**Social Network Analysis SOP**

Included in this folder is a document called “Social Network Analysis Overview” and another called “Social Network Analysis Using Gephi” which may be helpful to read prior to trying to implement this SOP.

1. **Data Collection and Preparation**

Method 1: Focusyn

You can access Focusyn through Ferrari and follow the steps outlined in FE/Assessments/SociogramDataWrangling and it will output and edge list to be imported into Gephi. This method required you to work with an in-language analyst to create a seed list of users that make up the network you want to visualize. A decent size seed list normally falls within the range of 10-20 users.

**Date Range:** You can specify any specific date range in the analysis.ipynb notebook

**Platforms:** Twitter

Method 2: JMWC Data Platfrom (JDP)

Make a request to the Science and Tech team to create a social network view of the collected date in Redshift. JDP uses already data collected from Pulse and imported into Redshift to create an edge list of interactions between authors and response authors. The interactions are continually calculated through a view, instead of a one-time collection with Focusyn.

**Date Range:** The edge list will only be created from the date that data was pulled into Pulse (through a campaign) to the current date.

**Platforms:** Twitter and Telegram

1. **Data Import into Gephi**

Method 1: Focusyn Edge List

The Focusyn edgelist needs to be transferred to whichever system you are hosting Gephi on.

Method 2: JDP RedShift View Import

Request a detailed guide on how to connect RedShift data to Gephi from Science and Tech team. A basic overview is to go to “Imports” and the configuration will look something like this:

Graphical user interface, text, application, email

Description automatically generated

(U) Gephi will then give options on the following parameters:

**Graph Type:** Determines whether or not to display the “direction” of interactions i.e. which user took the action to interact with another user creating an outgoing or incoming edge.

**Merge Strategy:** Determines how edge weights are calculated

**Create Missing Nodes:** Creates placeholders for missing nodes

**Include Self-Loops:** Show edges in which user interacts with its own content (heavy self-loop edge weight often indicative of a bot or organized effort to amplify content).

1. **Network Analysis using Gephi**

(U) Once the data is imported, you will need to choose a visualization layout. You can determine what type of layout is best based on the intended emphasis, the graph type and the number of nodes. The following decision tree can help:

Diagram

Description automatically generated

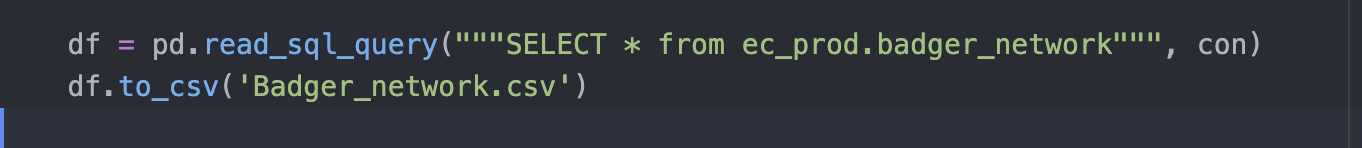
1. **Merging Gephi and Network Data for further Exploratory Data Analysis**

For further analysis of the data collected on Redshift and the Gephi graph

4.1: Export Data Lab as “{ProjectName}\_DataLab.CSV”

4.2: Export RedShift network view as “{ProjectName}\_network.csv” using TableCSV.py

Update query in TableCSV.py to the right view:

****

Run TableCSV.py in terminal:

**Text

Description automatically generated**

* 1. **Run SNA\_EDA\_Merged Data.ipynb on “datalab.csv” and “project.csv” to product export “m\_data.csv”**

This will give you the output of m\_data which includes the RS data and whichever columns from the Gephi datalab that you merged to it. The default is to assign the modularity class to the Redshift data.

The exploratory data analysis in this script provides the following:

* Count of Languages in Network
* Top users based on Eigenvector Centrality
* Key Influencers by Followership
* Time Series Analysis on Network Activity
* Time Series Analysis on Network Activity by project
* Count of content (volume) by modularity class
* Count of content (volume) by project
* Top co-appearances (edge list)

1. **Further NLP**
   1. Move to the NLP Folder
   2. If you want to run spacy (more complicated SNA)
      1. Move to Spacy folder in NLP
      2. Run spacy script on “m\_data.csv” to export entities and hashtags csvs
   3. Otherwise follow the NLP\_SOP.docx