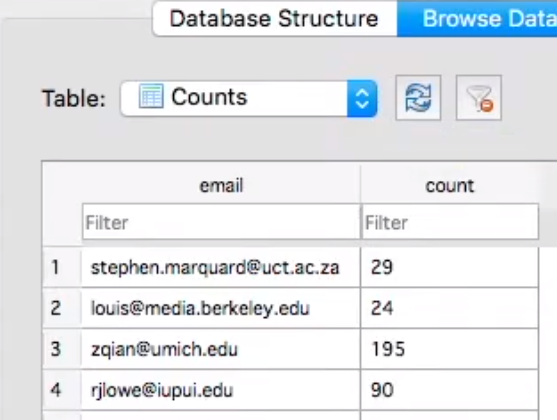
Question 1: We have previously worked with reading and writing into files. This task will be similar but instead you will be writing into a database. Use the mbox.txt file, which contains email messages, to count the number of times a certain email appears in that text and store it into a database file named (emai.db). [30 points]

Expected output when exploring the resulting database file you created should appear like this.



The table schema for the database is: CREATE TABLE Counts (email TEXT, count INTEGER)

**Note: Students have to use the SQLite database package to work with this assignment so that the outputted database can be read using SQLite browser database application. You can download the DB Browser for SQLite application at**[**https://sqlitebrowser.org/ (Links to an external site.)**](https://sqlitebrowser.org/)**to check your database output.**

Question 2: [40 points total. See the broken down points in parts]

In this question you’ll be writing code to get user information from Twitter and store it in a SQLite database. You will need a twitter account and a twitter developer access to query and collect your data.

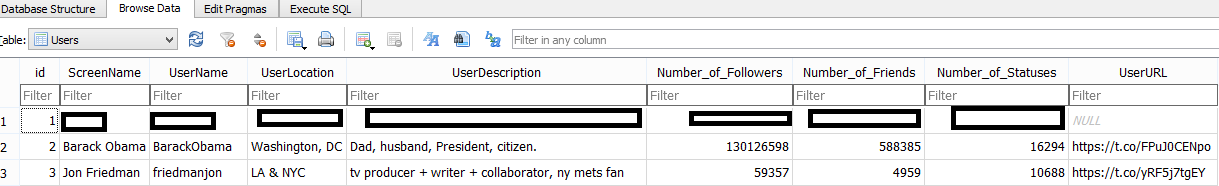
After setting up the connection parameter of Twitter (see lesson 1 interactive notebook or web services module’s interactive notebook to understand how to set up connection parameters to twitter).Create a database connection instance for sqlite db. Then, create cursor instance of that connection to run your sql queries.

Part A [10 points]

Create a table called Users with the following schema: An ID column which is of integer type, it is also auto-increment and a Primary key. 2nd column is ScreenName which is of text type. 3rd column is UserName which if type text. 4th column is UserLocation of type text. 5th column is UserDescription of type text. 6th column is Number\_of\_Followers of type INT. 7th column is Number\_of\_Friends of type INT. 8th column is Number\_of\_Statuses of type INT and the 9th column is UserURL of type text. [10 points]

Part B [20 points]

You will need to create a list of Ten users (theses are user names of the twitter account) to collect the data from. These can be any users of your choice (but make sure they have some followers to do part C of this question). You will then execute the INSERT query to insert information for the users. The information that are you are gathering is already mentioned in the column names of the table schema.



Your table in the sql db might look something like this.

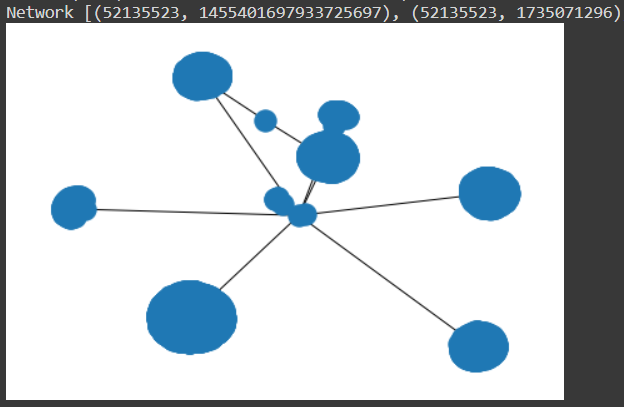
You will also execute SELECT \* query to present all the information in the User table in the output of your code cell.

Part C [10 points}

Network Analysis is at core of social network analytics. It gathers information about your reach potential and investigates your social network structure in the form of nodes and graphs. You can read more on Social Network Analysis and get some ideas on network graph here: <https://towardsdatascience.com/how-to-download-and-visualize-your-twitter-network-f009dbbf107b>

In this part you will be selecting **a user** to create a network graph of their twitter followers. Hints: Create an empty list to store network connections. Store connection pair as list of tuples. Use NetworkX library in python to draw the network graph. (NetworkX tutorial: <https://networkx.org/documentation/stable/tutorial.html>)

A screenshot of how your network graph might look like is shown below:



**Important: Make sure you use a counter variable to stop the loop after 10 connections. If you do not limit the node connection this can take very long time to execute and may even fail in execution. Also, you are making multiple request to Twitter API here, so once you run the cell with one user, running it again with another user (which is not required by the question) make cause rate limit error and you may see an error message like the following:   
RateLimitError: [{'message': 'Rate limit exceeded', 'code': 88}]**

**This can somewhat be remedied using additional argument while setting up your API. You can turn on the wait\_on\_rate\_limit flag to true.**

api2 = tweepy.API(auth, wait\_on\_rate\_limit=True)

Question 3:

Data visualization: Use the assignment\_4\_property\_tax\_report\_2019.csv file attached in the assignment portal for the questions below:

Part 1: I’m interested in finding out the rate of houses build per year after the year 1990. Present this information in a form of line chart with Years on X-axis and number of houses built on the Y-axis. Hint: Use dictionary to store number of houses built per year. [10 pts]

Part 2: I’m interested in finding out the percentages of houses built per zone category. Present this information in a pie chart that displays differing color and percentages per zone category. Hint: Use Dictionary to store number of houses per zone category. [10 pts]

Part 3: It is easier to detect trend in diagrams like scatter plots. I’m interested in finding out the rate of houses build per year after the year 1900. Present this information in a form of scatter plot with Years on X-axis and number of houses built on the Y-axis. Hint: Use dictionary to store number of houses built per year. [10 pts]

Note: Your visualization should contain X-axis label and Y-axis label and also a title for the plot.