Rockefeller College
University at Albany

Problem Set #3: Visualizing Networks

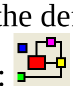
This problem set covers portions of the UCINET and related packages that are used to visualize network data. The following work focuses to some extent on multidimensional scaling (MDS) because MDS has grown prominent in both network analysis and other, more traditional forms of analysis. See Freeman's article in the *Journal of Social Structure* for a much wider overview of graphical procedures used by social network analysts, and for pointers toward available software for using them. As Freeman points out, multidimensional scaling focuses on the placement of the points (nodes in a network), and not on the depiction of linkages among the points.

However, we will also work with NetDraw, Mage, and Pajek. As computer technology has improved, the power of network visualization tools has grown accordingly. Until very recently (i.e., while I was writing my dissertation) the primary visualization tool was KRACKPLOT, a capable but very limited package. Now, the options are growing every day and include the ability to project networks in three dimensions (using Mage).

As I have said more than once, network analysis is at least as much art as science. Here, the scales tip even further toward art and away from science. The ability to gain insight from visualization is as much about using the tools and techniques as it is about being creative and inventive.

As before, we will work with the Florentine families from the "Padgett" dataset, "Games" data matrix extracted from the "Wiring" dataset, and your personal network. I will not go into great depth on all of the graphing and analytic options with the three visualization tools; you will need to do some exploration on your own.

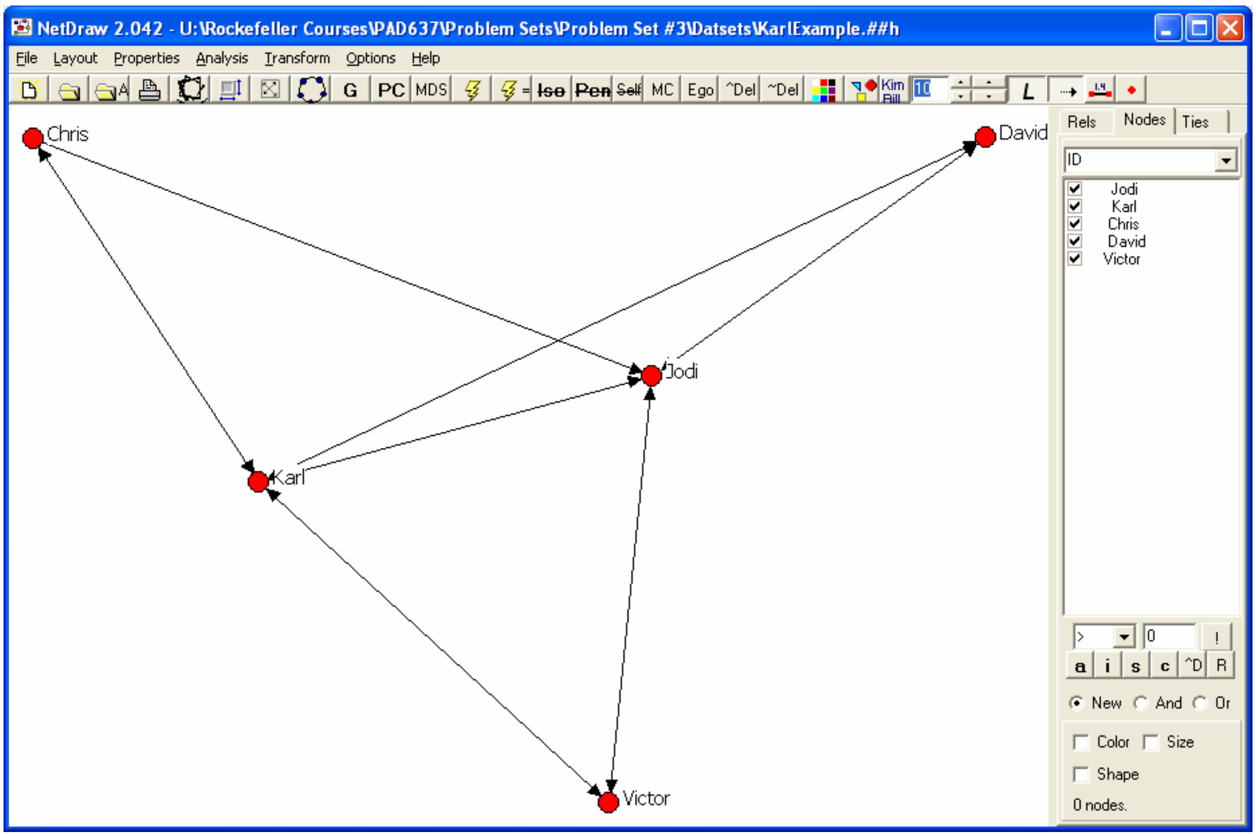
<u>Working With NetDraw</u>

NetDraw is the successor to KRACKPLOT that is now built into UCINET. To say that it is a big improvement over its predecessor is a vast understatement. Let's begin by looking at my personal network, as its visualization is quite simple. You should follow along using your personal dataset. Begin by making sure that you have set the default folder in UCINET to the location where the Wiring data is located. NetDraw, Mage, and Pajek all use the default folder that is set in the UCINET console. Then click on the NetDraw icon – this one:  – on the UCINET console. NetDraw launches as a separate window and program.

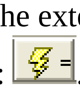
At this point you will probably want to have the 12-page NetDraw users guide at hand, as there are many options built into this program.

We now want to open my dataset in NetDraw using the File/Open/UCINET dataset/Network command. NetDraw can directly read a UCINET file. Once you navigate to a file, NetDraw will immediately display the network. A shortcut is to use this tool bar button: .

Here is what my screen looked like once I opened the file:



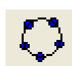
The nodes are placed using one of many built-in placement routines, which are available in the Layout menu. The primary placement options are:

Spring-embedding (toolbar: ) – This routine tries to get as much spacing as possible between nodes and to avoid having edges cross one another, to the extent possible. There are two icons for spring-embedding – the one shown above and this one: . The icon without and "=" sign causes NetDraw to execute a spring-embedded placement based on the options set in the Layout/Graph-Theoretic layout/Spring embedding menu. The icon with the "=" sign configures the algorithm to place nodes based on "node repulsion" and "equal edge length bias." This is the standard option. When so configured, the placement in space is based on forcing the nodes apart and tending to select placements that lead to equal edge lengths (i.e., equal length lines between nodes).

MDS (toolbar: MDS) – Executes a two-dimensional multi-dimensional scaling of geodesic distances and then uses that output to place the nodes – see below. (As of this writing, however, the output for this routine seems strange, because it depends on the previous positions of the points – even though MDS relies on the node's tie structure, not their spatial position. It may be that the MDS output is used to alter the relative position of the nodes, given where they were placed to start with.)

Gower (toolbar: G)– Executes as Gower metric scaling of dissimiliarity and then places the nodes. (We will not discuss this type of scaling, but it is interesting to compare the output of the MDS and Gower placements)

Principal components (toolbar: PC) – Executes a principal components analysis of the data and uses the output to place the point. (Think about factor analysis – this routine assumes that there is some set of unmeasured factors that make some points more similar than others.)

Circle (toolbar: )– Places the points in a circle.

Random – Places the points randomly in space.

One thing to remember about each of these layout routines: there is nothing that makes one layout analytically superior to another – hence the inclusion of a random routine. I have sometimes found running the Random placements over and over helpful, because you get to see the data without imposing one's particular biases on the placement.

You can create your own layout of the data by "dragging" any point around in space. For larger networks, this can be useful because it allows you to disentangle a clump of nodes.

Notice the right sidebar with tabs. The "Rels" tab lets one establish a criteria for the display of an edge between two nodes. If you have dichotomous data, this tool isn't very useful, but with valued data it can allow you to remove low-value ties. In my policy network data, this tool is very useful, because there are lots of "hangers-on" that communicate once or twice a year with other members of the network. What I really wanted to see were the core members, who communicate at high frequency. The Rels tab lets you set a criterion of greater than, equal to, greater than or equal to, etc. (just like the "dichotomize" command in UCINET proper).

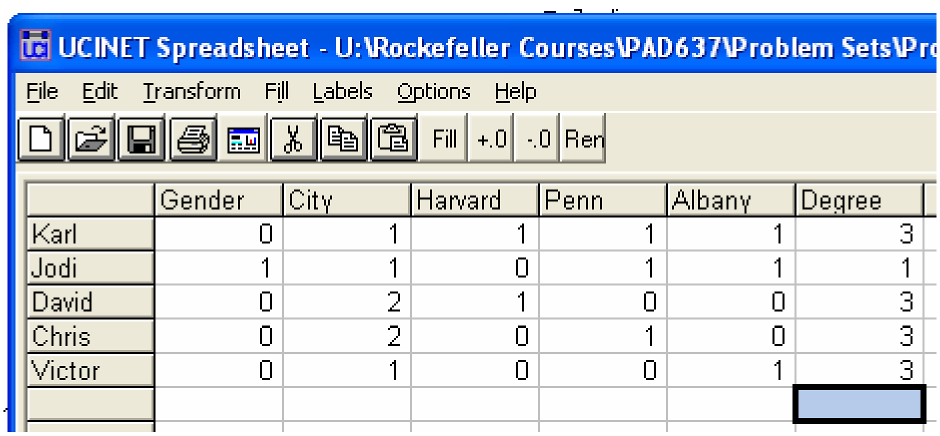
The "Nodes" tab is used for just that purpose – to turn on and off nodes. This can be very helpful when you want to see how a graph is affected by removing one member. To remove one member, "de-check" their box. In your personal network, declick yourself. Since your personal network is basically an ego-centric network, removing yourself is likely to remove many edges and leave a relatively poorly connected remnant. At the bottom of this menu are four buttons. The "a" button turns on all the nodes. The "i" button toggles between complements. Try turning off two or three nodes in your personal network and then click on "i". The nodes that were on will go off and the nodes that were off will go on. The "s" button turns on one node at a time. The "+" button adds nodes one at a time, starting with the top-most node that is turned off.

Notice at the top of the "Nodes" tab there is a "drop box" that says ID. In NetDraw files, you can add other characteristics of a node. This can be extremely helpful, because you can use these characteristics to make your network easier to interpret. For instance, you might want to make the female members of your personal network a circle and the male members a square.

In order to use this functionality, you must create a file that contains the attribute data. There are two ways to do this:

1. By creating a non-square matrix in UCINET.
2. By manually editing a NetDraw file, which uses a new file format, "VNA."

I will assume that most people will wish to use UCINET. There are many advantages to using a UCINET file, not the least of which is that you can create attribute data files in Excel, Stata, or SPSS and then copy and paste into the UCINET spreadsheet view; you may also import directly from an Excel file saved in Version 4,5, or 7 format. When using the UCINET option, the attribute data is stored in the familiar case-by-variable format that one would find in a standard statistical package. Here is the UCINET file containing the attribute data for my personal network:



	Gender	City	Harvard	Penn	Albany	Degree
Karl	0	1	1	1	1	3
Jodi	1	1	0	1	1	1
David	0	2	1	0	0	3
Chris	0	2	0	1	0	3
Victor	0	1	0	0	1	3

Attribute data is loaded into NetDraw by using the File/Open/UCINET dataset/Attributes command or this button on the toolbar: .

Once you have loaded the sociomatrix and the related attribute data, you are ready to begin work on the visualization. NetDraw allows you to save a visualization at any point in time by using one of two options: File/Save Diagram As… and File/Save Data As…. The "diagram" option allows you to save the current visualization as a picture in JPEG, Bitmap, or MetaFile format. These files are, of course, not manipulable; they are a snapshot of your visualization at that moment. Currently there is no direct copy and paste option. You can, of course, embed JPEGs, BMPs, WMFs, or EMFs in a Word document (using Word's Insert/Picture/From File…

command). However, these formats are subject to loss of resolution. The way to retain the image's quality is to "print" the file to a PDF or PostScript file and then embed it in a Word document. Dealing with PDFs and PostScript is much harder. It requires specialized software (the most common being GhostScript/GhostViewer, which are free) and, often, specialized expertise. If you are trying to produce publication-quality output from NetDraw, come see me for a few pointers – and cautionary tales about working with PostScript. For our assignments, embedded JPEGs are fine.

The other save option, Save Data As…, allows you to save the sociometric and attribute data for the visualization in several file formats (Pajek, Mage, UCINET, and VNA). Of the four, VNA is the most flexible. VNA files retain the sociometric data, attribute data, and the current state of the visualization (node placement, node colors and shapes, line colors and shapes, etc.).

When you save to VNA format, NetDraw will ask you if you want to save the data for the currently visible nodes or for all nodes. This option allows you to use the visualization to "prune" line or nodes in the network on the basis of visual characteristics. Unfortunately, the "save in UCINET format" command does not offer the option to save only the visible data – it always saves all the nodes. However, you can create a "reduced" UCINET file by (1) saving only the visible nodes to VNA format, (2) quitting the program, (3) restarting NetDraw, (4) opening the "reduced" VNA dataset you just created, and then (5) saving the data to a UCINET file.

VNA files are text files that you can open in Word or in NotePad. (If you use Word, be sure to save your changes as a plain text file, not a Word document.) NetDraw's user guide explains how to edit this file directly. In most cases, you will not want to manually edit the VNA file

You can use the attribute data to set the node shape. For instance, go to Properties/Nodes/ Shapes/Attribute-based. Select an attribute from the drop-down box at the top of the options dialogue. I changed the shape of female to "triangle" and male to "circle." Then using Properties/Nodes/Colors/Attribute-based I changed the color of male to red and female to blue. (The Nodes tab also lets you set colors, shapes, and sizes – see the check boxes at the bottom.)

For larger datasets where there are many different types of actors, setting each of these characteristics can help you – and your readers – understand the structure and shape of the network.

Using your personal network, open your file in NetDraw. Create a JPEG of this "default" image and include it with your write-up. Then add at least two attributes to your personal network dataset and use those attributes to determine the shape of the node markers and their color. Try at least three different layout options once you have "upgraded" your dataset. Include at least three versions of the network in your write-up. Which visualization do you find most useful/insightful? Why?

Using NetDraw for Analysis

NetDraw has some basic analytic routines built into it. We will briefly look at the Games relation from the Wiring dataset using NetDraw's K-core option. Begin by opening the "Games only" dataset you have previously completed. Then run the Analysis/K-cores command. NetDraw will not change the spatial positioning of the nodes, but it will change their formatting. There are two K-cores that are identified here. By default (I think), one is shown as black circles and the other is shown as red circles. The two social isolates are small blue circles. Go to the Nodes tab. You will see that NetDraw has added another attribute to the list – one called "*K-cores". The cores are identified by the minimum number of interconnections that are required to be a member. The "red" core has at least four internal connections; the "black" core has at least three; the social isolates each have zero. You may "turn on/off" each core by clicking on the checkbox next to the interconnections number.

Each of the other techniques in the Analysis menu work the same way: the network's members are partitioned, a new "attribute" is established in the Select nodes menu, and a formatting scheme is created and applied based on the analysis. The network will be formatted according to the last analysis you did, though the attribute partitions remain in the Select nodes menu.

After the analysis, you may still wish to try one or more of the layout options to get a clearer picture of the network's structure.

Using the Padgett marriage data, execute both a K-core and a blocks and cutpoints analysis. Include both network output in your write-up. What does the cutpoint analysis suggest regarding the important actors in the network? How does that compare with the K-core analysis?

Multidimensional Scaling

Adapted from original by Peter V. Marsden, Harvard University

Multi-dimensional scaling (MDS) is a widely used tool in both network analysis and other forms of social research (see the examples in the Bartholomew et al. reading). An MDS routine is included in NetDraw, but UCINET's procedure is more flexible, so we will experiment with it. The first step in multidimensional scaling is to obtain a matrix of distances or "proximities" among the nodes in a network. This may be done in many different ways, some of which have been introduced during the course already; more will be introduced as we go along. Multidimensional scaling is an approach for representing the similarities or dissimilarities among any set of objects, not just those related to social networks.

One straightforward way of doing this is to use the simple "adjacency" matrix for the network. *This is a "proximity" (or "similarity") measure because a high numerical value indicates that two nodes of the network should be placed close together in the spatial representation of the network.* If you are studying valued ties with ordinal or continuous measures of network links, these, too, can be scaled as proximity measures.

Another approach would be to study the path distance, or geodesic distance, matrix that is obtained via the Networks/Cohesion/Distance menu. *This is a distance (or "dissimilarity") measure because a high numerical value indicates that two nodes of the network should be far apart in the spatial representation.* We will encounter a number of other dissimilarity measures as well.

It's important that you maintain the distinction between a similarity and a dissimilarity measure. Weird results will be obtained if you try to scale a similarity measure as a dissimilarity measure, or vice versa. Peter Marsden calls this an "inside out" representation. Generally, if you attempt this, the resulting configuration of points will fit the data badly (as indicated by a large value of the "stress" measure). It should also be hard for you to interpret, since this would represent a rather substantial distortion of the data.

Sometimes it will be important to make an asymmetric dataset symmetric before submitting it for multidimensional scaling analysis. The routines implemented in UCINET assume that the distance from A to B is the same as the distance from B to A (some other multidimensional scaling software can relax this assumption, but such options are not implemented in UCINET and they involve some ancillary assumptions that you may not wish to make). You can create symmetry using the Transform/Symmetrize menu; recall from Problem Set #1 that it offers a number of choices about how symmetry can be obtained. This is not an issue with the Padgett marriage data or the Games data from the Wiring dataset, though, since the ties there are not directed. It may be an issue with your personal data and is an issue with the W&F children's network.

Another problem can arise if you have a network in which one or two points are isolates (having no links to others, and therefore "infinite" geodesic distance to the others). This can result in so-called "degenerate" solutions in a multidimensional scaling analysis, in which all of the connected nodes are piled up on top of one another, and widely separated from the isolate point(s). Such analyses are generally uninformative. They tell us what we already know, that some points are not part of the structure of social relations. It is usually best to exclude isolates using the Data/Extract menu before scaling the data, in order that connected points are separated from one another. For an example of this problem, try analyzing the Games data, in which two workers (I3 and S2) have no links to any of the others. The MDS solution you obtain will differentiate only modestly among the other 12 points. I imagine you will find that the picture obtained is more useful if those rows and columns are removed from the data matrix.

Executing an MDS. The main program within UCINET for obtaining multidimensional scaling results is in the Tools/MDS menu. We'll use the non-metric option. The difference between metric and non-metric has to do with what requirements are placed on the relationship between the proximities/distances in the data, and those in the geometric representation. Metric assumes that the scale of difference in two measures matters. Non-metric assumes that the order of measures matters but that the absolute differences may not be meaningful. For instance, in Likert scale data, we know that a "5" is more frequent communication than a "6", but the difference in frequency between "4" and "5" and between "5" and "6" may not be equal. Likert scales sometimes elicit S-shaped response curves. So we may wish to use the order data but not the cardinal properties of the data. Non-metric MDS recovers order only.

Within this menu, you must specify the UCINET dataset containing your proximity/distance measure, the number of dimensions you wish to obtain, whether your data are similarities or dissimilarities (it is crucial that this be specified correctly), and the name of an output file for the coordinates of your display (by default this is called NonMetricMdsCoord). Two dimensions is the default choice – and by far the most convenient from a presentational standpoint, but there is no *requirement* that MDS solutions be confined to two dimensions. In fact, we will generate a three-dimensional solution and submit it to Mage for display. There are also various technical options, but by and large the defaults on these are OK.

MDS output includes a plot of the points (this will be the first thing to appear on your screen), the coordinates for points as embedded in multidimensional space, and (optionally) various technical information. For presentational purposes, the plot will be the main thing, but the coordinates can be helpful if you want to engage in interpretational efforts, or if you want to obtain a prettier plot using other software – like Mage or NetDraw. It is possible to save the plot as a metafile, in order to insert it into a document you are working on. You also have options as to whether or not the points in the plot are labeled.

Interpreting and MDS. There are numerous ways to interpret an MD scaling solution, many of them ad hoc. I shall describe just a couple of them here.

The first is to determine whether the plot is technically acceptable. The relevant statistic here is the "stress" measure. Stress is a measure of fit. The statistic is the squared differences between the actual distance data and the "MDS" distance divided by the actual distances squared. This statistic goes to zero if the MDS solution perfectly reproduces the measured distances. The cutoff value for acceptable stress is ad hoc, but the general standard (and one that is promoted in the UCINET manual) is the following: below 0.10 is excellent; between 0.10 and 0.20 is acceptable; and above 0.20 is unacceptable.

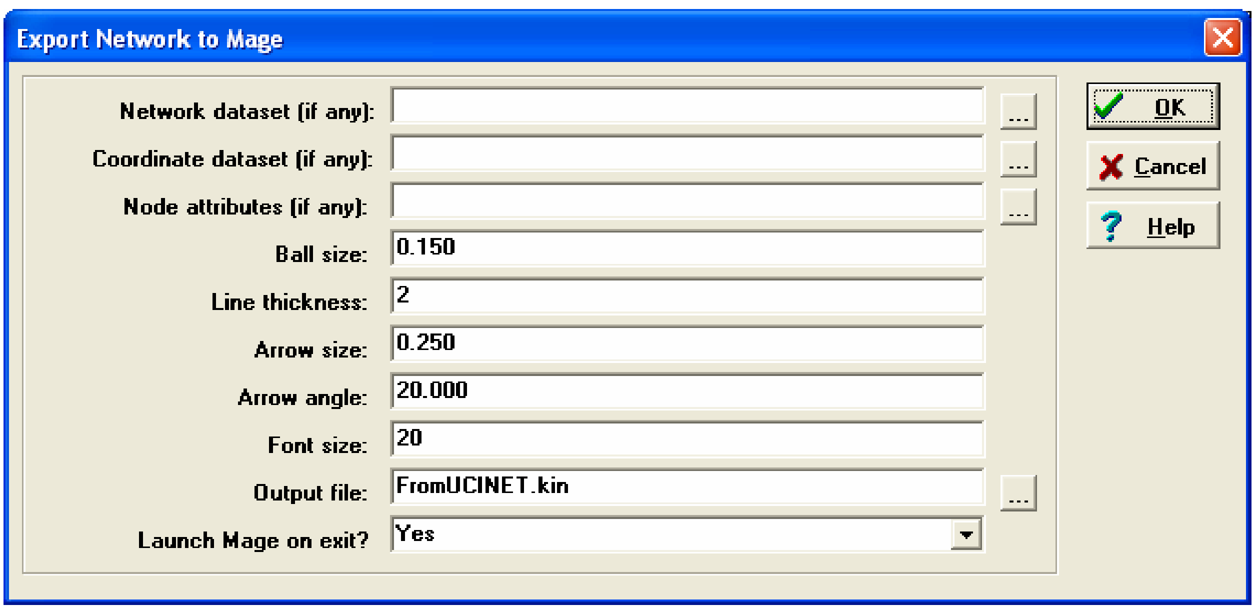
From a substantive perspective, the key problem is trying to determine what the dimensions mean. One approach is to try to identify attributes/variables that are associated with a node's position along each dimension. This can be done via correlating a variable with the coordinates that appear in the dataset generated by the MD scaling routine, or by simple regression. A more general version of the same idea is to conduct a multiple regression of a variable on the coordinates for <u>all</u> dimensions; this helps to locate the "best fit" projection of the variable into the multidimensional space. A variant on the regression procedure can help to locate "ideal points"; one version of this is a so-called "center/periphery" interpretation.

The best way to implement these kinds of efforts is to write out the coordinates dataset produced by UCINET as an ASCII file (Data/Export; use "Raw" format and be sure to set the number of decimal places to at least 2 or 3) and analyze it together with variable/attribute information in your own friendly regression program from the statistical software package of your choice. One way to start is by opening the coordinates dataset in the spreadsheet viewer and simply cutting and pasting into the spreadsheet view for Stata or SPSS.

A different way to proceed in interpretation is to look for clusters or regions in the space in which certain types of nodes are concentrated. This could be done using cluster analysis routines (we'll introduce these in the next few weeks) after forming a dissimilarity matrix using the coordinates dataset (Tools/Dissimilarities could do this–be sure to indicate that you want dissimilarities between the rows). Here, I mention briefly a less formal approach.

The basic idea is to create a categorical variable, probably one with a small number of categories. One then "stains" the MD scaling solution by highlighting the points that fall within a category; a regional interpretation is indicated if the highlighted points appear close to one another. NetDraw can be used to accomplish this (by inputing attribute data), but you can also do this by hand (!) if the number of points in your network is reasonably small.

Using the Games relation of the Wiring data, conduct both a 2 dimensional and a three dimensional non-metric MDS. Remember: MDS does not like social isolates included in the analysis. To simplify matters, create a new Games dataset that excludes the isolates. For the two-dimensional analysis, use NetDraw and its MDS routine. For the three dimensional MDS, use UCINET and the longer procedure described above. For the visualization, you will use Mage. Mage is invoked by clicking on the  ***icon on the console toolbar. UCINET will then display the following dialogue box:***



You need to specify two datasets. The "Network dataset" is the dataset you created by removing the two social isolates in the Games relation. The second is the dataset that contains the three-dimensional coordinates that UCINET's MDS routine generated. Click OK. Then UCINET will tell you that it is about to create a new dataset; this is a Mage dataset and has the file extension ".kin." Click OK. UCINET will then automatically launch Mage and display your MDS.

Spend some time looking at the graph. You can "twirl" the graph around by placing your mouse on a node "ball", holding down the left mouse key, and then dragging.

The only way to output a Mage picture is to first have it create a PostScript file using the File/Write Postscript File… command in Mage. When the Save dialogue comes up, keep the default settings and click OK. To insert the picture into your write-up, use Word's Insert/Picture/From File… command.

Compare the two graphs. The Wiring data came from analysis of work done in a single room where there were wirers, solderers, and inspectors. The Games relation shows a tie if two people engaged in "horseplay" during the course of a day. Does adding the third dimension to this data help us to interpret the relationships between these workers? How similar are the plots that result?