Center for Public Health Systems Science
GEORGE WARREN BROWN SCHOOL OF SOCIAL WORK
Washington University in St. Louis

Network Analysis from Start to Finish: Techniques, Tools, and Tips for Evaluating Your Network

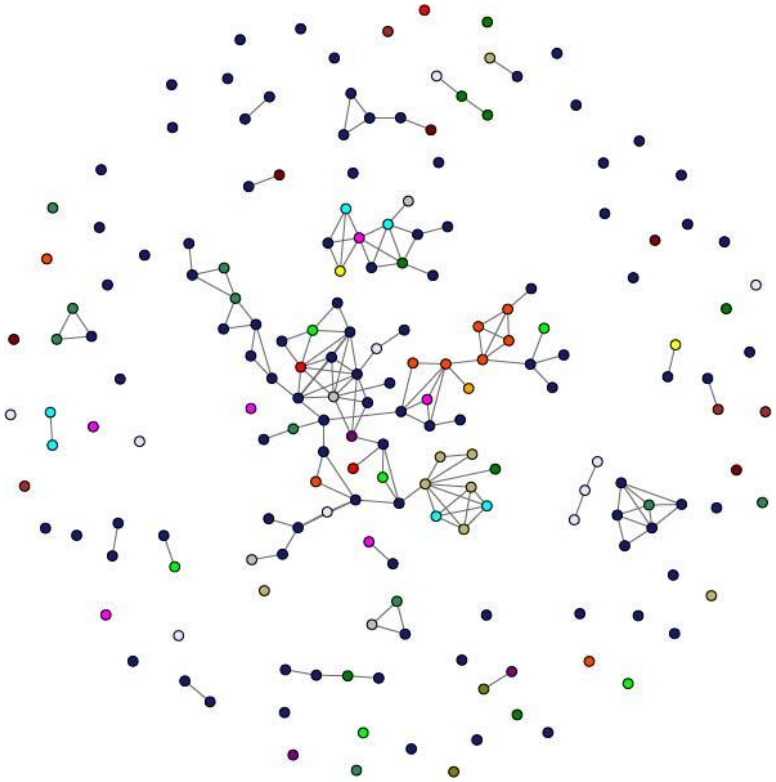
Bobbi J. Carothers

American Evaluation Association
Denver, Colorado
10/17/2014

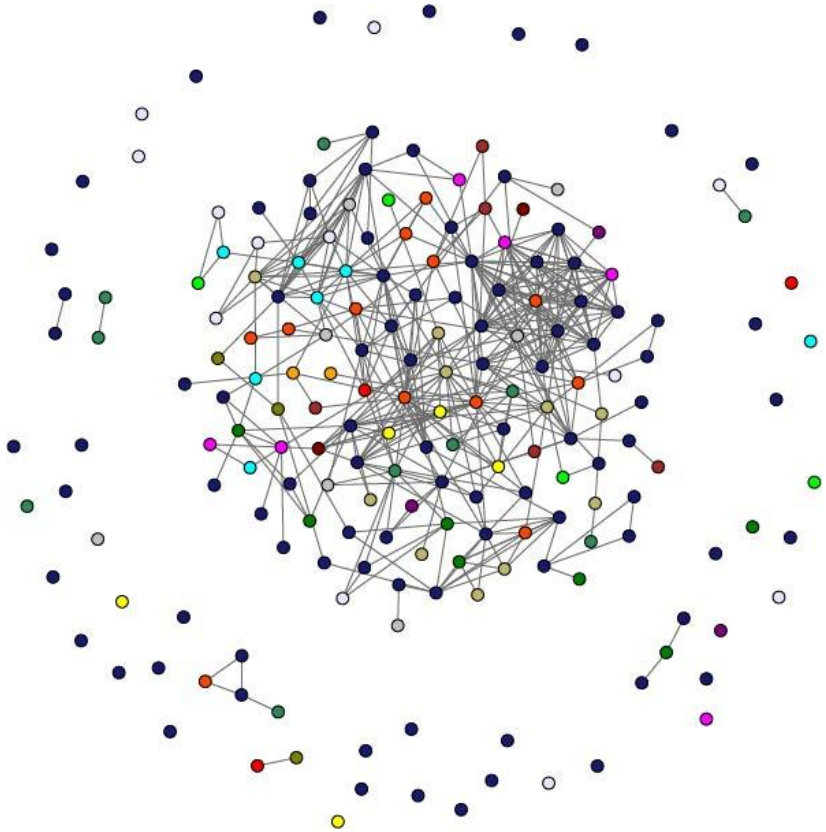
Example Graphic: What Is the Story?

- Grant submission collaborations
- Systems change over time

2007



2010



Example Statistics: What Is the Story?

- Numbers describe & confirm patterns in visualizations
  - Increase in density over time
  - Increase in cross-disciplinary collaboration over time

Year	Size	Density	Ave. Degree	Modularity	∆ Modularity
2007	186	.009	1.65	.140	
2010	193	.023	4.41	.054	- 61%

Steps to a Successful Network Analysis

1. Decide who is in the network
2. Decide on network measurements
3. Collect your data
4. Manage your data
5. Analyze your data
6. Visualize your data

Step 1: Network Boundary – Who is in the network?

Population vs. Sample

- Population of interest
  - All of the actors who really are part of the network
  - Examples
    - 9th grade students at Clayton High School
    - School of Social Work Faculty
- "Sample"
  - Including key actors is more important than size of the network
  - Shoot for at least 70% of possible respondents

Who to Include?

- Be guided by the relationships you want to measure
- Laumann criteria
  - Positional: formal membership
  - Reputational: knowledgeable person names members
  - Event: participation in activity of interest
  - Relational: contact with others in the network

Laumann, Mardsen, & Prensky (1983). Boundary specification problem in network analysis.

Reputational Example

Please list up to 10 individuals who work in Los Angeles County on *tobacco control* policy and advocacy. Please also indicate one or two of the people you who you would consider leaders in tobacco control policy and who are familiar with the work that others are doing in Los Angeles County. We will contact those leaders to learn about additional partners.

First Name	Last Name	Organization Name	Email (optional)	Leader?
				❑
				❑
				❑
				❑
				❑
				❑
				❑

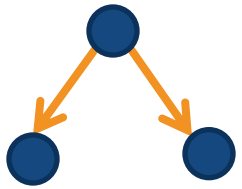
Unit of Analysis

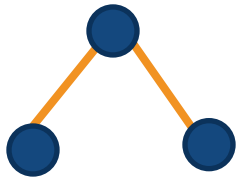
- Links between individual people
- Links between organizations/groups
  - Survey? Still need to talk to an individual to represent the larger group
  - Can survey a few individuals from each group and aggregate responses during data management
  - Consider how to phrase questions
    - How closely does *your organization* work with other organizations?
    - How closely do *you* work with other organizations?

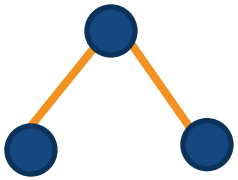
Step 2: Network Measurements – What relationships are you interested in?

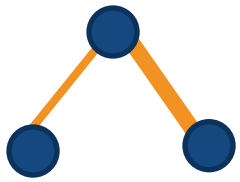
Characteristics of Ties

- Direction
  - Directed “Arcs”
    - A tie goes *from* one node *to* another
    - Patient referrals, flow of money, importance
  - Non-directed “Edges”
    - Inherently reciprocal
    - Co-authorship, collaboration
- Scale
  - Binary (dichotomous)
    - A tie is either present or absent (1, 0)
    - Awareness, friendship
  - Valued
    - The strength of a relationship can be rated on a scale
    - Level of collaboration, amount of contact









Awareness Example

- Are you aware of the following individuals' work in [area of interest]?

	Yes	No
John Smith	○	○
Tom Parker	○	○
Tina Jones	○	○
Bill James	○	○
Fred Myer	○	○
Etc…		

  - Is this directional or non-directional?
  - Is this binary or valued?
- Use as a filter for subsequent questions

Contact Example

- On average, how often have you had direct contact (e.g., meetings, phone calls, emails, faxes, or letters) with each of the following partners within the past year? (Do not count listservs or mass emails)

	No Contact	Yearly	Quarterly	Monthly	Weekly	Daily
Partner 1	○	○	○	○	○	○
Partner 2	○	○	○	○	○	○
Etc…	○	○	○	○	○	○

  - Directional or non-directional?
  - Binary or valued?
- How could you use this as a screener?

Activity Example

- What types of activities have you worked with each of your partners on [topic of interest during time frame of interest]? (Check all that apply.)

	Activity 1	Activity 2	Activity 3
Partner 1	☐	☐	☐
Partner 2	☐	☐	☐
Etc…	☐	☐	☐

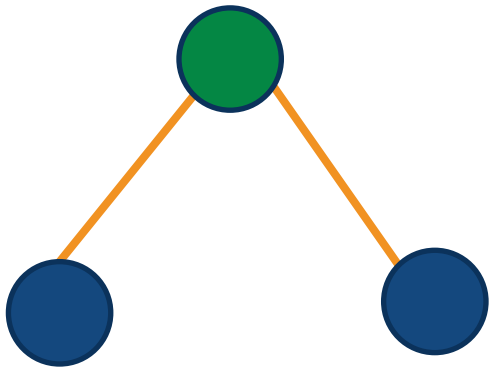
  - Directional or non-directional?
  - Binary or valued?
- Multiplex relationship

Other Possible Relationships

- Publication co-authorship
- Level of collaboration
- Flow of resources (money, information)
- Satisfaction with communication, collaboration, mentoring, etc.
- Barriers experienced with partners
- Dissemination
- Whatever people/organizations are *doing together*

Characteristics of Network Partners

- “Attributes”
- Can be collected with standard survey questions
- Displayed as different colors or shapes
- Gender, discipline, rank, socioeconomic status, etc.



Step 3: Data Collection – How can you obtain information about relationships?

Archival Possibilities

- Anything that links people directly or through a “mode”
- Social media
  - Facebook
  - Twitter
  - LinkedIn
- Institutional records
  - Grant submissions
  - Journal co-authorship (Scopus)
  - IRB applications
  - Classroom rosters

Online Survey

- Network-specific tools
  - Network Genie (https://secure.networkgenie.com/)
  - ONASurveys (https://www.s2.onasurveys.com/)
  - Partner Tool (http://www.partnertool.net/)
  - OpenEddi (coming soon!)
- General online survey platforms
  - Anything that allows display logic and text piped in from responses will work
  - SurveyMonkey (paid)
  - REDCap
  - Qualtrics

Network Survey Considerations

- Network questions ask participants to answer about their relationships with *all* of the partners they are linked to in the network
- If the network has 50 other partners…
  - Answering the same question 50 times
  - 4 network questions = 200 answers
- Keep size of network in mind when developing surveys

	Activity 1	Activity 2	Activity 3
Partner 1	☐	☐	☐
Partner 2	☐	☐	☐
Etc…	☐	☐	☐

Format: Free Recall

- Start with 1 or 2 *name generator* questions asking participants to list who they are connected to or aware of in the network
- Use the piped text feature of the online survey tool to display participant-generated names in subsequent network questions
- Benefits
  - Can “snowball” participants beyond original delineation
- Drawbacks
  - Cleaning creative spelling
  - Participants may be uncomfortable/unwilling to name partners
  - Recalling names → high participant burden
  - Contacting snowballed names → high researcher burden

Free Recall Examples

- Please identify up to __ people who you think are the *most important to [area of interest].*
- Please identify up to __ people who you have had the *most contact with* (*e.g.*, meetings, phone calls, faxes, letters, text/instant messages, or emails) regarding *[area of interest during timeframe of interest].*
- Please identify up to __ people who you have exchanged ideas or materials with most often regarding *[area of interest during timeframe of interest].*
- (In order for your information to be useful, you must include the names of individual people in the spaces for First and Last Name. Please include only one name per space.)

Free Recall Tips

- Can pipe in names from one name generator to be selected in a second

Select Previous Partners	And/Or Enter New Partners		
[Drop-down lists populated w/ text from previous generator]	First Name	Last Name	Organization Name
1.			
2.			
Etc…			

- Separate fields for first, last, and organization name → aids in data cleaning
- Consider *optional* field for contact email
- Consider linking to list of possible partners, if available → aids recall & reduces creative spelling

Format: Roster

- Present participant with a full list of network partners to answer about
- Benefits
  - Easy to clean & manage data
  - Easier for participants to recognize names than to recall them
- Drawbacks
  - Not feasible with very large networks
  - Comprehensive delineation essential

	Yes	No
John Smith		
Tom Parker		
Etc…		

Roster Tips

- Start with a screening question to filter out non-connected partners in later questions (online survey display logic)
- Order of names on roster questions = order of participant IDs
  - Data will export in an N x N matrix
  - Aids in later data management

	Yes	No
John Smith		
Tom Parker		
Etc…		

	John Smith	Tom Parker
John Smith		
Tom Parker		

Step 4: Data Management – How do you get network analysis programs to read your data?

Free recall vs. Roster formats

Data Management Goal

- Most network analysis programs can read files derived from an
  - Arc list

or

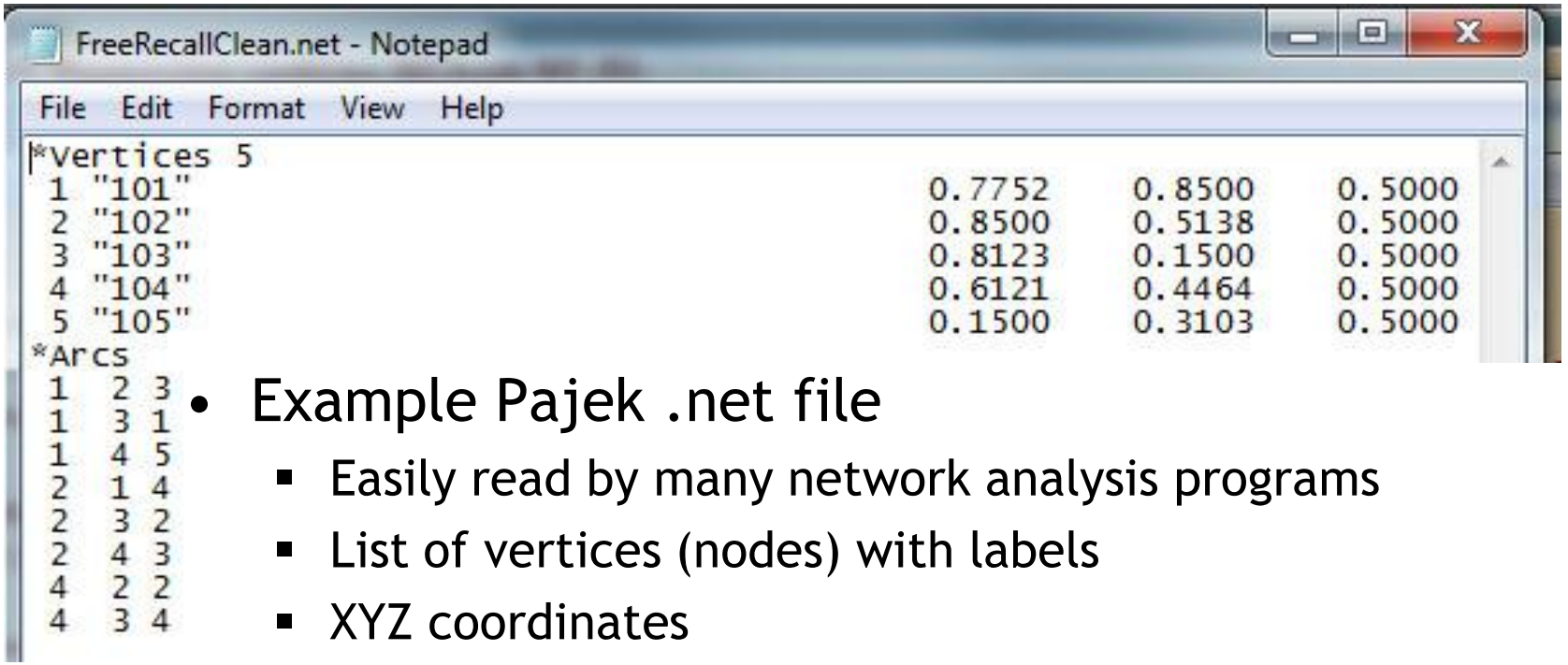
From	To	Value
John Smith	Tom Parker	3
John Smith	Tina Jones	5
Tom Parker	John Smith	4
Tina Jones	Tom Parker	2

  - N X N matrix

(gets converted to an arc list)

	John Smith	Tom Parker	Tina Jones
John Smith		3	5
Tom Parker	4		
Tina Jones		2	

Result to Aim For



- Example Pajek .net file
  - Easily read by many network analysis programs
  - List of vertices (nodes) with labels
  - XYZ coordinates
  - List of arcs (directional) or edges (non-directional)
    - From
    - To
    - Value (if applicable)

Handy Tools

- Pajek (pronounced “pie-yack,” Slovene for “spider”)
  - Network analysis software
  - Useful for fine-tuning network data & performing analyses
  - http://pajek.imfm.si/doku.php?id=pajek
  - Free!
- txt2pajek
  - Turns arc lists into Pajek .net files
  - http://www.pfeffer.at/txt2pajek/
  - Free!
- UCINet
  - Network analysis software, useful for converting matrix files to .net files, sorting .net files
  - https://sites.google.com/site/ucinetsoftware/home
  - Students: $40, Faculty & Government: $150, Others: $250
- Excel, SPSS/SAS/Stata

Data Management Tips

- Convert partner names to numeric IDs with a uniform number of digits
  - 101, 102, 103, etc.
  - Some programs don't recognize leading zeros (001, 002)
  - Some programs will otherwise sort like this: 1, 10, 11, 2, 21, 22… etc.
  - Different programs may not sort text strings consistently due to different handling of spaces and capitalizations
- Important to match order of network data with order of attribute data

Free Recall Data: Raw Format

- Data will look something like this:

ID	Name	AwareFirst1	AwareLast1	AwareFirst2	AwareLast2	AwareFirst3	AwareLast3	Con1	Con2	Con3
101.00	Smith, John	Thomas	Parker	Tina	Jones	William	James	3.00	5.00	1.00
102.00	Parker, Tom	bill	james	jon	smith	tina	jomes	2.00	4.00	3.00
104.00	Jones, Tina	Bill	James	Tom	Parker			4.00	2.00	.
105.00	Meyer, Fred							.	.	.

- Elements
  - Participant ID and Name, sorted by ID
  - First and last names of people participants listed in awareness name generator
  - Value for the level of contact for each partner
  - Some participants may not have nominated partners
- Strategy: create an arc list that can be converted to a .net file by txt2pajek

Free Recall Data: Transformation

- Convert to a rough arc list
  - Single columns for
    - Fist name
    - Last name
    - Contact value
  - Commands
    - SPSS: varstocases
    - SAS: proc transpose?
    - Stata: reshape long
  - Be sure to retain cases even when partner information is blank (isolate)
  - Sort by last name of nominated partners

ID	Name	ConFirst	ConLast	ConVal
104.00	Jones, Tina			.
105.00	Meyer, Fred			.
105.00	Meyer, Fred			.
105.00	Meyer, Fred			.
102.00	Parker, Tom	tina	jomes	3.00
102.00	Parker, Tom	bill	james	2.00
101.00	Smith, John	William	James	1.00
104.00	Jones, Tina	Bill	James	4.00
101.00	Smith, John	Tina	Jones	5.00
101.00	Smith, John	Thomas	Parker	3.00
104.00	Jones, Tina	Tom	Parker	2.00
102.00	Parker, Tom	jon	smith	4.00

Free Recall Data: Clean, Clean, Clean

- Clean nominated partner names so they are consistent
  - Concatenate last and first names, trimming extra spaces on the left and right
  - Fix creative spellings and capitalizations (recode)

ID	Name	ConFirst	ConLast	ConVal	Partner	PartnerClean
104.00	Jones, Tina			.	,	null
105.00	Meyer, Fred			.	,	null
105.00	Meyer, Fred			.	,	null
105.00	Meyer, Fred			.	,	null
102.00	Parker, Tom	bill	james	2.00	james, bill	James, Bill
104.00	Jones, Tina	Bill	James	4.00	James, Bill	James, Bill
101.00	Smith, John	William	James	1.00	James, William	James, Bill
102.00	Parker, Tom	tina	jomes	3.00	jomes, tina	Jones, Tina
101.00	Smith, John	Tina	Jones	5.00	Jones, Tina	Jones, Tina
101.00	Smith, John	Thomas	Parker	3.00	Parker, Thomas	Parker, Tom
104.00	Jones, Tina	Tom	Parker	2.00	Parker, Tom	Parker, Tom
102.00	Parker, Tom	jon	smith	4.00	smith, jon	Smith, John

Free Recall Data: ID Numbers

- Assign an ID number to partner names (recode)
  - Match w/ original ID if a participant or part of original delineation
  - Create new ID if not part of original delineation and you want to snowball
  - Add ID for null node

ID	Name	ConFirst	ConLast	ConVal	Partner	PartnerClean	PartnerID
104.00	Jones, Tina			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
102.00	Parker, Tom	bill	james	2.00	james, bill	James, Bill	103.00
104.00	Jones, Tina	Bill	James	4.00	James, Bill	James, Bill	103.00
101.00	Smith, John	William	James	1.00	James, William	James, Bill	103.00
102.00	Parker, Tom	tina	jomes	3.00	jomes, tina	Jones, Tina	104.00
101.00	Smith, John	Tina	Jones	5.00	Jones, Tina	Jones, Tina	104.00
101.00	Smith, John	Thomas	Parker	3.00	Parker, Thomas	Parker, Tom	102.00
104.00	Jones, Tina	Tom	Parker	2.00	Parker, Tom	Parker, Tom	102.00
102.00	Parker, Tom	jon	smith	4.00	smith, jon	Smith, John	101.00

Free Recall Data: Attribute File

- Goal: standard data file with node characteristics of original and snowballed partners
- Copy out a new file
- Transform
  - (varstocases/proc transpose/reshape long)
  - ID & Partner ID → Label (use in Gephi later)
  - Name & PartnerClean → Name
  - Drop null
- Sort by Label & remove duplicates
- Bring in attribute data later on

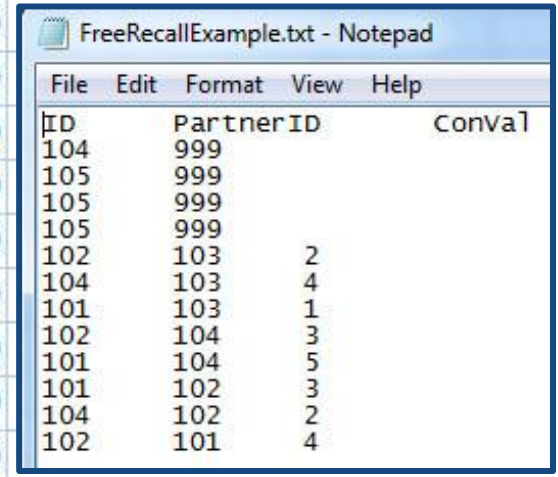
Label	Name
101.00	Smith, John
101.00	Smith, John
101.00	Smith, John
101.00	Smith, John
102.00	Parker, Tom
102.00	Parker, Tom
102.00	Parker, Tom
102.00	Parker, Tom
102.00	Parker, Tom
103.00	James, Bill
103.00	James, Bill
103.00	James, Bill
104.00	Jones, Tina
104.00	Jones, Tina
104.00	Jones, Tina
104.00	Jones, Tina
104.00	Jones, Tina
105.00	Meyer, Fred
105.00	Meyer, Fred
105.00	Meyer, Fred

Label	Name	Gender
101.00	Smith, John	Male
102.00	Parker, Tom	Male
103.00	James, Bill	Male
104.00	Jones, Tina	Female
105.00	Meyer, Fred	Male

Free Recall Data: Arc List

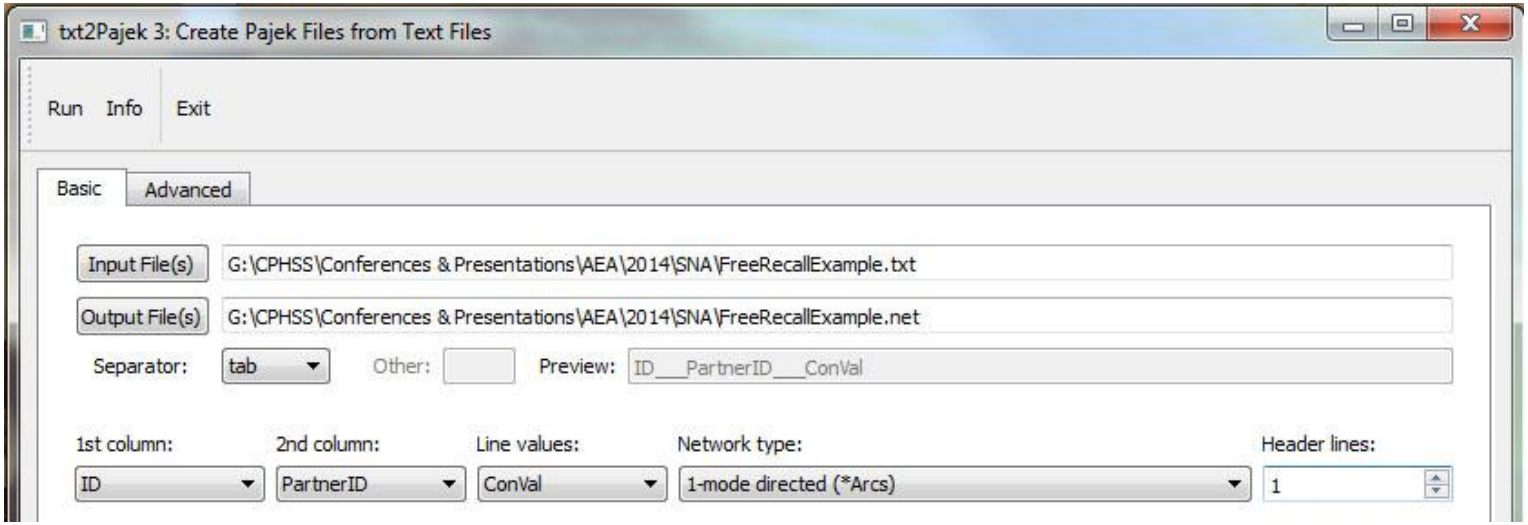
- Back to cleaned network data
- Save out as tab-delimited text file
  - Keep ID, PartnerID, and value only
  - Variable order is important
- Looks like lower part of Pajek .net file

ID	Name	ConFirst	ConLast	ConVal	Partner	PartnerClean	PartnerID
104.00	Jones, Tina			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
105.00	Meyer, Fred			.	,	null	999.00
102.00	Parker, Tom	bill	james	2.00	james, bill	James, Bill	103.00
104.00	Jones, Tina	Bill	James	4.00	James, Bill	James, Bill	103.00
101.00	Smith, John	William	James	1.00	James, William	James, Bill	103.00
102.00	Parker, Tom	tina	jomes	3.00	jomes, tina	Jones, Tina	104.00
101.00	Smith, John	Tina	Jones	5.00	Jones, Tina	Jones, Tina	104.00
101.00	Smith, John	Thomas	Parker	3.00	Parker, Thomas	Parker, Tom	102.00
104.00	Jones, Tina	Tom	Parker	2.00	Parker, Tom	Parker, Tom	102.00
102.00	Parker, Tom	jon	smith	4.00	smith, jon	Smith, John	101.00



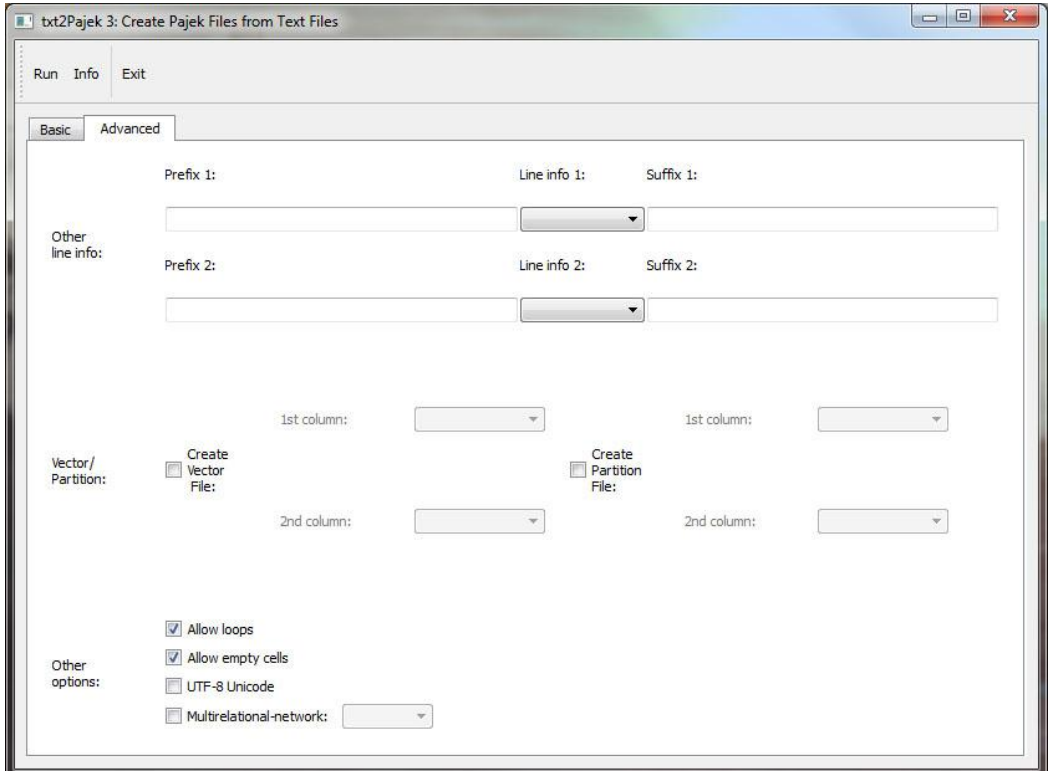
Free Recall Data: Convert to Pajek

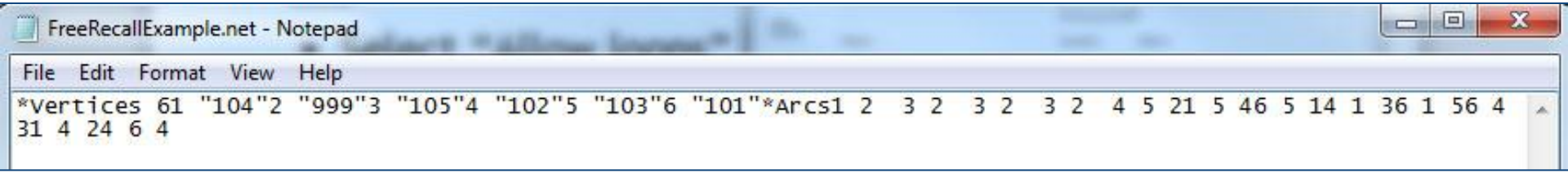
- txt2Pajek Basic tab
  - Select text file
  - Select appropriate separator (tab), 1st column (ID), 2nd column (PartnerID)
  - If network is valued, select appropriate column
  - Network type: 1 mode directed
  - Header lines: 1



Free Recall Data: Convert to Pajek

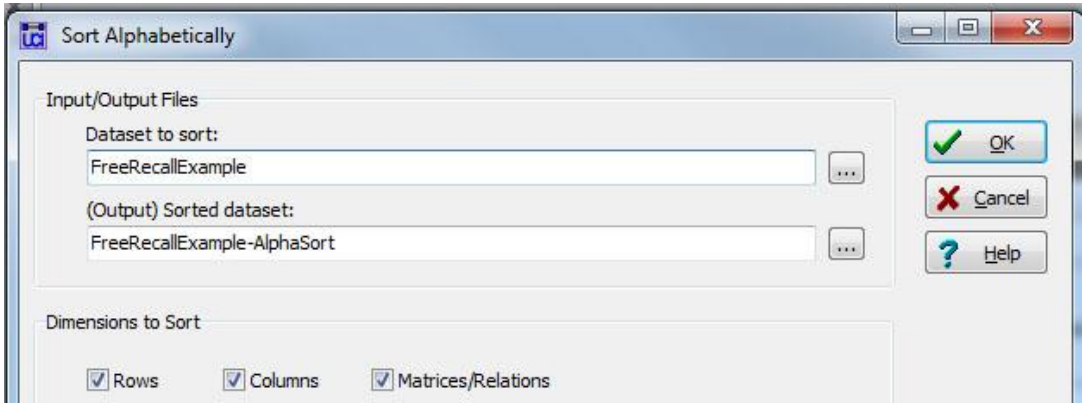
- txt2Pajek Advanced tab
  - Select “Allow loops”
  - Select “Allow empty cells”
- Run
- Hrm… still needs work

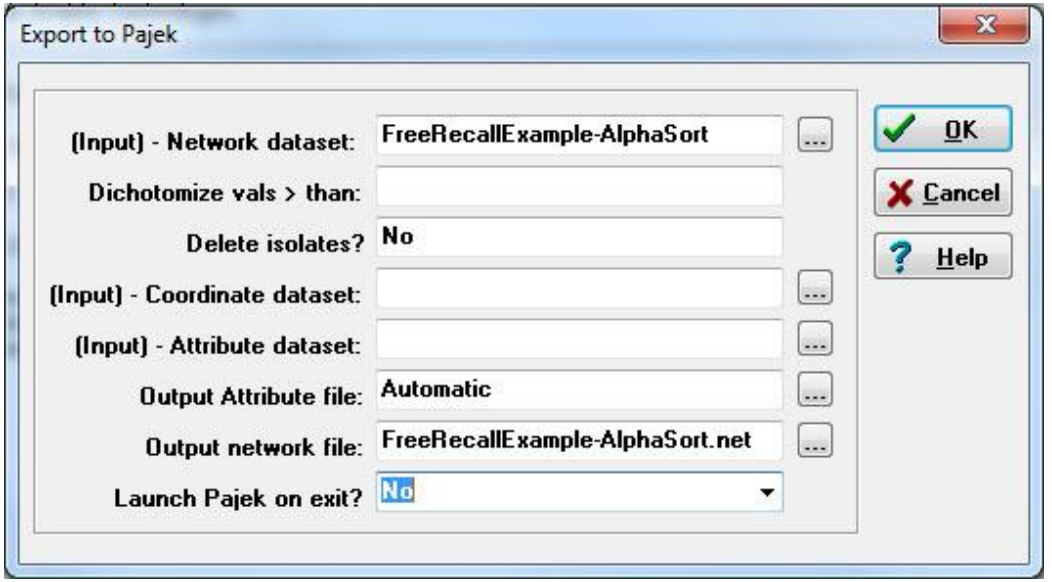




Free Recall Data: Sort Nodes

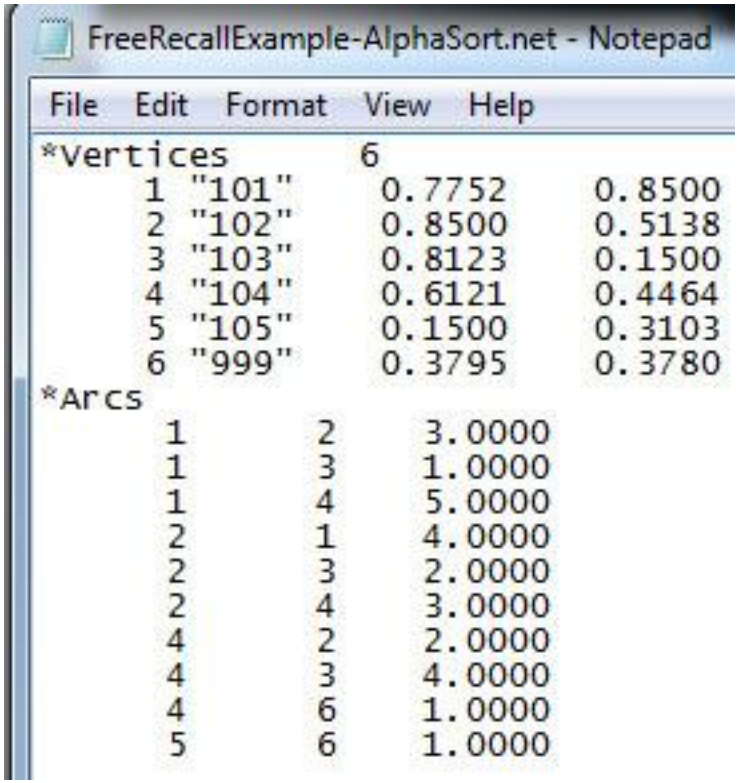
- UCINet
  - Data → Import text file → Pajek (select .net file)
  - Data → Sort Alphabetically
    - Select non-Crd ##h file
    - Keep Rows, Columns, and Matricies/Relations selected
    - Click OK
  - Data → Export → Pajek → Network
    - Select AlphaSort version
    - Do not launch Pajek (old version)
    - Click OK

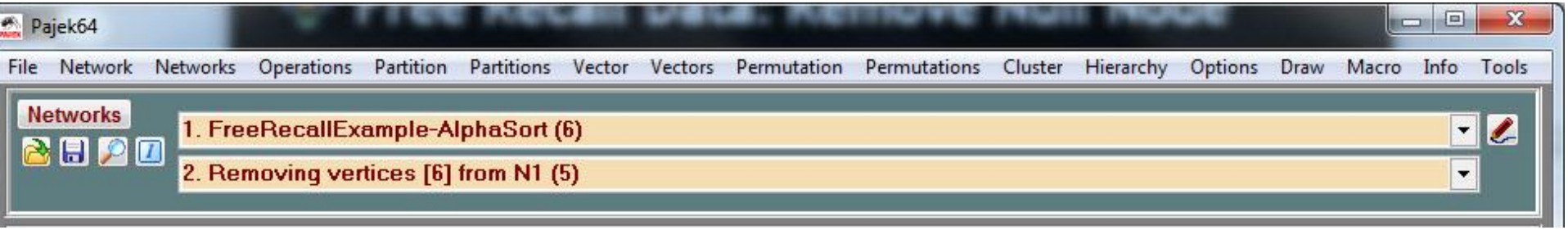




Free Recall Data: Remove Null Node

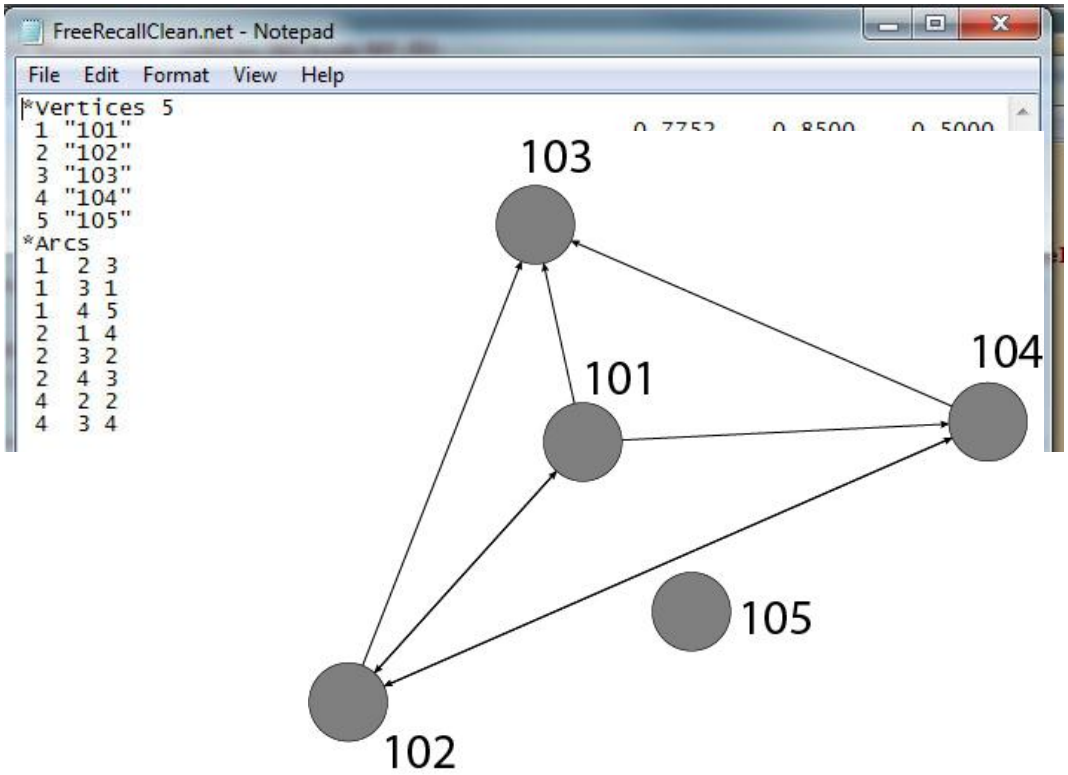
- Pajek
  - Drag & drop AlphaSort file into first network box
  - File → Network → Change Label to clean text
  - Network → Create New Network → Transform → Remove → Selected Vertices → enter appropriate # (in this case, 6)

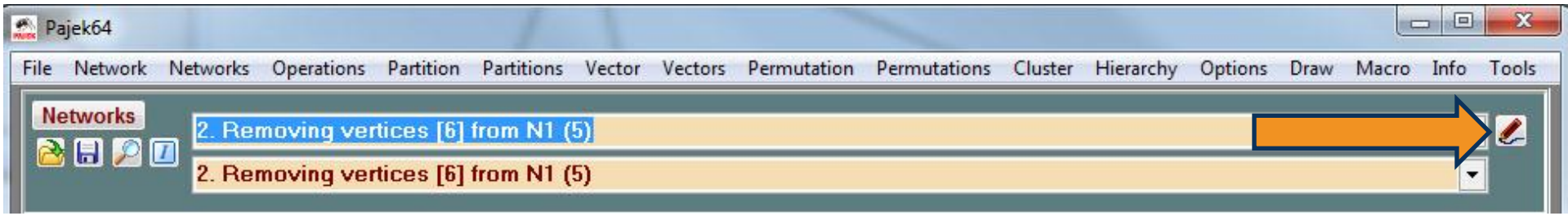




Free Recall Data: Draw Network

- Select clean network in first box
- Click Draw button





Roster Data: Raw Format

- Data will look something like this:

ID	Name	Con1	Con2	Con3	Con4	Con5
101.00	Smith, John	.	3.00	1.00	5.00	.
102.00	Parker, Tom	4.00	.	2.00	3.00	.
104.00	Jones, Tina	.	2.00	4.00	.	.
105.00	Meyer, Fred	.	.	.	.	.

- Elements
  - When sorted by ID, comes close to an N x N matrix
    - Con1 is everyone's contact rating for John Smith, Con2 is everyone's contact rating for Tom Parker, etc.
    - "From" is the ID column, "To" is each of the Con columns
- Strategy: create clean N x N matrix, use UCINet to convert to Pajek .net file

Roster Data: Insert Non-Respondents

- Add non-respondents in correct order

ID	Name	Con1	Con2	Con3	Con4	Con5
101.00	Smith, John	.	3.00	1.00	5.00	.
102.00	Parker, Tom	4.00	.	2.00	3.00	.
103.00	James, Bill	.	.	.	.	.
104.00	Jones, Tina	.	2.00	4.00	.	.
105.00	Meyer, Fred	.	.	.	.	.

- Aaannnd… that's all the cleaning you'll need!
  - (Way easier than free recall, eh?)

Roster Data: Attribute File

- Copy out new file
- Retain ID & Name
- Rename ID “Label”
- Bring in attribute data later on for visualizations

Label	Name	Gender
101.00	Smith, John	Male
102.00	Parker, Tom	Male
103.00	James, Bill	Male
104.00	Jones, Tina	Female
105.00	Meyer, Fred	Male

Roster Data: Export to Excel

- Back to network data
- Export as Excel file (remove Name)
- Clean
  - Clear ID cell
  - Find #NULL! & replace with 0
  - Copy ID numbers and Paste Special → Transpose

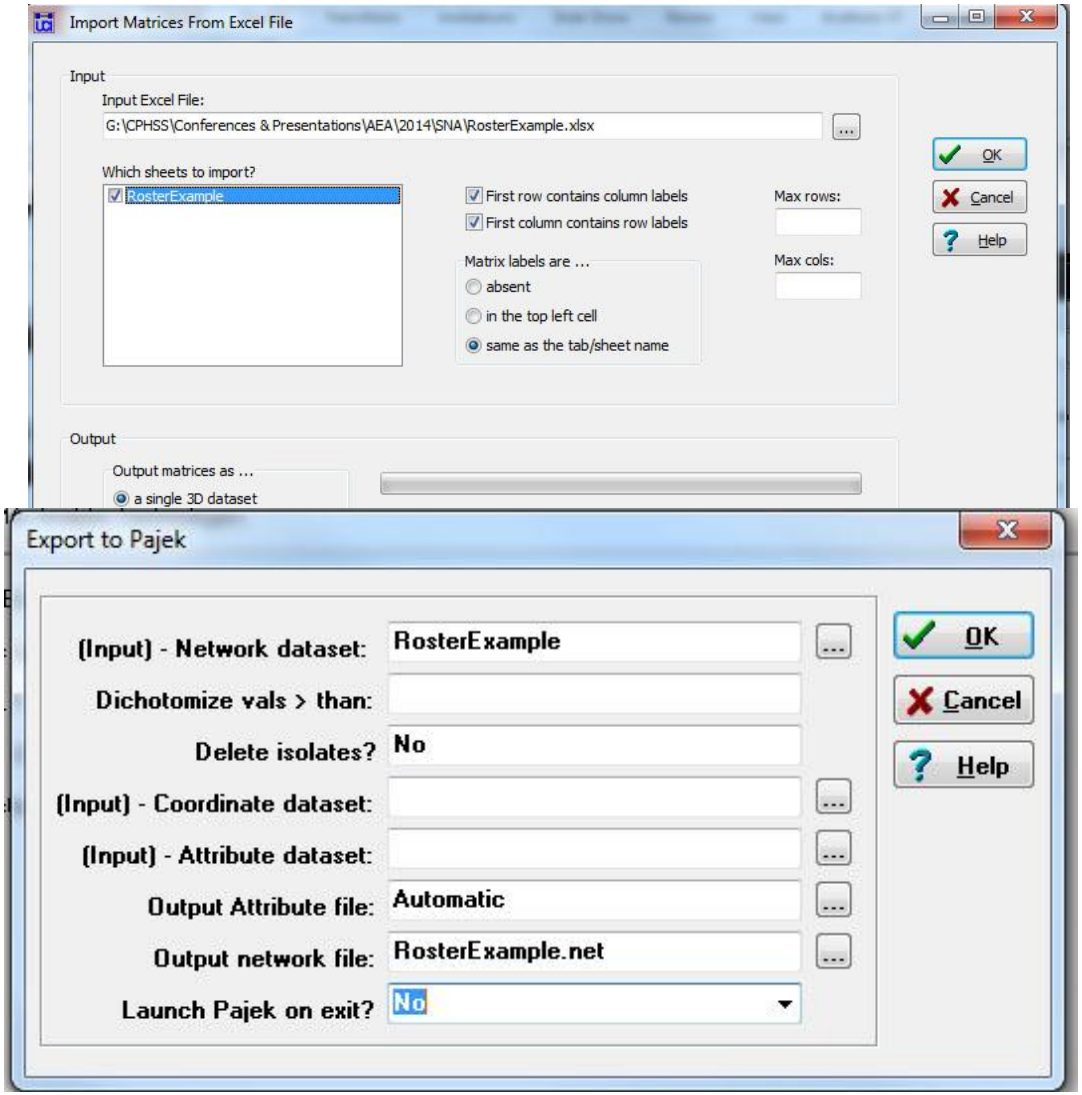
ID	Name	Con1	Con2	Con3	Con4	Con5
101.00	Smith, John	.	3.00	1.00	5.00	.
102.00	Parker, Tom	4.00	.	2.00	3.00	.
103.00	James, Bill	.	.	.	.	.
104.00	Jones, Tina	.	2.00	4.00	.	.
105.00	Meyer, Fred	.	.	.	.	.

A	B	C	D	E	F
ID	Con1	Con2	Con3	Con4	Con5
101.00	#NULL!	3.00	1.00	5.00	#NULL!
102.00	4.00	#NULL!	2.00	3.00	#NULL!
103.00	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!
104.00	#NULL!	2.00	4.00	#NULL!	#NULL!
105.00	#NULL!	#NULL!	#NULL!	#NULL!	#NULL!

A	B	C	D	E	F
	101.00	102.00	103.00	104.00	105.00
101.00	0.00	3.00	1.00	5.00	0.00
102.00	4.00	0.00	2.00	3.00	0.00
103.00	0.00	0.00	0.00	0.00	0.00
104.00	0.00	2.00	4.00	0.00	0.00
105.00	0.00	0.00	0.00	0.00	0.00

Roster Data: Convert to Pajek

- UCINet
  - Data → Import Excel → Matricies
    - Select file and appropriate sheet
    - Leave all other defaults as-is, click OK
  - Data → Export → Pajek → Network
    - Select file
    - Do not launch Pajek
    - Click OK



Roster Data: Final Product

- Look familiar?

RosterExample.net - Notepad

File Edit Format View Help

```
*Vertices       5
      1 "101"     0.2026     0.1500
      2 "102"     0.1500     0.5276
      3 "103"     0.2535     0.4722
      4 "104"     0.2112     0.8500
      5 "105"     0.8500     0.4938
*Arcs
       1       2     3.0000
       1       3     1.0000
       1       4     5.0000
       2       1     4.0000
       2       3     2.0000
       2       4     3.0000
       4       2     2.0000
       4       3     4.0000
```

Roster Data: If you don't have UCINet

- Pajek can also accept matrix formats
- Modify previous Excel file
  - Create vertex list with ID numbers
  - Matrix instead of arc list
  - Save out as tab-delimited text file
- Modify text file
  - Find “”” and replace with “
  - Change .txt extension to .mat

	A	B	C	D	E
1	*Vertices 5				
2	1	"101"			
3	2	"102"			
4	3	"103"			
5	4	"104"			
6	5	"105			
7	*Matrix				
8	0.00	3.00	1.00	5.00	0.00
9	4.00	0.00	2.00	3.00	0.00
10	0.00	0.00	0.00	0.00	0.00
11	0.00	2.00	4.00	0.00	0.00
12	0.00	0.00	0.00	0.00	0.00

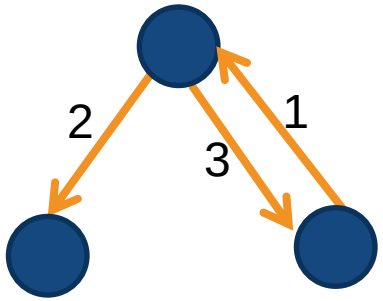
RosterExample.txt - Notepad

File Edit Format View Help

```
*Vertices 5
1          """101"""
2          """102"""
3          """103"""
4          """104"""
5          """105"""
*Matrix
0.00       3.00       1.00       5.00       0.00
4.00       0.00       2.00       3.00       0.00
0.00       0.00       0.00       0.00       0.00
0.00       2.00       4.00       0.00       0.00
0.00       0.00       0.00       0.00       0.00
```

All Data: Final Cleaning w/ Pajek

- Remove loops (if desired)
  - Network → Create New Network → Transform → Remove → Loops
  - Click “Yes” for “Create New Network?”
- Symmetrize
  - When relationship is inherently non-directional
  - Network → Create New Network → Transform → Arcs to Edges → All or Bidrected Only (usually All)
  - Create new network
  - Handle line values according to theoretical needs
    - Sum
    - Number
    - Minimum
    - Maximum
  - Export clean .net file



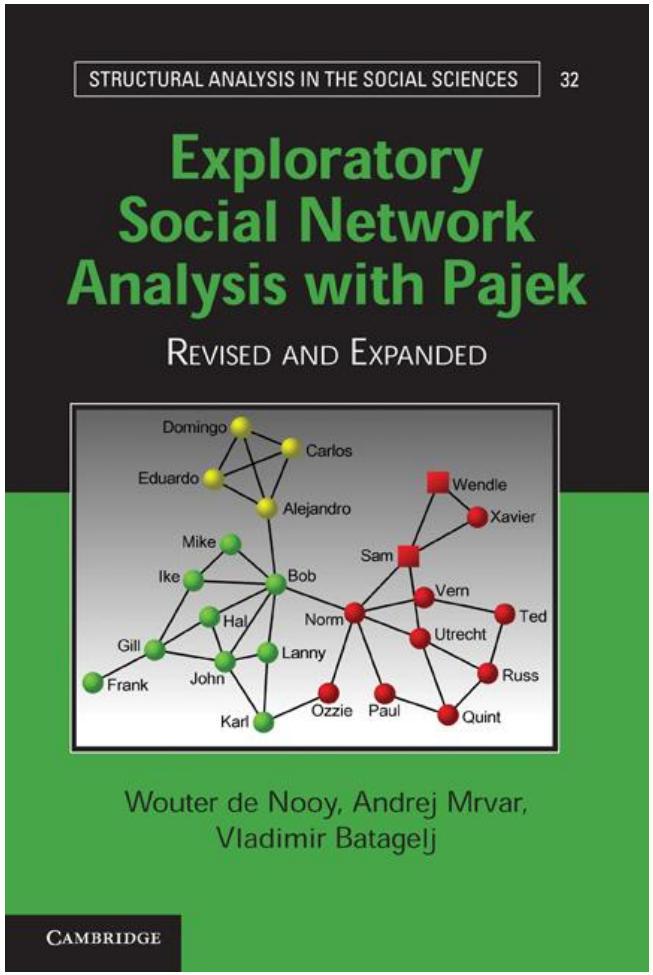
Step 5: Data Analysis – What Is the Structure of the Network?

Network Analysis Software

- Pajek
  - http://pajek.imfm.si/doku.php?id=pajek
  - Pros
    - Easy to learn
    - Transparent about what it does
    - Computes many standard network statistics
    - Free!
  - Cons
    - Can be difficult to produce attractive graphics
- Gephi
  - https://gephi.github.io/
  - Pros
    - Easy to learn
    - Easy to produce attractive graphics
    - Free!
  - Cons
    - Less transparent about what it does
    - Computes fewer network statistics
- Strategy
  - Perform analyses in Pajek
  - Transfer numbers to Gephi for visualizations

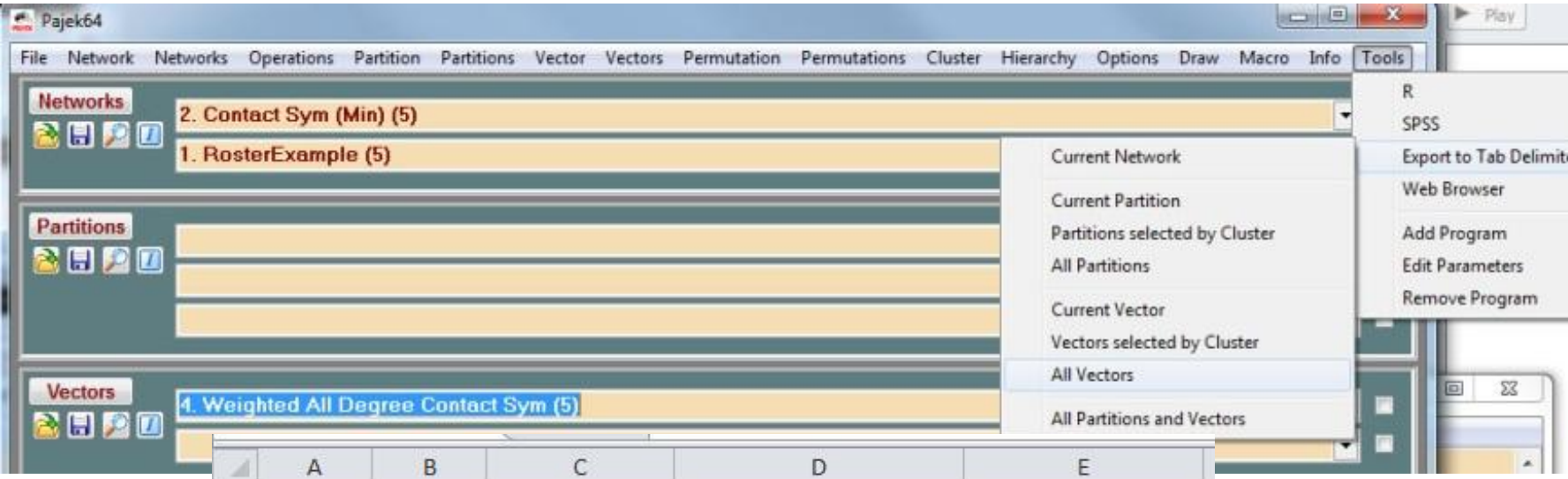
Getting the Numbers: Pajek

- Network-level statistics
  - Density, Average degree
  - Centralization (Degree, Betweenness, Closeness)
  - Modularity, VOS Quality
  - Blockmodeling
  - Many, many more!
- Node-level statistics
  - Centrality (Degree, Betweenness, Closeness)
  - Brokerage roles
  - Many more!



Exporting Node Characteristics

- From Pajek
- Tools → Export to Tab Delimited File → All Vectors (or whichever is most appropriate)



	A	B	C	D	E
1	Number	Label	All Degree Contact Sym (5)	Betweenness centrality Contact Sym (5)	Weighted All Degree Contact Sym (5)
2	1	101	3	0	9
3	2	102	3	0	7
4	3	103	3	0	7
5	4	104	3	0	11
6	5	105	0	0	0

Step 6: Network Visualization – What Does the Network Look Like?

or

How Do I Make Those Pretty Pictures?

Prepare Attribute File

- Attributes
  - Node characteristics (centrality, demographics, etc.)
  - Determine size & color of nodes in graphics
- Pull characteristic data from survey and network analysis into one SPSS, SAS, or Excel file

Label	Name	Gender	ID	Degree	Between	Weighted Degree
101.00	Smith, John	Male	1	3	0	9
102.00	Parker, Tom	Male	2	3	0	7
103.00	James, Bill	Male	3	3	0	7
104.00	Jones, Tina	Female	4	3	0	11
105.00	Meyer, Fred	Male	5	0	0	0

  - Change “Number” to “ID” if you’re planning to use Gephi for visualizations

Add Color Codes

- Hex values (safest – will later be exported to CSV)
- "Color" must be part of the variable name
- See http://colorbrewer2.org/ for colorblind, photocopy, & LCD compatibility

Label	Name	Gender	ID	Degree	Between	Weighted Degree	GenderColor
101.00	Smith, John	Male	1	3	0	9	#2b83ba
102.00	Parker, Tom	Male	2	3	0	7	#2b83ba
103.00	James, Bill	Male	3	3	0	7	#2b83ba
104.00	Jones, Tina	Female	4	3	0	11	#fdae61
105.00	Meyer, Fred	Male	5	0	0	0	#2b83ba

Export Attributes to CSV

- ID should be first column
- Label & Name are optional
- Gephi can only interpret one color variable at a time
  - Export different .csv files with different color-coded variables if needed

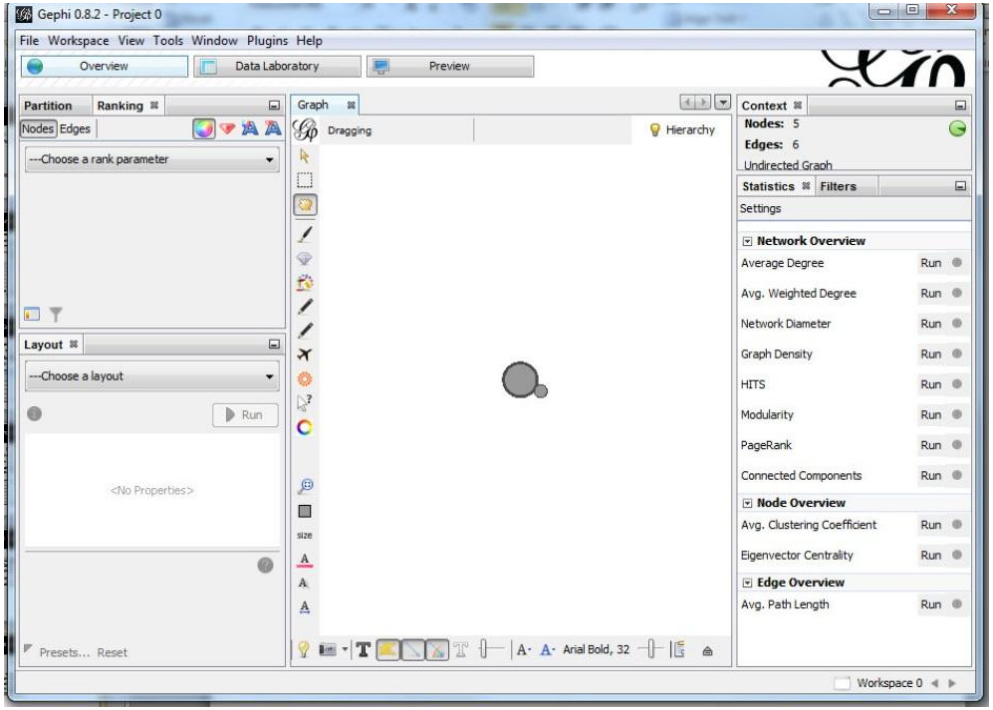
A	B	C	D	E	F	G	H
ID	Label	Name	Gender	GenderColor	Degree	Between	WeightedDegree
1	101	Smith, John	Male	#2b83ba	3	0	9
2	102	Parker, Tom	Male	#2b83ba	3	0	7
3	103	James, Bill	Male	#2b83ba	3	0	7
4	104	Jones, Tina	Female	#fdae61	3	0	11
5	105	Meyer, Fred	Male	#2b83ba	0	0	0

Gephi Resources

- Plugins
  - https://marketplace.gephi.org/
  - Give Color to Nodes: Allows Gephi to read hex color codes
  - Noverlap: Eliminates node overlap
  - Many other options available to browse!
- Tutorials
  - http://gephi.github.io/users/

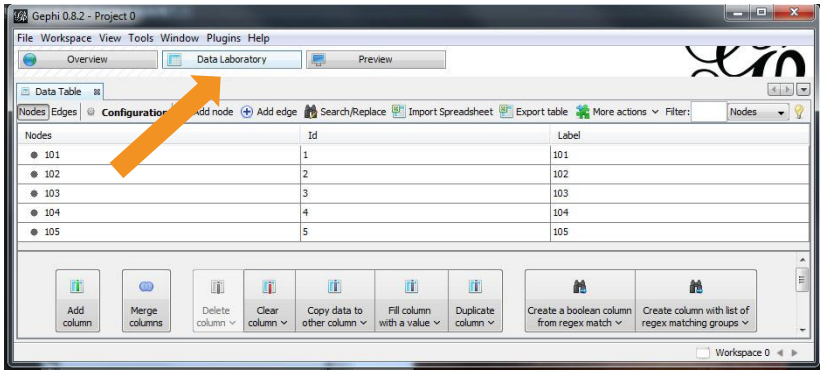
Import Network Data to Gephi

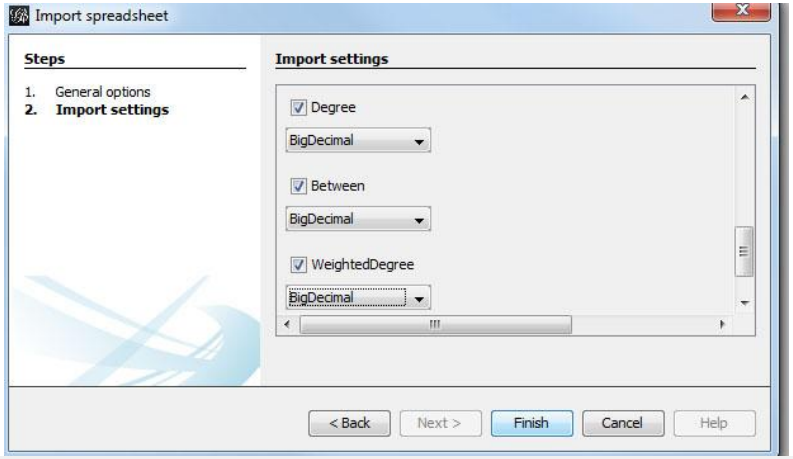
- Import clean .net file
  - File → open → select .net file
  - Select Directed, Undirected, or Mixed as appropriate



Import Attribute Data to Gephi

- Data Laboratory → Import Spreadsheet → select .csv attributes file
- Import Settings: change numeric variables from “String” to “Big Decimal”
- Finish





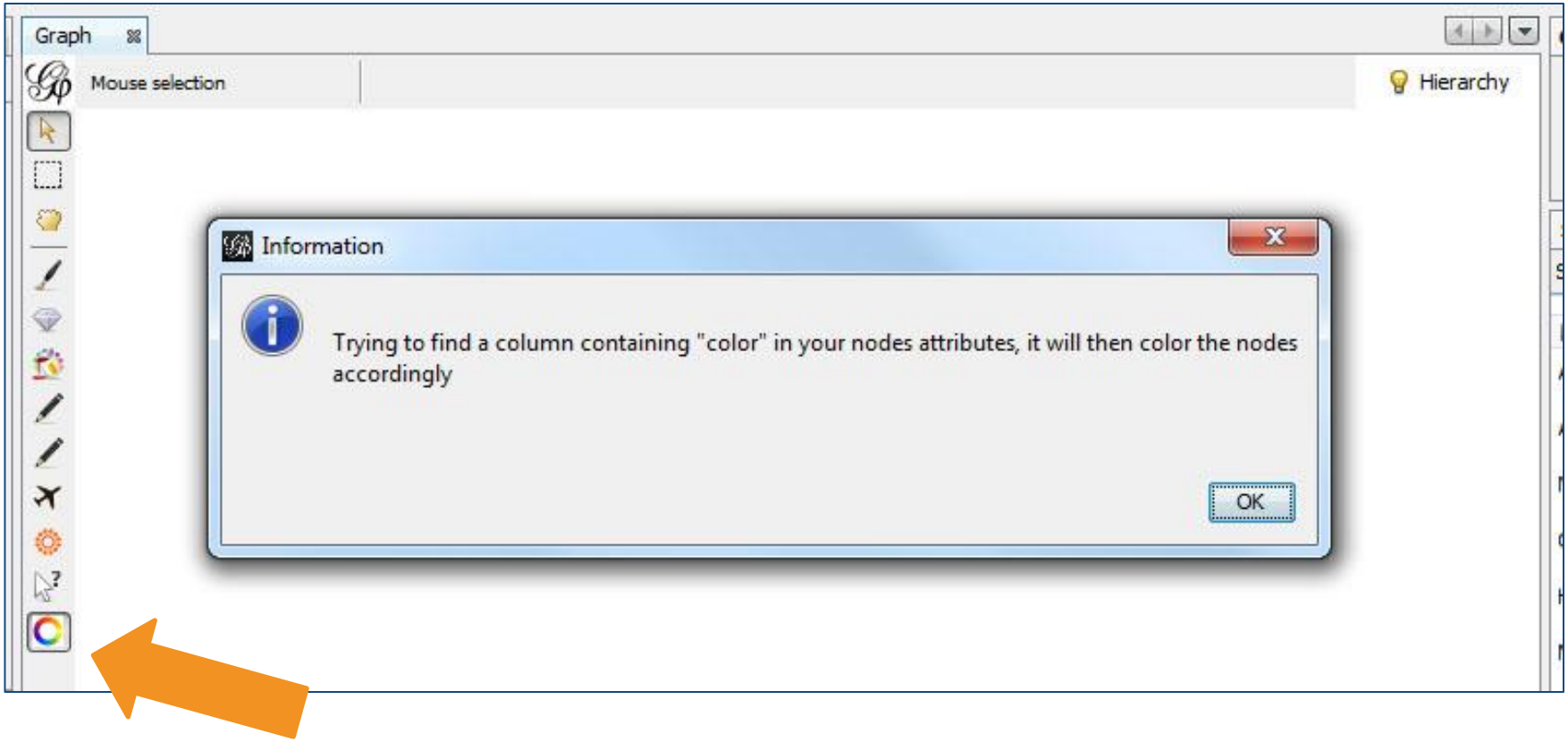
Data Table

Nodes | Edges | Configuration | Add node | Add edge | Search/Replace | Import Spreadsheet | Export table | More actions | Filter: | Nodes

Nodes	Id	Label	Name	Gender	GenderColor	Degree	Between	WeightedDegree
101	1	101	Smith, John	Male	#2b83ba	3	0	9
102	2	102	Parker, Tom	Male	#2b83ba	3	0	7
103	3	103	James, Bill	Male	#2b83ba	3	0	7
104	4	104	Jones, Tina	Female	#fdae61	3	0	11
105	5	105	Meyer, Fred	Male	#2b83ba	0	0	0

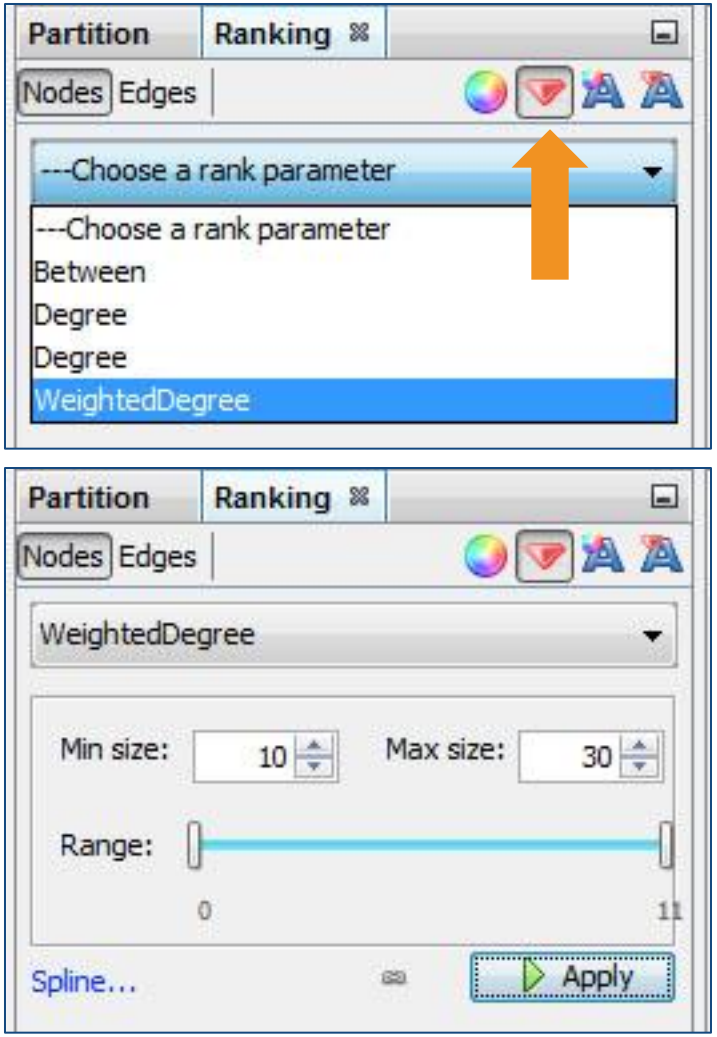
Apply Color to Nodes

- Overview tab
- Click on color wheel



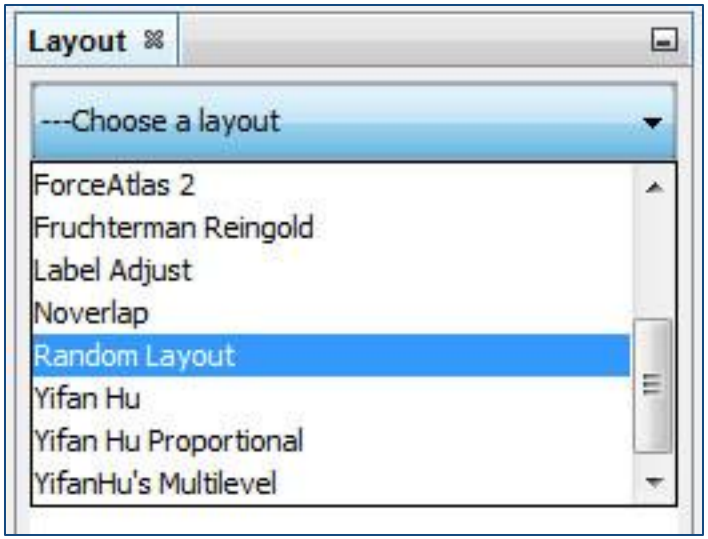
Apply Size to Nodes

- Left frame
- Click on diamond
  - Imported numeric attributes will appear
  - Select appropriate parameter
- Set min and max sizes
  - Best option depends on number of nodes and parameter range
  - Experiment and go with what looks best

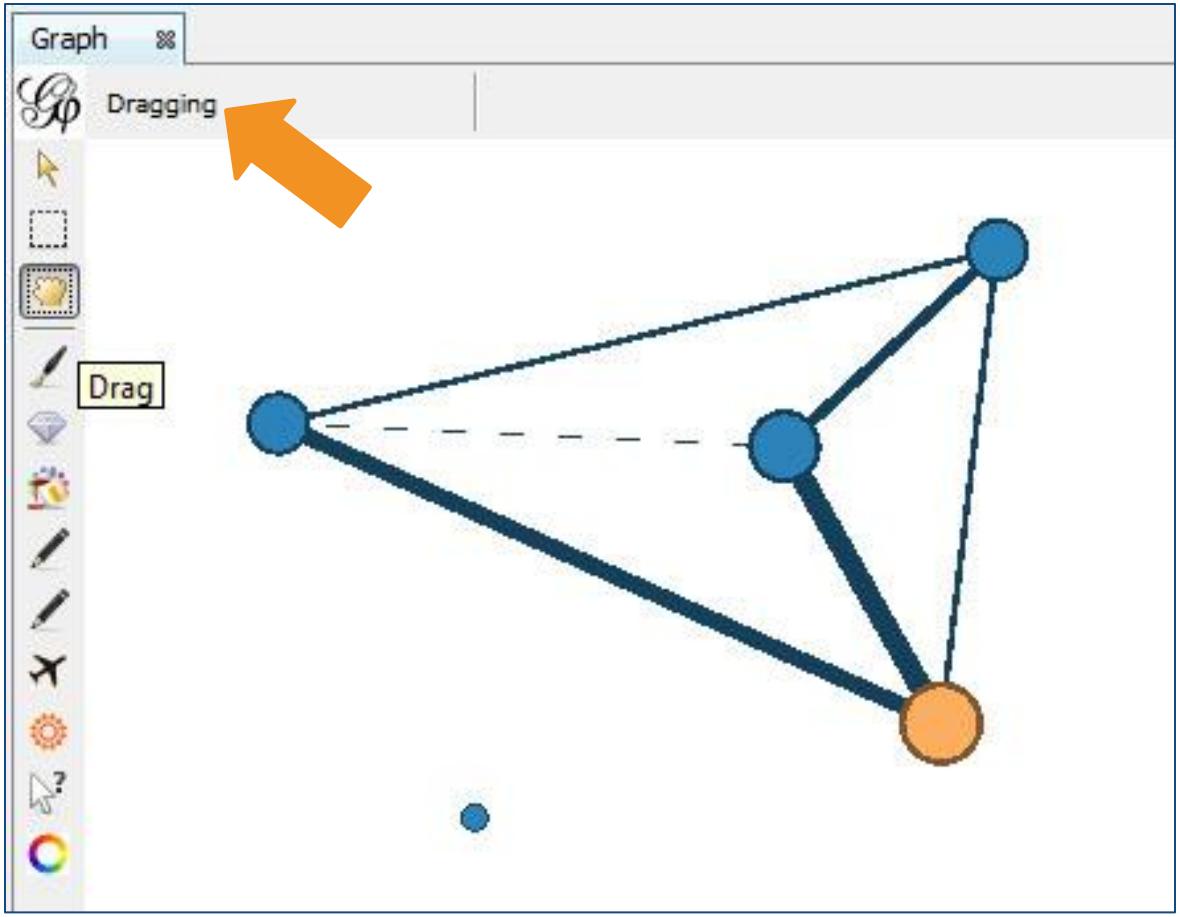


Choose a Layout

- Start with Random, end with Noverlap if required
- Experiment & see what works best
- Most layouts have settings you can fine-tune



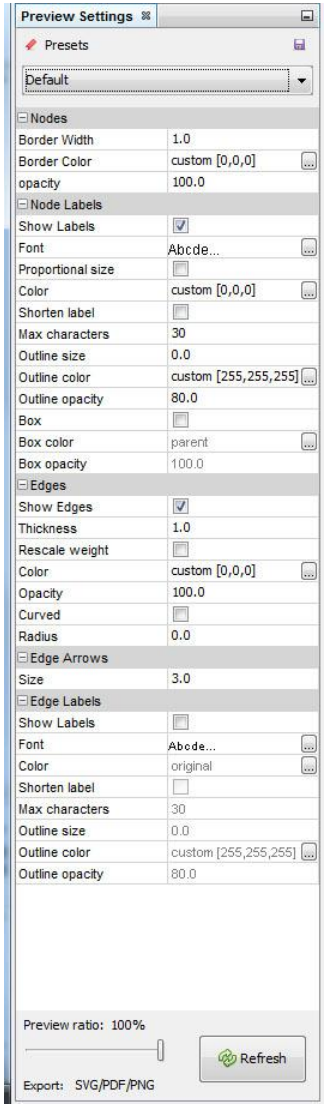
Re-Arrange Nodes Manually



- If needed
- Click on hand icon → allows you to click on nodes and move
- Click on “Dragging” to change the diameter of selection area

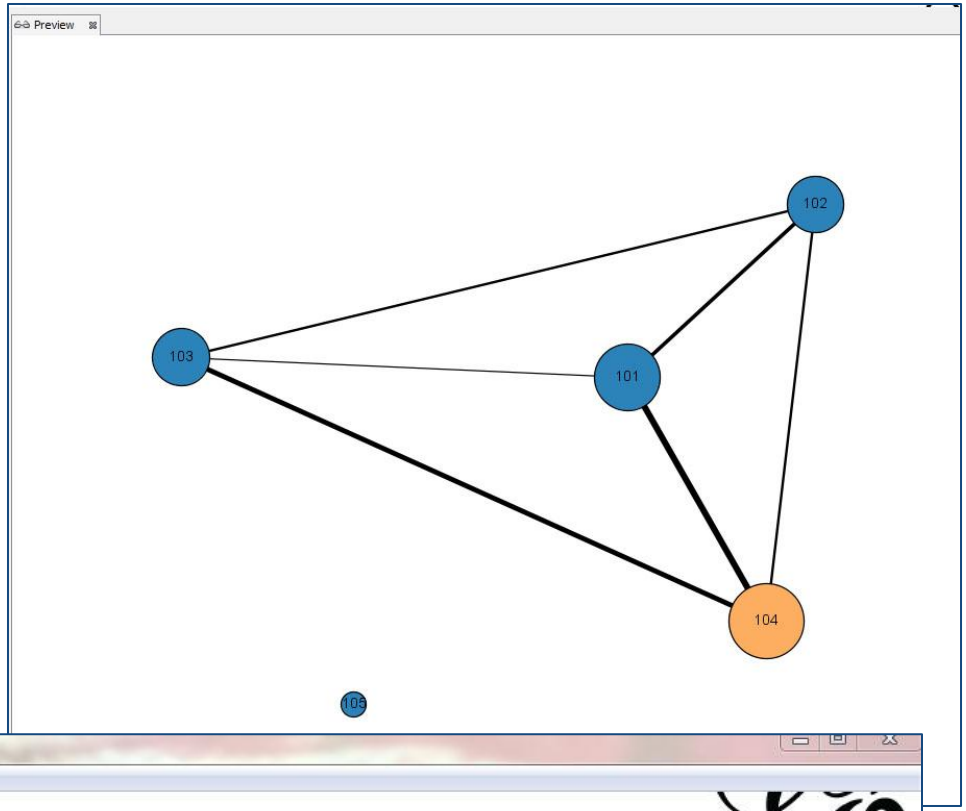
Adjust General Appearance

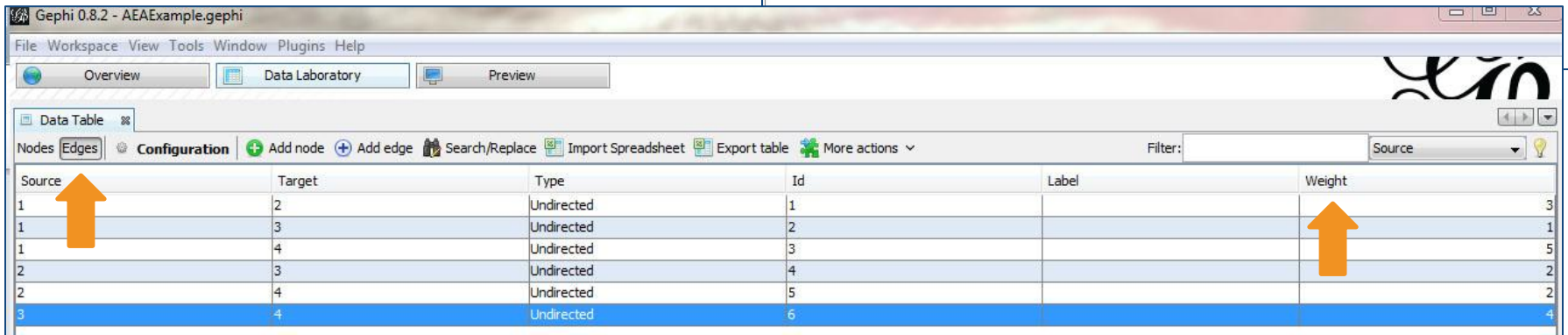
- Click “Preview” at the top, then “Refresh” at the bottom
- Labels, edges, arrow sizes (directional only)
- Click “Refresh” to show changes



If Lines Are Valued…

- Will display as varying thickness
- Options if you want lines w/ uniform thickness
  - Transform network in Pajek so all linevalues = 1, or
  - Export graphic in SVG and change in Adobe Illustrator, or
  - Data Laboratory → Edges, change weight values to 1

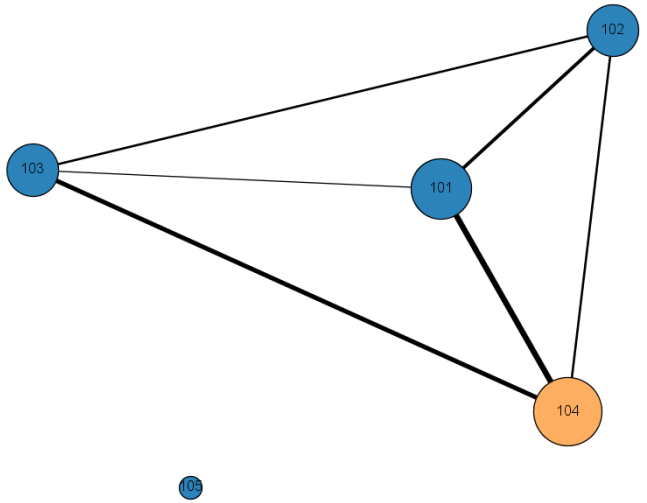




Source	Target	Type	Id	Label	Weight
1	2	Undirected	1		3
1	3	Undirected	2		1
1	4	Undirected	3		5
2	3	Undirected	4		2
2	4	Undirected	5		2
3	4	Undirected	6		4

Export Graphic

- WYSIWYG (What You See Is What You Get)
- SVG, PDF, or PNG options
- If you have Adobe Illustrator, saving to SGV will allow further fine-tuning



Questions?

Bobbi Carothers
bcarothers@wustl.edu

http://cphss.wustl.edu
@cphsswustl