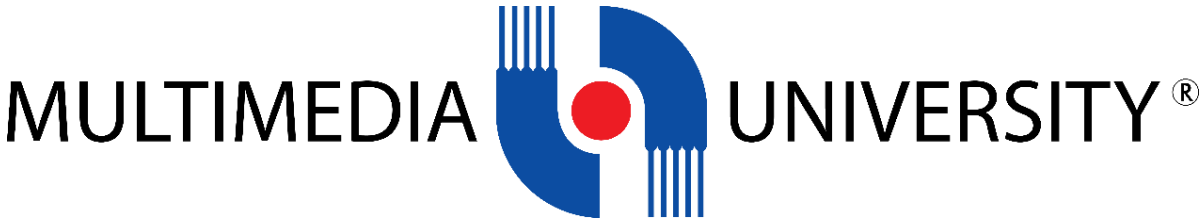

TDS3751

Social Media Computing

Social Network Analysis

Retail – Haha No Yume

Name	Student ID
Wong Zi Xiang	1142701102
Andy Yong Jun Jie	1142701565
Nazrin Zaidi	1141125301

Table of Contents

1 Introduction

Social networking has become one of the most important topics in recent times because of the several social networking sites that have now become part of our day-to-day living. There's a huge amount of information that we can gather from analyzing these social networks.

In this assignment, we are going to analyze the social network based on our target company, Haha no Yume's social media accounts. We will try to deduce information such as important people in the network and find groups of people or communities within the network.

2 Data Collection

For *Instagram* account, we will extract the information of Haha no Yume's followers and their followers as well (depth of 2) using API. Haha no Yume has a total of 9,647 followers at the time of extraction. There will be well over 10 million nodes and edges if we were to scrape all the follower data at depth of 2, which will take enormous amount of time to complete the betweenness centrality analysis (*O(nm)* time complexity, *n* = nodes, *m* = edges), not to mention our normal personal computer could not handle such humongous data. Thus, we decided to take a sample size of 30, which is the minimum sample size for the analysis to be statistically significant. A mere 30 followers and their followers amounted to around 130 thousand nodes and edges, imagine if we extracted all 9,647 followers at depth of 2.

For *Facebook* account, we could not extract the users' info of a page anymore since 5th February 2018, as the functions to extract user's info was changed to open to developers with the page's access token, which unfortunately we do not have. But, we could obtain the page like network of Haha no Yume through *netvizz*. The graph of page like network we obtained is comprised of 29 nodes and 51 edges, which is pretty small because Haha no Yume only like 3 pages.

3 Graph Statistics

3.1 Instagram

3.1.1 Basic Statistics

- **Average Degree** of 1.002. This means that most nodes are only connected to one node, which is normal since the depth is 2.
- **Graph Density** of 0.000007. This means that most of the nodes are not connected to each other.
- **Average Path Length** of 1.876. This means that on average, the entire journey from a start node to an end node would last nearly 2 nodes.

3.1.2 Centrality Measures

A. Degree Centrality

To measure degree centrality, we have chosen the in-degree as the metric. Number of in-degree is chosen because it represents the number of followers a node has. So, if we can find a node with the most in-degree, we found the person with the most follower. Also, out-degree is not considered in this assignment because for Instagram, people will not be able to see your updates and posts if only you followed them but they did not follow you. In *Figure 3.1*, we can see the top 10 most popular Instagrammer in Haha no Yume's network.

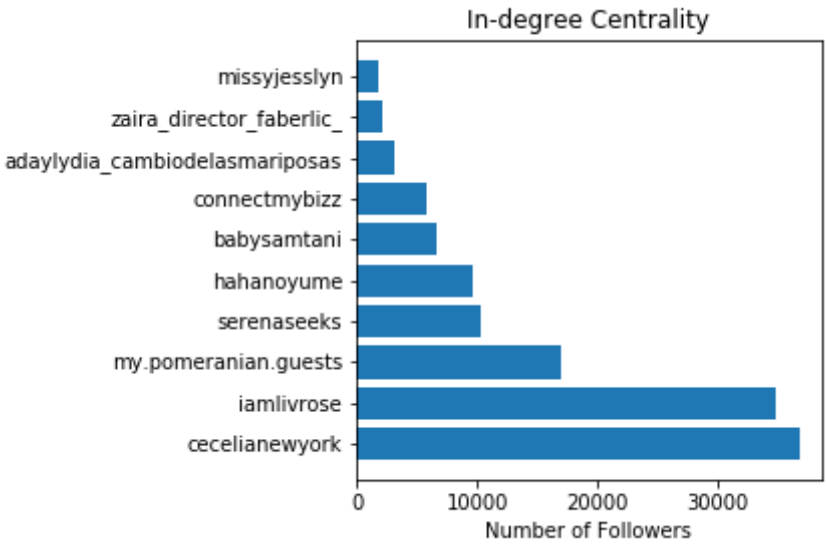

Figure 3.1 – Top 10 nodes with highest in-degree

B. Eigenvector Centrality

Eigenvector centrality appoints score to nodes based on how many other important nodes it is connected to. Through this, we can find out which node is connected to more influencers or whether they are the influencer themselves. In *Figure 3.2*, we can see the top 10 highest eigenvector centrality Instagrammer.

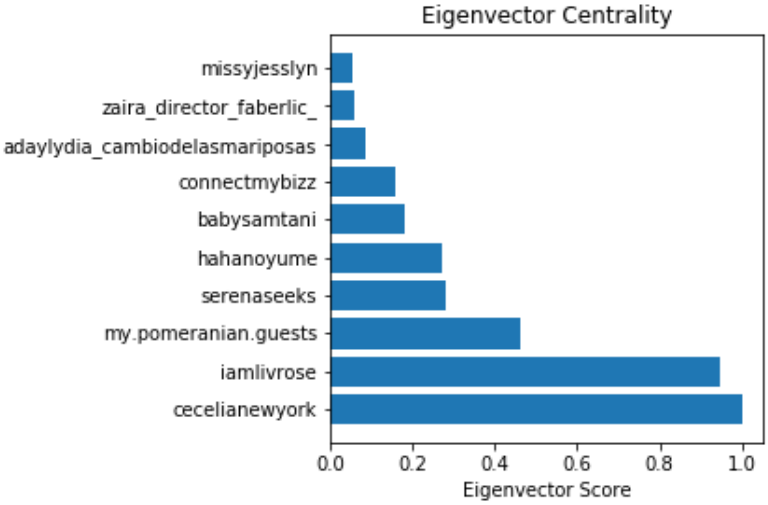

Figure 3.2 – Top 10 nodes with highest eigenvector score

C. Betweenness Centrality

Betweenness centrality measures how often does a node appears on other nodes' shortest traversal path. In other words, it measures how important nodes are in connecting other nodes. In *Figure 3.3*, we can see the top 10 highest betweenness centrality Instagrammer, Haha no Yume excluded.

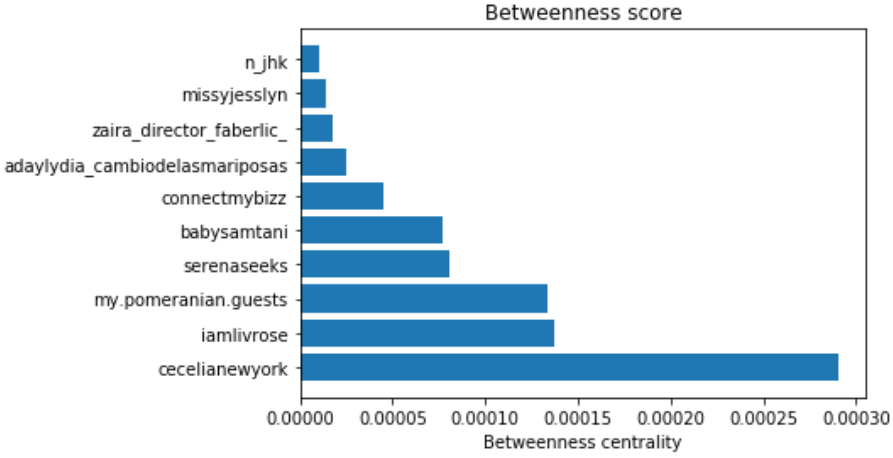

Figure 3.3 – Top 10 nodes with highest betweenness score

D. Closeness Centrality

To measure closeness centrality, we can use the Average Path Length algorithm. This will give us the average path length to be 1.876. Closeness centrality can be used to identify the people who can easily dissiminate information to other nodes. They are mostly sitting in the center of the network graph in which they have almost direct or short path to other nodes in the network. In *Figure 3.4*, we can see the top 10 highest closeness centrality Instagrammer.

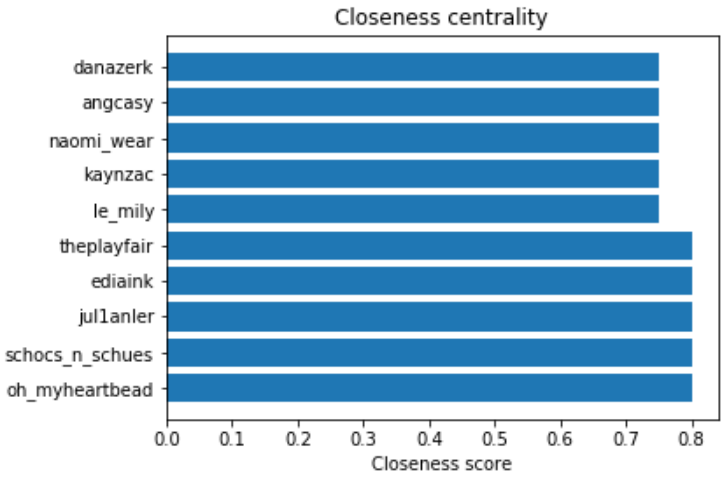

Figure 3.4 – Top 10 nodes with highest closeness score

E. High Degree against Closeness and Betweenness

Name	In-Degree	Closeness	Betweenness
cecelianewyork	36707	0.666667	0.000291
iamlivrose	34692	1	0.000137
my.pomeranian.guests	16962	0.666667	0.000134

Table 3.1 – Top 3 Instagrammers' centrality measures, sorted by in-degree

From *Table 3.1*, we can see that *cecelianewyork* has the highest degree, but it has low closeness. This means that he in a community that is far from the rest of the network. *iamlivrose* has high closeness but low betweenness, this means that its connections are more redundant because communication bypasses this node less.

F. High Closeness against Degree and Betweenness

Name	In-Degree	Closeness	Betweenness
iamlivrose	34692	1	0.000137
hahanoyume	9583	1	0.000386
babysamtani	6550	0.75	0.000077

Table 3.2 – Top 3 Instagrammers' centrality measures, sorted by closeness

From *Table 3.2*, we can see that *cecelianewyork* has low closeness. This means that he is near many nodes, but so are many other nodes. *Hahanoyume* is excluded, we should not analyse target company in this case because we are trying to find influencers. *babysamtani* has low degree, he maybe a key node connected to other important alters, but it has low betweenness at the same time. So, he is crossed-off our influencer list.

G. High Betweenness against Degree and Closeness

Name	In-Degree	Closeness	Betweenness
hahanoyume	9583	1	0.000386
cecelianewyork	36707	0.666667	0.000291
iamlivrose	34692	1	0.000137

Table 3.3 – Top 3 Instagrammers' centrality measures, sorted by betweenness

From *Table 3.3,* we can see that *cecelianewyork* has the highest betweenness, but as we have mentioned before, we should not analyse our target company. *cecelianewyork* has a low closeness, which means that he monopolizes the ties from a small number of people to many others.

H. Discussion

From the above analysis, it is found out that *iamlivrose* and *cecelianewyork* are the best influencers in Haha no Yume's Instagram network. *my.pomeranian.guests* can be considered to be approached.

3.1.3 Community

We used Gephi, a graph visualizing tools, to visualize the communities of Haha Yume's Instagram network. We calculated the modularity with 1 resolution. There are 19 communities found from this network, but we will only explore the Top 3 communities in depth.

Figure 3.5 – Haha no Yume's Instagram network graphs

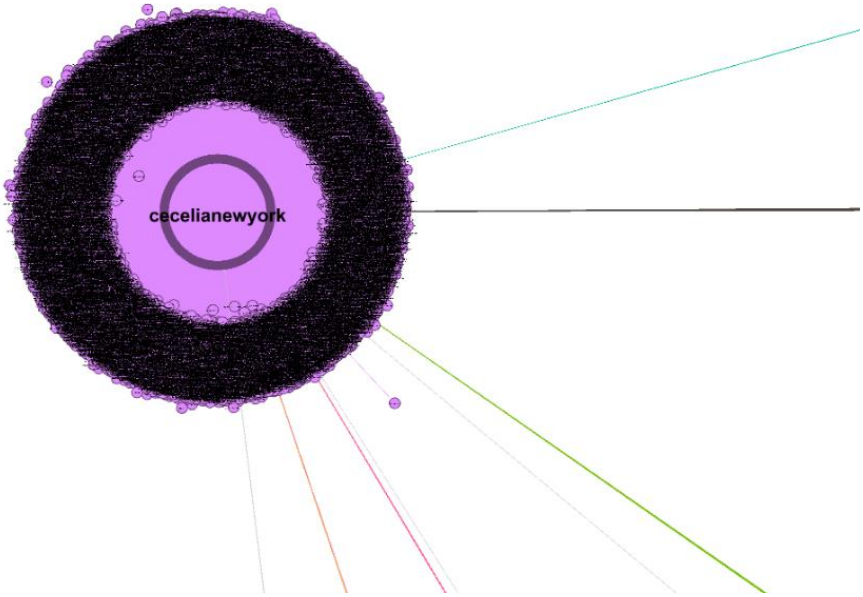

Figure 3.6 – Largest community

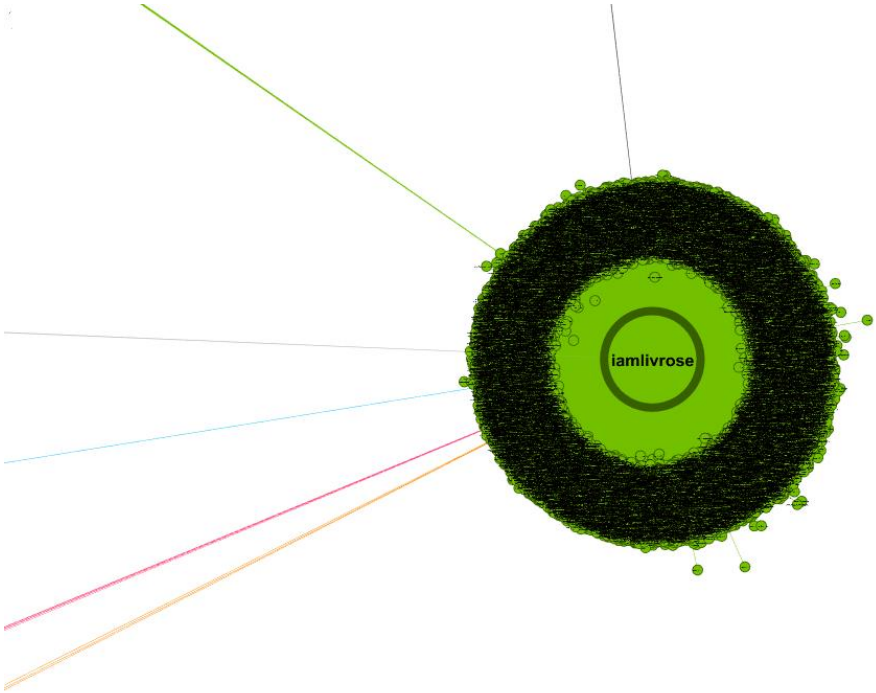

Figure 3.7 – Second largest community

The largest community (Purple, 27.76%) is focused around *cecelianewyork*, which is one of the Instagrammer we regarded as one of best influencers in the network. The second largest community (Green, 26.26%) is focused around *iamlivrose*, another best influencer on our list.

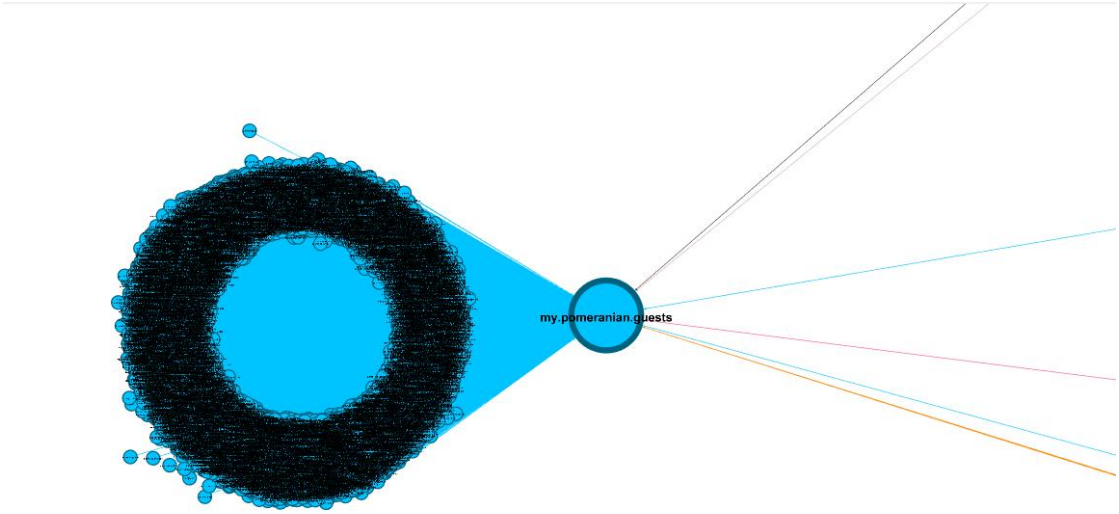

Figure 3.8 – Third largest community

On third largest community (Blue, 12.84%), we have a closer look on how a single could influence so many other nodes. We can clearly see how *my.pomeranian.guests*, which has the third highest degree in the network, formed its own community in the network.

3.2 Facebook Page Like Network

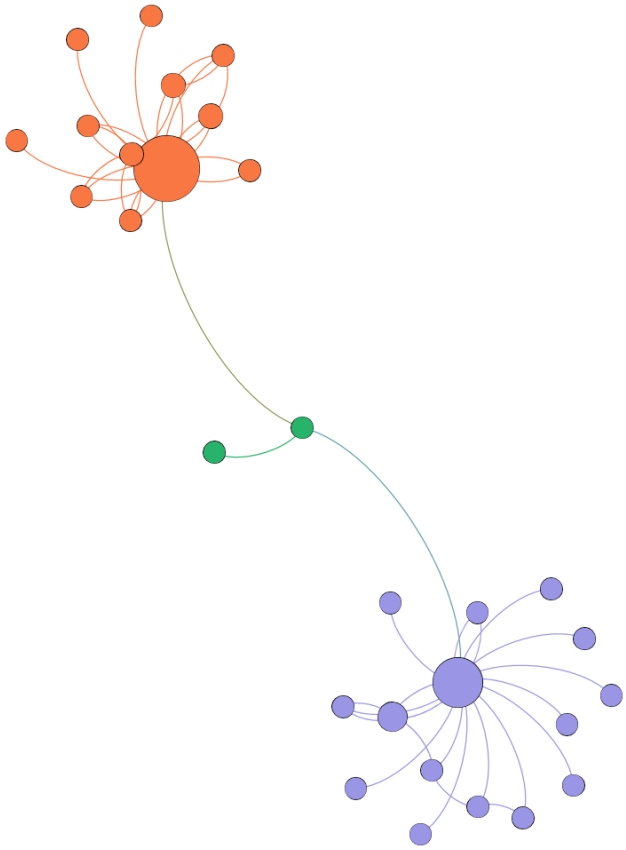

Figure 3.9 – Haha no Yume's page like network graph

As can see from the graph in *Figure 3.9*, there is not much information obtainable from Haha no Yume's page like network as they only liked 3 pages. But there are a few insights from this graph.

Firstly, from the zoomed in orange community in *Figure 3.10*, we can see that Haha no Yume's had interest in clothing, shopping and retail pages. Whereas in *Figure 3.11*, we can see that Haha no Yume's had interest in photography pages, most likely due to that they need photographer for their product's models.

Through this graph, Haha no Yume's can utilize the most influencing page such as BlueCicada Photography to spread job offers for photographers in other related pages for their products' models.

Figure 3.10 – Zoomed in orange community

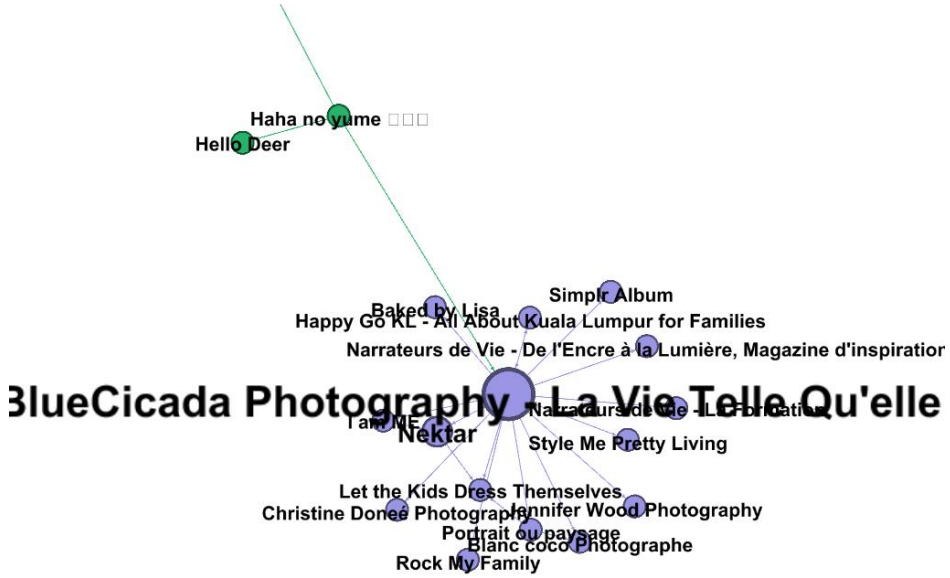

Figure 3.11 – Zoomed in purple community