Universidad Católica del Uruguay

Data Entry Lab

For this lab you will deal with data from two students who gathered data in the first week of an Undergraduate course in Social Networks and Organizations. **HOWEVER, AS WITH ALL LABS, IF YOU HAVE YOUR OWN DATA BY ALL MEANS USE IT HERE.** I am providing data, just for illustrative purposes (and to show the three most common forms of data entry you will likely use).

DATA

This lab uses two PDF files of surveys collected in the second week of an undergraduate social networks and organizations. The class was divided roughly in half and used each other as respondents. Both surveys are from the same half of the class, which had 17 students.

ThePerceptionOfHappiness

This survey asked one relational question – each student's perception of the happiness of the other students, in a simple yes/no format. In addition, the student asked three attribute questions of the student filling out the survey: Gender, Major/Concentration, and whether they were happy.

SchoolInteractionData

This survey asked one relational question about the respondent's level of interactions with students from each of Northeastern University's colleges, providing a rating scale ranging from 0 to 3. It also asked three attribute questions of the respondents: Major/Concentration (Field of Study), Gender, and Year in an open ended fashion.

EXERCISES:

1) Data Entry of Relational Data with **ThePerceptionOfHappiness**

   a) **In Excel:** Considering the nature of the relational data (i.e., perceptions of happiness) from this survey, determine what format we discussed (nodelist, edgelist, matrix and their variants) would be the most appropriate and efficient for capturing the data and enter the data into EXCEL in that format. (If you don't have Excel on your machine, just enter it directly into the DL editor in UCINET and skip step b.) Be sure to type the people's names exactly the same way every time you type them. UCINET cannot tell the different between "Andy" and "Andrew", or even – generally – between "Andy" and "andy" or "Andy " (e.g., where there is an extra space in one of those and different cases). Save the data in excel with whatever name you want so we can come back to it if we need to. (This is also an advantage of using EXCEL because you can share the data more easily and in a format more people can read this way.)

Universidad Católica del Uruguay

	A	B	C	D	E	F	G	H	I	J	K
1	Dan	Dan	Jacob	Olivia	Tim	Will					
2	Jacob	Majeed	Ziad								
3	Tim	Olivia									
4	Adi	Dan	Olivia	Ziad							
5	Olivia	Adi	Bella	Dan	Jaime	James	Marcella	Olivia	Siya	Tim	Will
6	Majeed	Majeed	Pranav								
7	Jaime	Adi	Bella	Jaime	James	Marcella	Olivia	Sofia	ziad		
8	Will	Dan	Olivia	Pranav	Tim	Will					
9	Pranav	Adi	Majeed	Will							
10											

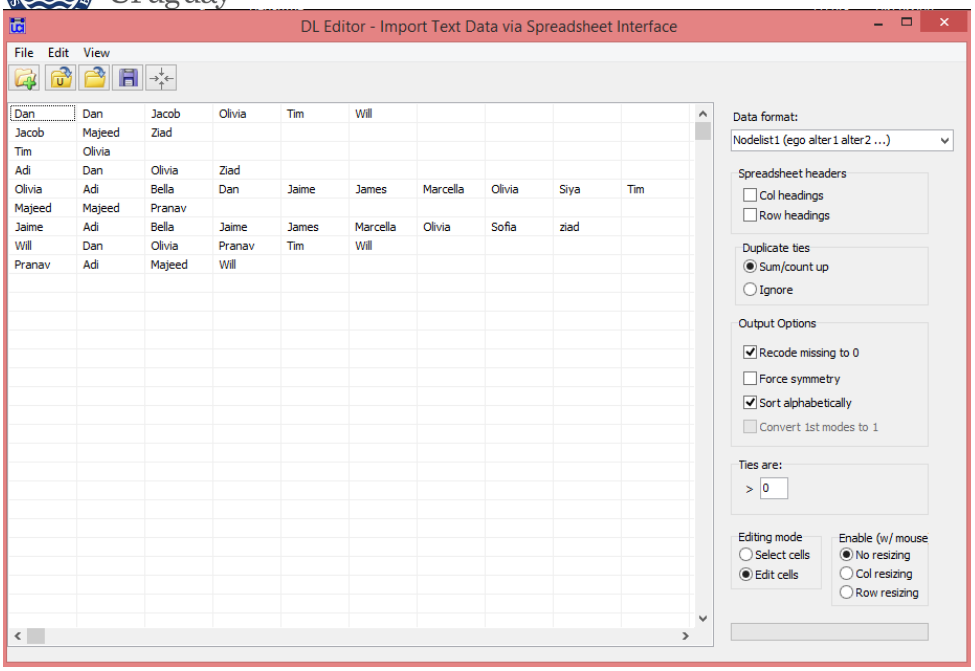
b) **In UCINET:** Select all your data in Excel and copy it to the clipboard using your preferred method (Edit | Copy, Control-C, Command-C on a Mac, etc.).

Open UCINET and open the DL Editor (in most recent versions, it is the fourth icon from the left, just to the left of the icon of the pencil, on the tool bar.) Paste the data (again, using your favorite method) into the DL Editor making sure that they get formatted into columns appropriately. (**MacNOTE:** If you are a Mac person, the keyboard shortcut for PASTE in UCINET uses the CONTROL key, **not** the Command key.) Very occasionally, the data will paste as single object in one cell, either all clumped text or a blue dot. Just delete that, click in the top left cell again and repaste the data and it should be fine.

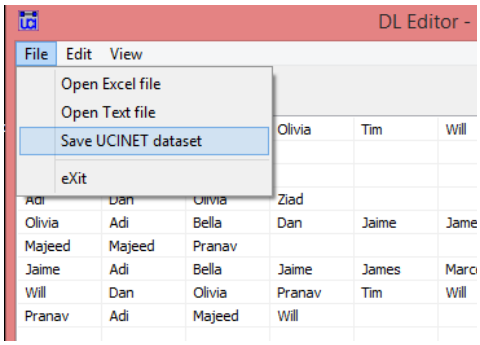
c) Now, on the right side of the DL Editor Window, select the appropriate format from the drop down list. Based on the format it should also default to certain choices regarding the presence or absence of column and row headers.

d) Now, go through each of the other options on the side of the dialog box and select the appropriate option for your data. In particular, for now, I want you to indicate that you DO want to recode missing data to zero. This is important because it means that if I did not specify a relationship, it will be coded as a zero, rather than missing data. That means, I do not have to make separate entries for all the relationships that people said DID NOT exist (which saves a lot of time).

Nodelist1 is the appropriate format because you have dichotomous (non-valued) data, as well as 1 Mode Data (Students by Students- not Students by Anything Else, which is 2 Mode Data). Nodelist1 also enables you to enter each ego *only once,* followed by the names of the alters listed by each ego (e.g., you do not have to make separate entries for each pair of relationships, which saves a lot of time).

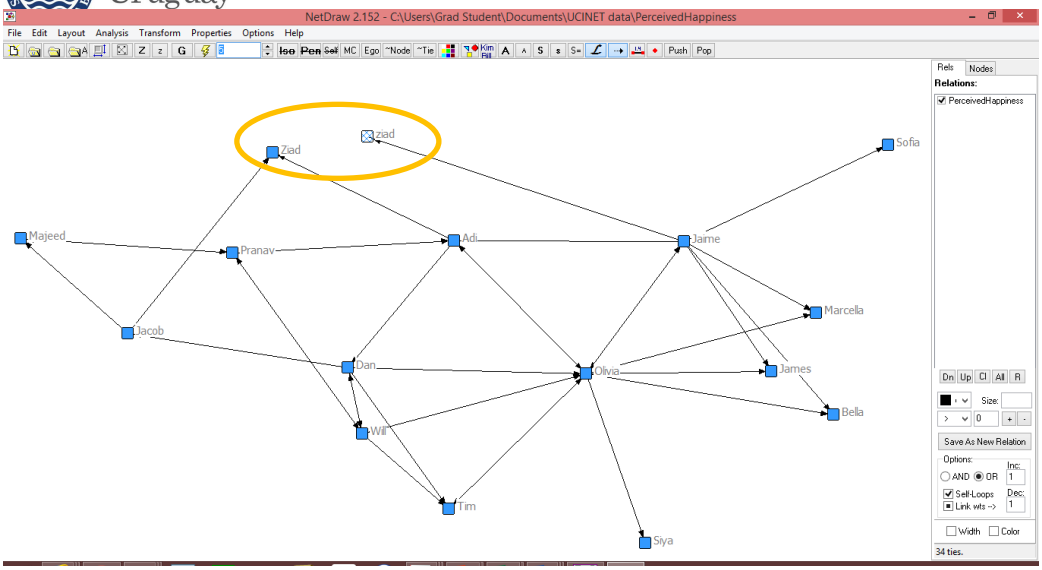


e) Save your data, by pressing on the disk icon at the top of the dialog box and giving the file the name **PerceivedHappiness.**



f) Congratulations, you have just made a UCINET dataset with Network Data. Close the DL Editor Window, and visualize your network in NetDraw. It is possible (even likely) that because of inconsistencies in the way names were entered, the diagram has some "duplicates" (e.g., you may see both "Andy" and "andy"). Track down any of these in the visualization, and go fix them in the Excel file, and repeat the steps above to copy & paste the data to the DL Editor, making sure your data replace the data that were there before saving the data on top of the previous files. You will have to Re-Open the file in NetDraw to see the effects of your changes in the visualization. Repeat this process until you have no duplicate names in your data.

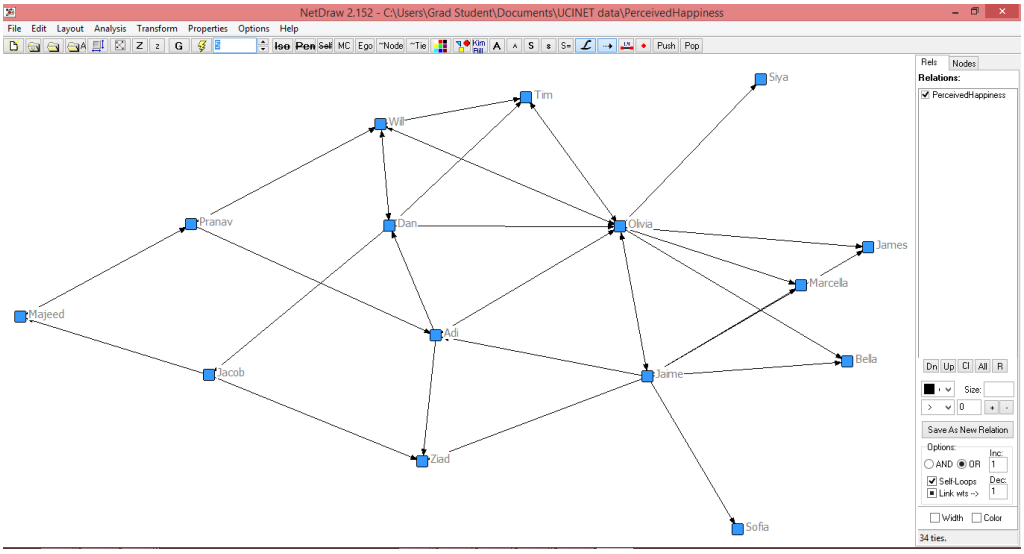
When opening PerceivedHappiness dataset in Netdraw, you may find that you created some "duplicates" due to inconsistencies in the way names were entered (e.g., you may see both "Ziad" and "ziad," as seen below).



Fix the inconsistencies in Excel (as shown below) for the indicated inconsistency/error:

	A	B	C	D	E	F	G	H	I	J	K
1	Dan	Dan	Jacob	Olivia	Tim	Will					
2	Jacob	Majeed	Ziad								
3	Tim	Olivia									
4	Adi	Dan	Olivia	Ziad							
5	Olivia	Adi	Bella	Dan	Jaime	James	Marcella	Olivia	Siya	Tim	Will
6	Majeed	Majeed	Pranav								
7	Jaime	Adi	Bella	Jaime	James	Marcella	Olivia	Sofia	Ziad		
8	Will	Dan	Olivia	Pranav	Tim	Will					
9	Pranav	Adi	Majeed	Will							

After fixing these inconsistencies, re-paste your data into the DL Editor, re-save the Data with the same file name (which will overwrite your file that contained duplicates/error, allowing you to avoid later confusion). Then visualize your corrected **PerceivedHappiness** dataset in Netdraw, as shown below.



Universidad Católica del Uruguay

2) Data Entry of Attribute Data with **ThePerceptionOfHappiness**

a) **In Excel:** Now, for each survey in the packet, we want to build attribute data. So, in Excel (or the DL Editor if you do not have excel), we are going to create a spreadsheet with the people's names (typed the same way as in Part 1) down the rows (but not the first row), and the attributes across the columns. So, the first row should contain labels, specifically "Name" in Column A, "Major" in Column B, "Gender" in Column C, and "Happy" in Column D. I would typically do this on another WORKSHEET in the same Excel Workbook.

b) Put the appropriate information for each person in each of the subsequent rows. The trick here is, other than labels (in this case, the people's names) which go in Column A, UCINET only deals with numbers. So, you have to build yourself a little code book to translate the values from text (e.g., "Male") to a numeric value (e.g. 2). I typically do this on a separate part of the worksheet (e.g., a few columns over from the actual data that I will cut and paste) in Excel (another advantage to using EXCEL for the first part of the data entry process). Save the data in Excel again with whatever name you will remember, or the same workbook if that's how you chose to do it.

	A	B	C	D
1	Name	Major	Gender	Happy
2	Dan	Marketing/Management	Male	Yes
3	Jacob	Management	Male	Yes
4	Tim	Management	Male	Yes
5	Adi	Management	Male	Yes
6	Olivia	MIS/Management	Female	Yes
7	Majeed	Marketing	Male	Yes
8	Jaime	Management	Male	Yes
9	Will	MIS/Management	Male	Yes
10	Pranav	Marketing/Management	Male	Yes
11				

Codebook - MAJOR	
1	Management
2	Marketing
3	MIS
4	Marketing/Management
5	MIS/Management

Codebook - GENDER	
1	Male
2	Female

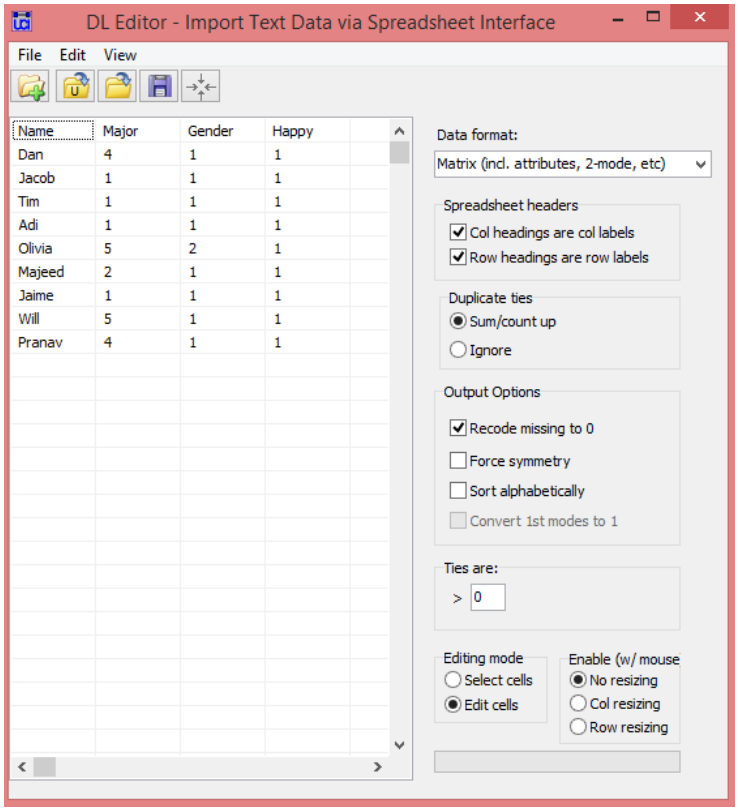
Codebook - HAPPY	
1	Yes
2	No

Recoded Data to be copied and pasted into the DL Editor in UCINET. We show both the text version (above) and the numeric version (below), but you do not have to build the text version at all. You can re-code as you enter if that is easier.

Universidad Católica del Uruguay

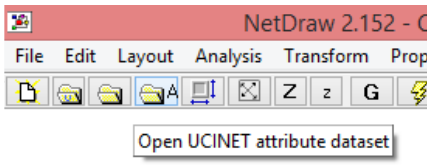
	A	B	C	D	E
1	Name	Major	Gender	Happy	
2	Dan	4	1	1	
3	Jacob	1	1	1	
4	Tim	1	1	1	
5	Adi	1	1	1	
6	Olivia	5	2	1	
7	Majeed	2	1	1	
8	Jaime	1	1	1	
9	Will	5	1	1	
10	Pranav	4	1	1	
11					

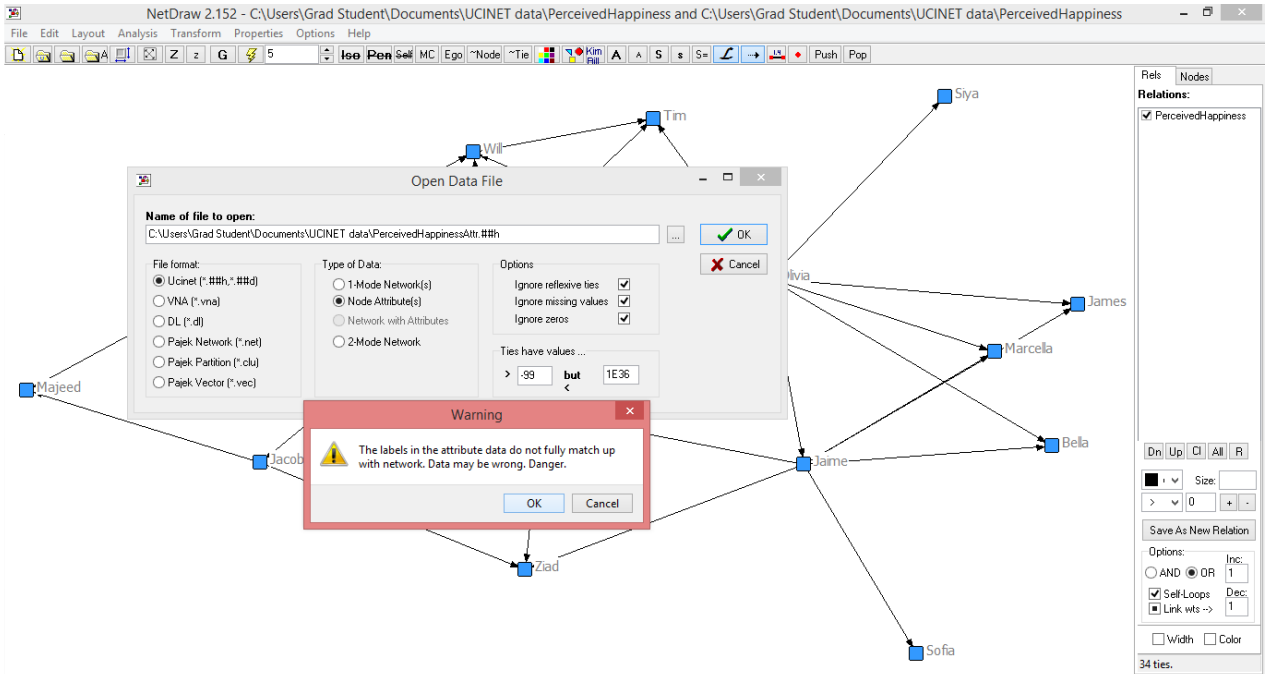
c) **In UCINET:** Once you have entered all the data into Excel (assuming you have), copy and paste it into the DL Editor the way you did in the previous section, being sure if you did create a code book that you do NOT include that in the data you copy and paste. In this case, the appropriate format is Matrix, and you should have row and column headings checked. (That should be the default when you pick Matrix format.) Make sure all the other options are correct, too. Then use the disk icon to save the data and give it the name **HappinessAttr**.

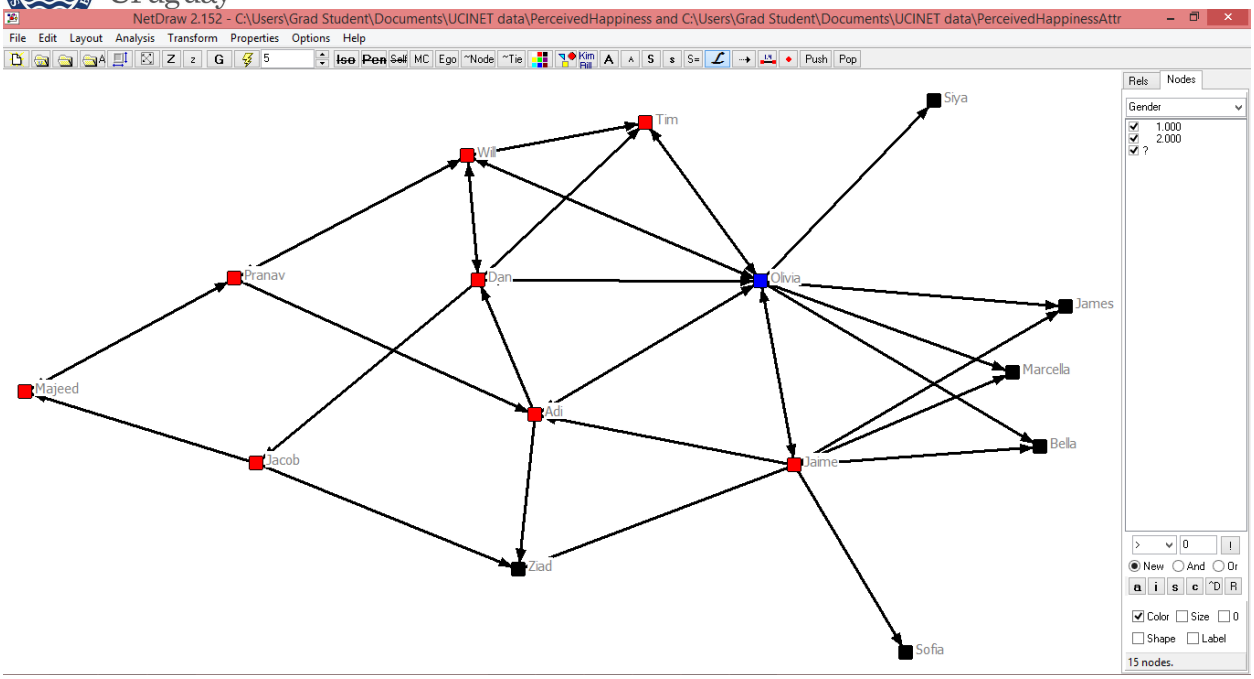


d) **In NetDraw:** If you closed NetDraw before, re-open the **PerceivedHappineness** network to visualize it. Now, using the folder with the A next to it, open the file you just created

HappinessAttr as an attribute file. You should get a message saying "The labels in the attribute data do not fully match up with network. Data may be wrong. Danger." This is because some people that were mentioned by others in the survey did not fill out the survey themselves. We will fix that in the next section. For now, just color the nodes by Gender using the control region, the color palette, or the properties menu.

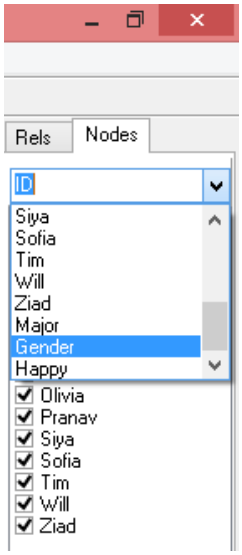


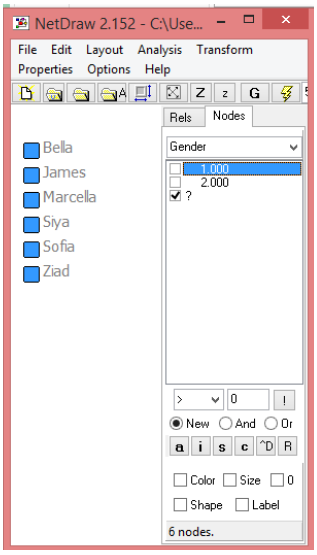




3) Dealing with missing attribute data.

a) **In NetDraw:** From the visualization produced in the last step, you should be able to tell which nodes do not have attribute data, because they will be a different color than those specified by the Gender attribute for men and women. (In the control region in NetDraw, on the Nodes tab, they should be indicated by the value "?". You can "Uncheck" the boxes next to the values you specified for male and female, and ONLY the nodes that are not in the attribute file should be left. (These nodes got into the network by being nominated by someone who did fill out the survey, even though they did not.) Make a list of these nodes.

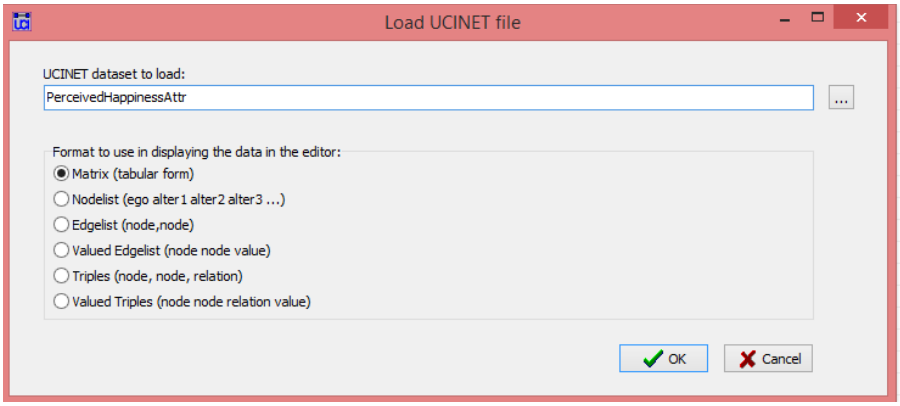




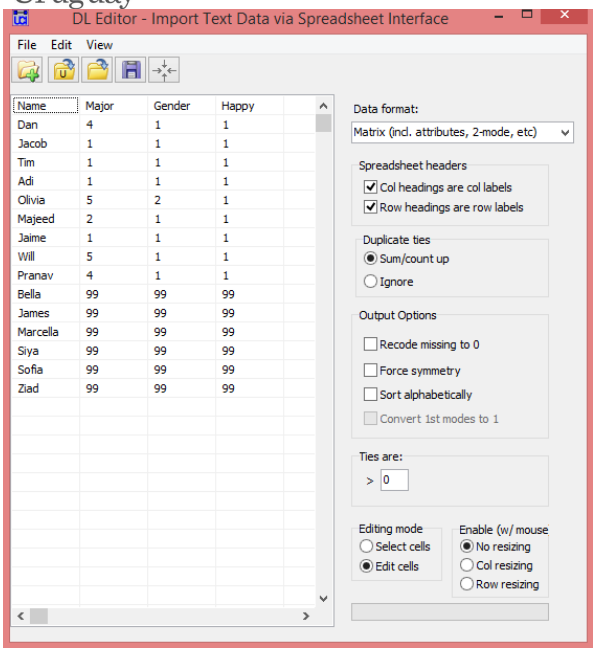
Nodes without attribute data (because they were nominated by someone who did fill out the survey, even though they did not fill out the survey):

b) In UCINET: If you are managing your data in Excel, you would normally want to go update the data there and copy and paste it again to create a new file. However, for the purposes of this lab, we are going to open the attribute file DIRECTLY in the DL Editor. To do this, open the DL editor. It may still have the data that you just saved. If it does, click on the folder with the green plus sign (+) to clear it out. (Although we want to work with this data, I want to have you open the UCINET dataset to see that process.)

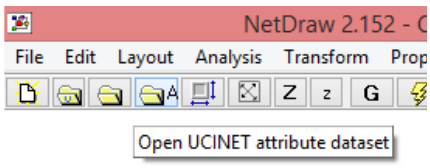
Now, click on the file folder with the "U" on it. Put in the filename "HappinessAttr" and say that you want to see the data in "Matrix (tabular form)" format. It should present the data very much like what it looked like before you saved it. (And maybe what it looked like when you opened this window again, before we cleared it out so we could use the open button.)

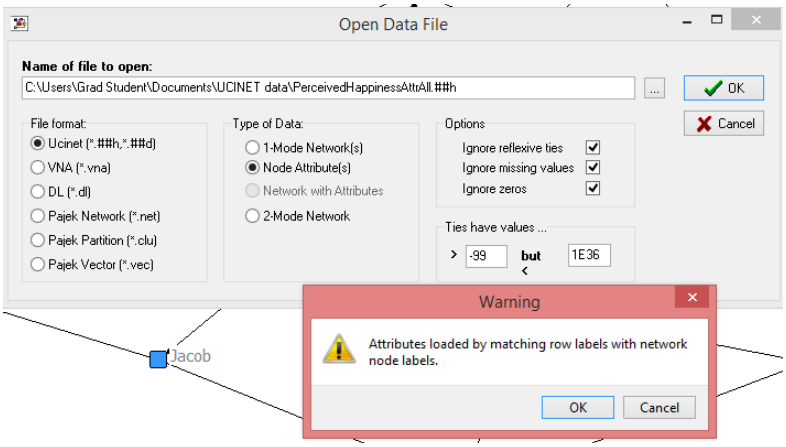


Now, starting in the row below all the data from the file, for each node in the list you made in step 3a, Enter the name in Column A exactly as it appeared in the network file in NetDraw. For each of their attributes use the value 99 to indicate missing data for the three attributes (assuming you did not use a value of 99 as a valid code for any of them, if so use a different value). Now, being sure the box that says "Sort alphabetically" is NOT checked (I am, again, demonstrating something), save the data with the name **HappinessAttrAll**.



c) **In NetDraw:** Now, load this new attribute file by clicking on the folder with the A next to it and typing in **HappinessAttrAll** in the filename box. You should get a different message from NetDraw that says "Attributes loaded by matching row labels with network node labels." This means, it found attribute data for all the nodes in the network, but they were NOT in the same order in the attribute file as in the Network file. NetDraw did the work of matching the labels to load attribute data for each node and it will display correctly. **HOWEVER,** UCINET does not do this. When we use attribute data in UCIINET, the nodes MUST be in the same order in both the network and the attribute file. Inside UCINET, you can use either Data | Sort alphabetically or Data | Match datasets | Match 1-mode net w/attrib data to create files that will match. (**NOTE:** The notation "Data | Match datasets | Match 1-mode net w/attrib data" is a convention for, starting on the Data menu, find the option Match datasets, which then has a submenu, on which you look for Match 1-mode net w/attrib data. The | delineates between levels in the windows menu structure.)





4) Dealing with missing response data

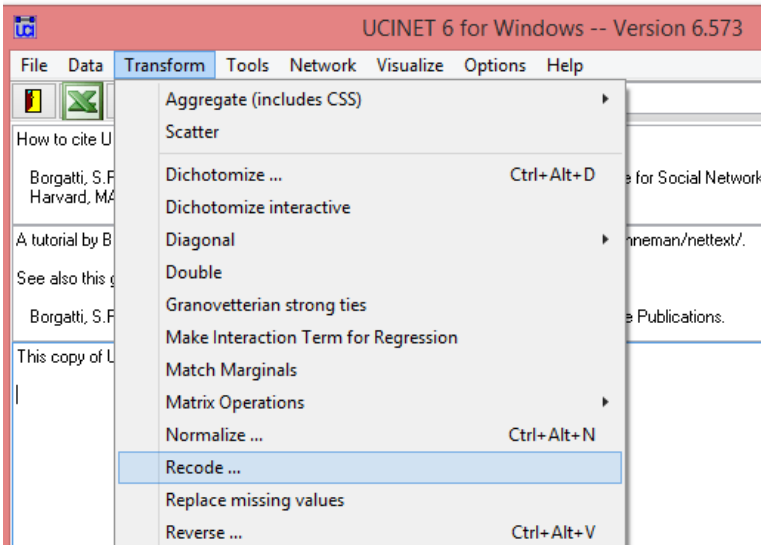
a) **In UCINET:** Revisiting the list you made in step 3a, that is a list of people whose survey response we do not have. But, presently, the data looks as if they said that they did not perceive any of their classmates to be happy (that is, they said "0" for everyone). To see what I mean, go back to UCINET and press the big D on the icon bar, and display the file "Perceived Happiness". This will present a matrix with ones and zeroes indicating if the person in that ROW said they perceived the person indicated by the column to be happy. But, if you look across the row for Marcella, you should see all zeroes even though you did not have any data to enter for Marcella.

		1 Adi	2 Bell a	3 Dan	4 Jaco b	5 Jaim e	6 Jame s	7 Maje ed	8 Marc ella	9 Olivia	10 Pran av	11 Siya	12 Sofi a	13 Tim	14 Will	15 Ziad
1	Adi	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
2	Bella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	Dan	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0
4	Jacob	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
5	Jaime	1	1	0	0	1	1	0	1	1	0	0	1	0	0	1
6	James	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	Majeed	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
8	Marcella	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
9	Olivia	1	1	1	0	1	1	0	1	1	0	1	0	1	1	0
10	Pranav	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0
11	Siya	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	Sofia	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
13	Tim	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
14	Will	0	0	1	0	0	0	0	0	1	1	0	0	1	1	0
15	Ziad	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

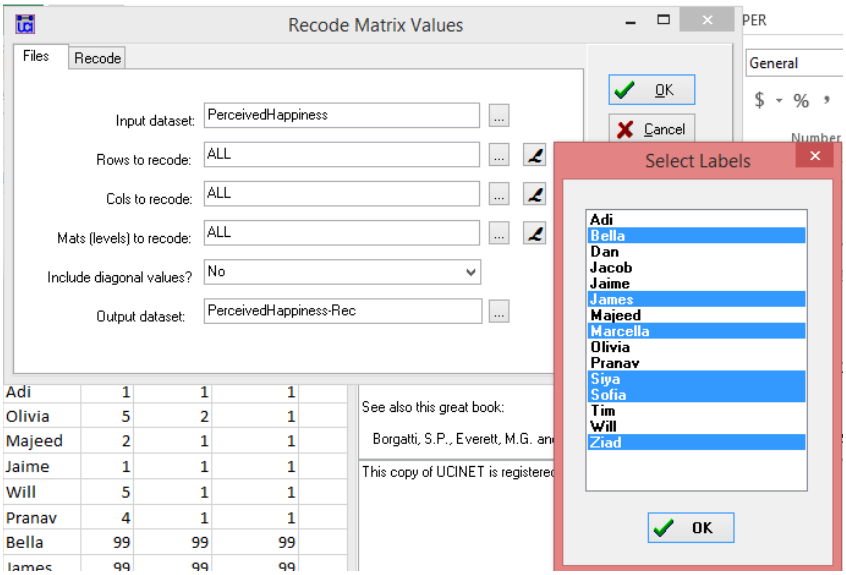
15 rows, 15 columns, 1 levels.

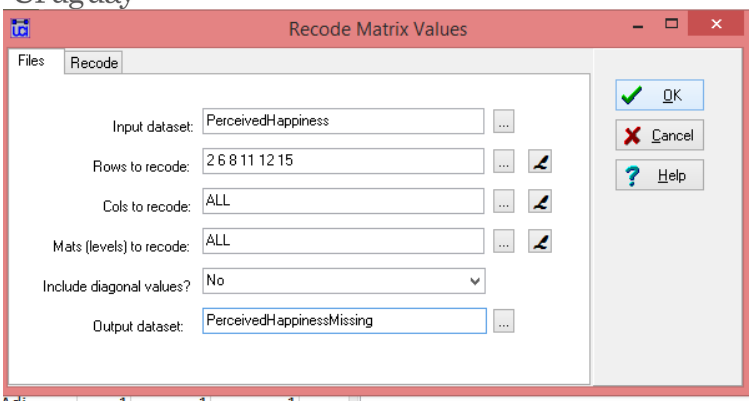
That happened because we specified that we wanted to "Recode missing to 0". This was actually good, because all the zeroes on rows for the people we did have surveys from came from this. But, these rows are technically misrepresenting the data and should be marked as missing data.

b) So, let's fix that using Transform| Recode in UCINET.

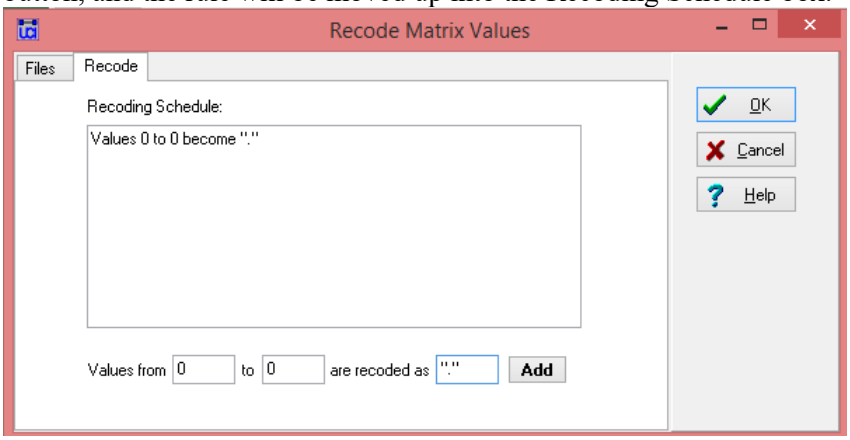


When you select this option, it brings up a dialog box with two tabs. It starts on the Files tab, and asks for the input file. In this case we are going to record the network **PerceivedHappiness**, so put that into the Input dataset box. Then, pressing the script L after the "Rows to recode" entry box, it will bring up the labels from your network. Select the ones from the list you made in step 3a (using Ctrl-Click so you can select multiple items) and press okay. It will translate this to the row numbers for those actors. The Columns to recode should remain "ALL" because we want to recode all those zeroes to missing data. The output file, by default will be PerceivedHappiness-Rec (the "–Rec" suffix added to indicate it is the output of the Recode routine). Change this to say **PerceivedHappinessMissing**.





c) Now click on the Recode tab, and on the bottom, put in a rule to recode values from 0 to 0 to be missing data. (Use a period, ".", to specify missing data in that box.) Then click the ADD button, and the rule will be moved up into the Recoding Schedule box.



d) Click okay, and it will display the new matrix, and for those rows, there should be spaces (indicating missing data) for the entries in those rows.

Universidad Católica del Uruguay

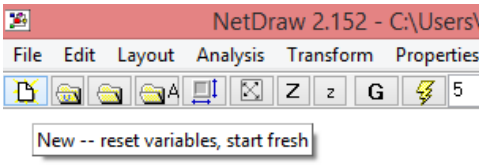
```
Recoding Schedule:

Values 0 to 0 become "."
```

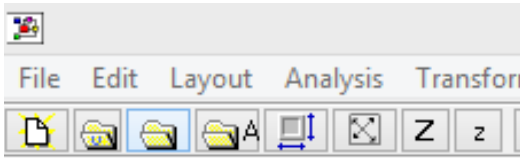
		1	1	1	1	1
		1	2	3	4	5	6	7	8	9	0	1	2	3	4	5
		A	B	D	J	J	J	M	M	O	P	S	S	T	W	Z
		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1	Adi	0	0	1	0	0	0	0	0	1	0	0	0	0	0	1
2	Bella		0													
3	Dan	0	0	1	1	0	0	0	0	1	0	0	0	1	1	0
4	Jacob	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1
5	Jaime	1	1	0	0	1	1	0	1	1	0	0	1	0	0	1
6	James					0										
7	Majeed	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
8	Marcella								0							
9	Olivia	1	1	1	0	1	1	0	1	1	0	1	0	1	1	0
10	Pranav	1	0	0	0	0	0	1	0	0	0	0	0	0	1	0
11	Siya											0				
12	Sofia												0			
13	Tim	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
14	Will	0	0	1	0	0	0	0	0	1	1	0	0	1	1	0
15	Ziad															0

e) In NetDraw: If you have a diagram open in NetDraw, click on the blank page icon (the first icon on the icon bar) to clear the image. Now open the network you just made, **PerceivedHappinessMissing**. The visualization should look exactly the same as when you visualized **PerceivedHappiness**, because NetDraw does not visualize missing data any differently than absence of tie data. But, UCINET will know the difference in calculating certain measures.

Click on the blank page icon:



Now open the network you just made, **PerceivedHappinessMissing:**



Universidad Católica del Uruguay