

Project I - Social Media Data Analysis

CS579 - Online Social Network Analysis

Project Report

Illinois Institute of Technology
February 2022

Kayenat Patil (A20495072)

Kajol Tanesh Shah (A20496724)

Contents

S.No	Title	Page. No
1	Abstract	3
2	Problem Statement	4
3	Introduction	4
4	Data Crawling	5
5	Data Miner	10
6	Data Cleaning	13
7	Feature Selection	13
8	Data Visualization	13
9	Network Measures	18
10	Team Efforts	29
11	References	29

Abstract

In this project, we have scraped and analyzed data from the social media platform ‘Instagram’ and plotted a relationship between the results obtained. Based on the visualizations we obtained, we have done the analysis of the networks and reported the observations. We have chosen Instagram as our social media platform as it is one of the most popular social networking sites which is used by many people and mostly youngsters, hence they are more prone to influencers and content produced by them. At the same time, Instagram doesn't support bots being used on their platform, which makes the process of gathering data much more difficult, since after the year 2020 traditional data scraping techniques do not work.

The goal of this project is to find out similarities between the users who were scraped (if any) and visualize the network formed. We mostly focused on ‘following’ data than ‘followers’ this was mostly done as a user can control who they are following whereas followers is a parameter with no direct control, especially with public and verified accounts having a follower count over one million, whereas their following count is much lesser and easier to work with.

At the same time. We considered users following in the range between 50 and 1000, the reason being that the possibility of having any nodes in common drastically reduced when nodes were less than 50. Whereas it was messy handling data and visualizing it when nodes were above 1000 (There is a case where we dealt with about 4000 nodes with the results attached below.)

Furthermore, we used NodeXL as the visualization tool and plotted histograms, Degree Distribution, Pagerank, Betweenness centrality, etc.

Problem Statement

The aim of this project is to crawl data from any social media platform like Twitter, Facebook, Instagram, Stackoverflow, etc and create a social network with at least 100 - 500 nodes. The social network created can be represented as a friendship network, co-authorship network, diffusion network, etc. After the data is fetched, the network should be visualized as a graph using graph analysis softwares like Networkx, NodeXL, etc. Then, using the software, the degree distribution of the graph should be plotted as a histogram. Choose two other network measures and plot them too.

Introduction

Today, social media is used by many people to share as well as discuss information and experiences with other people in efficient ways. Online Social networks are web-based services that allow people to connect with real-world friends online. Examples of such sites are Facebook, Instagram, Twitter, etc. Analysis of social networks can help us in understanding the connections between people and communities or organizations. Also, this can be useful to help businesses flourish by getting marketing insights and improving communication effectiveness. So, it is beneficial to learn how to analyze online social networks. There are various steps involved in this process. For our project, we have chosen the social media platform Instagram.

Data Crawling

Data scraping is the process of extracting data from a source. Scraping data from Instagram is not very easy and the official Instagram API only provides access to your own account which might seem acceptable but with the restriction of only scraping your own account details. At the same time, you need to provide your own username and password which isn't safe to have it exposed in a piece of code.

Furthermore, Instagram is a difficult platform to scrape data from as they have upgraded their policies to keep bots out of their platform. After further research, the terms of service of Instagram states that accounts cannot be altered by the means of an automated device, script, bot, spider, crawler or scraper, where 'alteration' refers to uploading a post, commenting, liking or sharing anything. It is against any and all forms of scripts that do not require to be overlooked by a user, making our task a bit more harder. These updates were made somewhere between 2019 and 2020.

But we still wanted to try methods that seemed to go against these policies, just to make sure that they weren't viable options. Since there are no legal repercussions to scraping public accounts or your own account, we tried running freely available code for this test.

1. Using package instaLooter

InstaLooter program can be used to download photos and videos from any Instagram profile or hashtag without using any API key. Below is the command to install the instaLooter package :-

```
$ pip install instaLooter
```

We downloaded the posts of a profile 'witty_souls' using the following command :-

```
$ instaLooter user witty_souls wittysouls
```

Here, we were able to scrap all the posts of the user 'witty_souls' in the folder named 'wittysouls' and no login details for Instagram were needed.



Image 1

In Image 1 are the posts which were extracted from the mentioned user profile ‘witty_souls’. While it was able to download images, it failed to download all images. Also, it was difficult to work with it as the package seemed to work for the first two times, but then failed to run on different machines.

We decided to try scraping data from public accounts rather than our own account. We decided to scrap hashtags that were popular and should generate a considerable amount of data.

\$ instaLooter hashtag foodislovee food-lover

Here, we were able to scrap posts which had the hashtag #foodislovee mentioned in it using the above command. But here, the command was not able to download all the posts under that hashtag and so, we couldn’t go ahead with this approach as we were not getting complete information. For eg. the above hashtag had more than 600 posts under it but instaLooter was able to download only around 200 posts.

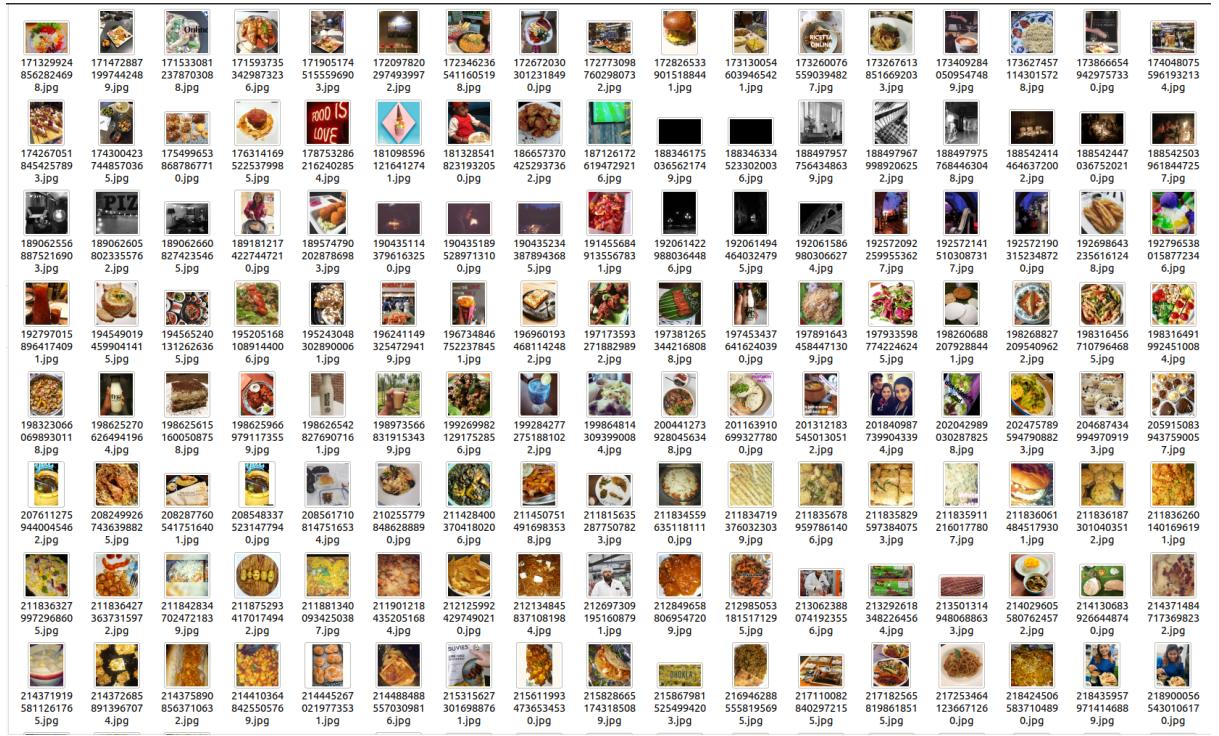


Image 2

Above is the screenshot of the images downloaded which were under the hashtag ‘#foodislovee’.

Again, while it was able to scrape some photos, it failed to scrape all photos. We also tried hashtags with about 500 posts and the package was only able to scrap 50. Also, it seemed to work on one system. It didn't work on the other, making it an impractical option.

2. Using package Instaloader

\$ pip install instaloader

Instaloader was able to download the profile details, captions, posts, etc. of the mentioned Instagram profile but after retrieving 12 posts, it needed login information. Even after using login information, we were not able to scrap all the data of the mentioned user making this an unsuitable option.

```
kajol@kajol-Lenovo-ideapad-310-15IKB:~$ instaloader profile witty_souls
JSON Query to profile/feed/: 404 Not Found [retrying; skip with ^C]
JSON Query to profile/feed/: 404 Not Found [retrying; skip with ^C]

HTTP redirect from https://www.instagram.com/web/search/topsearch/ to https://www.instagram.com/accounts/login/?next=/web/search/topsearch/
profile: Redirected to login page. Use --login.
Stored ID 34825794722 for profile witty_souls.
Hint: Use --login to download higher-quality versions of pictures.
[1/1] Downloading profile witty_souls
witty_souls/2020-12-17_19-22-21_UTC_profile_pic.jpg
Retrieving posts from profile witty_souls.
[ 1/79] witty_souls/2021-09-25_15-24-28_UTC.jpg [Fried Idli 😊❤️ PS: Made from...] json
[ 2/79] witty_souls/2021-09-18_14-20-50_UTC.jpg [Panipuri 😊 #healthy #food #fo...] json
[ 3/79] witty_souls/2021-09-17_13-35-17_UTC.jpg [Surat no locho 😊 PS: It is a g...] json
[ 4/79] witty_souls/2021-09-16_20-50-48_UTC.jpg [Crispy pakode 😊 #healthy #fo...] json
[ 5/79] witty_souls/2021-09-12_15-26-27_UTC.jpg [Misal pav 😊❤️ #healthy #food ...] json
[ 6/79] witty_souls/2021-09-10_02-45-34_UTC.jpg [Rice cutlets 😊 #healthy #foo...] json
[ 7/79] witty_souls/2021-07-11_14-53-52_UTC.jpg [👉👉 #poetry #hindi #hindipoetr...] json
[ 8/79] witty_souls/2021-05-30_17-18-23_UTC.jpg [My strength 💪 My backbone #yo...] json
[ 9/79] witty_souls/2021-05-17_15-22-17_UTC.jpg [👉👉 #picture #irony #life #ph...] json
[10/79] witty_souls/2021-04-24_15-33-34_UTC.jpg [A blessing in disguise ❤️ #y...] json
[11/79] witty_souls/2021-04-18_12-40-16_UTC.jpg [Reminiscence ❤️ PS: Rewrote o...] json
[12/79] witty_souls/2021-04-12_17-24-42_UTC.jpg [👉👉 #life #freedom #yourquote...] json

HTTP redirect from https://www.instagram.com/graphql/query to https://www.instagram.com/accounts/login/?next=/graphql/query
witty_souls: Redirected to login page. Use --login.

Errors or warnings occurred:
profile: Redirected to login page. Use --login.
witty_souls: Redirected to login page. Use --login.
(base) kajol@kajol-Lenovo-ideapad-310-15IKB:~$
```

Image 3

3. Using python script with selenium library

The selenium package is used in python to automate the interaction with browsers. Since selenium is a popular scraping tool and does work on scraping web pages, we decided to try it out with instagram.

With our script, we were able to only get details like number of posts, number of followers and number of followings. However, we were not able to download any posts or get specified information about the names of the followers or followings' list. It would return a numerical entry, not giving much value to the data scraped.

For eg. In the below screenshot, you can see the number of posts, followers and followings as the output of the script for the user '@aslisona'. But not much information about who the users follow, or any other relevant data.

```
(base) kajol@kajol-Lenovo-ideapad-310-15IKB:~/Desktop/temp_files$ python sel2.py

===== WebDriver manager =====
Current google-chrome version is 98.0.4758
Get LATEST chromedriver version for 98.0.4758 google-chrome
Driver [/home/kajol/.wdm/drivers/chromedriver/linux64/98.0.4758.80/chromedriver]
found in cache
/home/kajol/Desktop/temp_files/sel2.py:8: DeprecationWarning: executable_path has
been deprecated, please pass in a Service object
    driver = webdriver.Chrome(ChromeDriverManager().install())
/home/kajol/Desktop/temp_files/sel2.py:16: DeprecationWarning: find_element_by_*
commands are deprecated. Please use find_element() instead
    Posts = driver.find_element_by_xpath('//*[@id="react-root"]/section/main/div/he
ader/section/ul/li[1]/div/span').text

/home/kajol/Desktop/temp_files/sel2.py:18: DeprecationWarning: find_element_by_*
commands are deprecated. Please use find_element() instead
    Followers = driver.find_element_by_xpath('//*[@id="react-root"]/section/main/di
v/header/section/ul/li[2]/a/div').text
/home/kajol/Desktop/temp_files/sel2.py:20: DeprecationWarning: find_element_by_*
commands are deprecated. Please use find_element() instead
    Following = driver.find_element_by_xpath('//*[@id="react-root"]/section/main/di
v/header/section/ul/li[3]/a/div').text
1,638
21.2m followers
367 following
(base) kajol@kajol-Lenovo-ideapad-310-15IKB:~/Desktop/temp_files$
```

Image 4

Data Miner

Since crawling data on Instagram by running a python script was not possible, we looked at browser extensions to extract data. We used DataMiner which is a chrome extension and is free of cost.

- i) Installation: Use dataminer.io or simply search for DataMiner on your browser and click on Add to extension. This is a chrome exclusive extension. Click on Add to Chrome and the extension gets added.

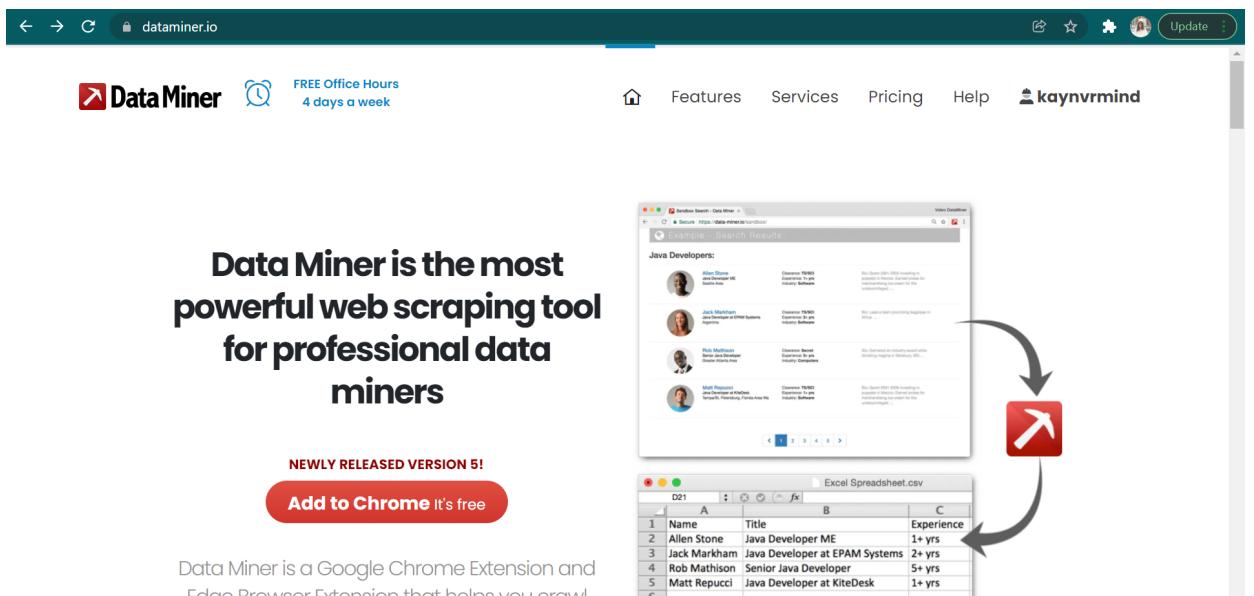


Image 5

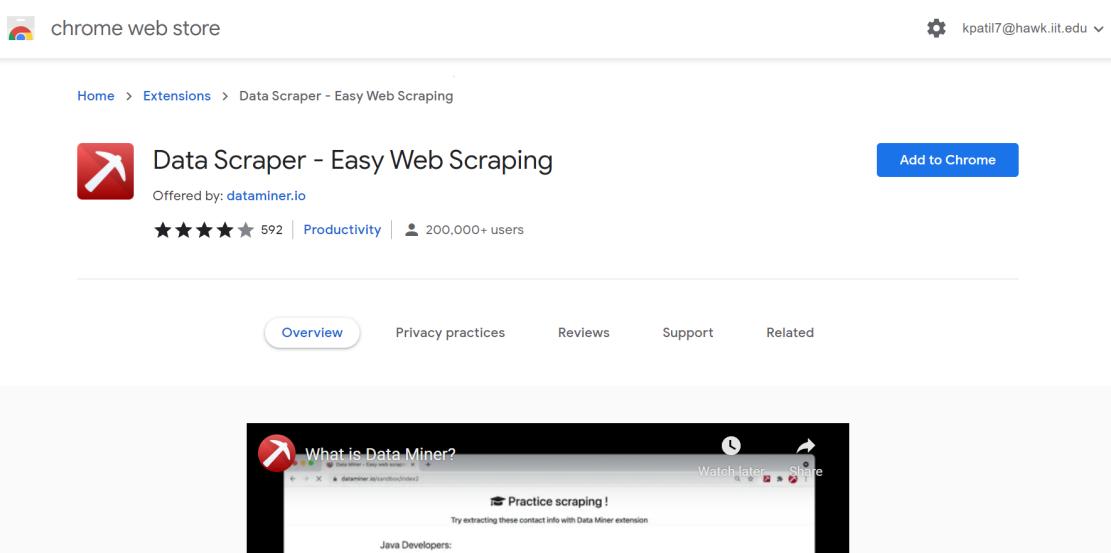


Image 6

- (ii) Working: After opening the profile that needs to be scrapped, click on the DataMiner browser extension. It should display a few pre-programmed recipes

(ie, scripts that only scrape comment, hashtags, followers, following etc.) Select the appropriate option before clicking on downloading and choosing the format needed.

The screenshot shows a browser window with the Instagram profile of Barack Obama (@barackobama). The Data Miner extension is active, displaying a sidebar with various scraping options:

- Public Recipes:**
 - test
 - IG - Get Emails
 - IG - Get Emails
 - Instagram from Search
 - IG Generic Recipe - Get All Links
 - Instagram Following
- My Recipes:** (disabled)
- Pinned:** (disabled)

At the bottom of the sidebar, it says "Logged in as kaynvmind@gmail.com". Below the sidebar, there are three thumbnail images of posts from Barack Obama's Instagram account.

Image 7

The screenshot shows the Data Miner application interface. On the left, there is a sidebar with navigation links like Home, User Manual, SCRAPE (with Page Scrape selected), and MY DATA (with Saved Results (11) and Uploaded Files (0)).

The main area is titled "Scraping Results" and shows the results for "Instagram from Search". It includes a preview section with the message: "This recipe scrapes 37 rows and 2 columns from this page". Below this is a table with the following data:

URL ↑	Link ↑
/barackobama/followers/	35.3m followers
/barackobama/following/	6 following
https://1.instagram.com/?u=obama.org/15years	Followed by jessicacapshaw,
/barackobama/followers/mutual/	Posts
/barackobama/guides/	Guides
/barackobama/reels/	Reels
/barackobama/channel/	Videos
/barackobama/tagged/	Tagged
/p/CZ0Tie0jNMF/	

At the bottom of the application, it says "You are scraping this tab: https://www.instagram.com/barackobama/" and "kaynvmind@gmail.com (Free)".

Image 8

Scraping Results → Instagram from Search

New Scrape Scrape and Append Next Page Automation

Scrape *Scrape and only keep the result data of 1 scrape.*

logs **Download**

saved a minute ago
2/12/22, 2:21 PM **37 rows** **Delete** **Download**

Download options:
If you have too much data, your download will be split into multiple files of 10000 rows.
CSV File **Excel File** **Copy To Clipboard**

Copy data to another file inside Data Miner
You can duplicate this result file with a specific name. After saving, you can go to Saved Results tab to find your data.
duplicate file name **Save As**

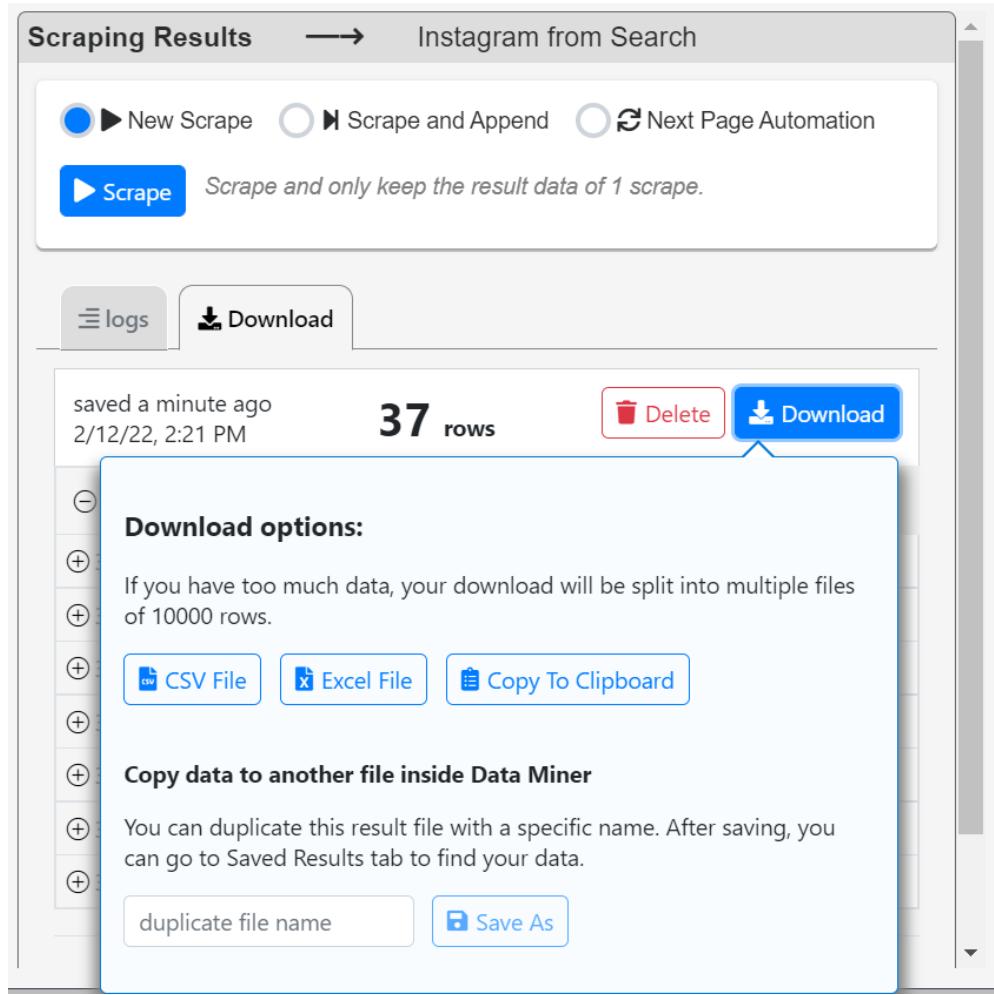


Image 9

For this project, we have scraped the ‘following’ data of some public profiles and analyzed their network.

We have scraped data of Top 10 most followed people on Instagram and a bunch of IIT clubs’ profiles and plotted their graphs.

Data Cleaning

Once data was collected, we had to clean the data to make sure it was consistent and didn't hamper our visualization.

These were what we anticipated:

- 1) Duplicate entries
- 2) Blank entries

Since our data was in a CSV format, it didn't take much time or effort to tackle these problems.

Feature Selection

Our focus was on the following count and the users that were followed by the profiles we selected. There were multiple ways in which this could have been done; by using URLs, usernames or full names.

- 1) URLs: Although there are no drawbacks to using URLs, it still made the node name lengthy especially since the last part of the URL was the only part that was needed whereas the primary part wasn't. For example in this URL '<https://www.instagram.com/zendaya/>' the https part would be the same for all users and only the part after '.com/' was needed.
- 2) Full names: The drawback with full names is that if a verified user's data is extracted, it is possible that the user chooses not to enter their name as their username was sufficient, so the data field was blank.
- 3) Username: Since Instagram has rules that ensure usernames are unique, bound by certain characters, this seems like the ideal field to use.

Data Visualization

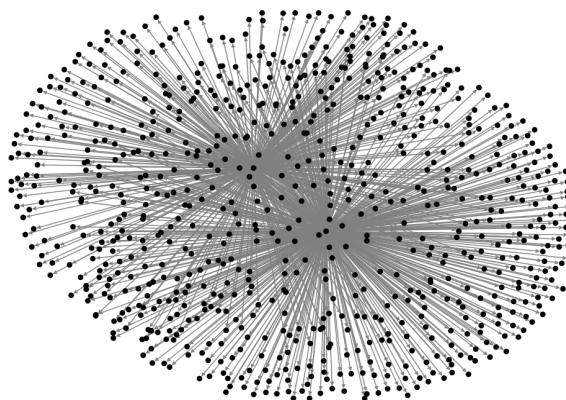
(i) Availability: The tool we used for visualization is NodeXL Basic. Since it's an excel template, it's pretty easy to use and doesn't require a lot of time to get used to. But at the same time, it's only available on Windows machines. Since it uses Excel, it doesn't work on a Mac or Linux system. If you don't have a windows machine, you could use AWS EC2 instances and get a virtual machine with Windows OS free of cost.

(ii) Choosing the right graph: We wanted to establish a friendship network between (say two) users based on the number of people they follow and what percentage of those are common between them. We can make a prediction that if the number of common following is high, that would mean that they are close friends and if it's low that could mean that they are just colleagues.

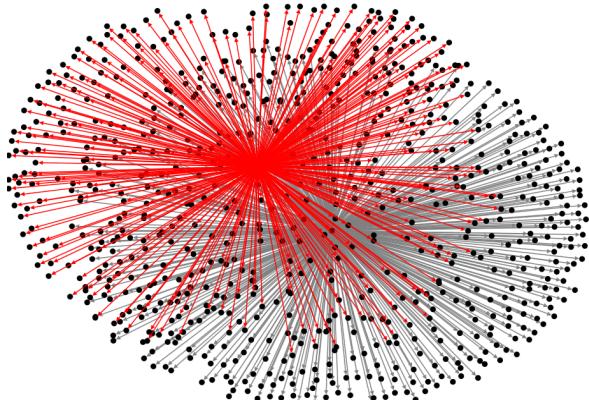
We picked two actresses, working in the same form of media and plotted their results.

NodeXL offers a variety of different graphing tools, it's important to identify which graph is the appropriate one.

Graph 1.1 doesn't return much information visually. In Graph 1.2 the red lines indicate a relationship between the main node and all the nodes the user follows. Which would mean that everything is jumbled up.

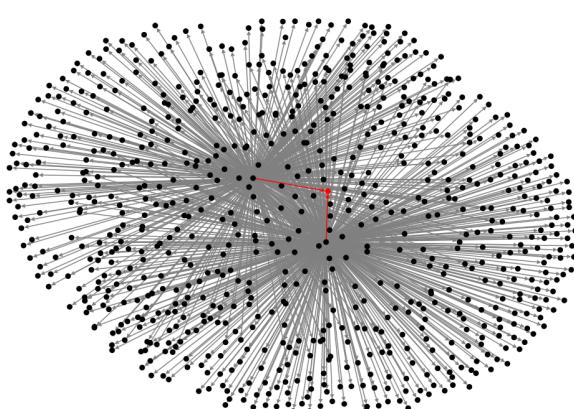


Graph 1.1

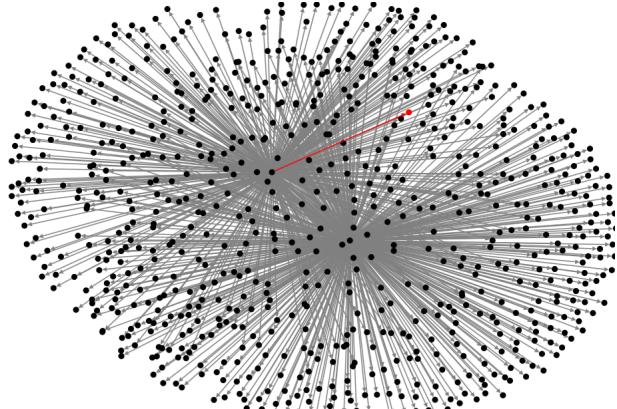


Graph 1.2

Graph 1.3 shows a user that has been followed by both the actresses, it could be a common friend, director or maybe a brand the two of them have worked for and Graph 1.4 shows a user that has been followed by only one actress.



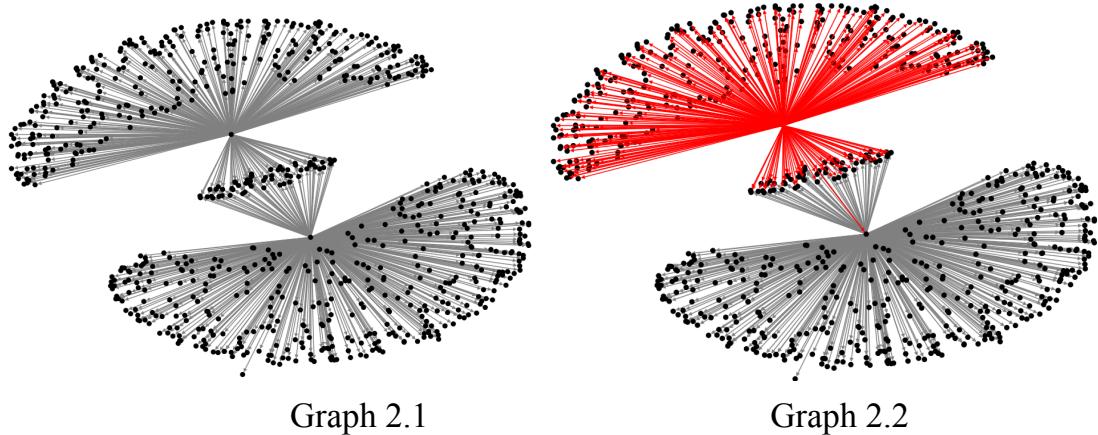
Graph 1.3



Graph 1.4

Since the graph is difficult to read it is difficult to draw any conclusions from it, which is why it is important to select the appropriate graph based. We found that the Harel-Korean Fast Multiscale graph had the best results.

No details were changed, yet visually it's clear that Graph 2.1 gives more information than any of the former graphs.

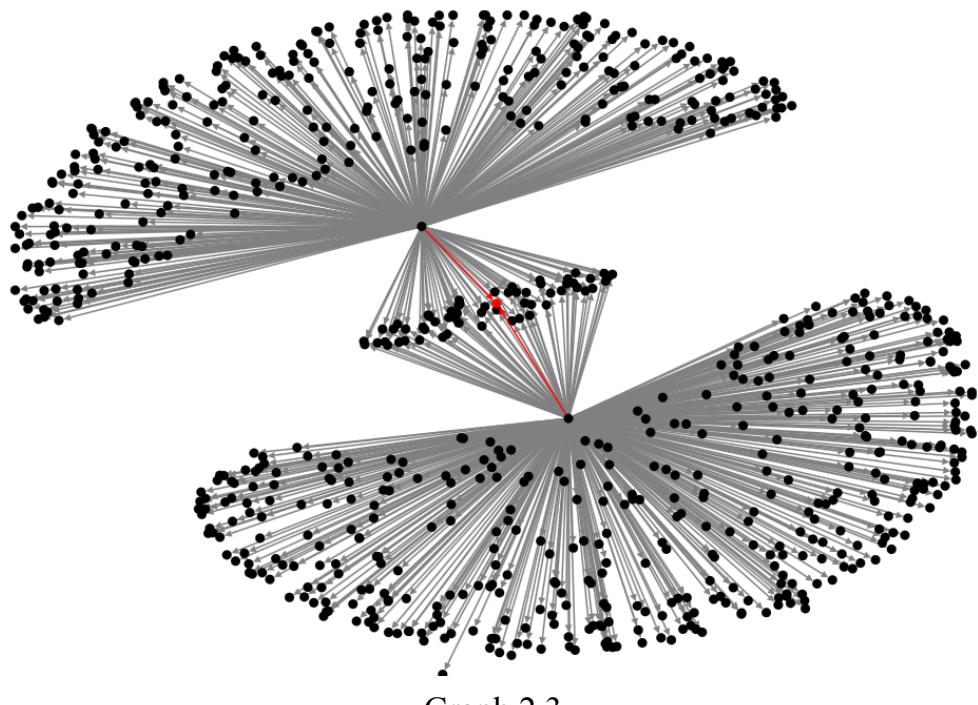


Graph 2.1

Graph 2.2

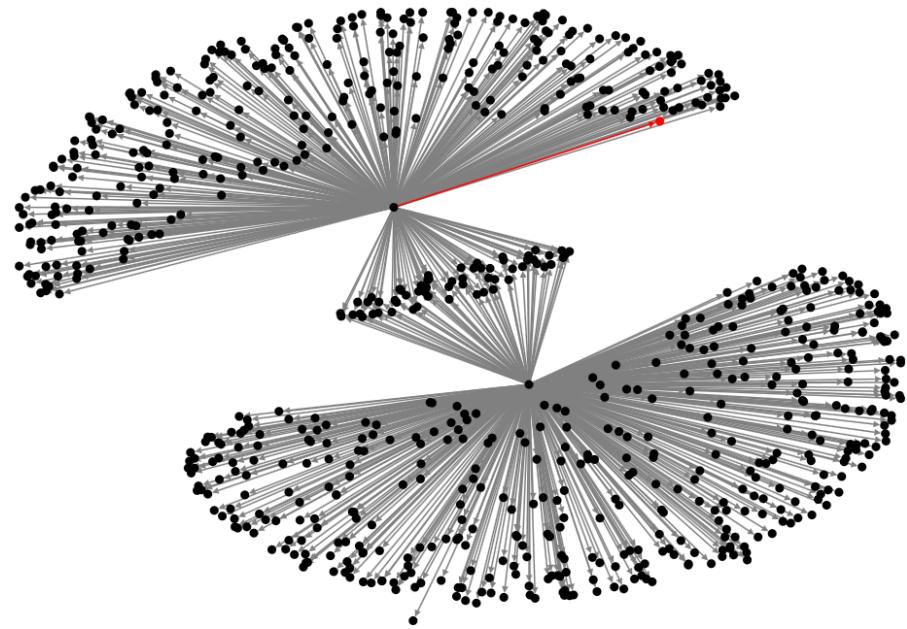
Graph 2.2 represents all the followers of Actress A. It's interesting to note that there is a red line going from one popular node to another, indicating that the two of them do follow each other.

Similarly we can generate all the other graphs that gives more information visually that Graphs 1.3 and 1.4.



Graph 2.3

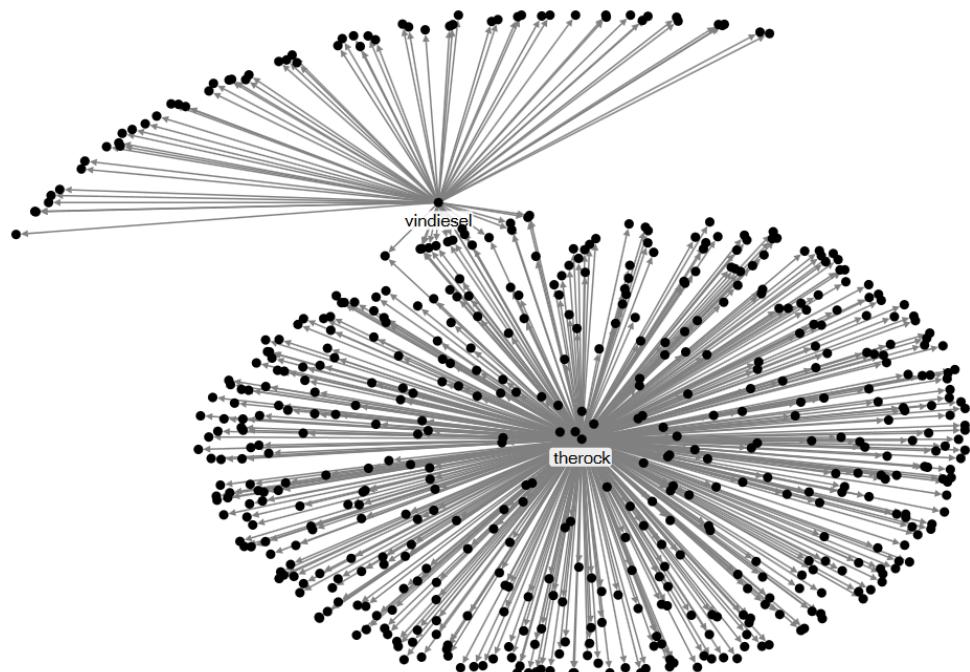
Here, we can conclude that the middle portion of the graph represents the followers the two actresses have in common. These nodes could represent other actors and actresses, maybe producers and directors that the two of them have worked with or popular brands that they represent.



Graph 2.4

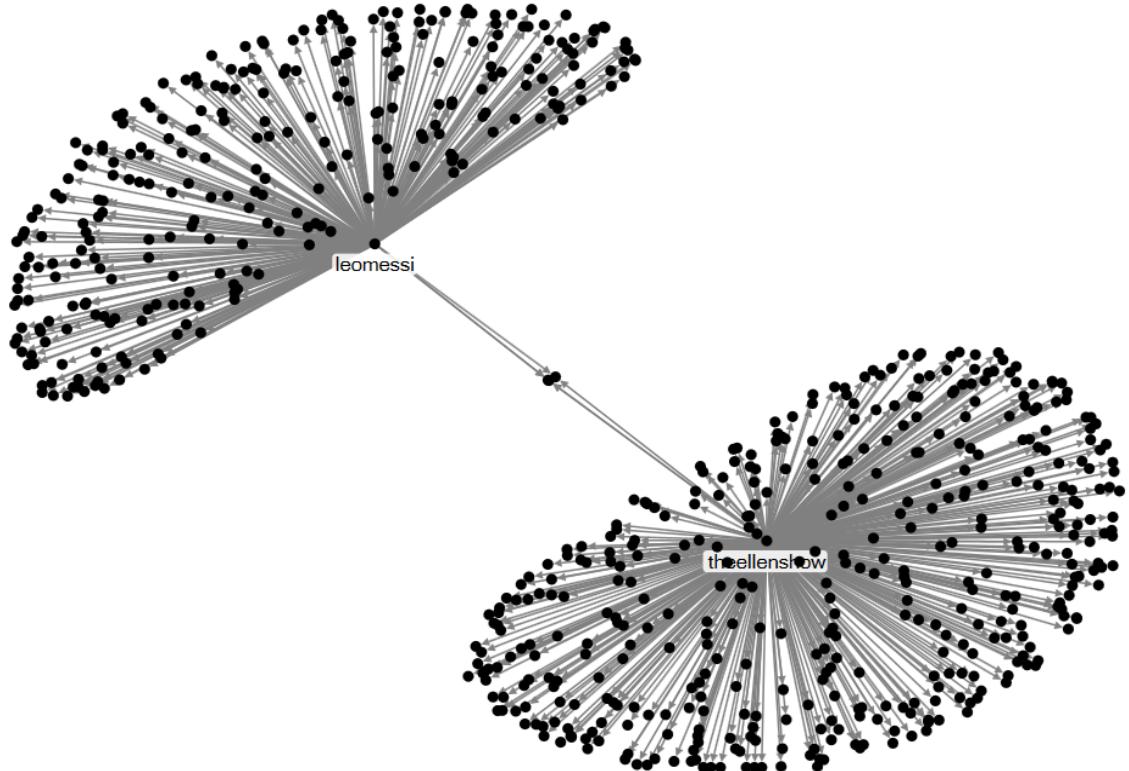
Even though Graph 2.4 is redundant since visually we can tell who follows whom as the graph has created a clear distinction between the nodes.

Another experiment we conducted was to build or try to simulate an enemy network, that is take two people who do not get along with each other and try to show a similarity in their followers. Also, we tried to establish similarities between two people who probably don't know each other personally, but might end up having some commonalities in their followers.



Graph 3

It is known that the famous actors Dwyane ‘the rock’ Johnson and Vin Diesel do not get along, to the point that they refuse to film with each other. In the movie franchise fast and furious, the scenes where the two of them were in together were filmed separately and then merged together. We plotted their follower graphs and found out that the two of them have friends in common that they follow. It doesn’t make up for a large percentage, but it’s still a significant number since Vin Diesel doesn’t follow many people. Most of their common followers include the cast and people who are a part of fast and furious along with a few late night hosts, musicians and artists.



Graph 4

We tried to eliminate all the factors that could contribute to two people having in common and landed on Lionel Messi the Argentine professional footballer who plays as a forward for Ligue 1 club Paris Saint-Germain and captains the Argentina national team and Ellen DeGeneres who is an American comedian, television host, actress, writer, and producer. She also hosts The Ellen DeGeneres Show, since 2003.

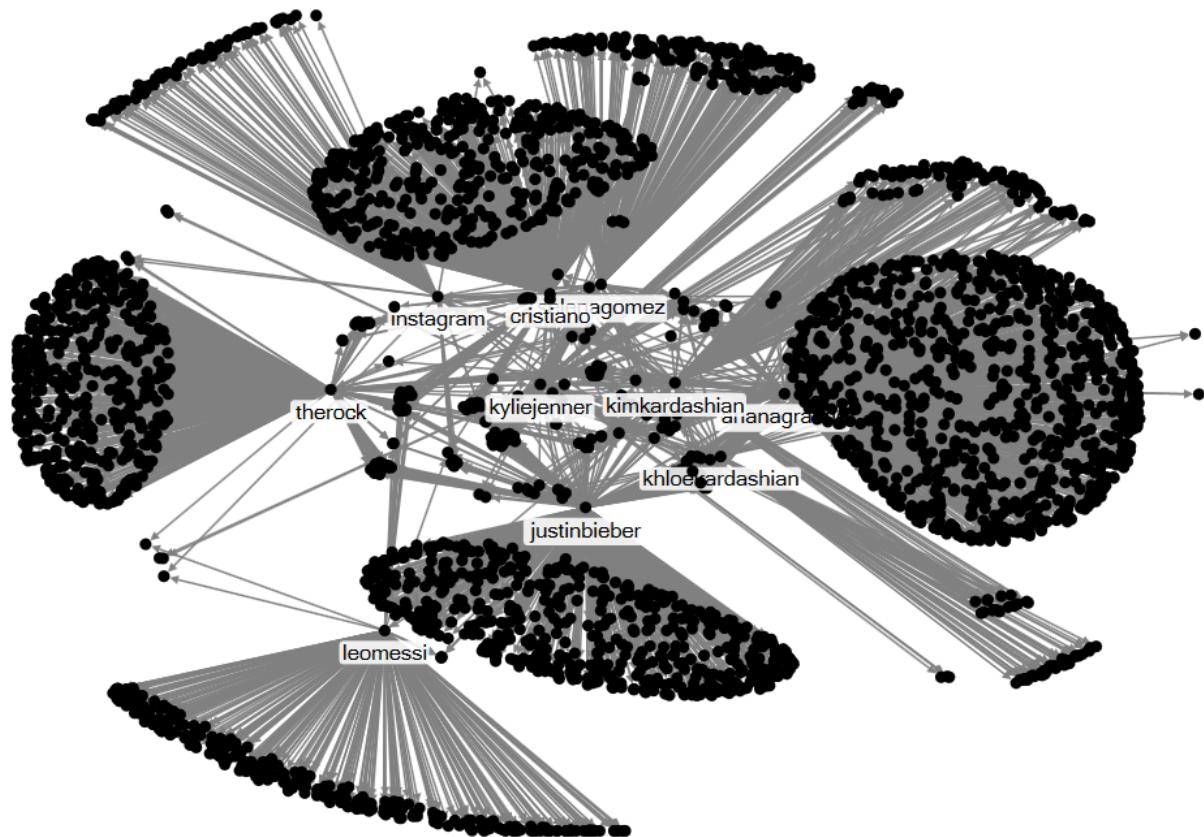
They have three nodes in common, namely, David Beckham, Ed Sheeran and Beyonce.

Network Measure Analysis

Top 10 most followed Instagram accounts

We scraped the ‘following’ list of the top 10 most followed accounts on Instagram. We wanted to analyze whether there are any profiles which are followed by more than one of the considered profiles. Below is the list of the top 10 most followed accounts:-

1. @instagram
2. @cristiano
3. @kyliejenner
4. @leomessi
5. @therock
6. @selenagomez
7. @arianagrande
8. @kimkardashian
9. @justinbieber
10. @khloekardashian



Graph 5.1

Above is a directed graph which is a friendship network representation of the Top 10 Instagram profiles. We can see that all the 10 profiles don't have the same profession. Leomessi and Cristiano are football players from Argentina and Portugal respectively whereas Justin Bieber, Ariana Grande and Selena Gomez are American singers. We have formed a friendship network where the users are the nodes and the edges show whether there is a friendship or any relation between them.

Conclusions:

1. While not all the top 10 Instagram users follow all the other users, there is still some notable following relationship established between them. There are some user accounts which are followed by more than one top 10 profile.
2. While they could be from different countries or have a different profession, there are profiles that are following some common accounts. For eg. Messi and The Rock have completely different professions; Messi is a football player whereas The Rock is a wrestler/actor and both are citizens of different countries. Still, there are some user accounts which both of them follow.
3. Profiles belonging to the same profession have some accounts in common which they follow. For eg, Ariana Grande and Selena Gomez, both are singers and follow some same accounts. As both have the same profession, it can be the case that the common users which they follow are also from the same profession or if they are friends, the accounts could be of their friends too. If they follow each other too, it could also be the case that they know each other or may have collaborated or worked together in the past or might do it in the future.
4. An association can be witnessed in the form of A follows B and B follows C, while A may or may not follow C. Justin Bieber follows Kim Kardashian and Kim Kardashian follows Sephora (Makeup brand) but Justin Bieber doesn't follow Sephora. At the same time Justin Bieber follows Kim Kardashian follows Britney Spears and Justin Bieber does follow Britney Spears as well.

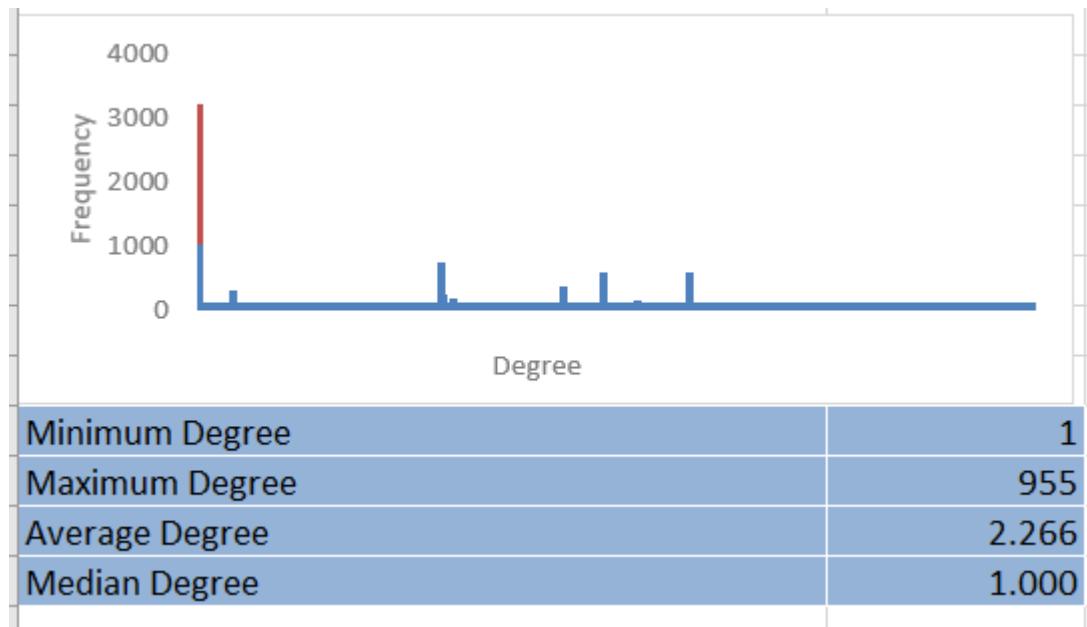
Graph Metric	Value
Graph Type	Directed
Vertices	3143
Unique Edges	3571
Edges With Duplicates	0
Total Edges	3571
Self-Loops	0
Reciprocated Vertex Pair Ratio	0.0028082
Reciprocated Edge Ratio	0.005600672
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	3143
Maximum Edges in a Connected Component	3571
Maximum Geodesic Distance (Diameter)	5
Average Geodesic Distance	3.116201
Graph Density	0.000361609
Modularity	Not Applicable
NodeXL Version	1.0.1.448

Graph 5.2

The above are the graph metrics of the friendship network established for the directed graph of the top 10 profiles' network with 3571 edges and 3143 nodes.

Network measures: Degree Distribution

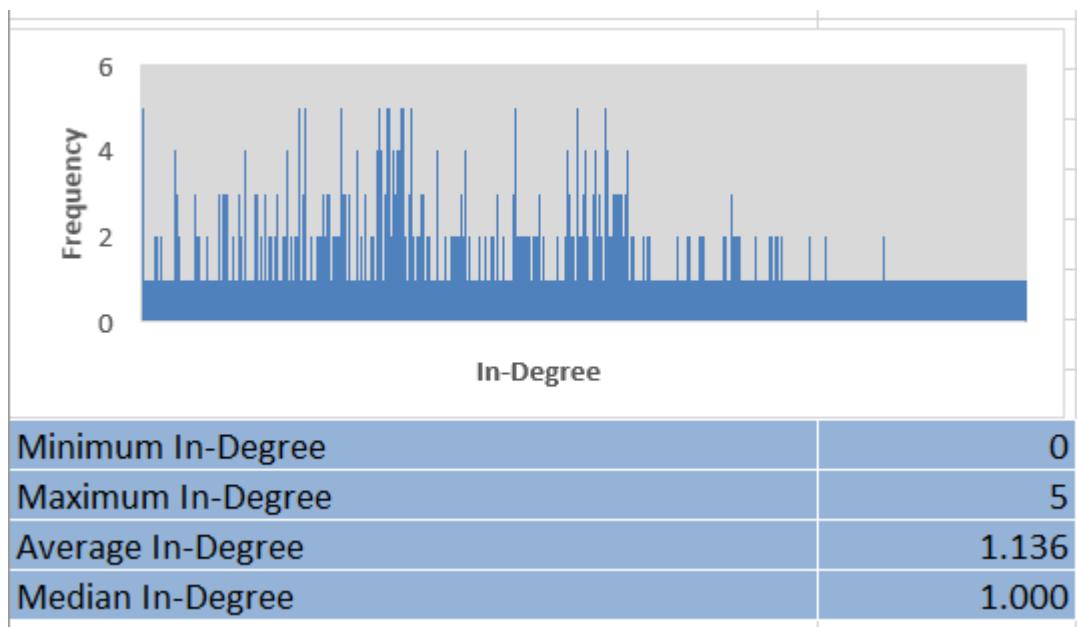
The degree of a node is the number of connections that it has to other nodes and degree distribution is the probability distribution of the degrees over the network. The degree distribution is plotted as a histogram where the minimum degree of a node in the network is 1 and the maximum degree is 955. The frequency of the nodes is 3143.



Graph 5.3

In-degree

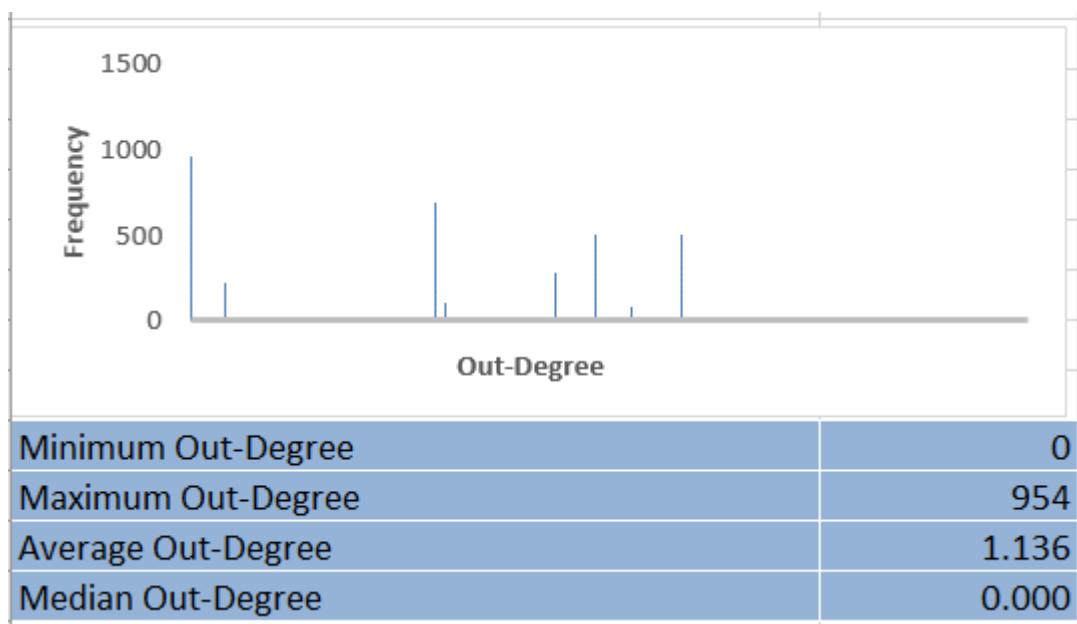
In-degree is the number of incoming edges directed into the nodes in a directed graph. The minimum in-degree for the graph is 1 as at least 1 user account which is followed by any 5 of the top 10 profiles.



Graph 5.3

Out-degree

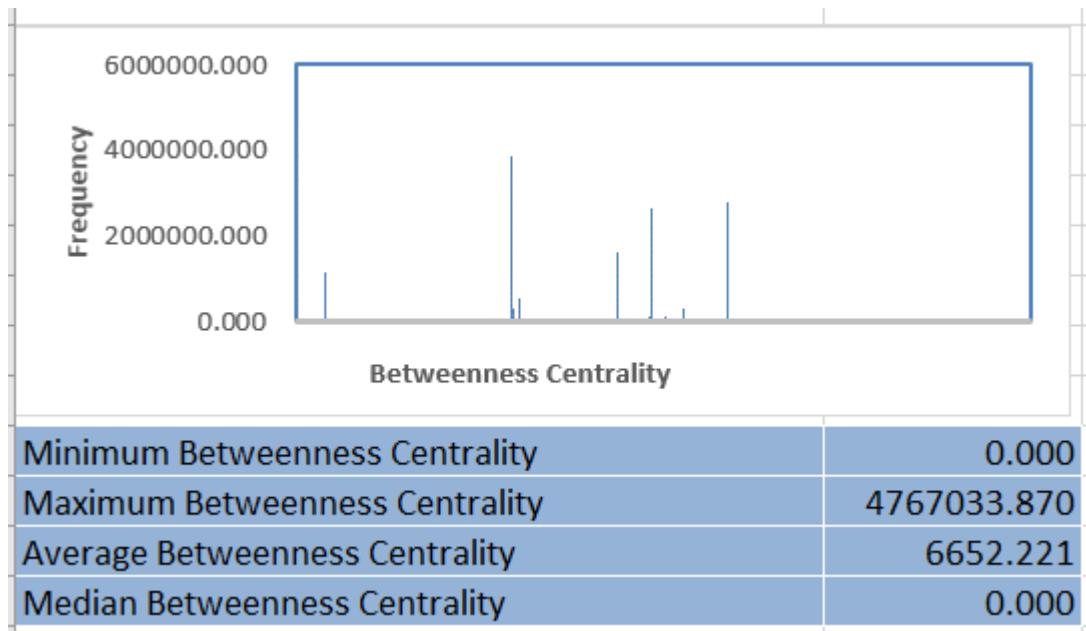
Out-degree is the number of outgoing edges from a node in a directed graph. The minimum out-degree in the graph is 0 and the maximum out-degree is the maximum following of one of the profiles, in this case it's Ariana Grande's following count.



Graph 5.4

Betweenness Centrality

Betweenness centrality is a measure of centrality in a graph.

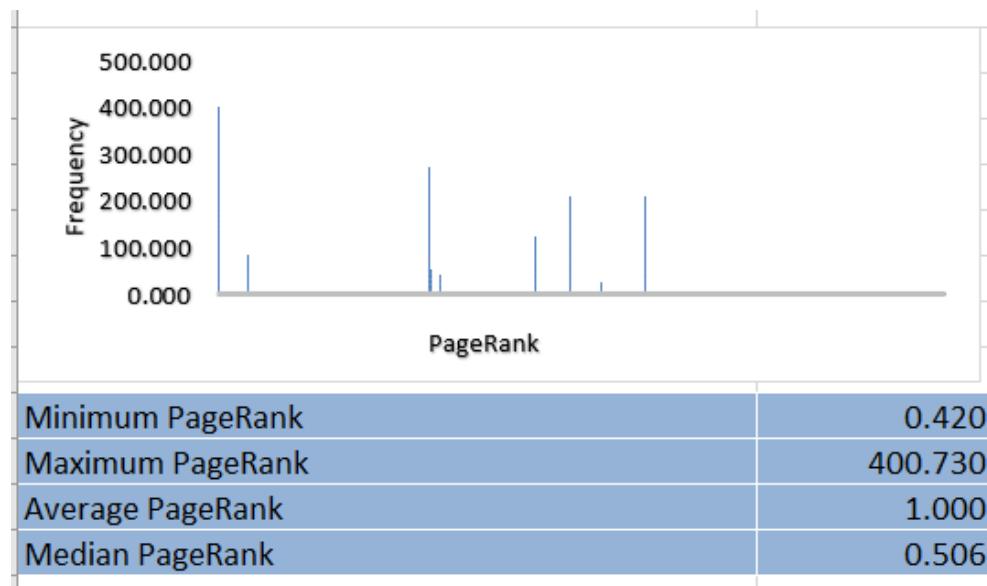


Graph 5.5

The nodes which act as a bridge for a node to reach another node have more betweenness centrality.

PageRank

Pagerank is the ranking of a node based on how important they are in a network.



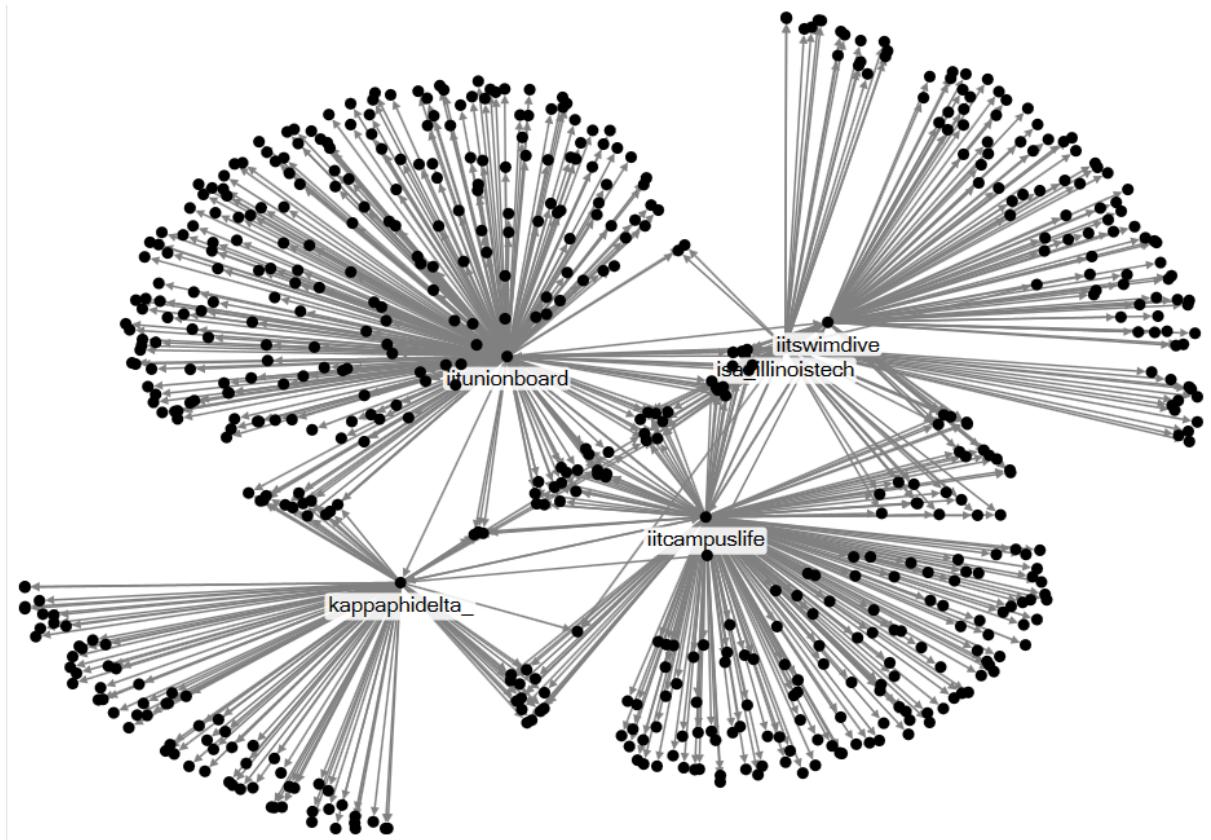
Graph 5.6

Here, mostly the nodes of the top 10 profiles are having higher page rank as they have many outgoing edges and incoming edges too in some cases.

Network Measure Analysis: 5 different organization/clubs at IIT

Here, we consider the results obtained from establishing a friendship network between different organizations at IIT. The organizations we picked are Union Board, Campus life, a sorority house, a sports team and a student social organization.

In theory, all the accounts are official accounts and should follow the school's official account along with a few similar ones. Like, student organizations may follow other similar organizations, sports teams may follow each other and etc.



Graph 6.1

The nodes in the center represent the strongest or the most general nodes like Illinois tech's official instagram account, Illinois tech sports and rec, etc. As the nodes stray away from the center, the fewer commonalities they share with the rest of the group. For example, there are a lot of nodes common between campus life and Kappa Phi Delta since the two of them look after on-campus

living. The commonalities between them could be that the two follow other official residents' houses like Gunsaulus hall, Kacek, MSV, etc.

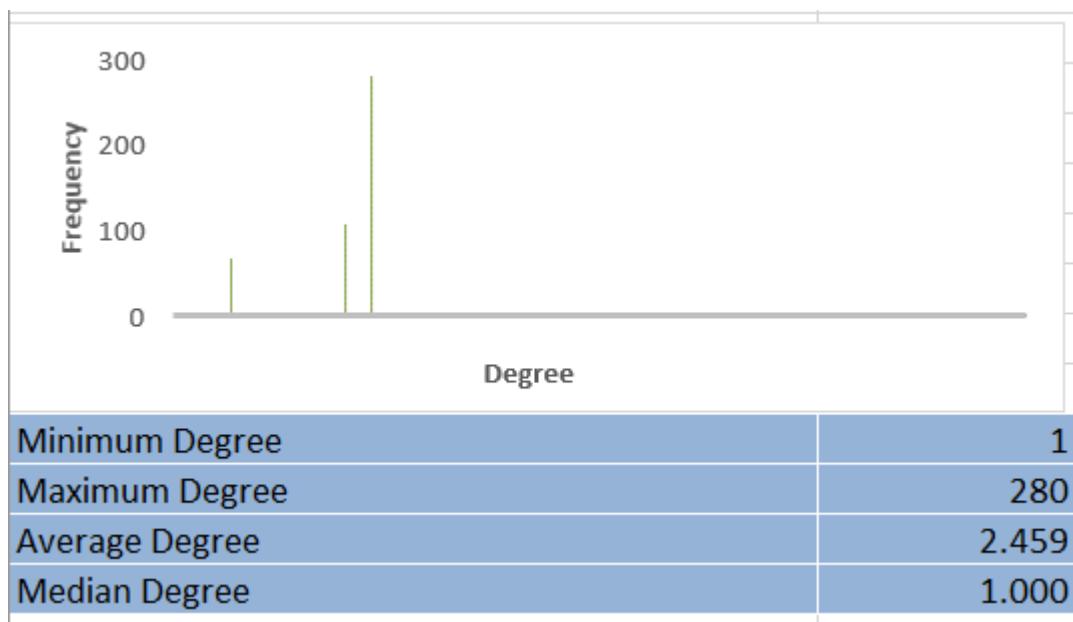
Graph Metric	Value
Graph Type	Directed
Vertices	623
Unique Edges	770
Edges With Duplicates	0
Total Edges	770
Self-Loops	0
Reciprocated Vertex Pair Ratio	0.005221932
Reciprocated Edge Ratio	0.01038961
Connected Components	1
Single-Vertex Connected Components	0
Maximum Vertices in a Connected Component	623
Maximum Edges in a Connected Component	770
Maximum Geodesic Distance (Diameter)	4
Average Geodesic Distance	2.653113
Graph Density	0.001987066
Modularity	Not Applicable
NodeXL Version	1.0.1.448

Graph 6.2

Graph 6.2 shows the metrics for the IIT network. Since we are only working with five accounts the number of edges is 770 with 623 nodes. A larger experiment could be run where all official accounts of IIT could be scraped to find out how well connected organizations are.

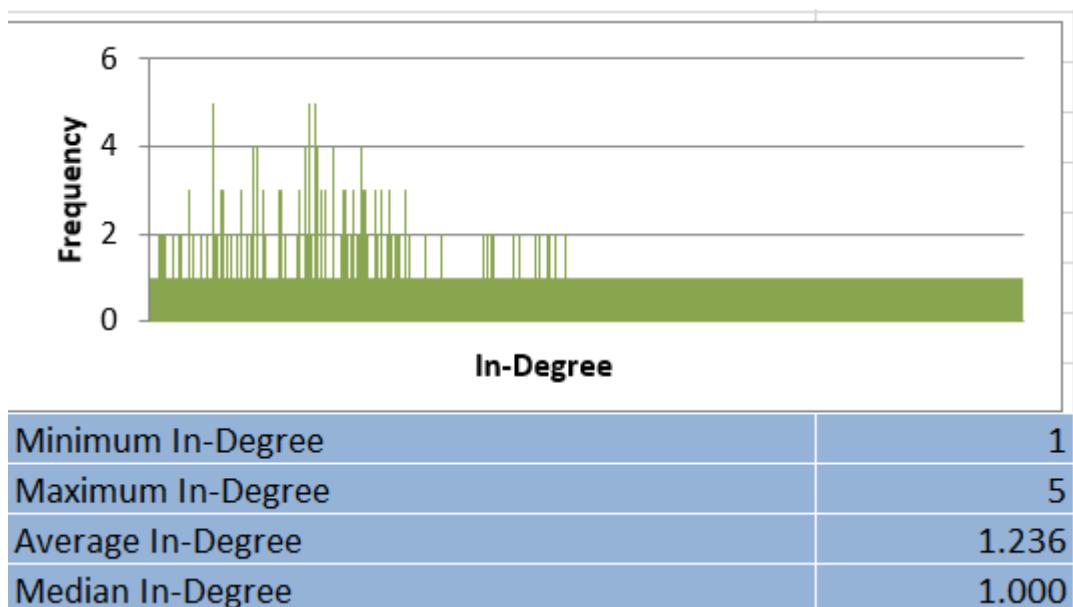
Degree distribution as histogram

Each node follows at least one account with 280 being the maximum followers. These could be other admin accounts or student body members of those organizations.



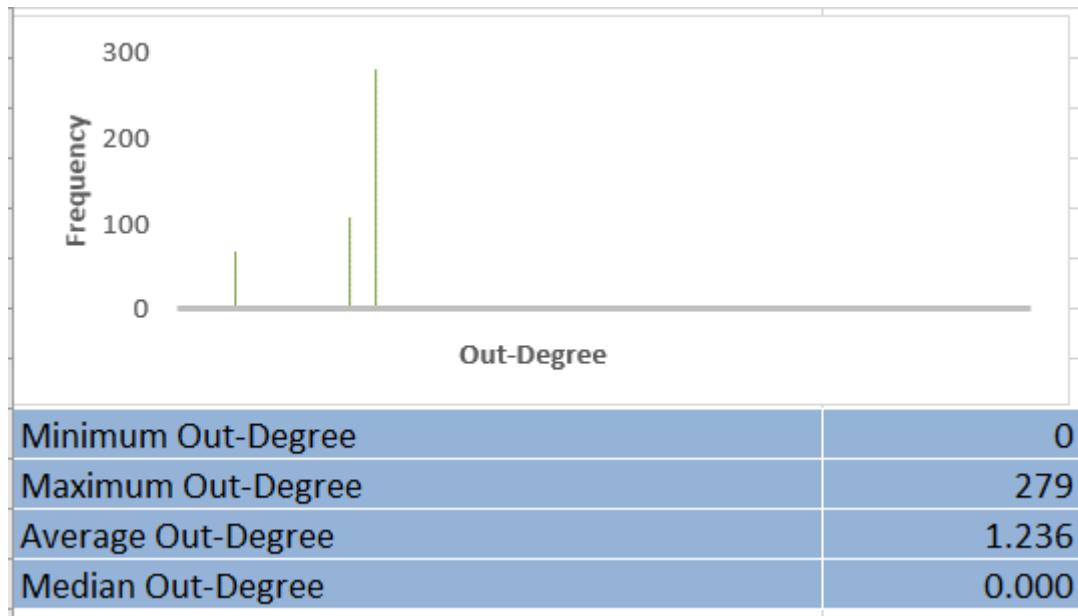
Graph 6.3

In-degree



Graph 6.4

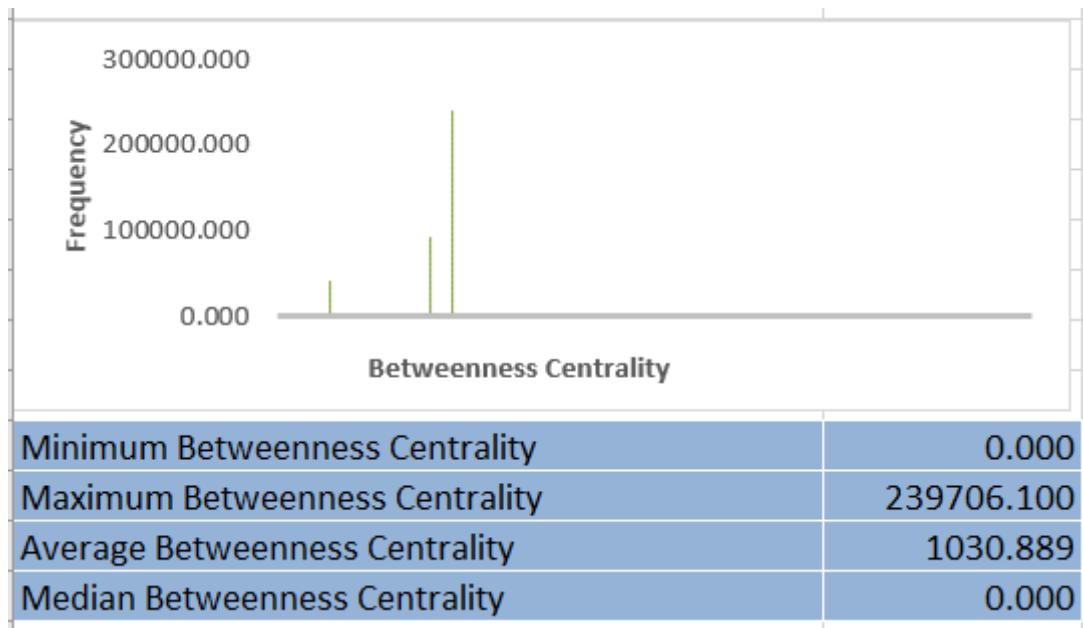
Out-degree



Graph 6.5

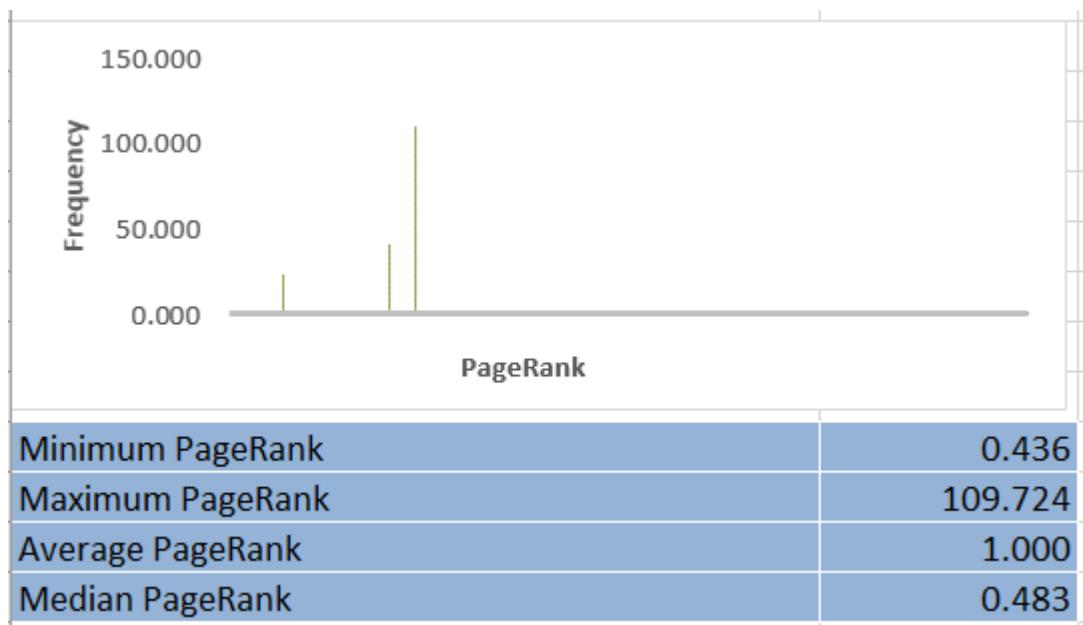
The out degree is zero since we are not considering all student accounts but only five. Say for example IllinoisTech is the official Instagram account of the university, but we haven't scraped its following, so the information we get here is how many of the 5 accounts follow IllinoisTech. Hence, it has indegree 5 but outdegree 0.

Betweenness Centrality



Graph 6.6

PageRank



Graph 6.7

Team Efforts

Tasks	Team-member name
Data scraping using python packages like instalooter and python script	Kajol
Data scraping of 5 IIT clubs' profiles and other celebrity profiles	Kayenat
Data scraping of Top 10 most followed Instagram accounts	Kajol
Data cleaning, feature extraction	Kajol
Graph plotting, histograms	Kayenat, Kajol
Network measures analysis	Kayenat, Kajol
Graph visualization using NodeXL	Kayenat
Report writing	Kajol, Kayenat
Data Mining using chrome extension	Kayenat

References

Instagram scraping:

[Easy Way to Crawl Instagram using InstaLooter | by Fahmi Nurfikri | DataSeries | Medium](#)

[How to scrape Instagram posts, comments, and photos](#)

[Scrape Instagram Data Legally via API - Followers, Comments, Posts, Hashtags & More](#)

[Easily scrape Instagram data in 5 minutes](#)

[Instagram Scraper: How to Scrape Data From Instagram \[2022\] | HackerNoon](#)

[GitHub - arc298/instagram-scrapers: Scrapes an instagram user's photos and videos](#)

[How to scrape and collect data from Instagram?](#)

[Data Miner](#)

Tutorials for NodeXL:

[NodeXL Tutorial \(part 1 of 3\)](#)

[Tutorial - How to Analyze Social Media Networks with Kumu and NodeXL](#)

[Walkthrough: Using NodeXL to Visualize Twitter Networks](#)

 [Unit 2.2 NodeXL Tutorial - Importing Twitter Data](#)