Social and Economic Networks
Gephi Instructions

Prof. Matthew Jackson

Network Analysis with Gephi Software (Version 0.9.1):

Downloading the Data Files : **When you are downloading the data files associated with the problem sets, then make sure that you Right click on a link that has the data set and then click 'save link as' and make sure it has the same file extension it started with. If instead you simply click on the file, some browsers will corrupt the file when saving it. In particular, many browsers will try to save a ".net" file as a ".txt" file (it will add the suffix and in some cases also corrupt the file). You can save a ".net" file properly using Chrome as follows. Go to the link. Then right click on the page and 'save as' - it will try to save it as text - so instead click on "all files" in the tab below. Then you need to remove the .txt that it will try to save with the name.**

In addition to the links on the course site, you can also find the data for the first problem set here: https://www.stanford.edu/∼jacksonm/26KeroNetwork.net

Note:

Download Gephi Gephi is free, available on its official website:
gephi.org/users/download.

It works for both Windows and Mac. (If you have problems opening Gephi, then it might be the version of Java - it seems to be sensitive to the precise version. If you are having troubles you can download a version that works here: Java update and you can read more about it here)

Gephi has undergone many changes, and may continue to evolve. Some of the instructions below may become obsolete, and so you may wish to explore updated tutorials if something seems not to work. If Gephi gives you headaches, you may prefer Pajek, which is somewhat simpler to use but is more cumbersome across operating systems.

You can find tutorials here

If you are downloading some of the data files associated with the problem sets, then make sure that you Right click on a link that has the data set and then click

'save link as' and make sure it has the same file extension it started with. If instead you simply click on the file, some browsers will corrupt the file when saving it.

Directions for the first problem set:

Download the data set (according to the directions above and remember where you save it!):

Open Gephi and load the data set:

- go to the "File" tab at the top
- select "Open" in the dropdown
- find the directory where you have stored the data, and open it. (If it gives you a problem, then instead go directly to the data set, right click on it, and when it say "Open with" find Gephi and open with it.)
- When opening a .net file, you can tell Gephi how to interpret the links; for the option "Graph Type", you can choose "Undirected" or "Directed" or "Mixed"
- **Make sure that when you open 26KeroNetwork.net that you change the entry of the Import Report's "Graph Type" option to "Undirected", as the default is "Mixed" and it will change some of the calculations Gephi makes.**

Using Gephi to calculate the Average Degree of the network

- find the window "Statistics" in your screen (if you cannot find it, you can open it by clicking "Window" tab at the top, then select "Statistics")
- under "Network Overview", find "Average Degree" and click "Run"
- it will generate a new window from which you can find the average degree and degree distribution
- Note: if the network is *directed*, Gephi will report the average total degree, which is the sum of in and out degrees.

Using Gephi to calculate Diameter of the network:

- under "Network Overview", find "Network Diameter" and click "Run"
- it will generate a new window from which you can find the Diameter, as well as other information about average path length (under "Results")
- record the diameter and average path length
- Note that these calculations again will depend on how Gephi is treating the data: for 26KeroNetwork.net make sure that it is being treated as "Undirected"

Using Gephi to calculate calculate Clustering coefficients for each node:

- find the window "Statistics" in your screen
- under "Node Overview", find "Avg. Clustering Coefficient" and click "Run"
- it will generate a new window from which you can find the distribution of clustering coefficients, and the average clustering coefficient
- the average clustering is listed by the "Run" button you just pushed
- record the average clustering
- if you want to find clustering coefficients for each node, click "Data Laboratory" at the top of the program
- choose "Nodes"
- you will find node-level data, including "Clustering coefficients"
- you can order the table based on clustering coefficients by clicking that tab

If you want to draw the network:

Go back to the overview window to draw the figure. Click the arrow at the bottom right to see the various sizing options.

To color nodes: click on the "Nodes" tab in the "Appearance" window on the left, then the "Unique" tab under "Nodes". Click and hold on the square next to the hexidecimal color code (e.g. #ff001f), hover over the color of your choice, and release. Then hit the "Apply" button.

You can rule various "Layout" algorithms and hit "Stop" - by playing with the "Optimal Distance" and "Repulsion" etc. settings, you can get the picture to fit better.

You can "size" nodes by clicking on the size graphic in the "Appearance" window when the "Nodes" option is selected (the size graphic consists of three circles, with the smallest circle filled in). Then, change the size value to the desired value and hit "Apply".

Directions for the second problem set:

Download the data set and remember where you save it (and also remember to see the note at the beginning of these directions about how to save the file so that it is not corrupted!):

In addition to the links on the course site, you can also find the data here:
https://www.stanford.edu/∼jacksonm/imports_manufactures.net

Open Gephi and load the data set:

- go to the "File" tab at the top
- select "Open" in the dropdown
- find the directory where you have stored the data, and open it. For the exercises involving the imports_manufactures.net graph, choose "Graph Type" as "directed".

Using Gephi to calculate some basic centrality measures:

- click the "Overview" tab
- find the window "Statistics" in your screen
- under "Network Overview", find "Network Diameter" and click "Run"
- choose "directed", and check "Normalize Centralities in [0,1]"
- click the "Data Laboratory" tab at the top of the program
- choose "Nodes"
- find "Closeness Centrality"

  - you can order the table based on clustering coefficients by clicking that tab
  - find the node with highest Closeness Centrality, and please ignore the four nodes with CC=1 (Barbados, Romania, Trinidad Tobago, and Tunisia)
- what is the highest level of "Betweenness Centrality"?

Using Gephi to generate Poisson random networks:

- go to the "File" tab at the top of your screen
- first pick new project
- then go back to the "File" tab and choose "Generate" and "Random Graph..."
- input "400" when asked "How many nodes"
- input the "wiring probability" you want
  - choose a wiring probability around $ave.degree/400$
  - note Gephi only generates directed network (unless you use other plugins), but you can treat it as undirected by ignoring the directions of the links.

Examine whether the network is connected:

- find the window "Statistics" in your screen (back under "Overview")
- under "Network Overview", find "Connected Components", choose "undirected" and click "Run"
- it will report the number of connected components

To redo the above, do not forget to click "New Project" under the "File" tab first.

Directions for the third problem set:

Download the data set (and remember where you save it, and also remember to see the note at the beginning of these directions about how to save the file so that it is not corrupted!)::

In addition to the links on the course site, you can also find the data here:
https://www.stanford.edu/∼jacksonm/Ring25.net

Using Gephi to add a link

- click the "Data Laboratory" tab at the top of the program
- click "Add edge" (on top of the current window, "Data Table")
- click "undirected"
- choose "Source node" and "Target node" as desired (since you are adding an undirected link, either node can be the source node)
- click "Ok"

Using Gephi to delete a link

- click the "Data Laboratory" tab at the top of the program
- change from "Nodes" to "Edges" under "Data Table"
- click on the edge you wish to delete
- either hit delete on keyboard, or right click and choose "Delete"

Download the data sets (and remember where you save them, and also remember to see the note at the beginning of these directions about how to save the file so that it is not corrupted!):

In addition to the links on the course site, you can also find the data here:
https://www.stanford.edu/∼jacksonm/NetScience.net
https://www.stanford.edu/∼jacksonm/YeastS.net

Open Gephi (or start a new project) and load the data set YeastS.net:

- go to the "File" tab at the top
- select "Open" in the dropdown
- find the directory where you have stored the data "YeastS.net", and open it.

- while opening, please choose graph type as "undirected", and unclick "create missing nodes" under "More options"

Using Gephi to examine the component size distribution:

- find the window "Statistics" in your screen (if you cannot find it, you can open it by clicking "Window" tab at the top, then select "Statistics")
- under "Network Overview", find "Connected Components" and click "Run"
- "undirected" will be your only choice
- it will generate a new window from which you can find the size distribution, as well as the number of weakly connected components.

To find the size of the giant component:

- find the window "Appearance" and choose "Nodes"
- choose "Attribute"
- choose "Component ID" in "—Choose an attribute"
- report the percentage for the largest component

To import a spreadsheet, start with the "File" tab and click "New Workspace"

Next, in the "Data Lab" click on "nodes" and then on "Import Spreadsheet" and make sure that "Nodes Table" is clicked on the popup window (and the file is a csv file). Import the file.

Next, do the same for the "edges" and **be sure that the "Edges table" is clicked in the popup window**.

Have fun!