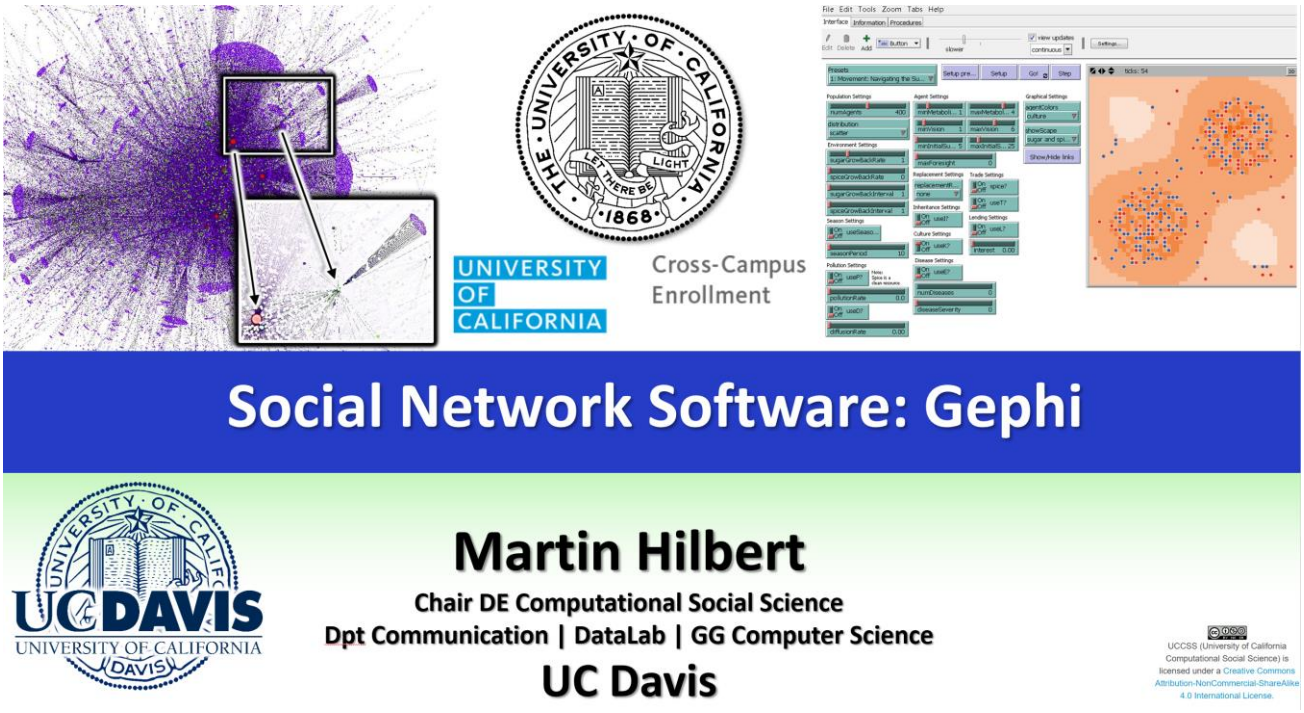


Contents

1. Data wrangling: cleaning and preparing the Data for Gephi 2

2. Installation (Gephi) 5

3. Visualize the Networks 6

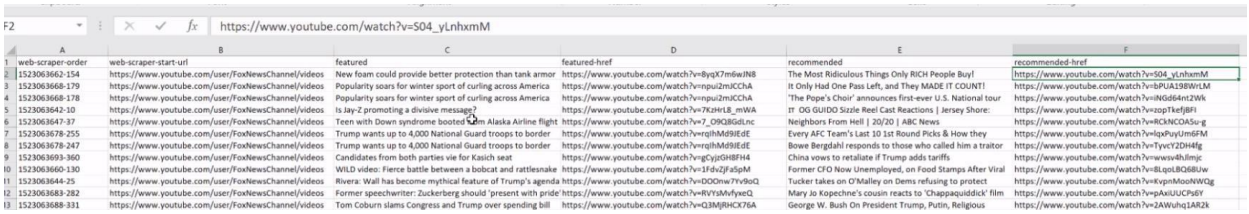
4. Analyze the Networks 10

A prerequisite for this lab is that you have web-scraped two YouTube channels (which is a previous lab).

1. Data wrangling: cleaning and preparing the Data for Gephi

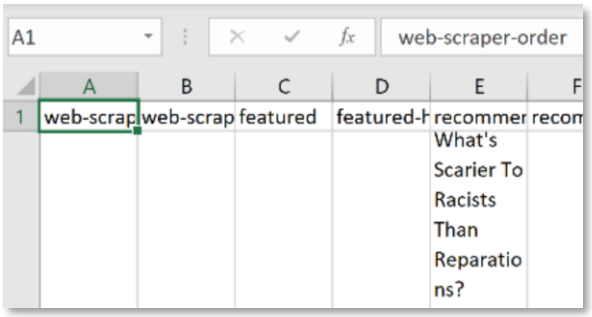
1. Once you have completed webscraping for two channels, open your spreadsheet. Your spreadsheet should have six columns:

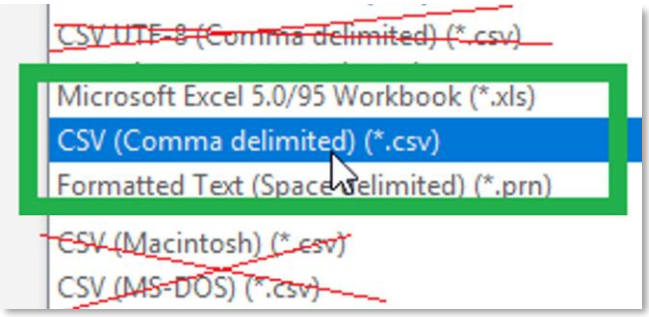
   web-scraper-order | web-scraper-start-url | featured | featured-href | recommended | recommended-href



	A	B	C	D	E	F
1	web-scraper-order	web-scraper-start-url	featured	featured-href	recommended	recommended-href
2	1523063662-154	https://www.youtube.com/user/FoxNewsChannel/videos	New foam could provide better protection than tank armor	https://www.youtube.com/watch?v=8yqX7m6wJN8	The Most Ridiculous Things Only RICH People Buy!	https://www.youtube.com/watch?v=S04_yLnhxmM
3	1523063668-179	https://www.youtube.com/user/FoxNewsChannel/videos	Popularity soars for winter sport of curling across America	https://www.youtube.com/watch?v=npui2mJCChA	It Only Had One Pass Left, and They MADE IT COUNT!	https://www.youtube.com/watch?v=bPUA198WrLM
4	1523063668-178	https://www.youtube.com/user/FoxNewsChannel/videos	Popularity soars for winter sport of curling across America	https://www.youtube.com/watch?v=npui2mJCChA	'The Pope's Choir' announces first-ever U.S. National tour	https://www.youtube.com/watch?v=iNGd64nt2Wk
5	1523063642-10	https://www.youtube.com/user/FoxNewsChannel/videos	Is Jay-Z promoting a divisive message?	https://www.youtube.com/watch?v=7KzHrL8_mWA	IT OG GUIDO Sizzle Reel Cast Reactions | Jersey Shore:	https://www.youtube.com/watch?v=zopTkefjBFI
6	1523063647-37	https://www.youtube.com/user/FoxNewsChannel/videos	Teen with Down syndrome booted from Alaska Airline flight	https://www.youtube.com/watch?v=7_O9Q8GdLnc	Neighbors From Hell | 20/20 | ABC News	https://www.youtube.com/watch?v=RCkNCOA5u-g
7	1523063678-255	https://www.youtube.com/user/FoxNewsChannel/videos	Trump wants up to 4,000 National Guard troops to border	https://www.youtube.com/watch?v=rqIhMd9JEdE	Every AFC Team's Last 10 1st Round Picks & How they	https://www.youtube.com/watch?v=lqxPuyUm6FM
8	1523063678-247	https://www.youtube.com/user/FoxNewsChannel/videos	Trump wants up to 4,000 National Guard troops to border	https://www.youtube.com/watch?v=rqIhMd9JEdE	Bowe Bergdahl responds to those who called him a traitor	https://www.youtube.com/watch?v=TyvcY2DH4fg
9	1523063693-360	https://www.youtube.com/user/FoxNewsChannel/videos	Candidates from both parties vie for Kasich seat	https://www.youtube.com/watch?v=gCyjzGH8FH4	China vows to retaliate if Trump adds tariffs	https://www.youtube.com/watch?v=wwsv4hJlmjc
10	1523063660-130	https://www.youtube.com/user/FoxNewsChannel/videos	WILD video: Fierce battle between a bobcat and rattlesnake	https://www.youtube.com/watch?v=1FdvZjFa5pM	Former CFO Now Unemployed, on Food Stamps After Viral	https://www.youtube.com/watch?v=8LqoLBQ68Uw
11	1523063644-25	https://www.youtube.com/user/FoxNewsChannel/videos	Rivera: Wall has become mythical feature of Trump's agenda	https://www.youtube.com/watch?v=DOOnw7Yv9oQ	Tucker takes on O'Malley on Dems refusing to protect	https://www.youtube.com/watch?v=KvpnMooNWQg
12	1523063683-282	https://www.youtube.com/user/FoxNewsChannel/videos	Former speechwriter: Zuckerberg should 'present with pride'	https://www.youtube.com/watch?v=RVYsMvfyxeQ	Mary Jo Kopechne's cousin reacts to 'Chappaquiddick' film	https://www.youtube.com/watch?v=pAxiUUCPs6Y
13	1523063688-331	https://www.youtube.com/user/FoxNewsChannel/videos	Tom Coburn slams Congress and Trump over spending bill	https://www.youtube.com/watch?v=Q3MjRHCX76A	George W. Bush On President Trump, Putin, Religious	https://www.youtube.com/watch?v=2AWuhq1AR2k

   If your spreadsheet does not look as nicely organized as this one, the default in your spreadsheet software might be that the text in the cells is "wrapped". Find out how to deal with this online, for example: https://www.google.com/search?q=excel+wrap+text Also, make sure you save the file as plain "CSV", when saving it to your computer ("save as"):





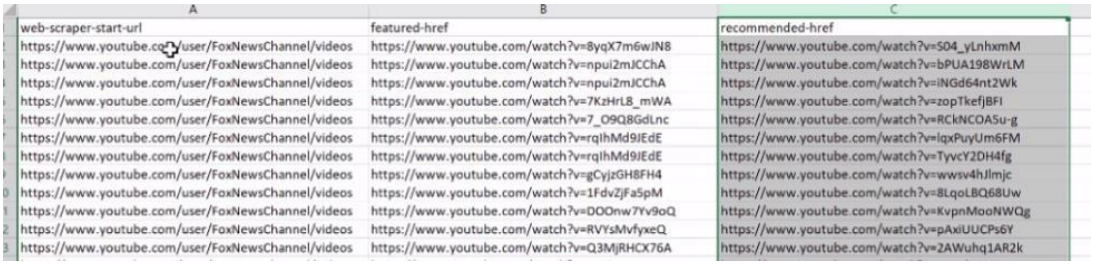
   In general, as always when struggling with a computational task, simply do an online search where you describe your problem: almost certainly someone else already found a solution for it!

2. Next, it's time to clean your data (spreadsheet) for Social Network Analysis. Since we are interested in featured and recommended videos – we will delete the **web-scraper-order** column. If you're unsure how to delete columns, see: https://www.google.com/search?q=excel+delete+column or https://www.google.com/search?q=delete+columns+spreadsheet

3. Next, we will also delete the **featured** and **recommended** columns since it's more solid to work with URLs (no two videos will have the same URL but they can have the same title).

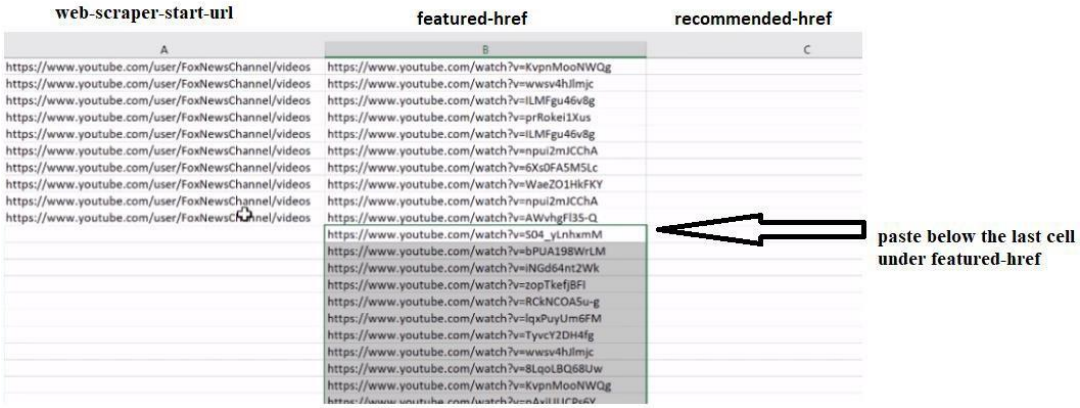
4. At this stage, your sheet should have three columns

   web-scraper-start-url | featured-href | recommended-href



	A	B	C
	web-scraper-start-url	featured-href	recommended-href
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=8yqX7m6wJN8	https://www.youtube.com/watch?v=S04_yLnhxmM
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=npui2mJCChA	https://www.youtube.com/watch?v=bPUA198WrLM
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=npui2mJCChA	https://www.youtube.com/watch?v=iNGd64nt2Wk
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=7KzHrL8_mWA	https://www.youtube.com/watch?v=zopTkefjBFI
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=7_O9Q8GdLnc	https://www.youtube.com/watch?v=RCkNCOA5u-g
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=rqIhMd9JEdE	https://www.youtube.com/watch?v=lqxPuyUm6FM
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=rqIhMd9JEdE	https://www.youtube.com/watch?v=TyvcY2DH4fg
	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=gCyjzGH8FH4	https://www.youtube.com/watch?v=wwsv4hJlmjc
0	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=1FdvZjFa5pM	https://www.youtube.com/watch?v=8LqoLBQ68Uw
1	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=DOOnw7Yv9oQ	https://www.youtube.com/watch?v=KvpnMooNWQg
2	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=RVYsMvfyxeQ	https://www.youtube.com/watch?v=pAxiUUCPs6Y
3	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=Q3MjRHCX76A	https://www.youtube.com/watch?v=2AWuhq1AR2k

5. Since we are interested in what users are exposed to on this YouTube channel, we will combine the featured and recommended videos under the same column. To do this, **cut** (*e.g. select all the cells under recommended-href and CUT function in your spreadsheet software*) all the URLs under the **recommended-href** column and **paste** (*PASTE function in your spreadsheet software*) below the last cell under the **featured-href** column.



6. The **web-scraper-start-url** (the source) will be empty for the copied cells. Since all these videos also come from the same channel (the same scraped news channel), we will copy the source URL from the **web-scraper-start-url** column and paste it in all empty cells.

7. Once you have done that, you should have two columns in your spreadsheet. First the **web-scraper-start-url** column with the channel link (in this case, https://www.youtube.com/user/FoxNewsChannel/videos) and the **featured-href** with all the video URLs from the particular channel

   Note each cell under the **web-scraper-start-url** column should have the **same channel URL!**

8. Gephi needs the data to be prepared in a certain format. So next, we will rename **web-scraper-start-url** column as **Source**, the **featured-href** column as **Target,** and **recommended-href** as **Link**. It should look something like the figure below!

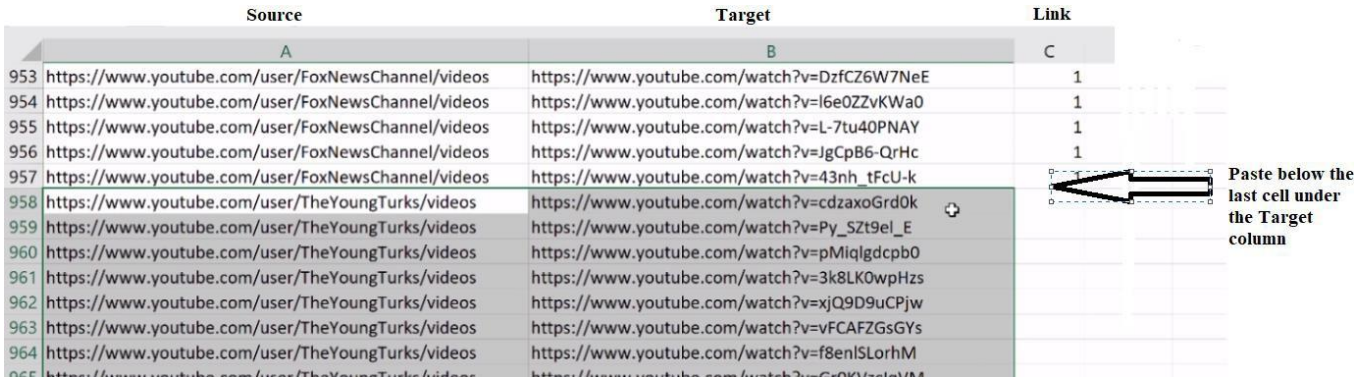
	A	B	C
1	Source	Target	Link
2	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=8yqX7m6wJN8	
3	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=npui2mJCChA	
4	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=npui2mJCChA	
5	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=7KzHrL8_mWA	
6	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=7_O9Q8GdLnc	
7	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=rqlhMd9JEdE	
8	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=rqlhMd9JEdE	

9. Next, we will input a value of 1 to **ALL** cells under the **Link** column (in this case from row 2 until your last row).

950	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=eqceHviNBC4	1
951	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=Q3MjRHCX76A	1
952	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=T8y1Q_4tDUU	1
953	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=DzfCZ6W7NeE	1
954	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=l6e0ZZvKWa0	1
955	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=L-7tu40PNAY	1
956	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=JgCpB6-QrHc	1
957	https://www.youtube.com/user/FoxNewsChannel/videos	https://www.youtube.com/watch?v=43nh_tFcU-k	1
958			
959			

Congratulations! The data sheet for one of your channels is ready for Gephi! Now, follow the same procedure from **step 1 so step 7** from this section and prepare the data spreadsheet for the other channel. Then combine both channels (the order doesn't matter).

10. Copy all the cells from the **web-scraper-start-url** and **featured-href** columns **of one of your** prepared data spreadsheet. Open the data spreadsheet you prepared for the **other channel**, and paste the copied cells in the data spreadsheet you prepared for the second channel. Paste the copied cells under the last row (in this case).



11. Make sure that the value of 1 is in ALL the blank cells under the **Link** column.

12. Your data spreadsheet is now cleaned and prepared. You have a list of videos, approximately half of them should be from the first channel you scraped last week and the other half from the channel you scraped this week. Go ahead and save the file on your desktop as a **CSV (comma delimited)(*.csv)** file. You can name the file as "youtubelinks"

Congratulations, we are done with preparing your spreadsheet for our analysis in Gephi!

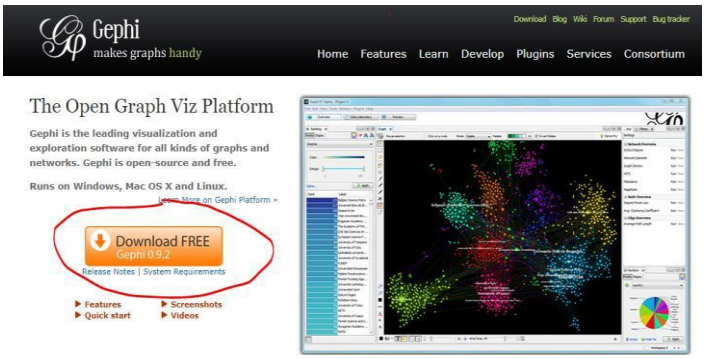
2. Installation (Gephi)

Gephi is the leading visualization and exploration software for all kinds of graphs and networks. Gephi is open-source and free. It runs on Windows, Mac OS X and Linux.

Please download and install the software

1. Go to the link: https://gephi.org/

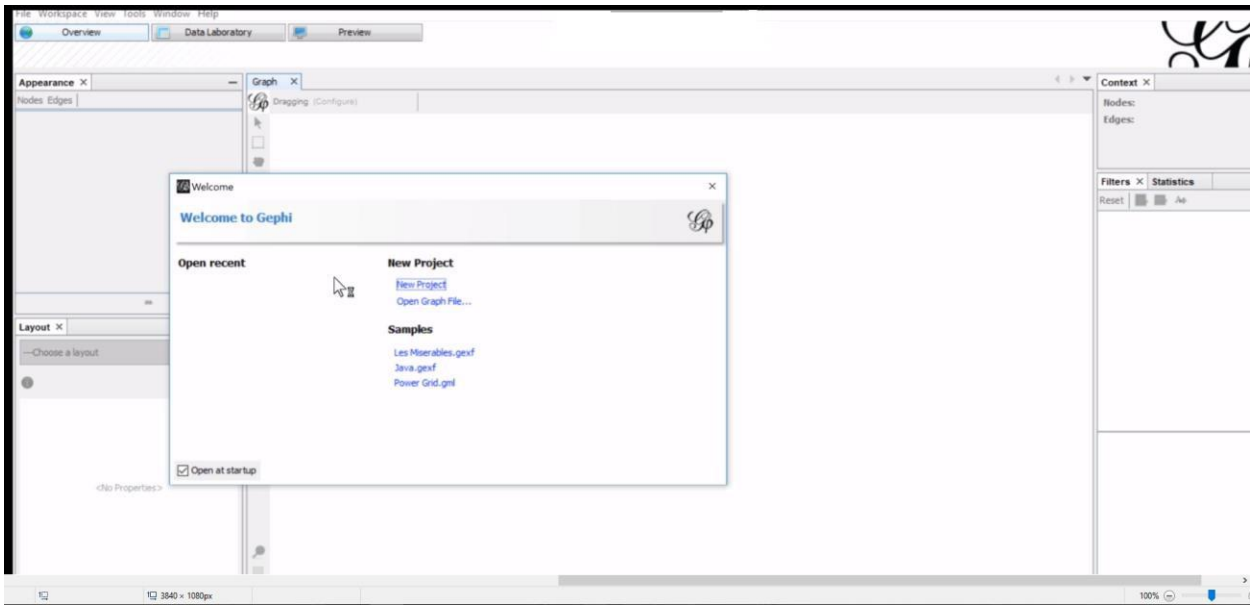
2. Download and follow the instructions to install the software from "**Download Free Gephi …**".



   This may or not be applicable to you. If there's any missing file (in this case, a Java file). Please go to **Google** and download the patch you are missing.

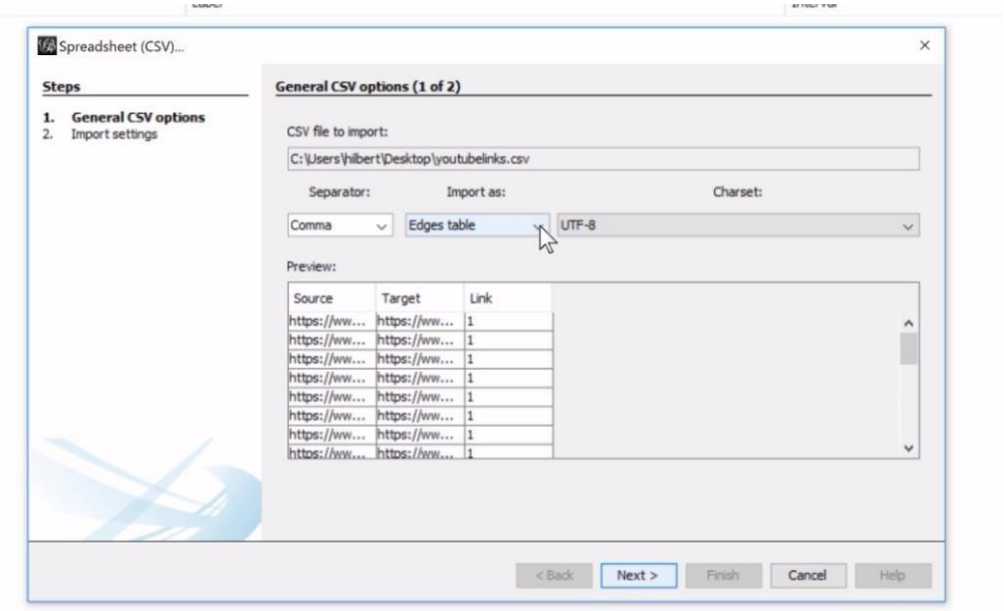
   In case you fail to install Gephi, you can also use the online version of Gephi available at www.rollapp.com/app/gephi . We suggest that you work with the downloaded version of Gephi rather than the cloud version, but both are fine.

3. Once you have installed **Gephi** this is how the interface should look:

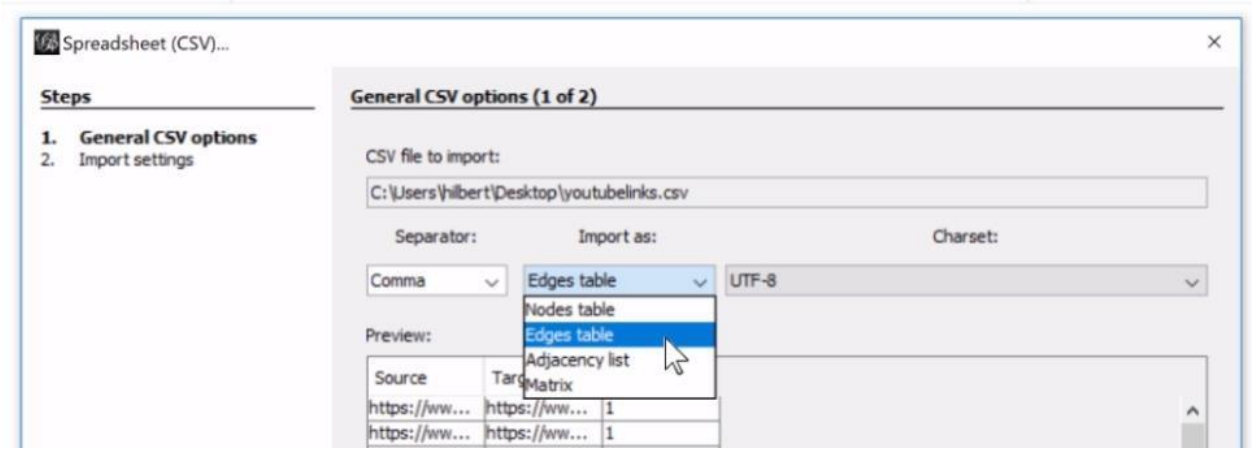


3. Visualize the Networks

1. Click on **New Project**
2. Click on the **Data Laboratory Tab**
3. Click on **Import Spreadsheet**
4. Import the "nameyouchose.csv" spreadsheet you have created for this assignment. Once you import, you should get a popup as the figure below

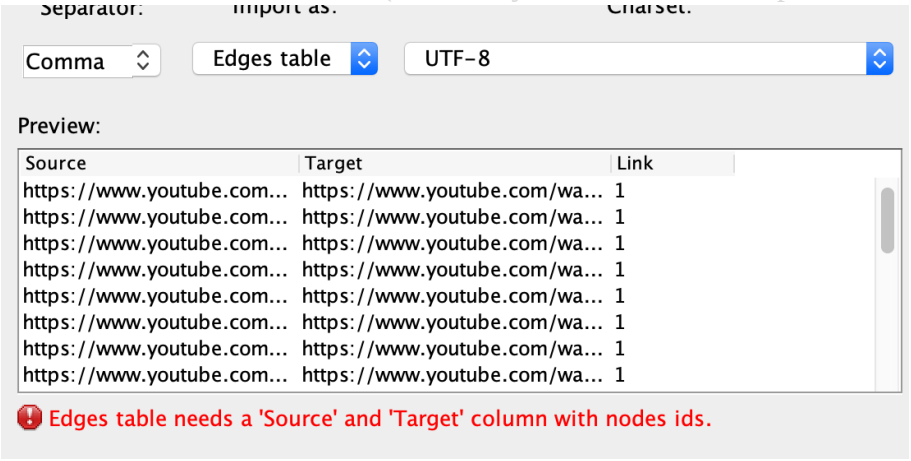


5. Make sure, the Import as dropdown is set to "**Edges table**"



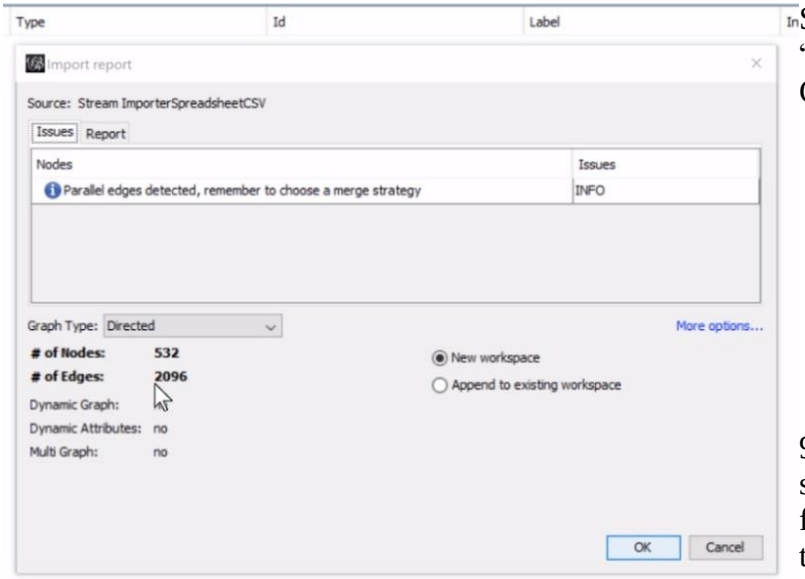
6. Click **Next**
7. Click **Finish**

If you run into problems, such as when "Source and Target are not recognized importing csv", just search online for whatever problem is reported to you, together with the word "Gephi" to find the solution. This is an important aspect of doing computational social science: search online for particular help you might need… you'll find the answer!



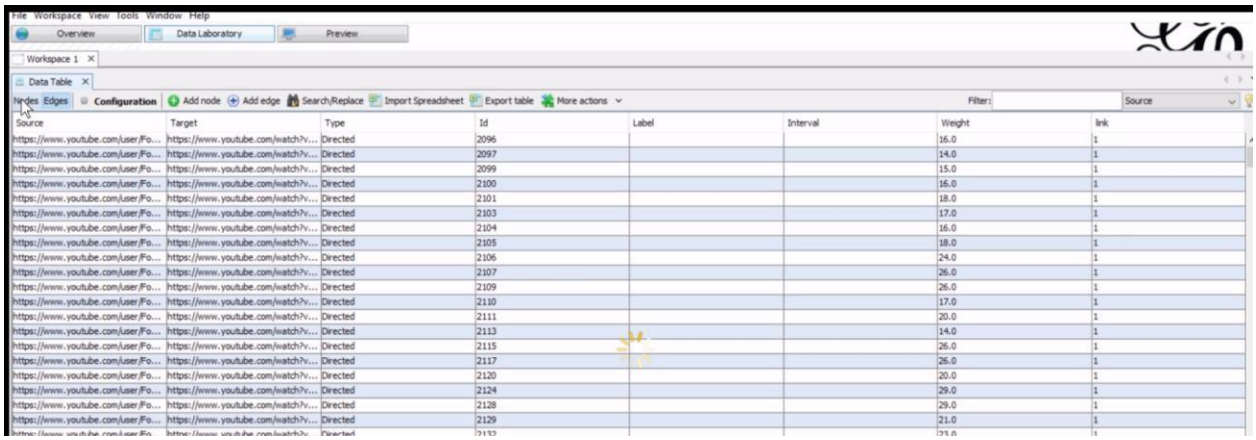
=> "Let me google that for you"

8. Gephi should show you a **summary** of imported spreadsheet. It provides you with number of egdes (in this case 2096) and nodes (in this case 532)

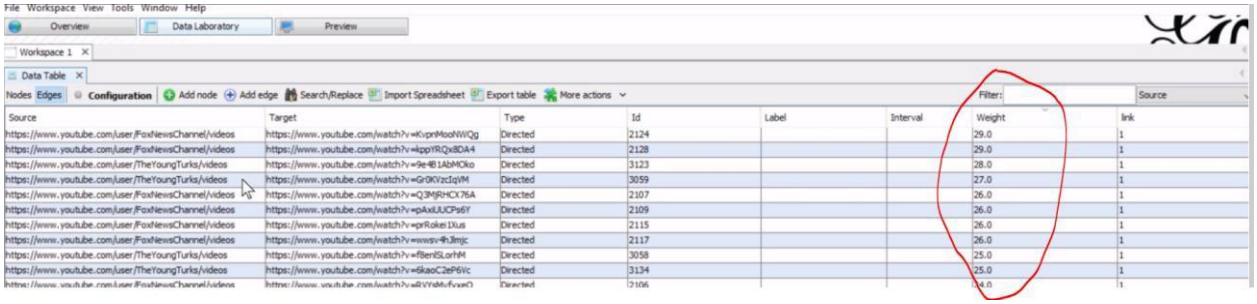


Set Graph Type to "**Directed**", Click on "**Append to existing workspace**" and Click **OK**

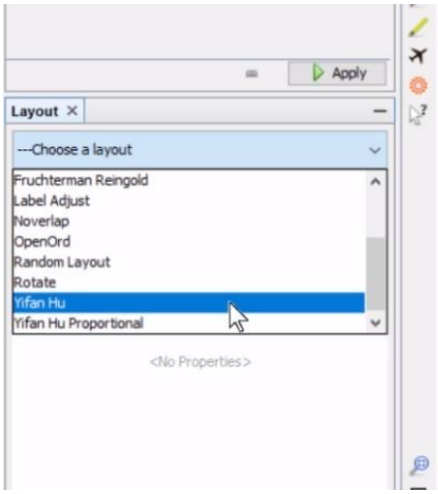
9. Your workspace should be similar to the following figure (pay attention to which tab window you are in: see top left corner).



You can sort your worksapce by "**weight**" to see the weighted ties. Note: even if you don't sort if by "**weight**" it will not make any difference to your network/analysis.



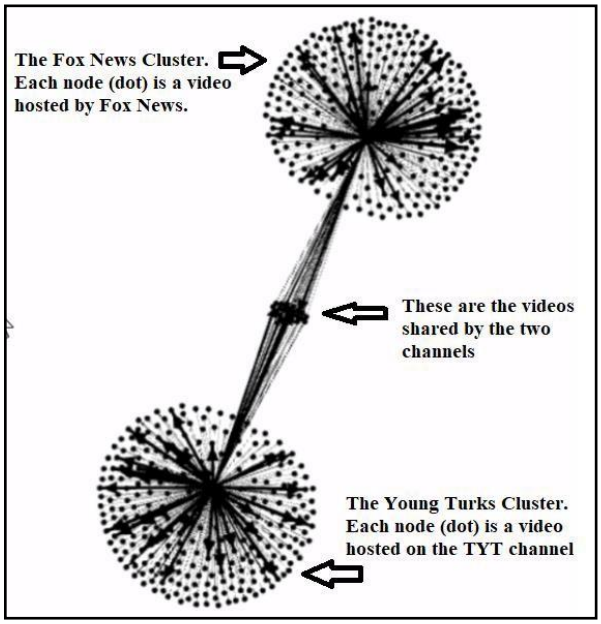
11. Click on the **Overview** tab next to the Data Laboratory



12. Click on the **Layout** dropdown and choose **YifanHu**

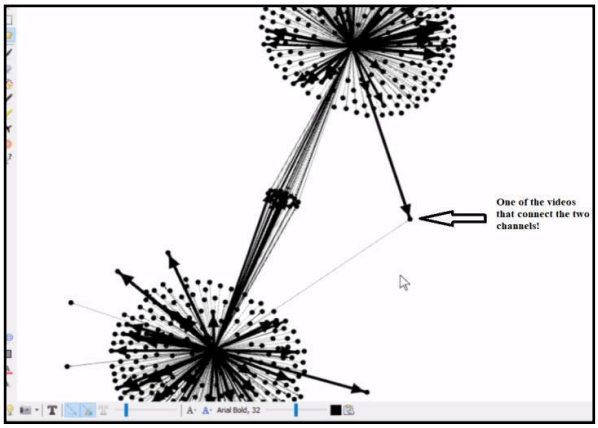
13. Click **Run**

The algorithm will run and your network should look similar (somewhat similar if not the same to the figure below



You can move around the displayed network with the help of your mouse (drag left or right, or simply zoom)

14. You can also check for more information on the videos that connect the two channels. To do so, click and drag one of the nodes (see the figure).



15. Next, click **right-click** on your mouse and select "**Select in Data Laboratory**"

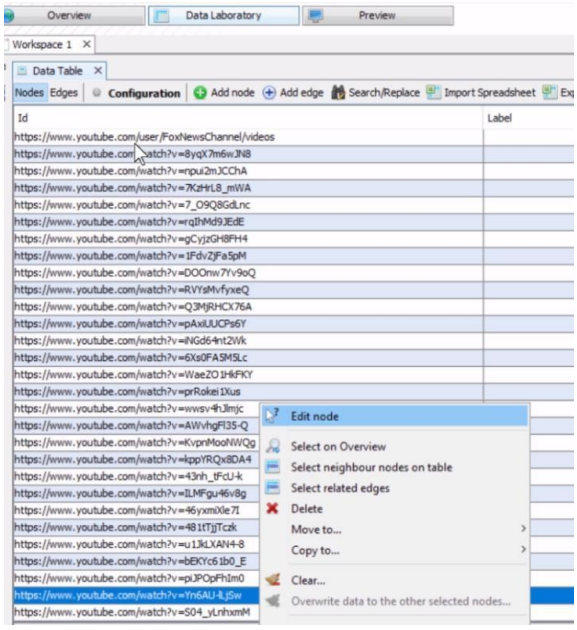
16. Click on the **Data Laboratory tab** and you can check which video connects both the YouTube channels

17. In order to watch the video –

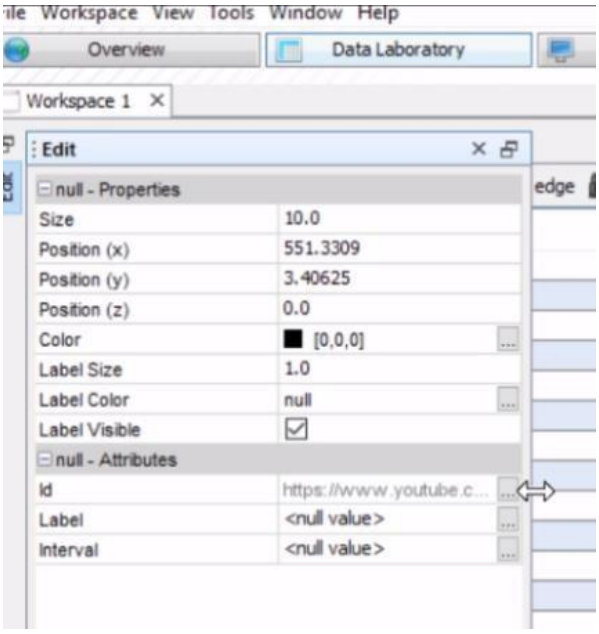
In the Data Laboratory mode

a) right-click on node
b) select edit node
c) click on node ID
d) a pop up window will open up with the node ID
e) highlight/select the node ID and use ctrl+c or command+c to copy
f) go to a browser, ctrl + v or command+v, and watch the video.

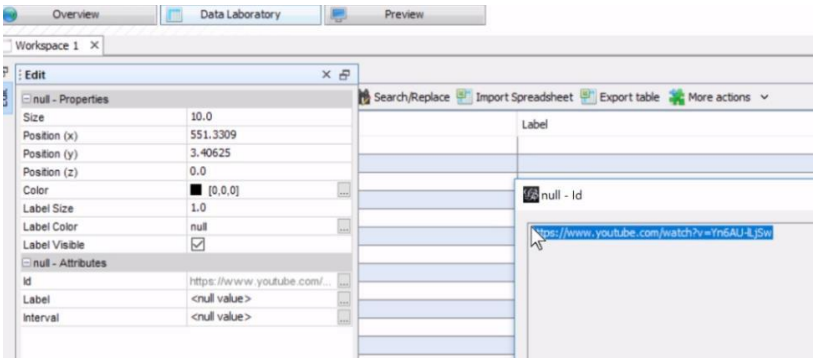
In this case, it was the **Fox News video** - https://www.youtube.com/watch?v=Yn6AU-lLjSw



step a - b



step c

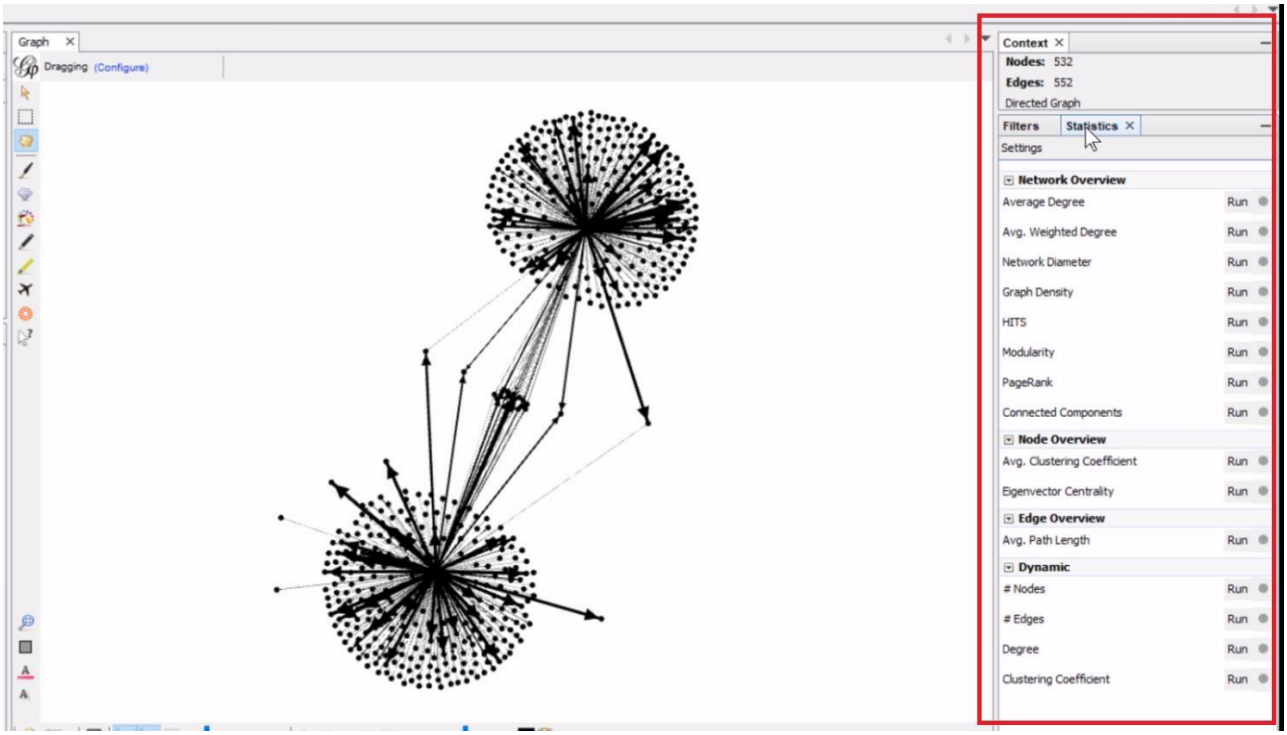


step d - e

To Know More: In order to know more about another video(s) (nodes), which connects the two channel, you can repeat the previous steps

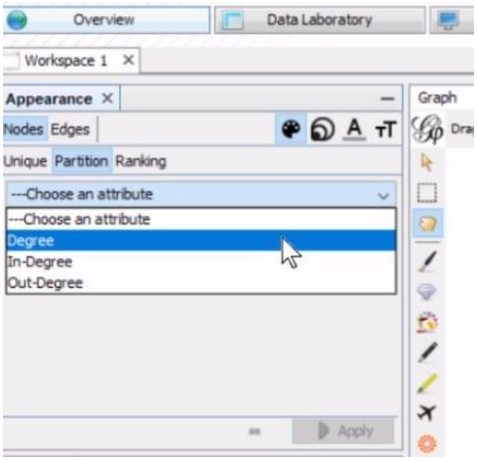
4. Analyze the Networks

1. Click on the **Statistics tab** on the right side of the page. Here you will find the range of statistics that Gephi can calculate for you.

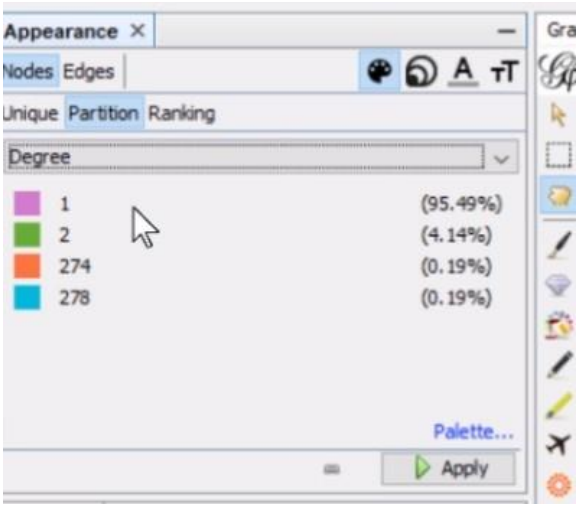


2. Click on Run to calculate the corresponding metric. For e.g., in this case, click on **Run** next to **Average Degree** to calculate the metric. The value will be shown next to the metric and a popup will show up with the metric details

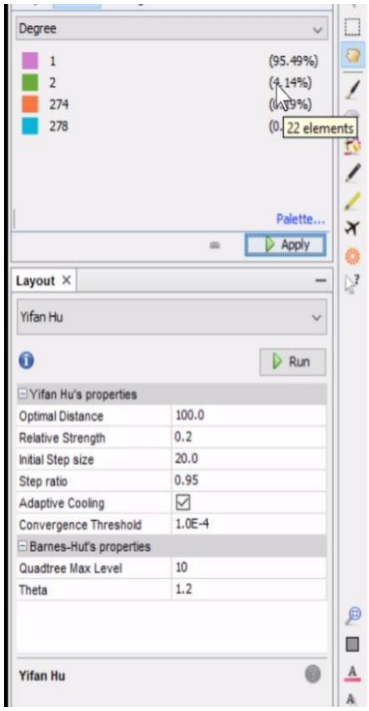
3. Go to Apperance tab, and choose an "**Attribute**". Different nodes will have different attributes. In this case, we will see how many nodes have different degree attributes so we will choose the "**degree**".

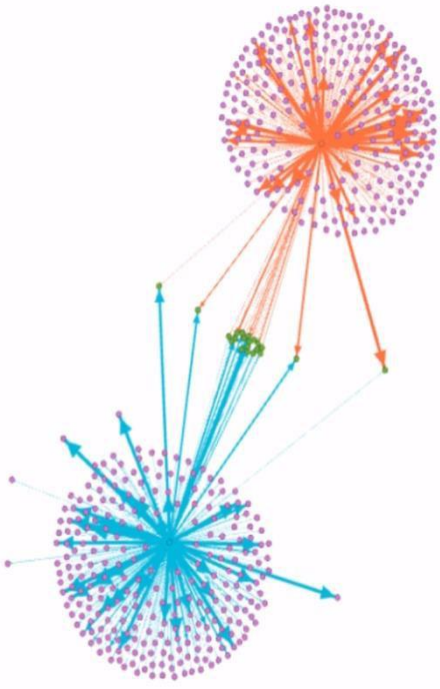


4. In this case, we can see the **degree attributes** (figure below)



5. You can click on the **Apply button** to apply the respective color to the nodes. For e.g, in this case a node colored in pink (95.49% of the nodes) have 1 degree, a node with color green (4.14% of the nodes) have 2 degrees and so on.

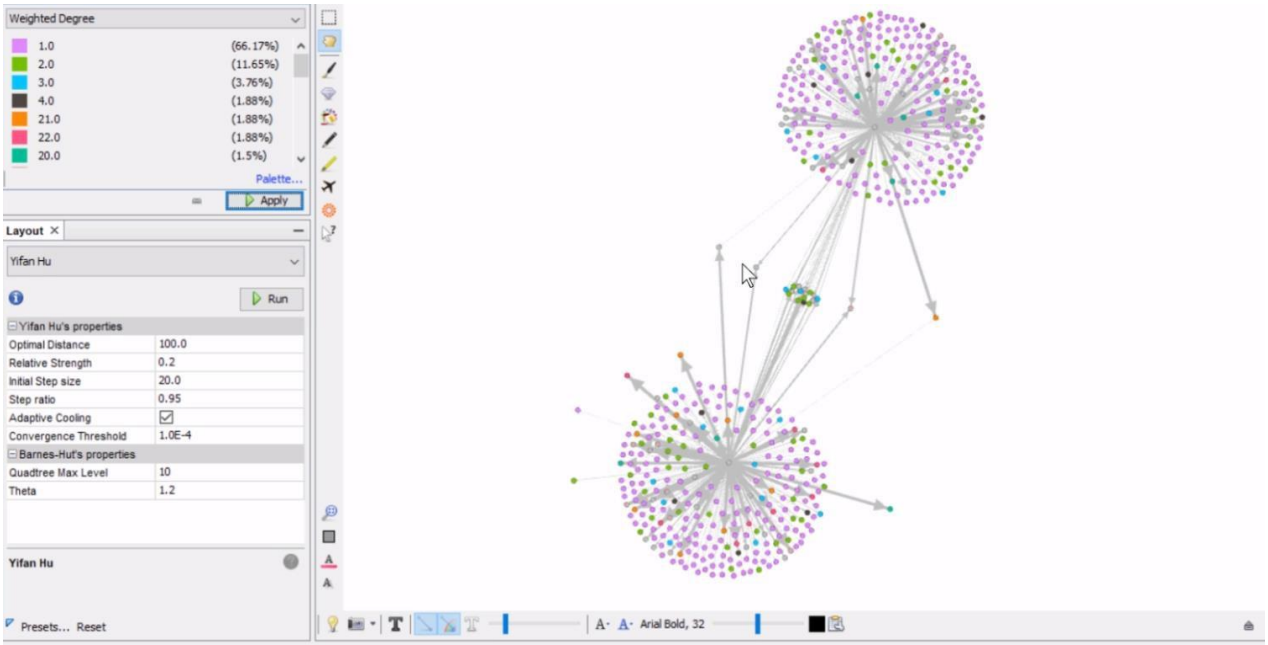




6. To check the metric associated with **Average Weighted Degree** (where nodes are weighted according to the links); click **Run** next to Average Weighted Degree on the right bar (as you did for Average Degree in step 2).

7. You can now partition your network based on Weighted Degree. To do so, follow step 3 to 5 and choose **Average Weighted Degree** in the dropdown.

   In this case, the network looks like the following figure



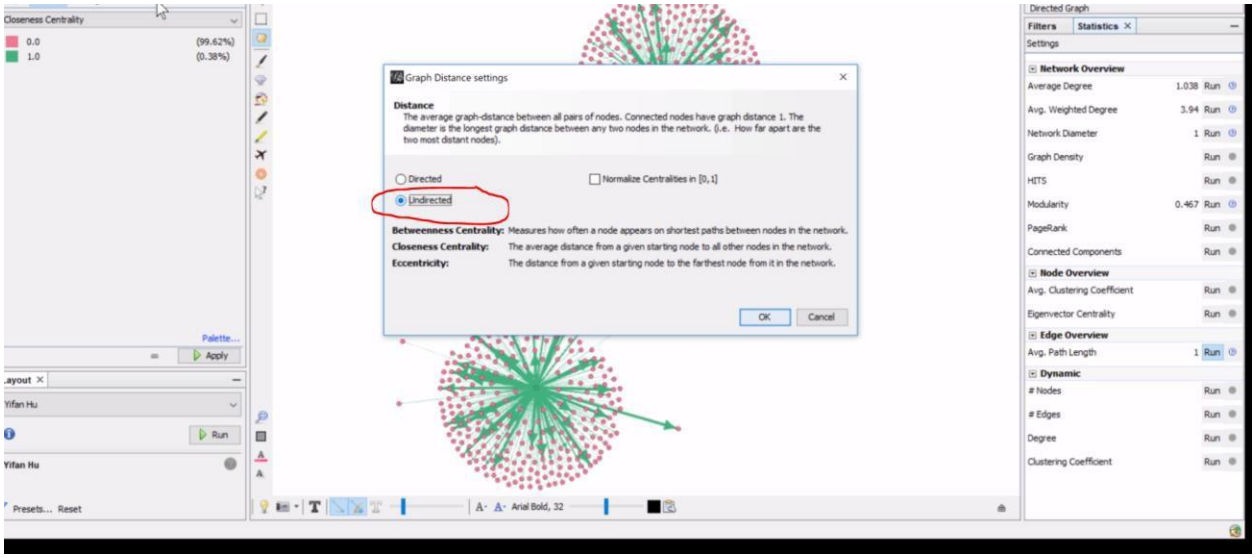
8. You can also classify your network into different classes. Use **Modularity** to do so. Click Run next to Modularity and then partition your network based on calculated Modularity (as you did for Average Degree in step 2 – 5). In this case, the graph looks like the following figure

Here you can see two communities, green group (Fox News) and pink group (The Young Turks)

9. You can also calculate and visualize your network based on **Average Path Length** that calcuates the centrality meaures (betweeness and closeness centrality). Click Run next to Avg Path Length and then partition your network based on calculated **Closeness Centrality** (as you did for Average Degree in step 2 – 5). In this case, the graph looks like the following figure:

Note: This graph is for a directed network →

For an **undirected network** you can **re-run Avg Path Length** and choose **undirected** in the popup (see below)



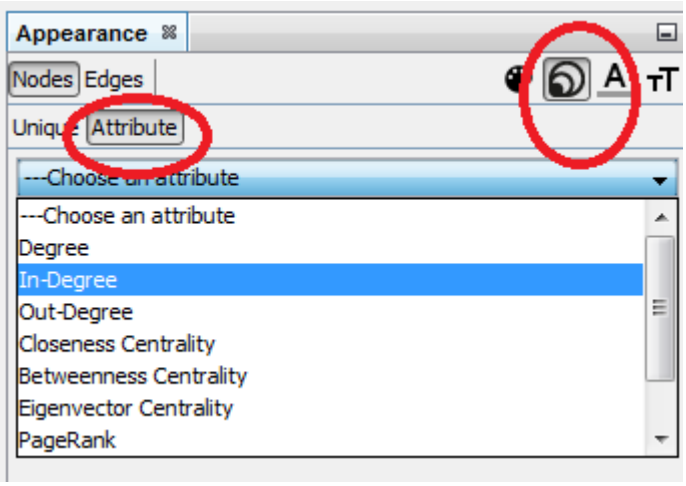
To plot the network with the new cacluted metrics (undirected avg path lengths); partition your network based on calculated **Closeness Centrality (**as you did for Average Degree in step 2 – 5).

This time, the network would look like the following figure:

10. So, you can also calculate and visualize your network based on **any attribute** mentioned on the right panel. Just click Run next to that attribute and then **partition your network** based on calculated metrics for that attribute to visualize your network. These include all the attributes highlighted in the following figure

EXTRA:

To increase the node size based on an attribute – choose the attribute under the dropdown and then click on the circled button in the figure.



Congratulations! You've successfully analyzed your first network!