeim school; it is unclear whether this had a noticeable effect on the functioning of Noordin's network (or, for that matter, Jemaah Islamiyah) because it is relatively easy to set up another school somewhere else. A nonkinetic strategy of building alternative schools that teach traditional subjects (e.g., reading, writing, arithmetic, and moderate forms of Islam) near the more central schools in Noordin's network might prove to be a more effective strategy in the long run, because it would most likely facilitate the integration of Indonesians into civil society (Roberts and Everton 2011; Tilly 2005).

Of course, the School network is not the only network analysts would want to consider. For example, the communication network is relatively dense and highly centralized, suggesting that it might be vulnerable to the removal of a few actors (i.e., a kinetic action) or the diffusion of disinformation through the network (i.e., a nonkinetic action). In both cases, centrality measures might prove useful in identifying key actors as could Borgatti's (2006) key player algorithms. We take up the former in Chapter 7 and the latter in Chapter 8.

Finally, in this chapter we have only considered these metrics in terms of a single snapshot in time. In reality, as we noted in the first chapter, dark networks are in a constant state of flux, with actors entering and leaving almost continuously. In other words, the density, centralization, and fragmentation of a network will almost certainly change from one time period to the next. Thus, we also need to monitor these networks over time to see whether they are becoming more or less dense or more or less centralized. Such information will (or at least should) inform the crafting of strategies as well. We will explore how to do this in Chapter 10. For now, we turn to examining how to use SNA to identify the subgroups of particular networks and how this information might be used in the disruption of dark networks.