

Student Name: Muhammad Yousuf Amray

Student Number: 501038218

Course Title: ITM 795: Social Media Analytics

Please attach this cover page as the first page of your completed assignment before submitting it.

Assignment #	3
Due Date	See D2L

**Return all assignments through the Course Assignment Submission Page
(Unless otherwise specified by the instructor).**

I hereby certify that I am the author of this document and all sources used in the preparation of this assignment have been cited in accordance with Ryerson's Code of Student Conduct directly or paraphrased in the document. Sources are properly credited according to accepted standards for professional publications. I also certify that this paper was prepared by me (all group members if it is a group paper) for this purpose.

Section 1: Executive Summary (approximate length: 1 page)

- 1) Recap what dataset you analyzed for this assignment. What was your final search query used to collect tweets? Indicate if you re-used the dataset from A1/A2 or if you collected a new dataset for this assignment and why.
- 2) There are many methods that can be used to analyze a Twitter dataset. In the context of this assignment, explain what research questions the **network analysis** approach helped you answer and how you answered them. Be as specific as possible!
- 3) Provide information about the type(s) of communication network you examined (reply, mention, retweet and/or quote-type network) and why. How many nodes and ties were there?
- 4) Summarize your results and implications (very briefly!).

1) Twitter Search Query:

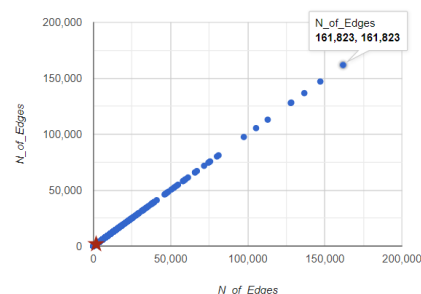
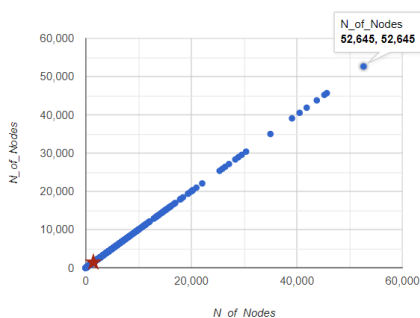
(Mahsa OR Amini OR Iran) (#Mahsaamini OR #iranprotest) lang:en lang:en

My final search query is as mentioned above. This is the same search query I used in assignment 2. Therefore, I am using the same dataset for this, Assignment number 3. However, it is different from assignment 1. So, in this query I have selected tweets that contains these words, “Mahsa”, “Amini”, or “Iran”. As these are the words that are most used in tweets regarding my topic. Furthermore, to be more specific about my data, I have selected two main hashtags that are trending according to the topic. These hashtags are #Mahsaamini and #iranprotest. Moreover, to understand the tweets collected I have selected the English language.

2) In order to execute my network analysis efficiently I used these research questions:

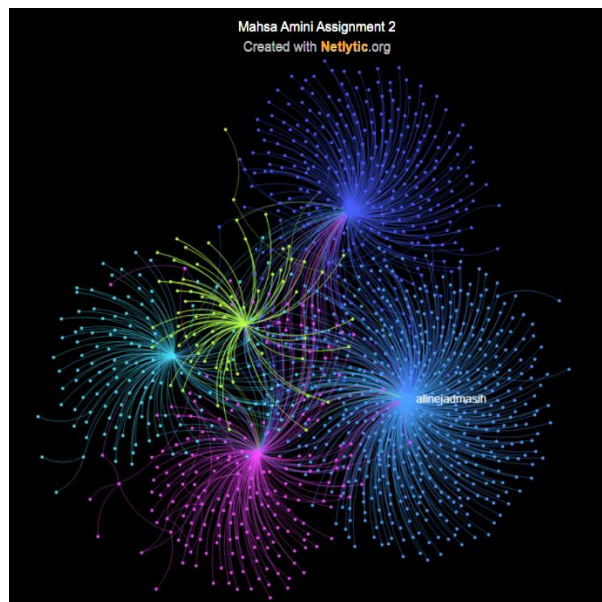
1. Who are the most popular tweets that people are retweeting the most and why?
2. Who are the authors of those tweets?
3. Is there tweet creating any impact on how the case is going on?

3) The types of communication networks I examined are replied to mentions, retweets, user A replied to user B, and user A quoted user B. So, there are 52,645 nodes and 161,823 ties.

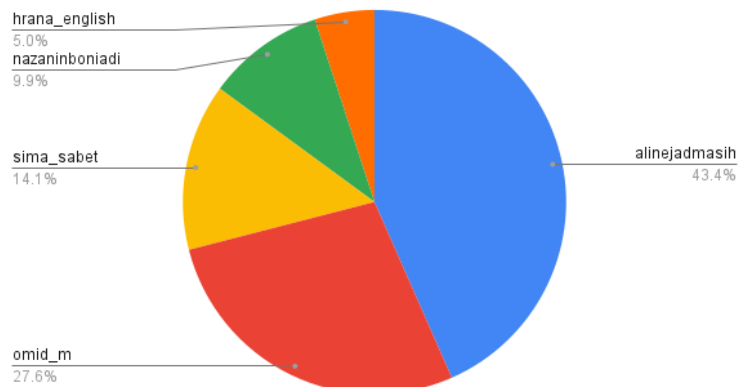


Section 2: Influential Accounts (approximate length: 2 pages)

- 1) List and describe Top 5 most connected accounts in your datasets based on the **in-degree** centrality and Top 5 most connected accounts based on the **out-degree** centrality.
- 2) For each account, provide their username, name (if available) and a short explanation (in your own words) of why they are influential in your dataset.
- 3) To support your analysis, include relevant screenshots and sample tweets as needed.

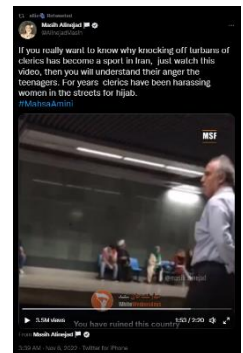


Top 5 Connected Accounts (In Degree)

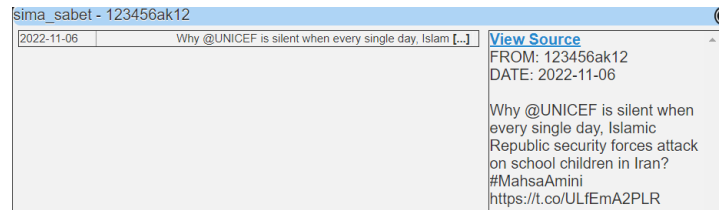


The top 5 most connected accounts in my dataset according to network analysis based on in-degree centrality are alinejadmasih, omid_m, sima_sabet, nazaninboniadi, and hrana_english.

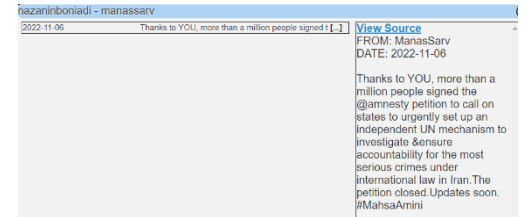
- 1) The twitter account alinejadmasih is a female named **Masih Alinejad**. She is an Iranian Journalist and activist. Who left Iran in 2009 and has been living in New York, United States in exile since 2014. She raises her voice against the women rights in Iran and her tweets are the most, replied, mentioned, retweeted, and quoted. Which means her tweets have 43.4% impact on the ongoing case of Mahsa Amini. Which is the highest share. On the right side there is a screenshot of one of her popular tweets.
- 2) The second most connected account is omid_m who is named **Omid Memarian**. He is also a journalist born in Iran. The reason for him being top 2 on connected accounts is because he also raises voice against the Iran's government, and he have been trending lately with 27.6% of top tweets. I am attaching screenshots of his tweet as a reference.



- 3) The third top account of a female named Sima Sabet and her twitter username is sima_sabet. She is the host of a talk show. Which is aired on the most popular news channel of Iran which is called Iran International. Which means she is already famous, and her voice is heard by Iran. So, her most popular tweet is in which she targets UNICEF for being quiet against what is going on in Iran. Here is the screenshot



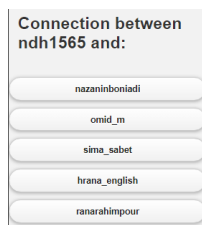
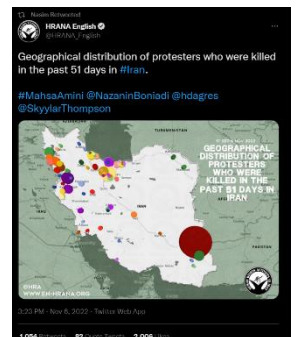
- 4) The fourth top account is nazaninboniadi who's full name is Nazanin Boniadi. She is an actor and Human Rights Advocate. Her most popular tweet is about why is UNICEF quite against what is going on Iran. Here is the screenshot of her tweet



- 5) The fifth top account is hrana_english. It is a non-profit organization called Human Rights Activists News Agency. The most popular tweet by this organization is attached on the right, and the tweet is about protester getting killed in Iran.

The top 5 most connected accounts in my dataset according to network analysis based on out-degree centrality are mercilezpython, ndh1565, 295nwz6745, iranian16356883, and moezziah.

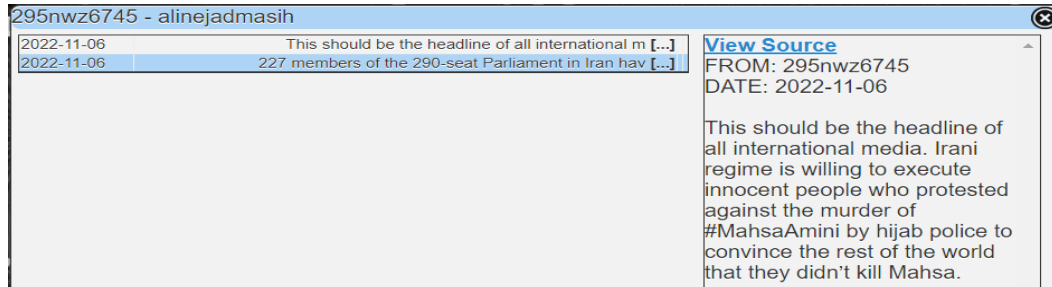
- 1) The first account is mercilezpython who has named his account as “**SHWATHAM** ➡”. He has the most retweets of tweets of the famous twitter accounts mentioned above such as omid_m and alinejadmasih. This user belongs to Madras; India and he retweets anything against Iran from the famous twitter accounts. Here is an example of his retweet. Therefore, I didn't find him to be influential at all as he has 285 followers.



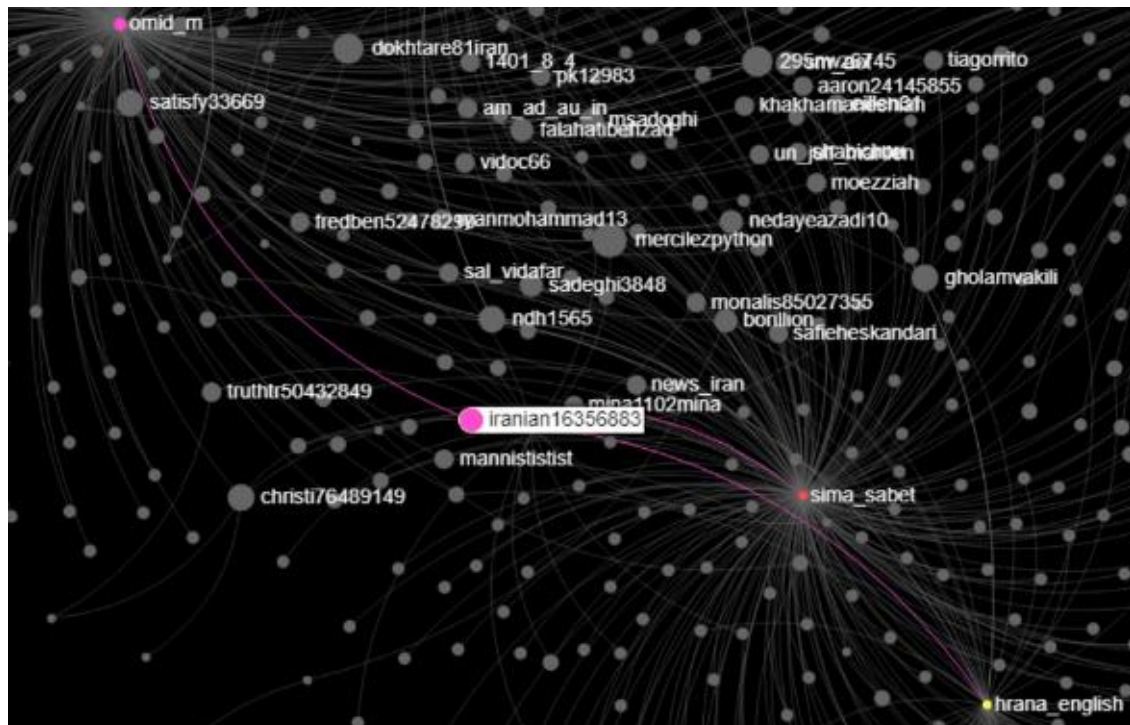
- 2) The second most connected account based on out-centrality is ndh1565. He has connection with top 4 twitter accounts that are, nazaninboniadi, omid_m, sima_sabet, and hrana_english. On the left is the connections of this account.



- 3) The third most connected account is 295nwz6745, out-degree is 6 and the major connections with omid_m, alinejadmasih, and hrana_english. This users name is Luis Fierro. Below is an example of his retweet with alinejadmasih.

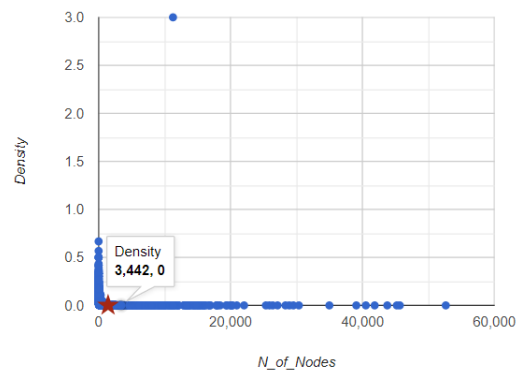
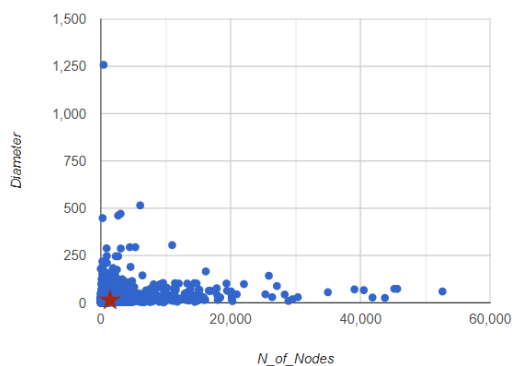
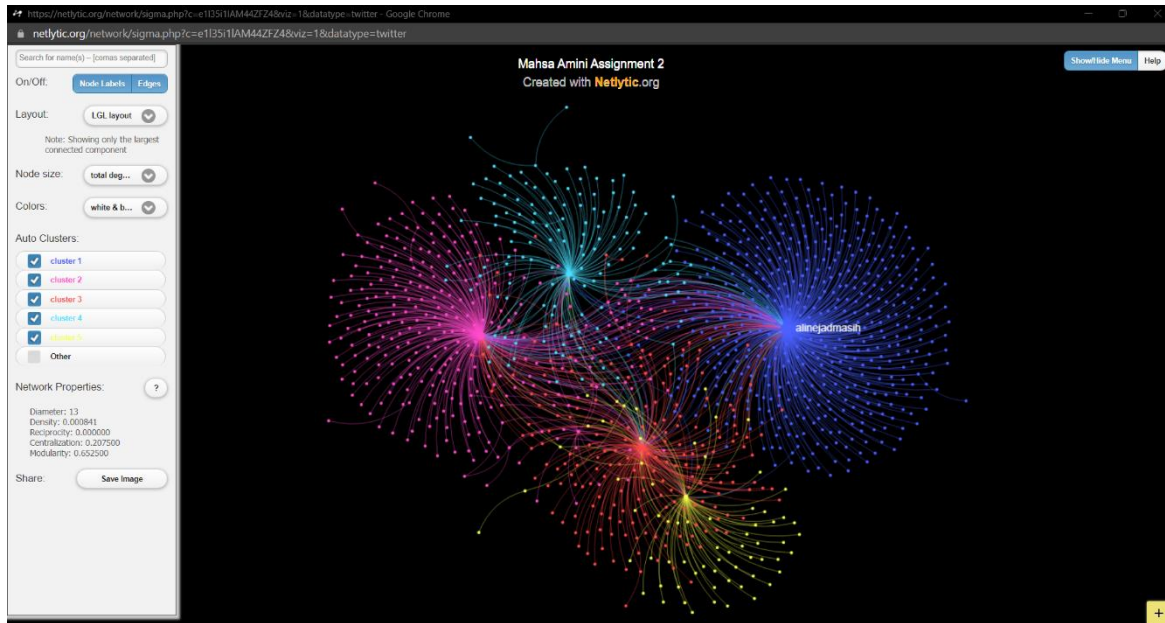


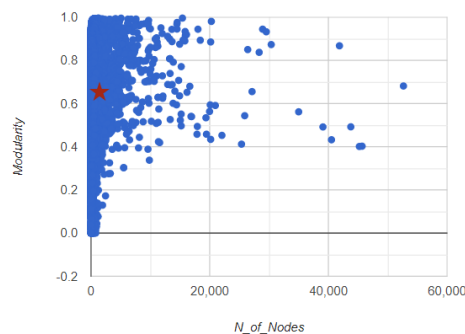
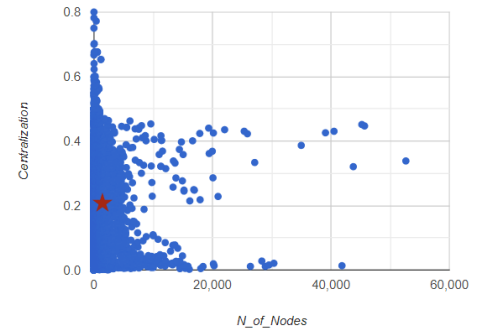
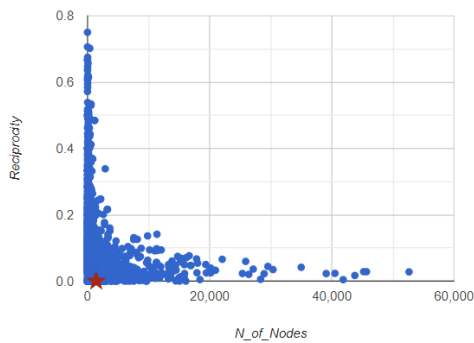
- 4) The fourth top connecting account is iranian16356883. This account doesn't have a proper name. However, it has 4 out-degree connections. On the right is the screenshot of its node.
- 5) The last top connecting account is moezziah. The username of this account is AMMo, and it has many retweets of the tweets posted by omid_m, alinejadmasih, and sima_sabet. Here is the example of retweets and I have circled the account.



Section 3: Community Detection (approximate length: 2 pages)

- 1) List the values of the following network properties for your dataset and **interpret** them in the context of your network using your own words: diameter, density, reciprocity, centralization, and modularity.
- 2) Summarize the three largest clusters in your network as detected by Netlytic.
- 3) Based on your manual examination of accounts and sample posts in the network visualization, explain the nature of each cluster.
- 4) To support your analysis, include relevant screenshots and sample tweets as needed.





Diameter: According to network properties on my dataset the Diameter of my network analysis is 13 which means it takes 13 nodes to reach from one network to the other. In simpler terms Diameter is the distance between two networks.

Density: The density of my network analysis is 0.000841 (0.08%). Density indicates the overall connectivity of network. It is calculated by dividing the **Total Existing Connections** by **Total Possible Connections**. So, if everyone is connected to everyone the density is 1. However, in my case as the density is 0.000841, this means there is a very minor number of tweets connected to each other. Therefore, density isn't strong in my dataset.

Reciprocity: Indicates how many online participants are having **two-way conversation** which means replying to tweets. However, in my case the average reciprocity is equal to zero (0). Which means no online participant was replying to each other. According to my topic I believe most users are retweeting and making important tweets viral. To make the protest trending and no one is having an actual conversation or discussion.

Centralization: When the network is dominated by few central participants the values are closer to 1. However, when more people are contributing to discussion and information dissemination values are closer to 0. Centralization value in my dataset is equal to 0.207500. Which means it is closer to 0 this tells us that more people are contributing to the discussing, and it is not dominated by few participants.

Modularity: provides an estimate of whether the participants are engaged in the same conversation (value is close 0) or are engaged in different conversations and communities with weak overlap (Value is close to 1). The modularity of my dataset is equals to 0.652500. Which means the network consist of different conversation.

In the first cluster reciprocity is equals to zero because there is barely any two-way conversation but mostly the users are retweeting the tweets of alinejadmasih. However, this cluster is centrally dominated by alinejadmasih. Everyone is just retweeting her tweets that's all.

The second cluster is dominated by omid_m and reciprocity are also bare minimum. Which means there are more retweets and no replies or discussions.

The third cluster is dominated by sima sabet, again her messages are being retweeted and no such conversations found.

References

Note: If relevant, reference external sources that you cited in the report.

Appendix

Note: If you have more than 3 relevant figures or tables, include them in the Appendix and reference them in your report.