**The Big “Book” of Using Gephi**

**1. Installation**

Just double-checking, you’ve installed Gephi and updated as per the software installation instructions. Right? Right.

**2. Network Types**

You can have a network with either one of two node or edge types.

Node types:

* Unipartite – all nodes are of the same type (i.e. every node is a person, every node is a word, etc.)
* Bipartite – there are two types of nodes (i.e. there are both people and jobs as nodes in your network, both words and article titles in the network, etc.)

Edge types:

* Directed – an edge goes from one node to another node in a specific manner (i.e. the 8:40 am flight goes from O’Hare to Charlotte)
* Undirected – an edge applies to both nodes (i.e. Tom and Jan are friends, the friendship is bi-directional and the edge goes both ways)

So now let’s break this down.

**First a digression and very serious warning**

Bipartite graphs have been very popular, (I believe) because their construction is very literal (i.e. connecting actors if they’ve been in the same movie together, you show both the actors and movies). It’s also possible to see, without using any advanced features of Gephi, some very basic node attributes (think how many actors are in a movie or how many movies has an actor be in).

The reality is that deeper analysis of Bipartite graphs is actually very challenging, far more than unipartite graphs. As scientists, i.e. the people at the edge of analyzing this topic, we frequently skip this added complexity by simply analyzing the unipartite version of the network (i.e. the network of actors alone, where two actors are connected if they have been in the same movie).

Adding to this challenge, there are not tools built into Gephi to layout a bipartite graph naturally or analyze network metrics beyond the degree (which could easily be done in Excel). To identify the communities of a bipartite graph properly, you actually need to import the unipartite graphs of both sides and run the modularity algorithm on each one separately to even get close to the real community structure.

So my thoughts: **do not build or analyze a bipartite graph for your homework assignments unless you are intentionally taking on a more difficult, complicated analysis project.**

I know that it seems odd that a unipartite network (which seems more unnatural) would be easier to analyze and understand—but that’s the magic of network analysis. Even though it **seems** like we’ve discarded a large amount of information about the problem, all of these relationships taken together can tell us a huge amount that would be hidden otherwise (and couldn’t be discovered through Excel!)

**Now which type of Unipartite network are you interested in building?**

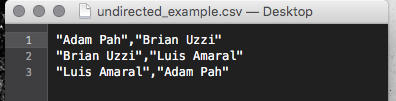
* Are you interested in how a quantity (information, memes, money, etc.) may flow through a network? **You want a directed network**.
* Are you interested in identifying groups within a network? **You want an undirected network.**

**2. Loading a network**

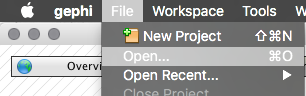
I will discuss how to load a network from the menu instead of from the splash screen.

**2.1 Undirected networks**

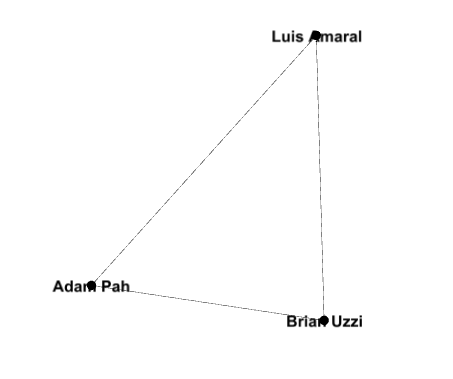
The simplest file format for an undirected network is an edgelist stored in a CSV file. A CSV file will have two columns separated by a comma, with a node name going on either side of the comma. Gephi does not expect any spaces or punctuation in the node name, so each node name should be enclosed within quotation marks as shown below.



From within Gephi, go to the `File` menu and then `Open`. Load the CSV file that you have created.



You should now see your network loaded into Gephi.



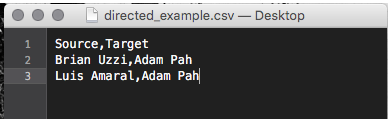
Luis, Brian, and myself are all connected because we form the director board of the Northwestern Institute on Complex Systems.

**2.2 Directed networks**

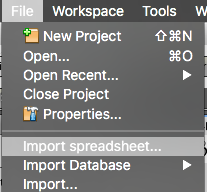
The simplest manner for a directed network is similarly a CSV file. There are three major differences:

* Your first row must be ‘Source,Target’
* You do not need the quotation marks around a node name
* Edges are directional, the edge goes *from* the node in the ‘Source’ column *to* the node in the ‘Target’ Column

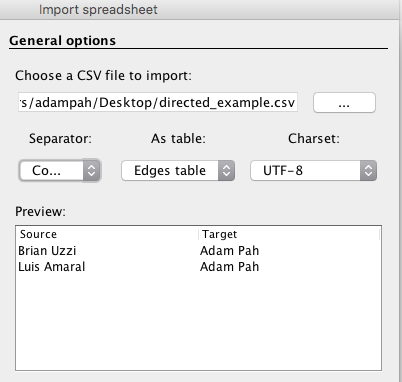
Again, I use the directors of NICO to illustrate this. Brian Uzzi and Luis Amaral are the co-Directors, while I am the Associate Director. If I wanted to illustrate how work is assigned, I would craft the network like so:



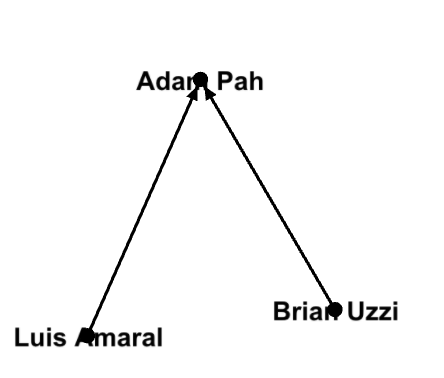
In Gephi, go to `File` and then `Import spreadsheet`



Click on `…` and choose the CSV file you have created. Make sure that `Separator` is put to `Comma` and `As table:` is put to `Edges table`. The `table` selector will determine if the loaded data goes into creating edges or nodes (make sure you are creating nodes).

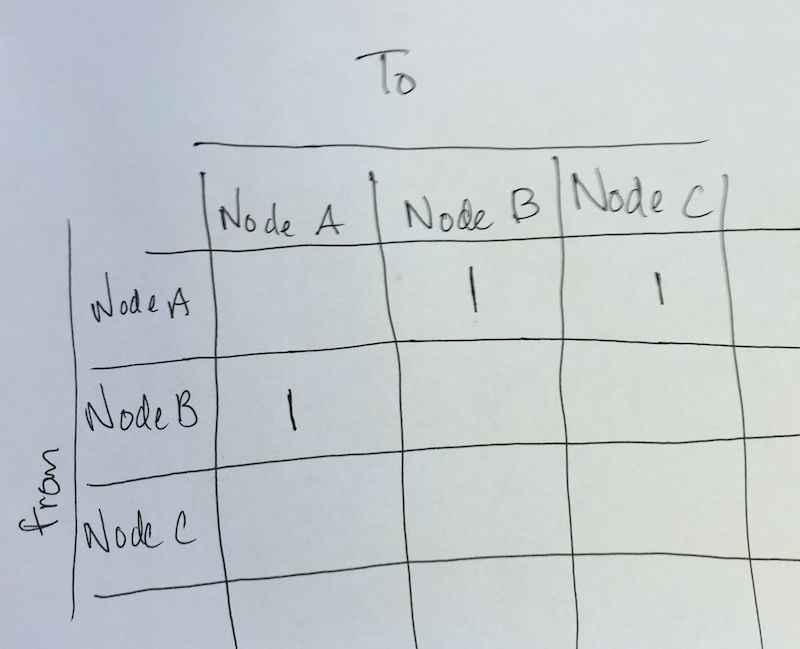


Click `Next` and then `Finish`. You should see your graph now.

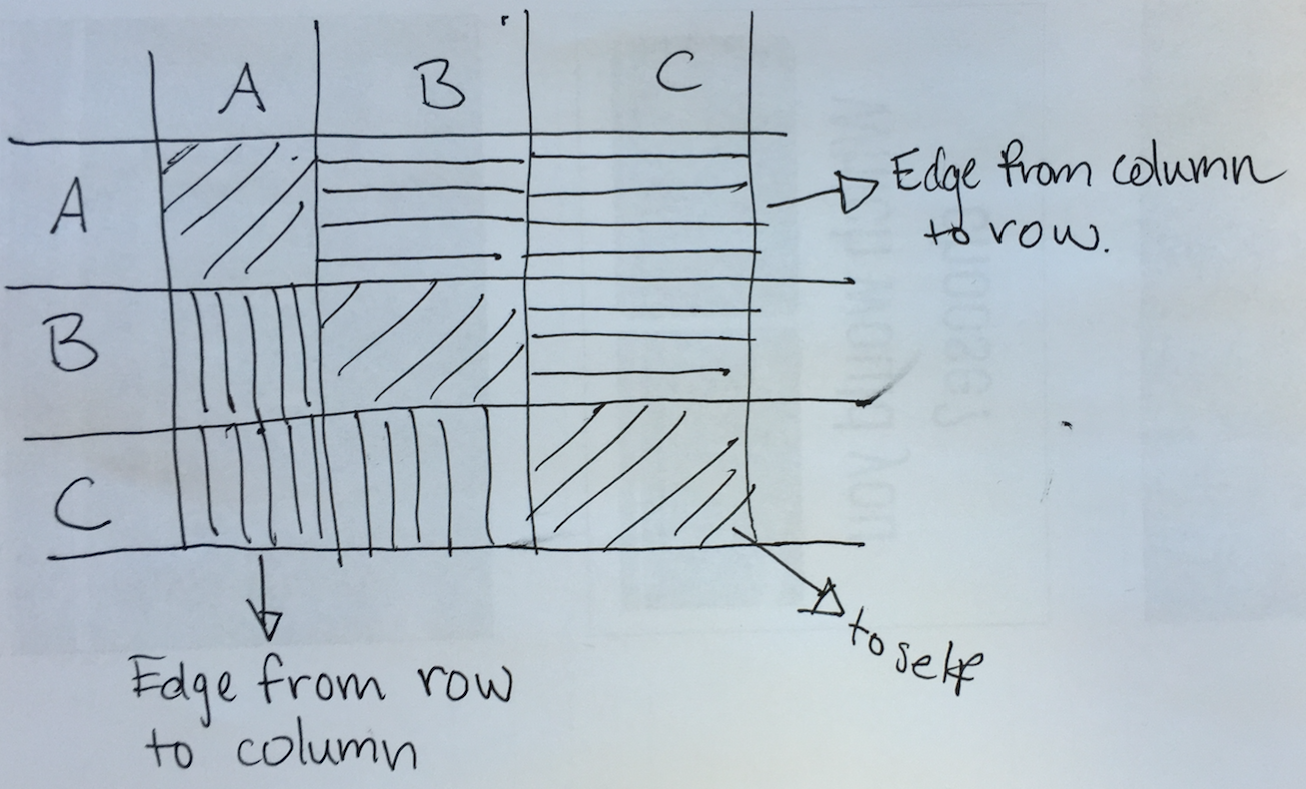


**2.3 Adjacency Matrices**

Adjacency matrices can handle either a directed or undirected network. An adjacency matrix is must be balanced (i.e. the same node is in the same position as both a row and column), like so:



This is called a symmetric matrix, whenever there should be an edge between two nodes we put a value of 1 (otherwise the value should be 0 or the cell left blank). The directionality of edges can be assigned like so:

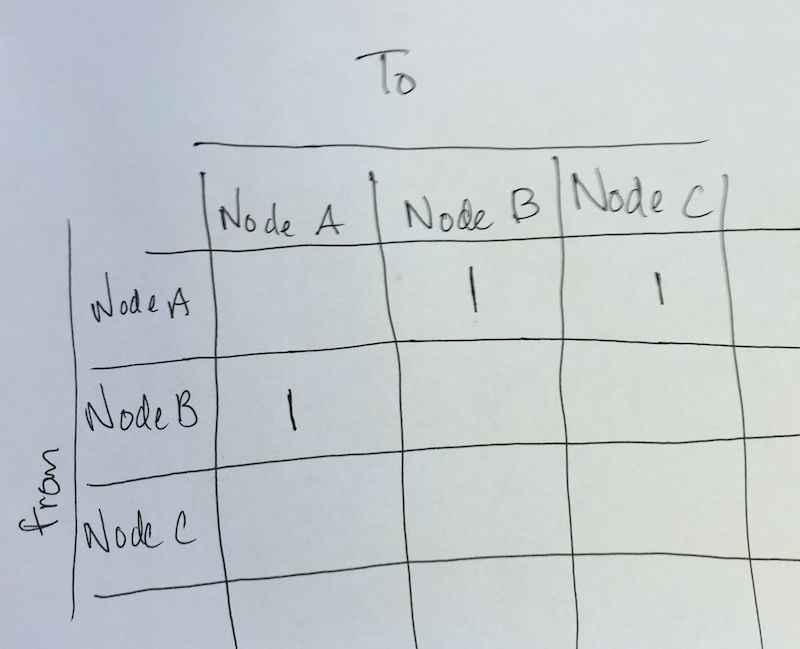


Putting a positive value on the diagonal means that it is a self-edge, I am connecting to myself (self-edges typically make sense in transactional networks-- payments to one-self, linking to your own web-page on-line, etc.).

Putting a positive value on the lower triangle (the bottom part of the matrix) means that an edge will go from the node in the row to the node in the column.

Putting a positive value in the upper triangle (the top part of the matrix) means that edge will go from the node in the column to the node in the row.

If you want an edge to be undirected (to point to both the node in the row and column), then there should be a positive value in the matrix on both sides of the diagonal. So if we go back to our original matrix:



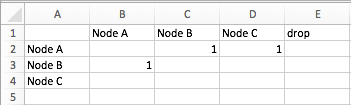
What we have are the following edges (starting from the upper left corner, reading every row in stream):

* Node A to Node B
* Node A to Node C
* Node B to Node C

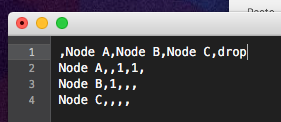
We can build this network in Excel, but we have to be mindful of a few things. Similarly, to the undirected edgelist, we cannot have spaces in a node name unless we start the name with a quotation mark.

To import an adjacency matrix into Gephi a `;` must separate every cell instead of a ‘,’. Also, every row must end with a `;`. Excel cannot save a file like this, so we must modify the Excel CSV in a text-editor to prep it for loading into Gephi.

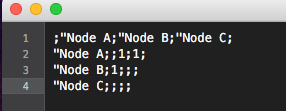
To start create a matrix in excel, make sure to add a column at the end (you will delete the name in a later step)



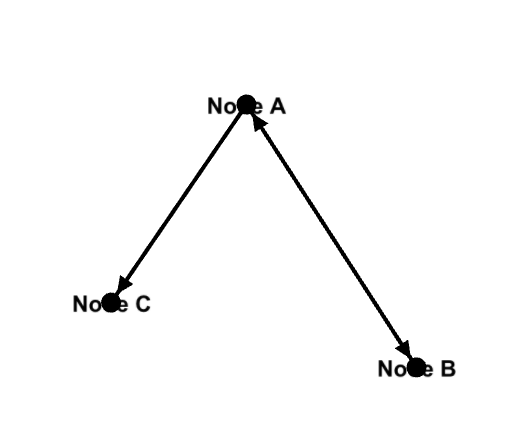
I have recreated our initial network here, now save the file as a CSV. Open the file in your text editor of choice like below:



Delete the word `drop` and use `Find and Replace` to change all commas to semi-colons. If there are spaces in your node names, escape each node name with a quotation mark.



Open Gephi, go to `File` and then `Open`. Load the adjacency matrix you have created.



You can see when we turn up the edge thickness that the edge goes to both Node A and B, while the other edge only goes from Node A to Node C.

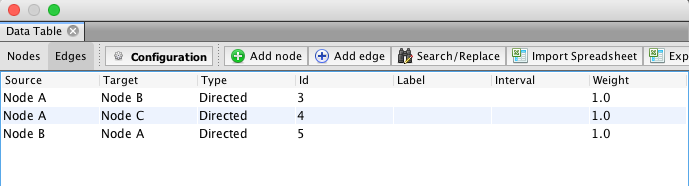
**3. Adding attributes to nodes or edges**

**3.1 Edge weights**

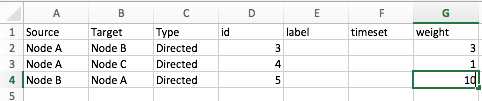
There are times when you want to convey that some relationships are stronger than each other. To do this you will want to construct a ‘weighted’ network.

**3.1.1 Weighted, Directed and Undirected Networks**

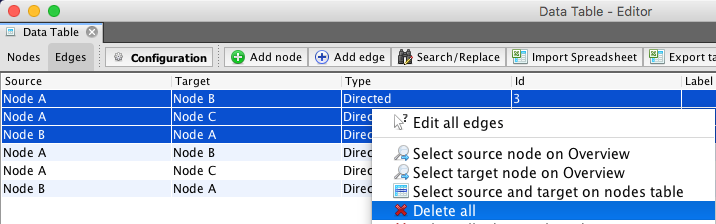
Load the network as a CSV file as instructed above. Go to the `Data Laboratory` and make sure to click on `Edges` in the upper left.



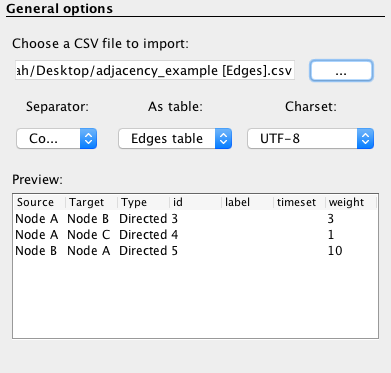
Export the edges to a CSV file using the `Export table` button. Open the file in Excel and change the `Weight` value for each edge to reflect the weight.



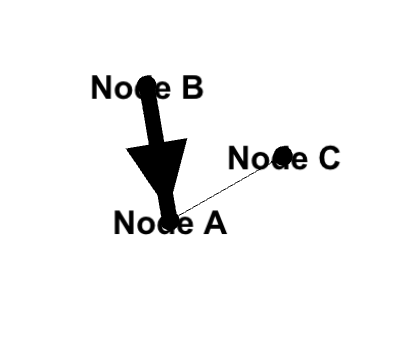
Save the file. Go back to Gephi and from the `Data Table` window. Select all the edges, then right-click and choose `Delete all`



Now click `Import Spreadsheet`.

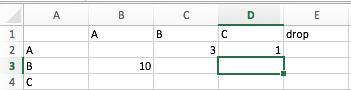


Choose the edited file to load, make sure that `As table:` is set to `Edges table`, then click `Next` and `Finish`. You will now have edges with weights.

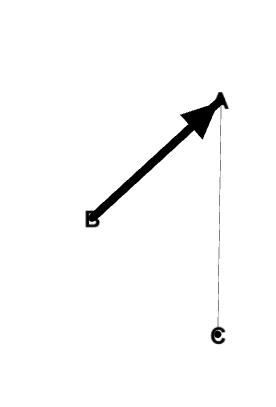


**3.1.2 Adjacency Matrices**

You can add weights to an adjacency matrix directly. Instead of putting a `1` in the matrix, put a different value to represent the weight strength.



Load the matrix as instructed above and you will see the edge-weights.



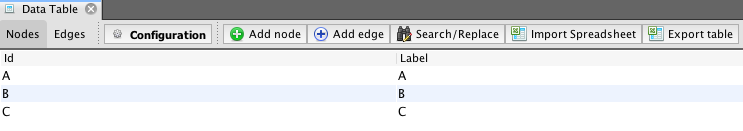
**3.2 Other attributes**

You may also want to add other attributes, such as a categorical or numerical variable that you want to visualize in addition to the raw network. Categorical variables would be something like gender for people and a numerical variable would be something like income.

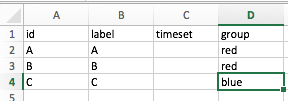
**3.2.1 Categorical variables**

Categorical variables can be used to color nodes (but not size) from the appearance tab.

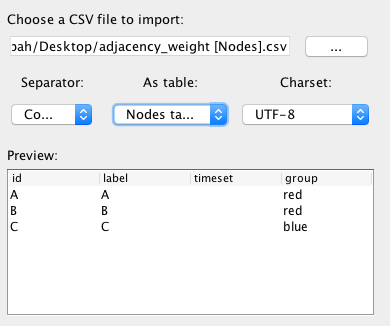
To add a categorical variable to your nodes go to the `Data Laboratory` tab and click `Export table`. Make sure that you are in the `Nodes` tab in the Upper left-hand corner.



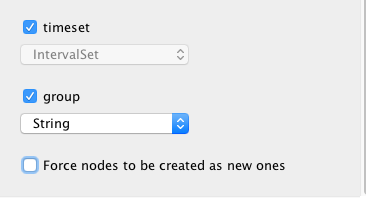
Open the exported CSV file in Excel and add your categorical variable.



Import the modified CSV file in Gephi using the `Import spreadsheet` button. Make sure that `As table:` is changed to `Nodes table`. Click the `Next` button.

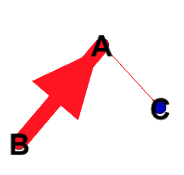


Scroll down to the bottom of the summary, uncheck the `Force nodes to created as new ones` box.



Click `Finish`.

Go back to the `Overview` pane in Gephi. Using the appearance menu select the `Nodes`, then the Palette symbol, then `Attributes`. You can choose the newly created variable from the dropdown and color the nodes using the `Apply` button.

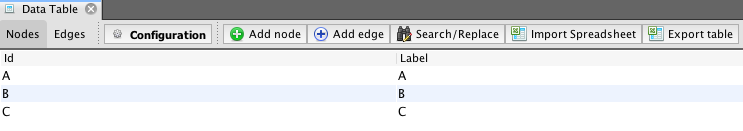


This would be the resulting image of coloring the nodes by the group colors.

**3.2.2 Numerical variables**

Numerical variables can be used to color or size nodes from the appearance tab.

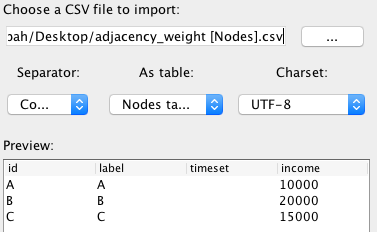
To add a numerical variable to your nodes go to the `Data Laboratory` tab and click `Export table`. Make sure that you are in the `Nodes` tab in the Upper left-hand corner.



Open the exported CSV file in Excel and add your categorical variable.

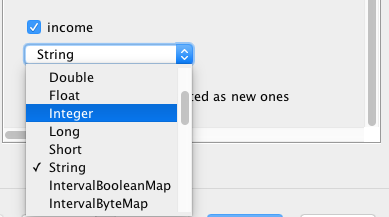


Import the modified CSV file in Gephi using the `Import spreadsheet` button. Make sure that `As table:` is changed to `Nodes table`. Click the `Next` button.



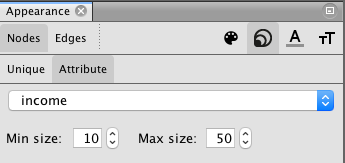
Scroll down to the bottom of the summary, uncheck the `Force nodes to created as new ones` box.

**Now make sure to change the data type of your new variable.** You should change your data type to either `Integer` (if your variable is 1, 2, 3, etc.) or `Float` (1.5, 2.5, 3.5, etc.)---Make sure to scroll up if you can’t find these options.

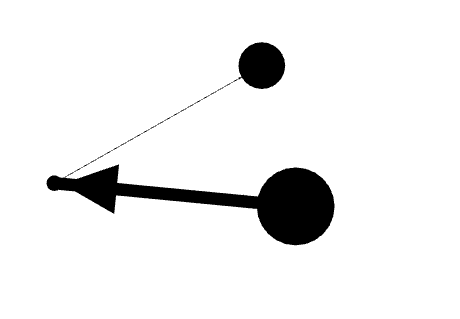


Click `Finish`.

Go back to the `Overview` pane in Gephi. Using the appearance menu select the `Nodes`, then the Size symbol, then `Attributes`. You can choose the newly created variable from the dropdown and color the nodes using the `Apply` button.



This would be the resulting image of sizing the nodes by `income`.

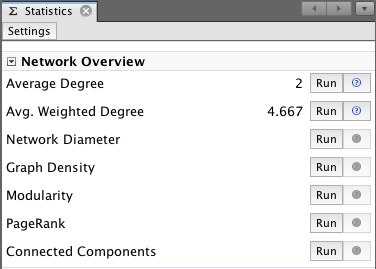
****

**3.3 Analysis on attributes outside of Gephi**

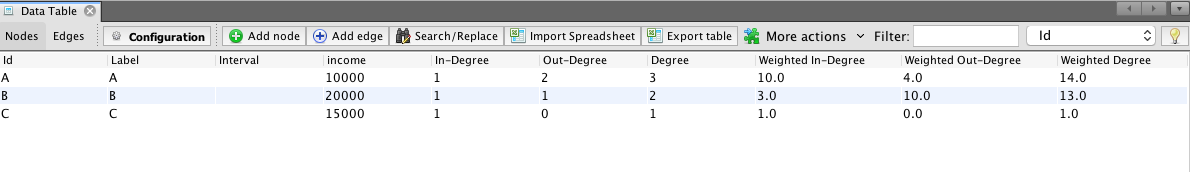
Using additional programs (such as Excel, Stata, R, Python, etc.) to perform further statistical analysis is always welcomed. The workflow for this is:

Load Network > Run Statistic > Export Node table > Perform analysis

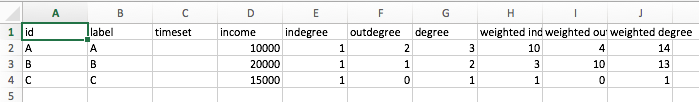
As an example, if you wanted to use node degree in a model here is the workflow. From the `Statistics` pane click the `Run` button next to `Average Degree` and `Avg. Weighted Degree`.



From here go to the `Data Laboratory` pane

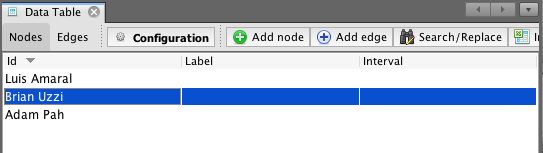
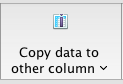
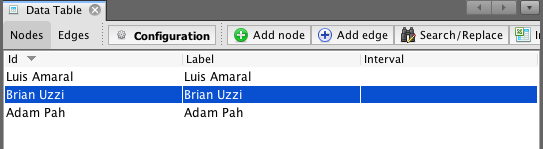
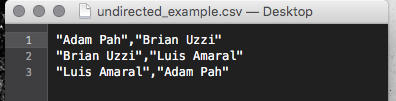


Click the `Export table` button, now use this excel spreadsheet to add in further analysis.



From there you can put the data into any other program that you want for further analysis.

**FAQ**

1. ***I ran the Modularity algorithm and there aren’t any groups? I ran the Modularity algorithm and it returned a value of zero?***  
   The modularity algorithm does not work with directed edges (this type of information isn’t a part of how the algorithm is implemented in Gephi). To run the modularity algorithm, import the network as an undirected network.  
     
   If you also want to study the flow using the directed network with groups, follow this workflow:  
   \* Import the network as an undirected one  
   \* Run the modularity algorithm  
   \* Export the node table (which will now have community assignments)  
   \* Load the network as a directed one  
   \* Import the node table with the community assignments
2. ***I imported my network, but I don’t see the graph in the window?***Did you update Gephi? Double-check and try to update it (I have not had this be an issue with Gephi after it was updated).   
     
   If there are no updates and you followed the steps in this guide to create the network, then contact either the current Teaching Assistant or myself.  
     
   When you contact one of us, include your network file so that we can diagnose your issue.
3. ***I don’t see labels on my nodes when I turn on the labels***This is because Gephi only has a node id for the node and not a label (Gephi displays labels when you turn on text). To add a label:  
   \* Go to the `Data Laboratory`  
   \* Click on `Nodes` (Notice it in the Upper Left. Make sure you are not on the `Edges` table)  
     
   \* Go down to the bottom and click on the `Copy data to other column`  
     
   \* Select `Id` after clicking the button and then choose `Label` from the dropdown menu. You should now see the same name in both `Id` and `Label`  
   
4. ***I’m having an issue with my node names.  
   There are more nodes in Gephi than in my spreadsheet.  
   The node labels are incorrect.***  
   This issue is likely related to having a space in the node name and importing through the `File` -> `Open` method  
     
   You can check if this is an issue by looking at all of the node names in the `Data Laboratory`   
     
   If you have this issue, edit the original network file and enclose each node name in quotation marks, like below:  
   
5. ***Gephi won’t load the file I’ve created with non-English names/irregular symbols***  
   So there isn’t actually a way around this. The trick is to just use `Find and Replace` in the program of your choice and remove the punctuation/non-English characters.