Egocentric network data – sometimes known as “personal” network data or “survey” network data – assembles information on the local social environments surrounding individuals. Generally the analysis of egocentric network data is not as involved as that of complete network data. Usually it amounts to the construction of index- or scale-like measures. Some of these are additive, though others have more complex forms. Once indices or scales are obtained, they can be linked to other data that are thought to be antecedents or consequences of network structure – for example, Burt links egocentric measures of structural autonomy to industry cost-profit margins or to job promotions. Usually, these relationships are explored within standard statistical packages. However, it is also possible to learn quite a bit from ego-centric analysis of complete network data. The first part of this problem set focuses on using UCINET to analyze a complete network dataset; the second part suggests ways that one may use relational ideas in more traditional forms of analysis.

**Ego-centric analysis of complete network data**

First, egocentric measures may be constructed on the basis of complete network data. That is, one may look at the structure of local networks within a complete network. This means successively treating each actor as “ego” and calculating various ego network properties. Within UCINET 6, this is done in the Network/Ego Networks menu. This has entries for “Density”, “Structural Holes” and “Brokerage”.

The Density command calculates a variety of ego-centric measures. In essence, this command isolates the portions of a complete network dataset that the ego can reach and calculates a series of measures for the data, including:

**Size:** The number of actors (alters) that ego is directly connected to.

**Ties:** The total number of ties in the ego network (not counting ties involving the ego). In other words, how many ties exist between the alters in the ego network.

**Pairs:** The total number of pairs of alters in the ego network – i.e., potential ties. This is calculated using the formula N(N-1), where N is the number of alters in ego’s network (i.e., N is the size of the ego network).
**Density:** The number of ties in the ego network divided by the number of pairs, times 100.

**Avgdist:** The average geodesic (graph-theoretic) distance between pairs of alters. This is only computed for networks in which every alter is reachable from every other. In many cases, this measure is blank because ego networks are often not fully connected – think about the small personal datasets you have used for earlier problem sets.

**Diameter:** The longest geodesic distance within the ego network (unless infinite because the dataset is disconnected).

**NweakComp:** The number of weak components in the ego’s network.

**PweakComp:** The number of weak components as a percentage of the number of alters.

**2StepReach:** The number of alters that are within 2 links of ego. (My copy of UCINET seems to give the wrong answer for this.)

**ReachEffic:** 2-step reach as a percentage of the number of alters plus the sum of the their network sizes. (My copy of UCINET seems to give the wrong answer for this.)

**Broker:** Number of pairs not directly connected. If pairs are not connected, then the ego acts as broker between the members of the pair.

**Normalized Broker:** Broker divided by number of pairs. Percentage of pairs in which ego may act as broker.

**Ego Betweenness:** Betweenness centrality of ego in ego’s network

**Normalized Ego Betweenness:** Betweenness of ego in ego’s network, normalized by the size of the network.

The Structural Holes command calculates measures of effective size, network efficiency, constraint, and hierarchy as outlined by Burt in *Structural Holes* (chapter 2). It also constructs dyad-level indicators of redundancy (*Structural Holes*, p. 51) and constraint (p.54). The measures constructed do not take account of data on “secondary” structural holes.

The Brokerage entry calculates measures of brokerage as defined by Fernandez and Gould in a 1989 article in *Sociological Methodology* (see the help file for the full reference). This routine requires a partition vector indicating the membership of an actor in a position. The partition vector is usually based on a structural or abstract equivalence analysis of the data and then a clustering run to develop a partition matrix.
These routines produce output data sets that can be edited and later merged with other data in order to explore statistically the links between ego network properties and other measured properties of an actor.

For the following exercises we will use the data from the “PRISON” dataset. My copy of UCINET could not find the data description in the “help” command, so I have put the description at the end of the problem set.

Run the Network/Ego Networks/Density command on the Prison data. Which four or five prisoners have the largest egocentric network? Which four or five prisoners broker most extensively in their ego network? If friendship in prison is the basis for creating social capital needed to mobilize and coordinate an insurrection, which prisoners would the warden fear most, solely based on the network data found here (since a knowledge of the person’s intent and organizational skills would also be very useful)?

Now run the structural holes analysis on the Prison data. Based on the structural holes analysis, which two or three prisoners are likely to know the most about the willingness of members of the tier to participate in an insurrection?

Now think about the constraint concept found in Structural Holes. What would constraint mean substantively in this case? What is constrained? Which two or three actors face the least constraint (use the “Structural Holes” table in the output to answer this question)?

For the ego that faces the least constraint, which two alters in their network contribute the most to their total measure of constraint? (you will need to look at the constraint matrix carefully to answer this question.) Why do these actors seem to contribute to constraint? Use NetDraw to visualize the Prison data and then see if you can make sense intuitively why these two actors tend to constrain the least constrained ego.

Now run an equivalence partition of the Prison data. You may choose any of the three versions of equivalence, but your choice should make sense in light of the fact that we are interested in helping the warden avoid an insurrection. Include a paragraph in your write-up justifying your choice. After you run the equivalence routine, use a complete link clustering to find the partition. Choose a partition from the output and use it to run the “Brokerage” routine in Network/Ego Networks. Look carefully at the legend for the types of brokers (its printed at the bottom of the first output matrix). Which actor has the most brokerage positions in the network? Given the context, which type of brokerage might be the most dangerous (i.e., likely to help create a large insurrection)? Does your analysis of brokerage match your earlier analysis from the “density” routine?

Overall, which actors in this network would you advise the warden to monitor most closely? Why?
Ego-centric analysis of survey data

Note: No items are due from this section, but you may wish to try these procedures.

Often, however, one will wish to work with egocentric data alone, rather than in context of a complete network data set. In this case, UCINET is not an appropriate way to proceed.

Ordinarily, egocentric data are studied using conventional statistical packages like SPSS, SAS, or Stata. Usually the data are organized in a “flat” respondent-by-variable format, where variables describe different properties of alters, respondent-alter links, or respondent-respondent links. Various combinations of the data are then made to describe properties of the egocentric network for each respondent.

For example, egocentric network data in the General Social Survey (GSS) are organized in a respondent-by-variable format. A large number of variables is required to describe the elements of each respondent’s egocentric network. In the 1985 GSS, which includes the most extensive egocentric data, examples of egocentric network variables include the following:

- NUMGIVEN (# of alters cited)
- CLOSE12 (closeness of alters 1 and 2) [similar variables appear for other pairs of alters]
- SEX1, RACE1 (sex and race of alter 1) [similar variables for other alters and attributes]
- PARENT1 (indicator variable for whether alter 1 is a parent) [similar variables for other alters and role relations]
- TALKTO1 (frequency of discussion with alter 1) [similar variables for other alters and other properties of R-alter links such as closeness and duration]

Ego network measures are constructed by performing various aggregations of these variables. For example, network density is the mean value of CLOSE12-CLOSE45 (there are 10 possible pairs of alters for each respondent; the mean is taken across those pairs that are valid for any given respondent). In Stata, for example, this might be achieved as follows:

```stata
egen density=rmean(CLOSE12 CLOSE13 . . . CLOSE45)
```

Measures of network composition are means of attribute variables; for example sex composition of the egocentric network could be indexed as the mean of SEX1-SEX5 (again taking the mean across only valid values). A suitable Stata command might be:

```stata
egen sexcomp=rmean(SEX1 SEX2 SEX3 SEX4 SEX5)
```

(One may want to recode the SEX variables before this aggregation such that “sexcomp” is interpretable as “proportion female” or “proportion male.”)

Measures of ego network heterogeneity can be defined using dispersion measures for alter attribute variables like SEX1-SEX5 or RACE1-RACE5. For examples of some of these sorts of measures for the GSS data, see the Marsden article in the American Sociological Review, 1987. For example, if one wished to measure age heterogeneity as the standard deviation of ages across
alters within each respondent’s egocentric network, the following Stata command would be suitable:

\[
\text{egen agessd=rsd(AGE1 AGE2 AGE3 AGE4 AGE5)}
\]

Not all functions one might wish to use are prepackaged; programming is of course more involved if one has to develop one’s own functions.

3. Occasionally it will be useful to view egocentric data not as a “flat” respondents-by-variables data set but instead as a hierarchical data set in which there is a record for each alter, which is nested within a respondent record. Various multilevel analysis techniques can then be used in studying the egocentric data. Within the statistical package Stata, the “reshape” command is very helpful in switching between a “flat” organization of the data and a hierarchical organization.

Description of PRISON Dataset

GAGNON & MACRAE PRISON

<table>
<thead>
<tr>
<th>DATASET</th>
<th>PRISON</th>
</tr>
</thead>
<tbody>
<tr>
<td>DESCRIPTION</td>
<td>One 67x67 matrix, non-symmetric, binary.</td>
</tr>
<tr>
<td>BACKGROUND</td>
<td>In the 1950s John Gagnon collected sociometric choice data from 67 prison inmates. All were asked, &quot;What fellows on the tier are you closest friends with?&quot; Each was free to choose as few or as many &quot;friends&quot; as he desired. The data were analyzed by MacRae and characterized by him as &quot;less clear cut&quot; in their internal structure than similar data from schools or residential populations.</td>
</tr>
</tbody>
</table>