

We will be handing out flash drives. If you would like to follow along, please install Sci2 and Gephi on your machines:

You may refer to the print handouts for more detailed installation instructions. In the LOEX 2013 folder on the flash drive:

-  gephi-0.8.2-beta.dmg ← Mac users
-  gephi-0.8.2-beta.setup.exe ← Windows users
-  gephi-0.8.2-beta.tar.gz ← Linux users
-  instruction data.csv
-  instruction.properties
-  sci2-N-1.0.0.201206130117NGT-linux.gtk.x86_64.tgz ← Linux users
-  sci2-N-1.0.0.201206130117NGT-macosx.carbon.x86.tgz ← Mac users
-  sci2-N-1.0.0.201206130117NGT-win32.win32.x86.zip ← Windows users

Remix Your Data: Visualizing Library Instruction Statistics

Brianna Marshall

David Edward “Ted” Polley

Introductions

Brianna Marshall

MLS/MIS Candidate, Indiana University

Graduating May 2014

Ted Polley, MLS/MIS

Research Assistant

Cyberinfrastructure for Network Science Center, Indiana University

Spreadsheets

You love
them, right?



Why visualization?

- Exposes patterns not apparent in raw data (ahem, spreadsheets)
- Helps the information you've been collecting tell a story
- Provides an even deeper analysis of your instruction—the big picture—that can shape:
 - Assessment
 - Departmental goals
 - Use of human resources

It's ACRL-approved!

*The Value of Academic Libraries:
A Comprehensive Research Review and Report*
(2010) emphasized the need:

“[T]o build librarians’ capacity to document, demonstrate,
and communicate library value”

It's ACRL-approved! (x2)

The 2012 report from the *Value Of Academic Libraries Summit: Connect, Collaborate, and Communicate*, reiterated, calling for:

“[M]ultiple replicable approaches for documenting and demonstrating library impact on student learning and success”

If you're still not convinced...

- It's impressive!
 - To your boss
 - On your department walls
 - In your portfolio
- A new tool in your toolkit
- Why not?

We'll try

- ▣ To provide context—what we're doing and why
- ▣ To explain the terminology—but if you get confused, just reference the handout
- ▣ To keep it simple—though if this is brand new, it may be overwhelming. And that's okay.

Remember

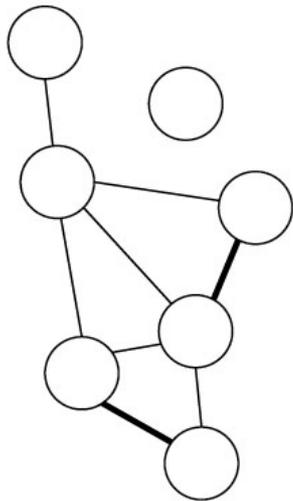
- ▣ There's a lot of documentation available—you don't have to remember everything now
- ▣ We're available to answer questions about your data after the session

On the agenda

- Brief introduction to networks
- Visualization using qualitative data (example)
- Visualization using quantitative data (walkthrough)

Introduction to Networks

Undirected Networks



Nodes:



Edges:



Node Degree:

Number of edges connected to nodes

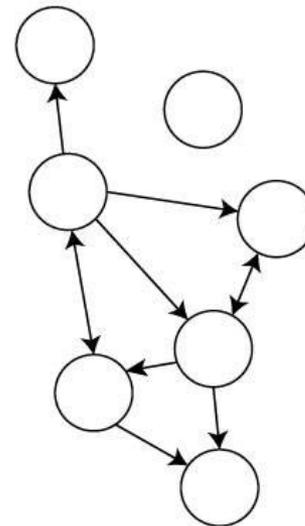
Isolates:

Nodes that are not connected to the rest of the network

Edge Weight:

Demonstrates relative importance of relationships

Directed Networks



Edge Direction:

Directional relationship is represented by arrows

In-Degree:

Number of incoming edges

Out-Degree:

Number of outgoing edges

Visualizing qualitative data

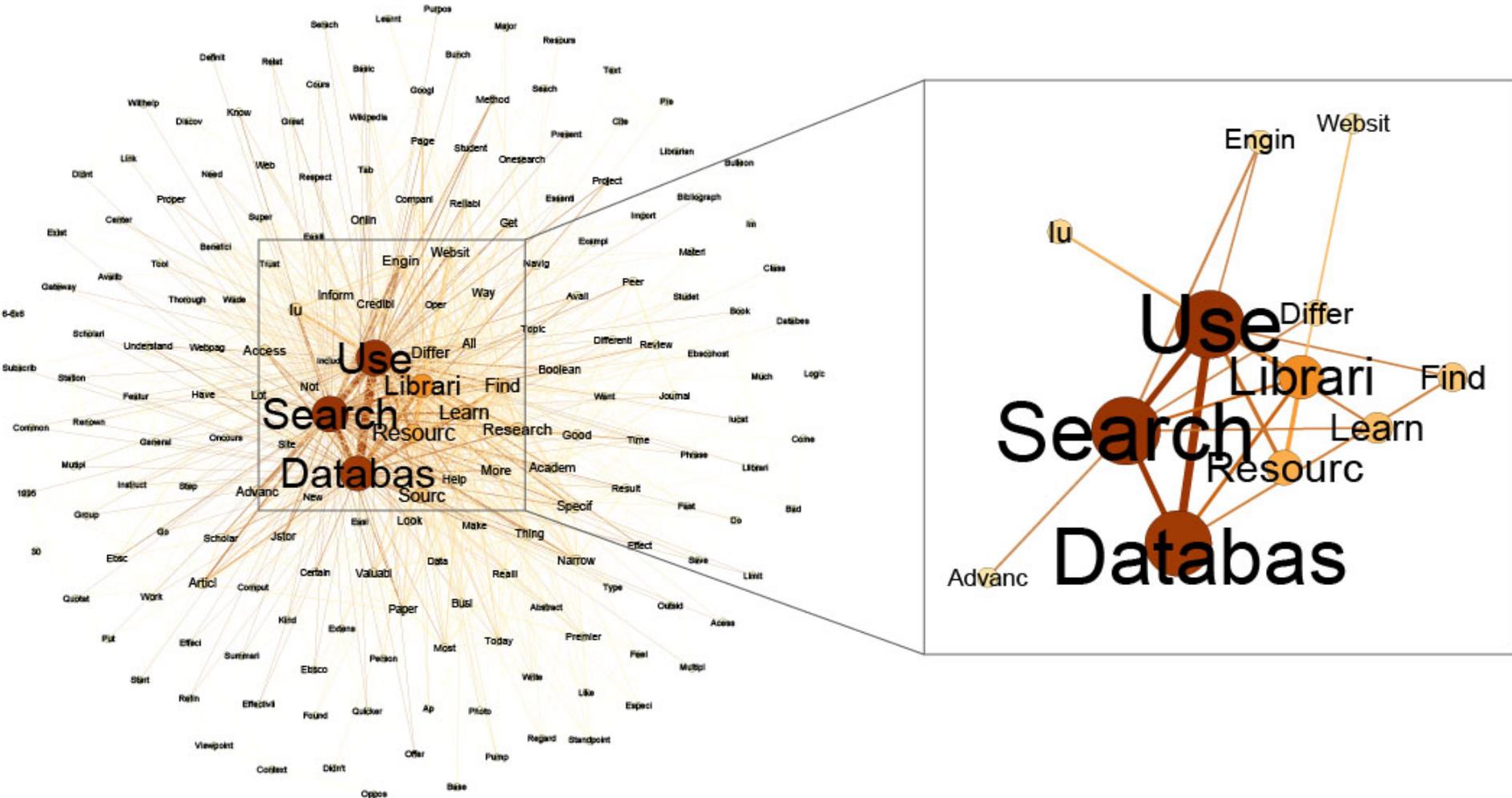
- Visualizing qualitative data can present a real challenge
- A lot of qualitative data comes in the form of free-text responses to post-instruction surveys and evaluations
- OBJECTIVE

Create a word co-occurrence network to identify any emerging themes from the responses to open-ended questions on the evaluations of an individual instructor

- DATA

The data come from the 138 student responses to the evaluation of one instructor. This particular network visualizes responses to the question, “What was the most valuable thing you learned today?”

Word co-occurrence network of library instruction data



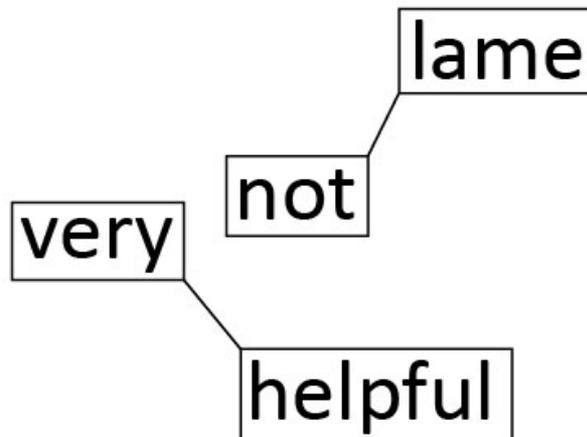
Word co-occurrence networks vs. tag clouds

- Tag clouds display the frequency with which certain words appear in a body of text
- This word frequency can be represented using font size and/or color

helpful
very
lame
not

Word co-occurrence networks vs. tag clouds

- Word co-occurrence is helpful in determining meaning
- Word co-occurrences are displayed by making explicit connections between words



Visualizing quantitative data

■ OBJECTIVE

To compare the number of instruction sessions taught to courses (and departments) over a particular period of time. The number of students reached will also be shown.

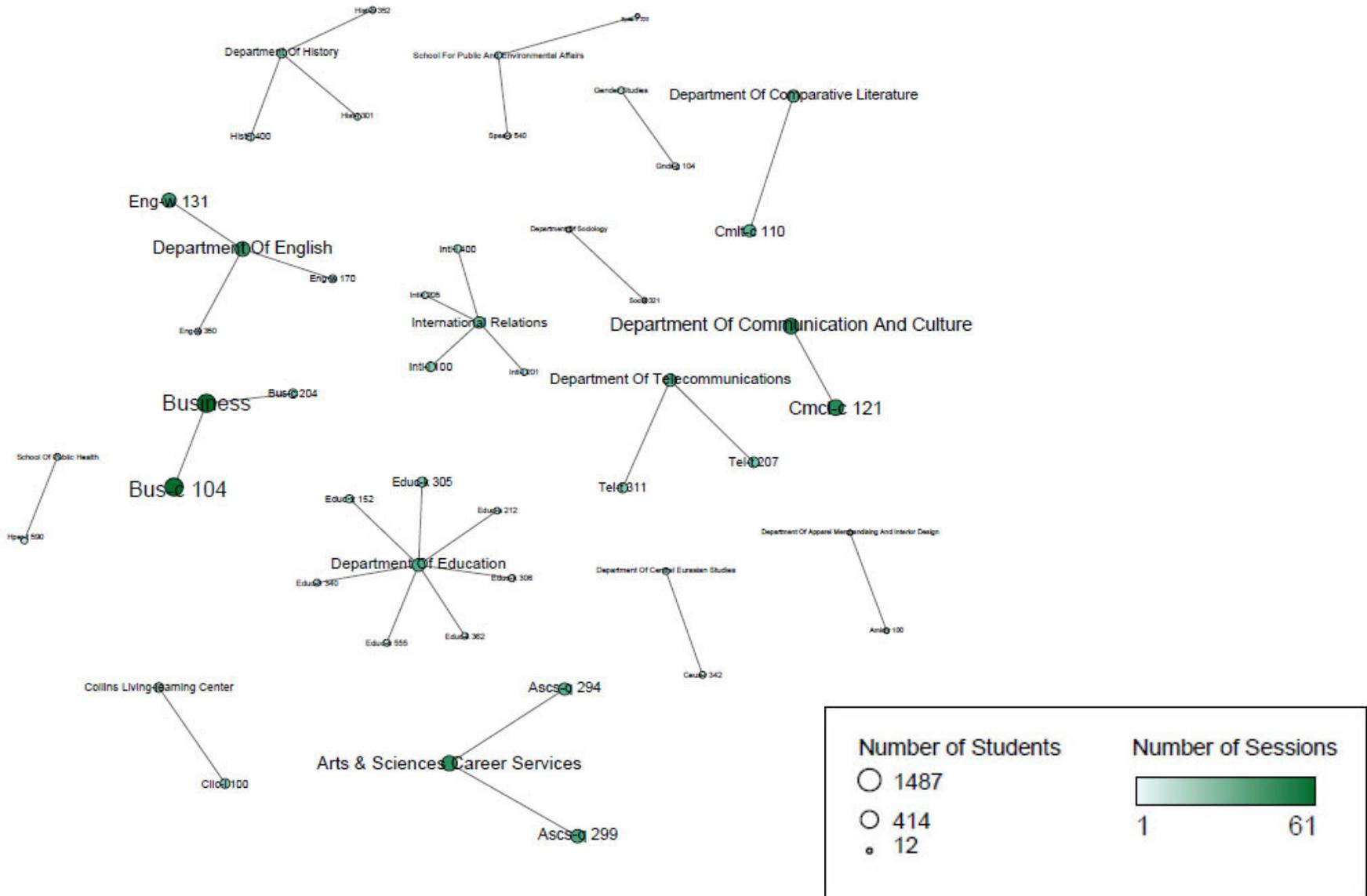
■ DATA

Today, we're using real data from the Indiana University Teaching & Learning Department (Thank you Carrie Donovan!)

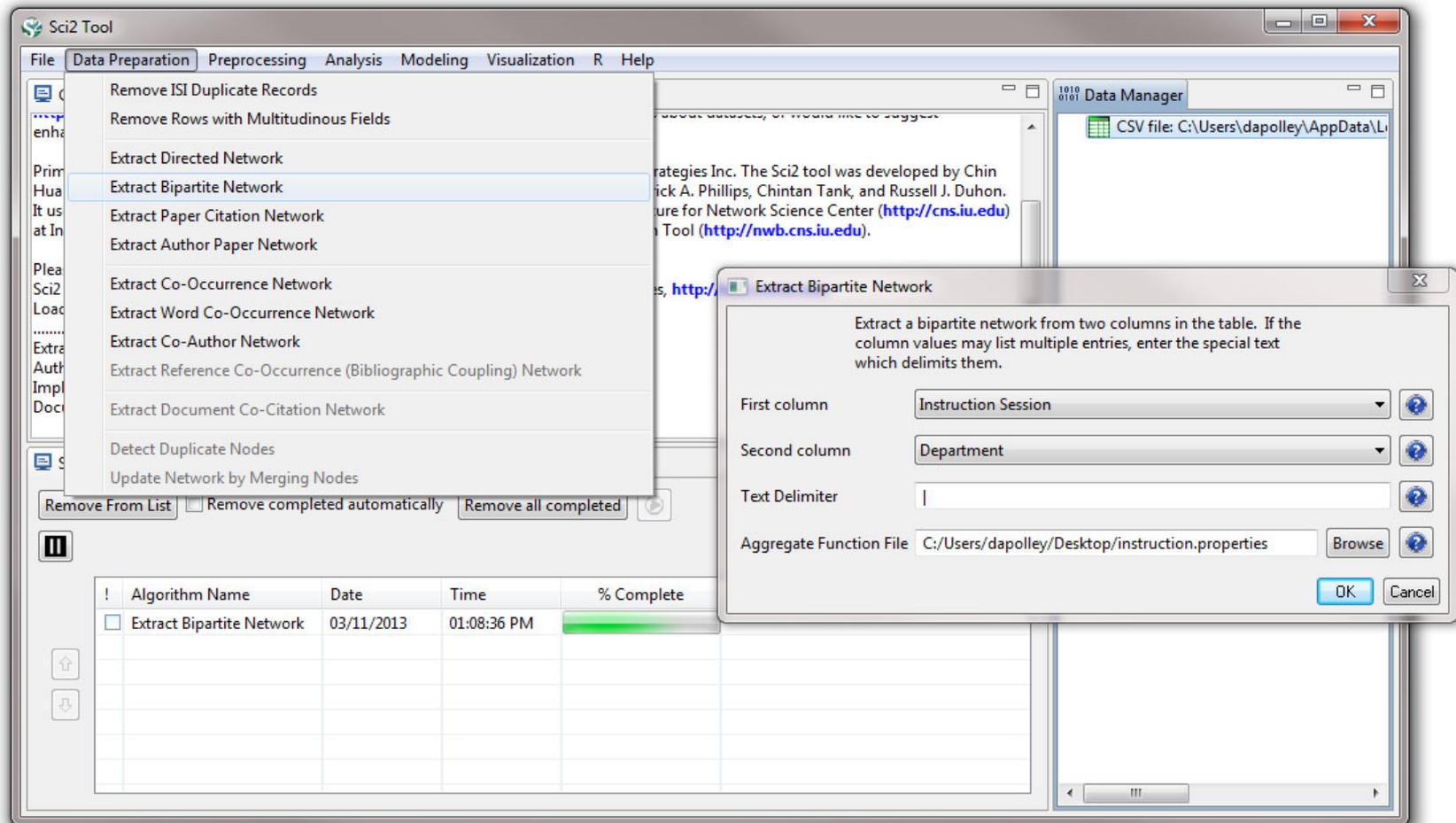
In a nutshell

1. Start with a spreadsheet of library instruction data
2. Load it into the **Science of Science (Sci2) Tool**
3. Create a network that connects the departments with the instruction session taught
4. Load the network into **Gephi**
5. Clean it up
6. Voila! Your visualization is complete

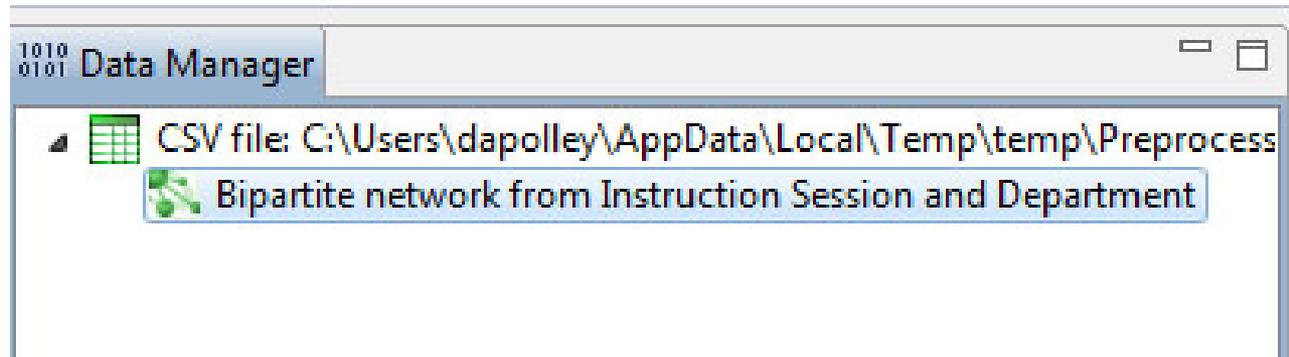
Instruction Sessions: Academic Year 2012



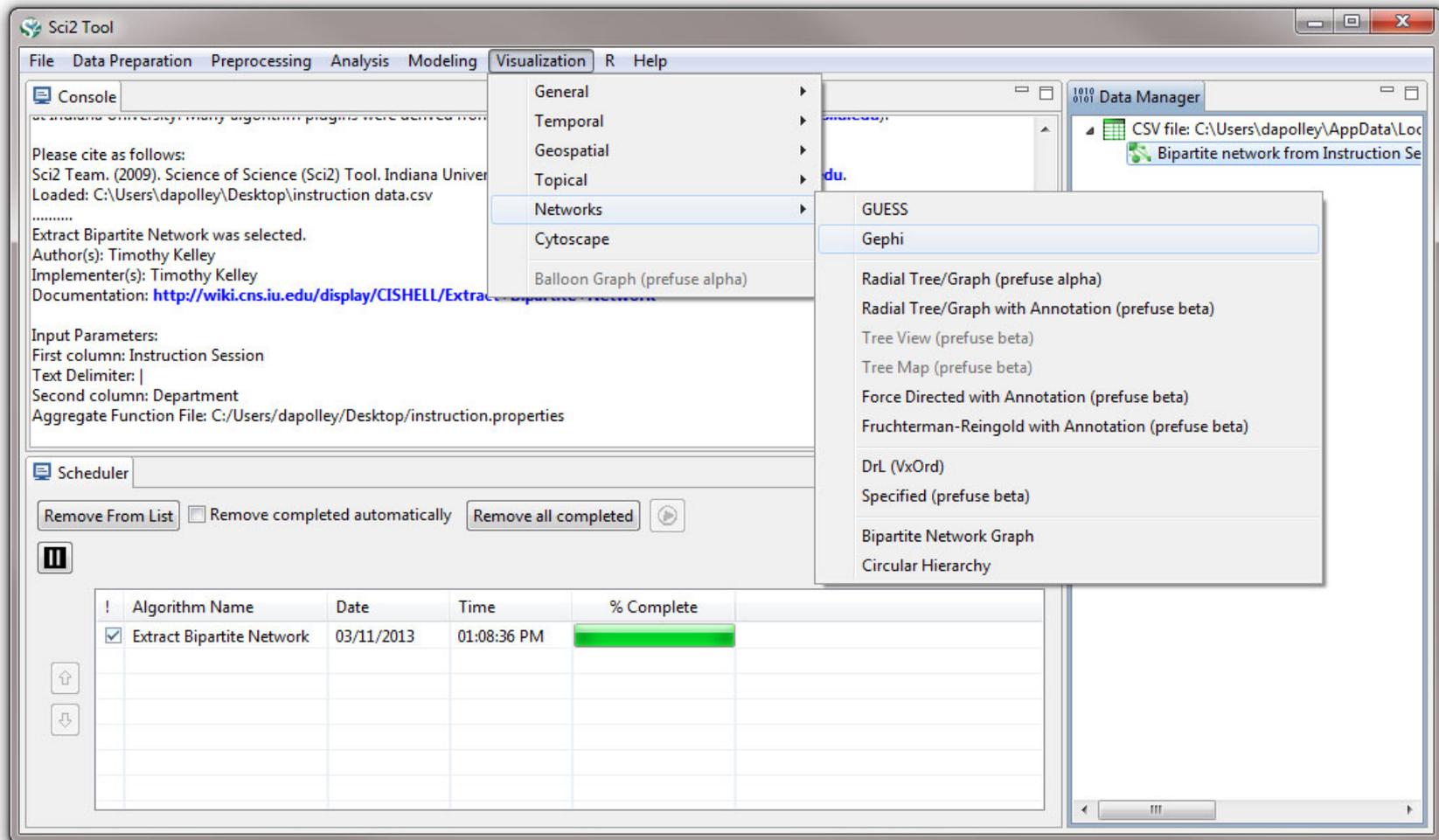
From there you will need to extract a bipartite network connecting the instruction sessions to the departments for which those classes are a part of. Run '*Data Preparation > Extract Bipartite Network*' and enter the following parameters:



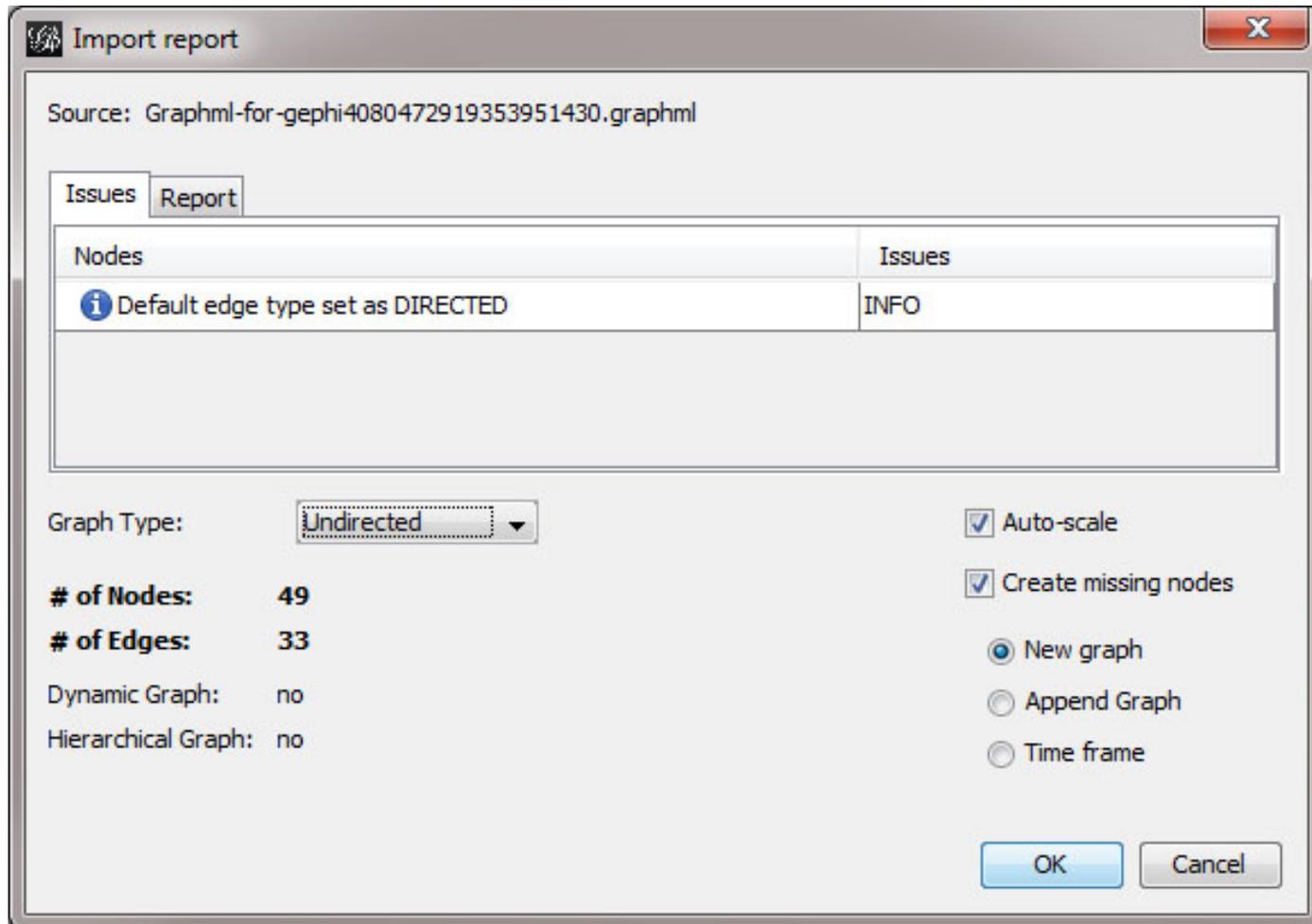
This will result in a bipartite network in the data manager:



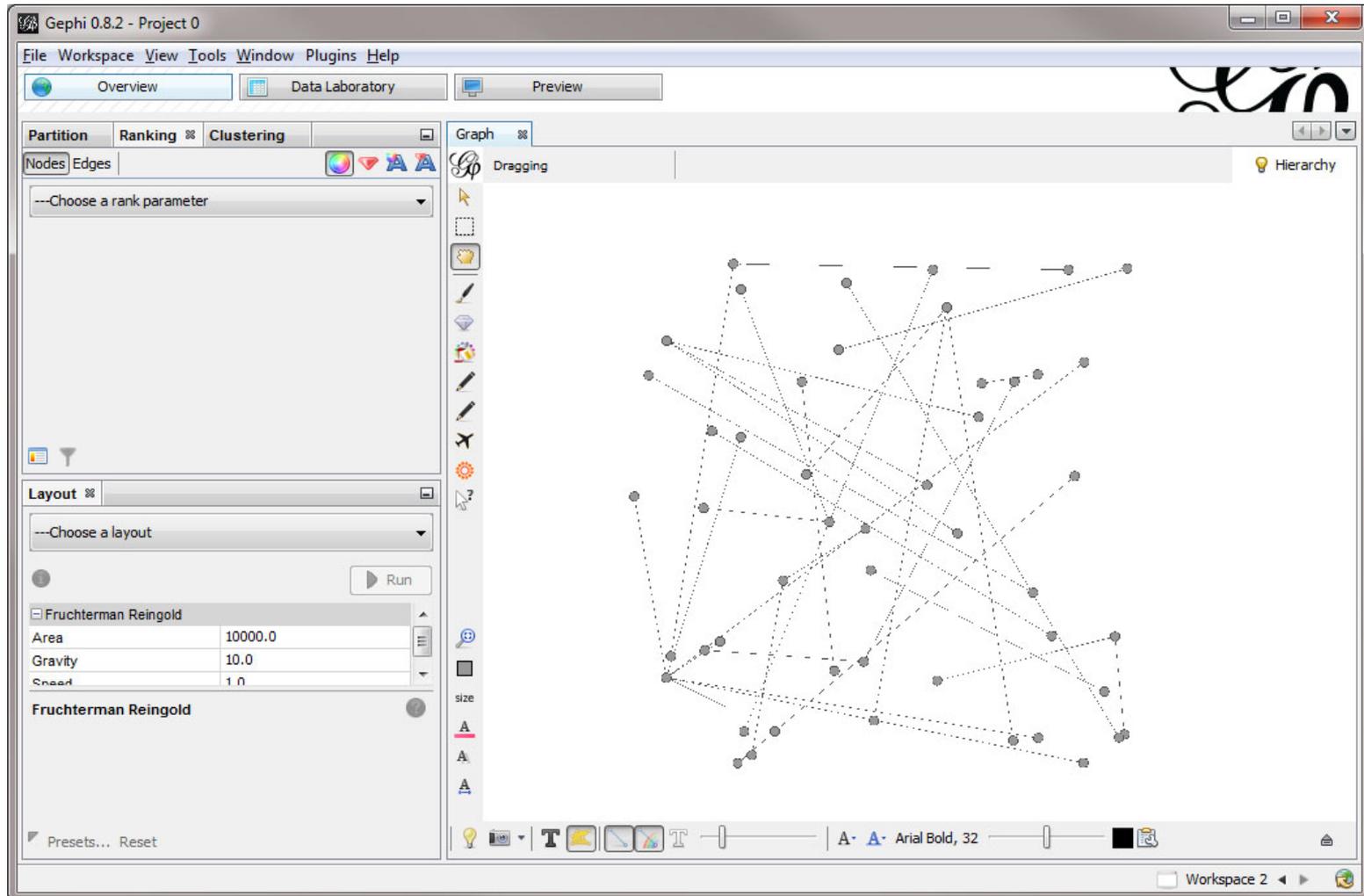
Next, open this network in Gephi, '*Visualization > Networks > Gephi*':



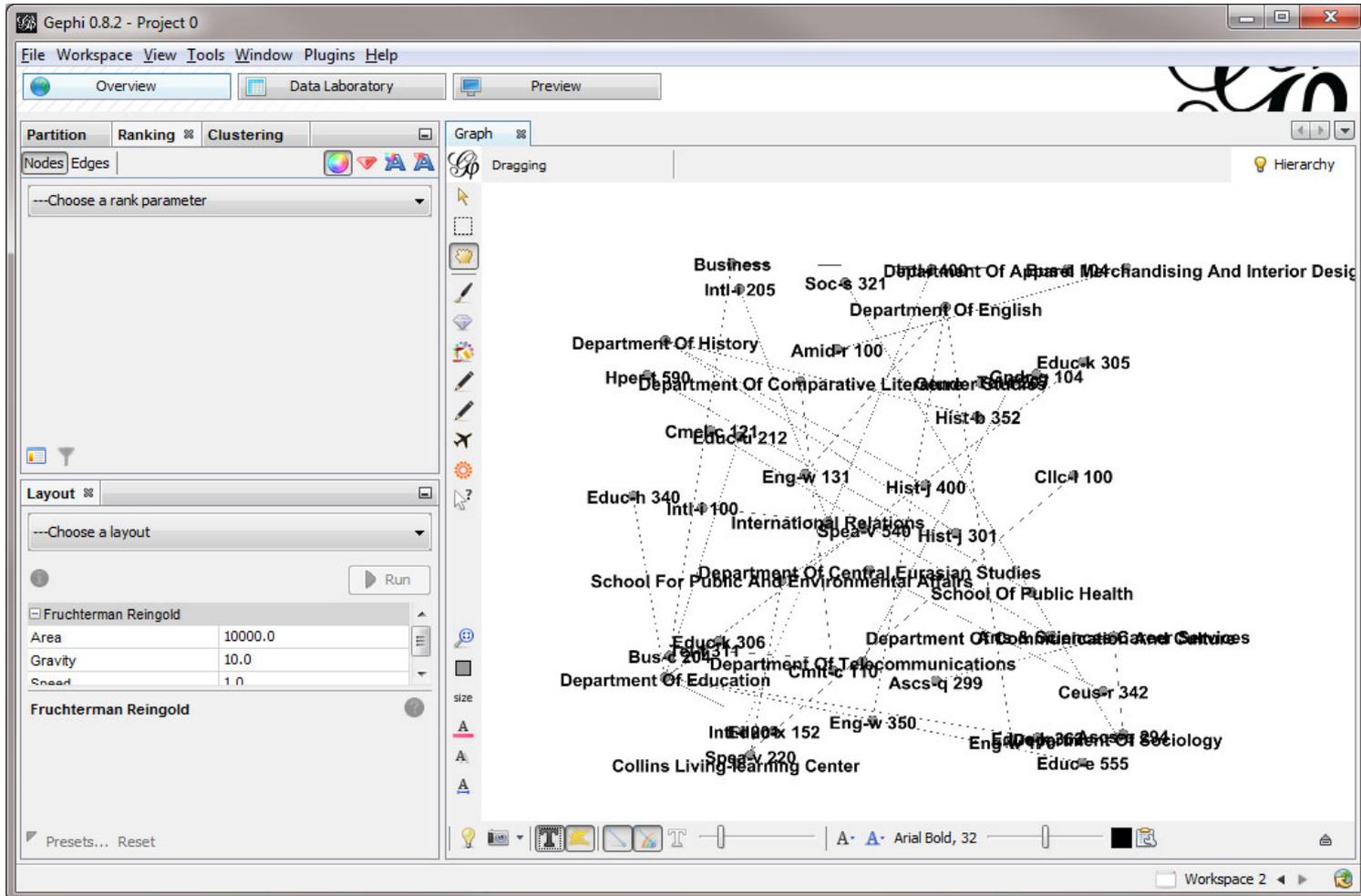
When you open the network in Gephi you will be shown a loading report that tells you the number of nodes and edges in the network and asks you to specify whether it is a directed or undirected network. Load the file as an undirected network:



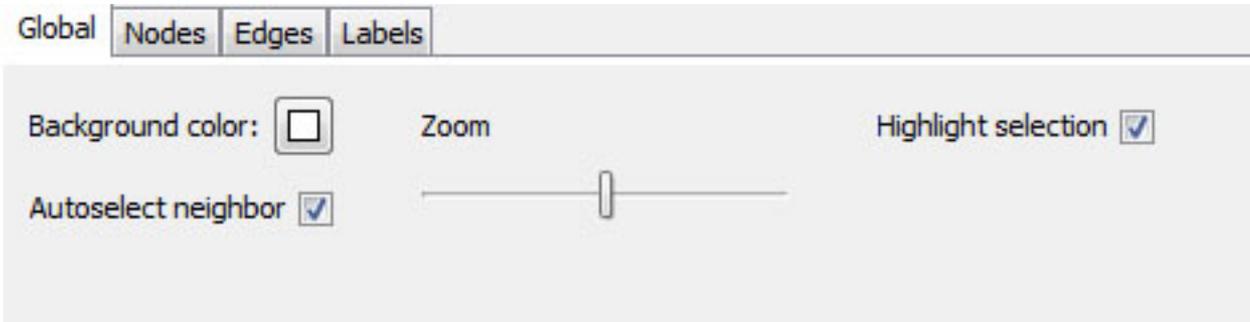
When the network is opened in Gephi it will be laid out randomly:



If you do not see this select the Overview tab at the top of the Gephi window. To see the labels for this network, click the **T** button in the menu. By default, Gephi will label the nodes:



This will make it easier to move the nodes around to achieve your desired layout. You can select one of the preset layouts provided by Gephi, but for this particular network it is almost easier to read the nodes if you manually perform the layout. To manually adjust the nodes select the drag tool:  and click on the node you wish to move, dragging it to the desired location. To center the graph in the window you can use the  tool, or right-click just off the graph in the white space and dragging the graph to the center of the window using the drag tool. To zoom-in, you can use the scroll wheel on your mouse, or you can click on the  button in the bottom right-hand corner of the window and this will open the menu that will allow you to zoom in.



Gephi 0.8.2 - Project 0

File Workspace View Tools Window Plugins Help

Overview Data Laboratory Preview

Parti... Ran... Graph

Clustering al... Settings

Dragging Hierarchy

Arts & Sciences Career Services Business School Of Public Health
 Ascs-q 299 Bus- 204 Hper-t 590
 Ascs-q 294 Department Of Apparel Merchandising And Interior Design
 School For Public And Environmental Health 301 Eng-w 350
 Department Of English Amer- 100
 Department Of History 131 Eng-w 170
 Spea-v 220 Spea-v 540 Hist-b 352 Hist-j 400 Educ-h 306 Gender Studies
 Department Of Comparative Literature Educ-k 362 Educ-u 212
 Department Of Communication Department Of Education 104
 Cmlt-c 110 Cmlt-c 121 Educ-k 305 Educ-x 152
 Department Of Telecommunications Learning Center Intl- 205
 Intl- 205
 Tel-t 311 Tel-t 207 Cllc- 100 International Relations Intl- 400
 Intl- 100
 Department Of Central Eurasian Studies Intl- 201
 Department Of Sociology Intl- 201
 Ceus-r 342
 Soc-s 321

Reset Run

Layout Noverlap Run

Noverlap	
speed	3.0
ratio	1.2
margin	5.0

Noverlap Presets... Reset

size A A Arial Bold, 32

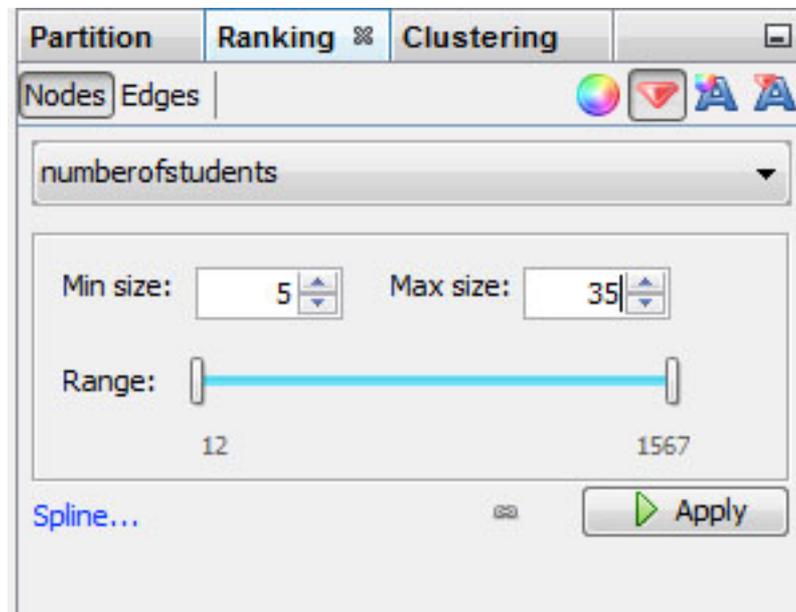
Global Nodes Edges Labels

Background color: Zoom Highlight selection

Autoselect neighbor

Workspace 0

Now we want to change the size of the nodes based on the number of students in those instruction sessions. By default the number of students will also be aggregated to the department nodes, so you will also be able to determine the total number of students in all the instruction sessions related to a particular department. To resize the nodes switch to the rank tab in the upper left-hand corner of the window and choose the node size icon:  Then set number of students as the rank parameter, and set the size range from 5 to 35:



Click apply and the nodes will be resized from 5 to 35 based on the number of students in the instruction sessions. By default Gephi will apply a linear scaling to the size of these nodes. This is fine if your data are normally distributed, but if your data are skewed to one end or the other, you might want to consider a different scale.

For example, these data are skewed to the upper-end of the range. The largest node in terms of the number of students is BUS-104 with 1487 students and the next closest node is CMCL-121 with 636 students. The majority of the nodes have fewer than 50 students. This means that it will be more difficult to see the differences between the sizes of the nodes at the lower end of the range.

We can apply a logarithmic scale to the sizes of our nodes that allows us to see the differences in the sizes of nodes with fewer students more clearly. To do so, click on the link the "Spline..." link below the size range and this will bring up the Interpolate window. You can see that the default scale is linear, but if you choose a logarithmic scale, such as the third option, you will be able to see the differences in nodes with fewer students more clearly:

Gephi 0.8.2 - Project 0

File Workspace View Tools Window Plugins Help

Overview Data Laboratory Preview

Partition Ranking **Clustering** Graph

Nodes Edges

numberofstudents

Min size: 5 Max size: 35

Range: 12 1567

Spline... Apply

Layout

Noverlap

Run

Noverlap	
speed	3.0
ratio	1.2
margin	5.0

Noverlap

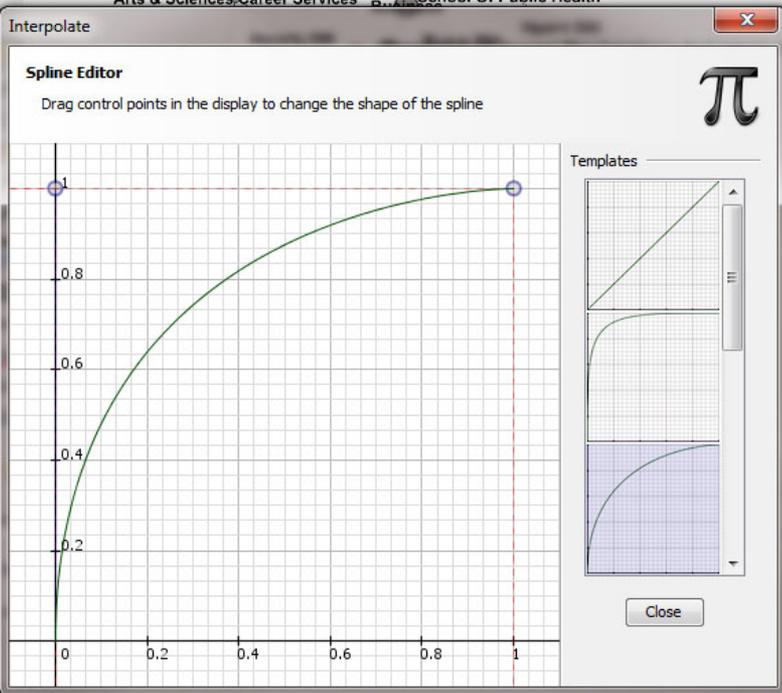
Presets... Reset

Arts & Sciences Career Services School Of Public Health

Interpolate

Spline Editor

Drag control points in the display to change the shape of the spline



Templates

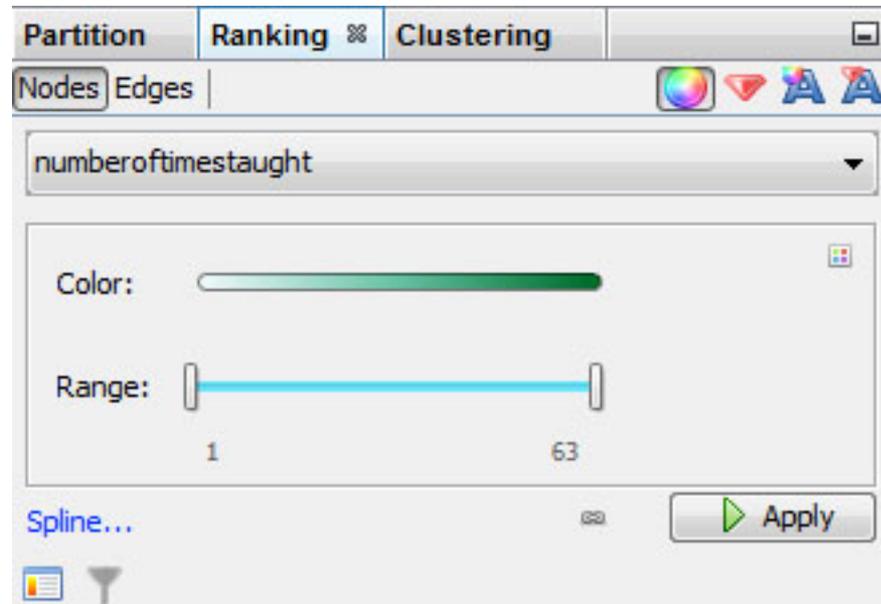
Close

Workspace 0

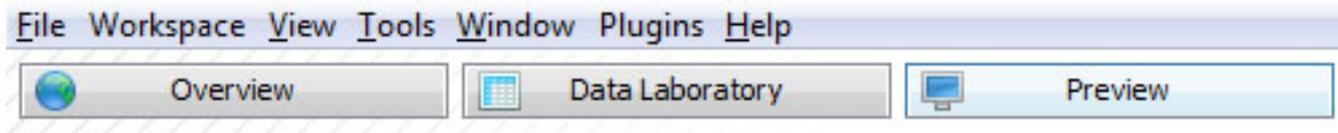
Once you have selected the desired scale, click Apply. Now the nodes will be resized in a way that allows you to more clearly see the differences in the sizes of the nodes with fewer students:

The screenshot displays the Gephi 0.8.2 interface. The main window shows a network graph with nodes of varying sizes, representing different departments and their student counts. The nodes are connected by edges, forming a complex network. The interface includes a menu bar (File, Workspace, View, Tools, Window, Plugins, Help) and a toolbar with various icons. The left sidebar contains several panels: 'Partition' (Nodes, Edges), 'Ranking' (numberofstudents, Min size: 5, Max size: 35, Range: 12 to 1567, Apply button), and 'Layout' (Noverlap, Run button, table with speed: 3.0, ratio: 1.2, margin: 5.0). The bottom panel shows 'Global' settings (Background color, Zoom, Highlight selection, Autoselect neighbor) and 'Nodes' settings (size, color, font). The graph itself shows nodes labeled with department names and student counts, such as 'Department Of English' (170), 'Department Of History' (131), 'Department Of Education' (152), and 'Department Of Central Eurasian Studies' (100). The nodes are connected by edges, forming a complex network.

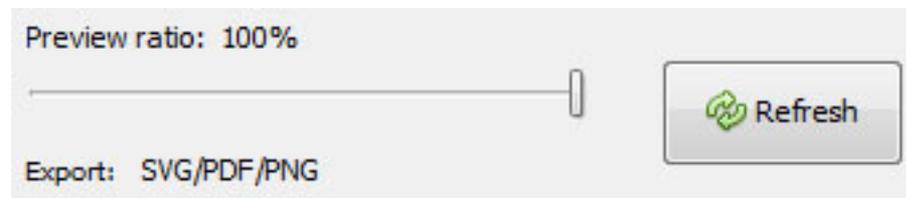
Now you will want to color the nodes based on how many times an instruction session has been taught, and again the department nodes will have an aggregation of the number of times each of the instruction sessions have been taught for courses in a department. To color nodes select the color tool:  and set the ranking parameter to number of times taught:



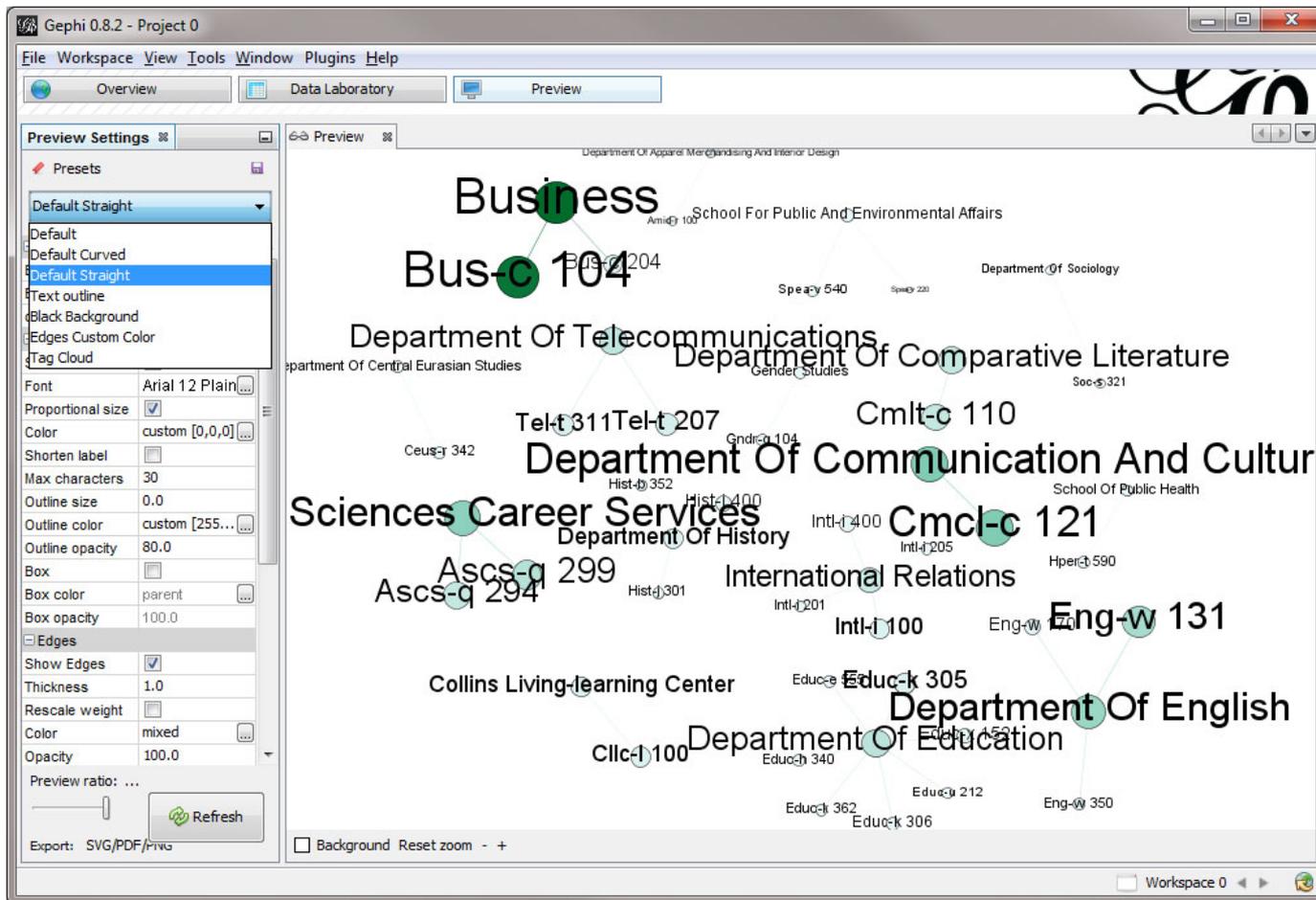
Now to see how the visualization will look when it is exported, go to the Preview tab at the top of the Gephi window:



You will probably need to hit the refresh button in the bottom left-hand corner of the Gephi window in order to see the visualization in the graph window:



You may want to choose a different preset layout. Here, the Default Straight layout has been applied:



This particular layout will display the node labels proportional to their size, which makes the graph easier to read and allows viewers to immediately identify important areas of the graph. Now to export the graph, select the export option in the bottom left-hand corner of the Gephi window:



Finally you can create a legend for your visualization in a visual editor such as Adobe Illustrator or Inkscape. To see how to use Inkscape, an open source vector image editing program, to create a legend for your visualization, see [2.4 Saving Visualizations for Publication](#). The final visualization could look something like this:

Instruction Sessions: Academic Year 2012

