TDS3751
Social Media Computing

Assignment (Part 1&2)
Tutorial Session: TT1V
Lecturer: Soo Wooi King

Student Name	Student ID
Lim Yixen	1171103542
Bryan Yeap Yee Kuong	1171103792
Wong Jit Chow	1171103534

Table of Contents

Introduction

The domain chosen is cinematic domain. The three brands selected are Golden Screen Cinemas (GSC), Tanjong Golden Village (TGV) and MMCineplex (MMC).

History

Malaysia has 169 cinemas in the cinematic domain that are open to the public. Perlis and Kelantan are the two states without cinemas. Golden Screen Cinemas is the largest movie chain. TGV Cinemas, and MMCineplexes are among the larger operators. Aside from these companies, there are a few smaller companies that undertake screening on a modest basis. As for our project this time around, we have chosen Golden Screen Cinemas and TGV Cinemas and mmCineplexes as our brands for comparison.

Golden Screen Cinemas (GSC), Malaysia's largest cinema exhibitor, is a wholly-owned subsidiary of the PPB Group (a Kuok Group member), which is a movie and content exhibitor and distributor in Malaysia. Through a cooperation with Galaxy Studio, it operates approximately 600 screens in 70 places across Malaysia and Vietnam, with 504 screens in 55 locations in Malaysia and 108 screens in 18 locations in Vietnam.

Tanjong Golden Village (TGV), a joint venture between Tanjong of Malaysia and Village Roadshow of Australia, was formed with the help of Orange Sky Golden Harvest, a theater operator based in Hong Kong. Tanjong acquired full ownership of the remaining interests. In November 1994, the first TGV cinema in Malaysia opened at Bukit Raja Shopping Centre, and by 2005, more than 30 TGV cinemas had opened around the country. The first English film to be aired was King Kong, followed by Anniyan, a Tamil film. TGV began theatrical distribution of films in 2013.

mm2 Asia Ltd., founded in 2008, is a significant film and television/online content producer in Asia. mm2 Asia is a producer that offers services throughout the filmmaking process, from financing to production to marketing and distribution. The parent company for mm2 Entertainment Pte Ltd (mm2 Singapore) and mm2 Entertainment Sdn Bhd is mm2 Asia (mm2

Malaysia). Through their Group companies and/or strategic working partnerships, the Group maintains representative offices in Malaysia, Hong Kong, Taiwan, and China. In seven years, they produced, co-produced, and/or distributed over 50 films thanks to their established multi-market presence. Cathay Cineplexes Sdn Bhd business operations at two locations in Malaysia - Cathay Cineplex Damansara and Cathay Cineplexes City Square - were bought by mm2 Asia in 2015.

Other Social Media Channel

Twitter, Facebook, and Instagram are the three social media profiles used by GSC Cinema. Their social media accounts have 185935, 2182415, and 349957 followers, respectively. GSC Cinema posts different types of content on its social media channels, with the exception of blockbuster movie posters and trailers. GSC Cinema is supposed to do this to attract social media users to follow all of their accounts on various social media sites. Instead of seeing the same content on all three platforms, GSC Cinema followers might see separate content from GSC Cinema.

TGV Cinema holds 3 social media accounts which are Twitter, Facebook and Instagram. Their followers count for the social media accounts are 80521, 837583, 239818 respectively. TGV Cinema uploads different contents on their social media accounts excluding blockbuster film poster and trailers. It can be said that TGV Cinema does this to encourage social media users to follow all their accounts on different social media platforms. Followers from TGV Cinema can then view different contents from TGV Cinema instead of repetitive contents on all 3 platforms.

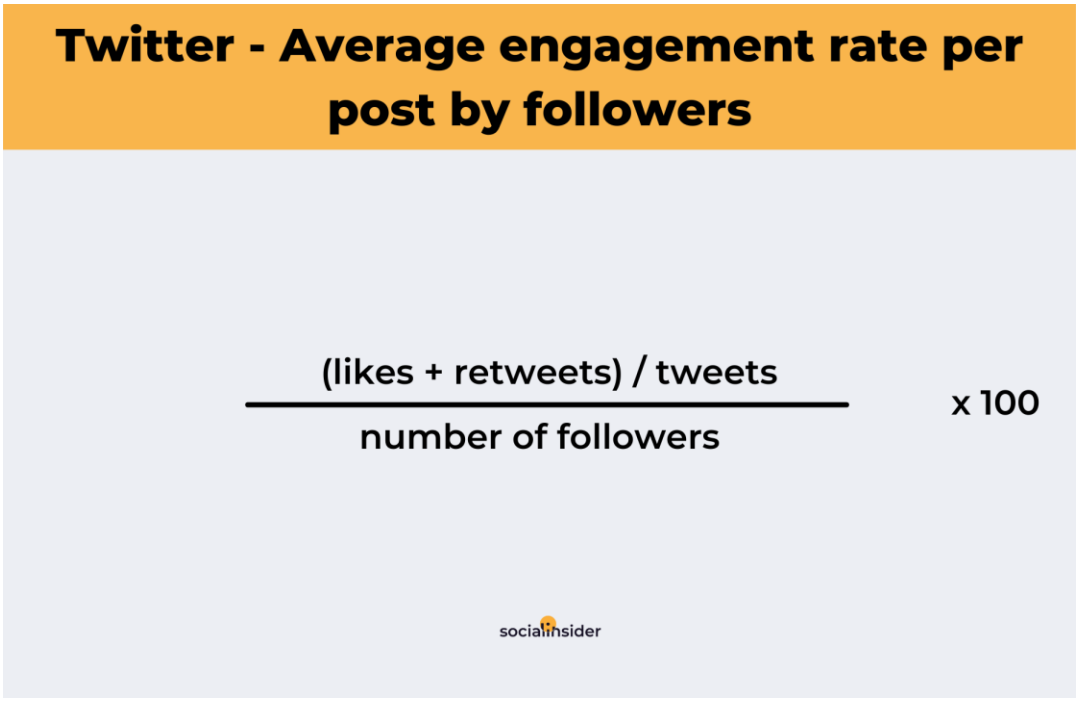
MMCineplexes holds 3 social media accounts which are Twitter, Facebook and Instagram. Their followers count for the social media accounts are 3811, 79684, 13030 respectively. MMCineplexes uploads the same contents on their social media accounts including blockbuster film posters and trailers. As for followers of MMCineplexes, they may find mmCineplexes content to be repetitive as the same content can be seen on three different social media platforms.

Part 1

Metrics

- Followers growth rate
    - This metric is used to determine whether the cinema has a positive follower growth rate which can increase their exposure of information.
    - As Twitter does not store the historical changes of followers count of the accounts, daily data collection is needed to record the number of followers each day within the two weeks period. Then, a line chart is shown by using this dataset. The percentage of the increasing or decreasing of the followers will also be calculated and displayed.
- Active rate
    - The active rate of a twitter account is important as it indicates whether the cinema is displaying enough information about what movies they are or they will be screening, so that the user can always get to know which cinema they need to go to for the movie that they like.
    - The tweets posted per day by each of the cinemas need to be collected. All the tweets whether the cinema created themselves or retweeted will be taken into consideration. A line chart is prepared for the active rate.
- Responsiveness
    - Another significant aspect that affects the audiences to choose which cinema to go to is the responsiveness of it. Sometimes, the audiences have many questions to ask about the movies such as the screening time, the ticket price, the type of hall screening the movie etc. The audiences need a fast response or it might affect their choice of cinema and interest.
    - From the tweets posted by each cinema collected above, a filter is applied to make sure the tweets are replied to someone. Then, the reply ID of the tweets are used as a parameter to search for the original tweet by using the Twitter API. The reply time is added up and averaged by the total replies made. A bar chart is used to show the average response time of each cinema.

- Engagement rate
    - Engagement rate is the key to understanding social media. In other words, this is the way to measure how people interact with the content you publish. This somehow also represents whether the content tweeted actually attracts the target audiences.



    - In Twitter, the engagement rate is calculated as the formula shown above. After the engagement rate of each cinema is obtained, the information is shown in a bar chart to compare.
- Top 5 hashtags used
    - For tweets that involve hashtags, the audience can easily search for it or it has a higher chance to be exposed to the audience. Thus, using more hashtags is good for a Twitter account.
    - From the tweets posted by the cinema data collected above, the count of hashtags used was extracted. The name and the count are then analyzed to determine which of the three cinemas used the hashtags efficiently.

Data Collection

Details

- The data collection period starts from 7/3/2022 and ends on 20/3/2022 (14 days).
- The collected data are the followers count each day and the tweets posted by each cinema within the period.
- The tweets data is collected by using the Twitter API on the Google Colab and the library functions such as user_timeline, get_user.

Scripts

- Collect all tweets from the cinema:

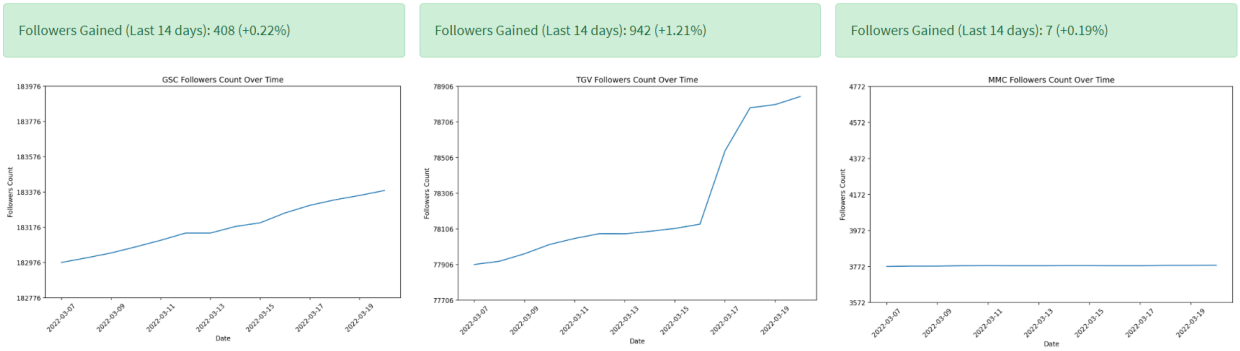
```
tweets = []
count = 100
startDate = datetime.datetime(2022,3,7) #The date contains year, month, day, hour, minute, second, and microsecond.
endDate = datetime.datetime(2022,3,21)

# This section is for you to learn about how the arrays in python function when using Tweepy
def queryTimeline(username,filename):
  result = api.user_timeline(username,count=count)
  for t in result:
    if t.created_at < endDate:# and t.in_reply_to_status_id == None:
      tweets.append(t)

  while(result[-1].created_at >= startDate):#
    # time.sleep(3)
    print('Last tweet collected was created at',result[-1].created_at,'Retrieving more')
    result = api.user_timeline(username,count=count,max_id=result[-1].id-1)
    for t in result:
      if t.created_at >= startDate:# and t.in_reply_to_status_id == None:
        tweets.append(t)

  #json file
  with open(filename,'w') as f:
    for t in tweets:
      f.write(json.dumps(t._json)+'\n')

  files.download(filename)
  print(username+'\'s all tweets downloaded.')
```

- Collect data of daily followers count:

```
def getFollowersNum(screen_name,filename):
  userDetails = api.get_user(screen_name)
  user = {}

  dt_string = datetime.datetime.now().strftime("%Y-%m-%d")

  followers_count = userDetails.followers_count

  user['date'] = dt_string
  user['followers_count'] = followers_count

  dt_string2 = datetime.datetime.now().strftime("%Y%m%d")
  with open(filename+'_audience_growth_rate_'+dt_string2+'.json','w') as f:
    f.write(json.dumps(user)+'\n')

  files.download(filename+'_audience_growth_rate_'+dt_string2+'.json')
```

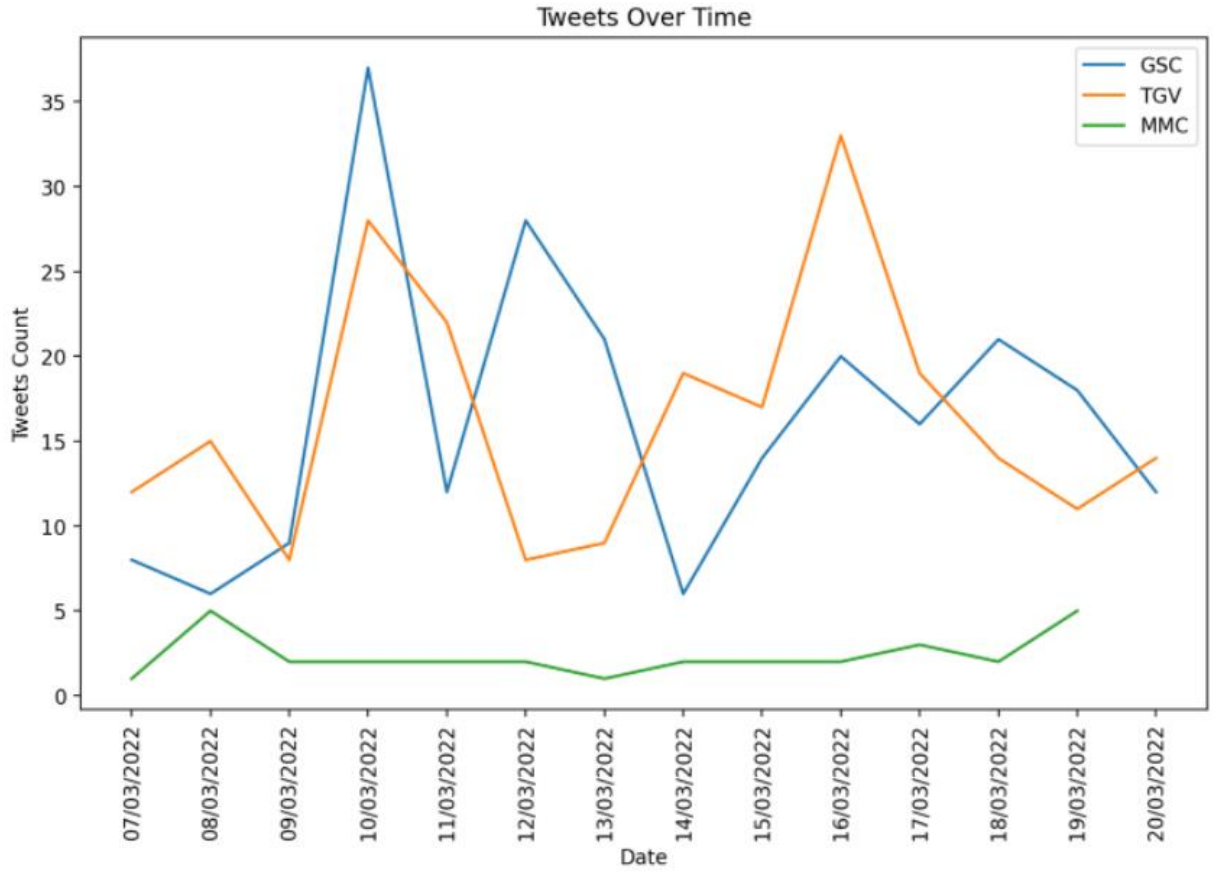
Outcomes

Followers Growth Rate



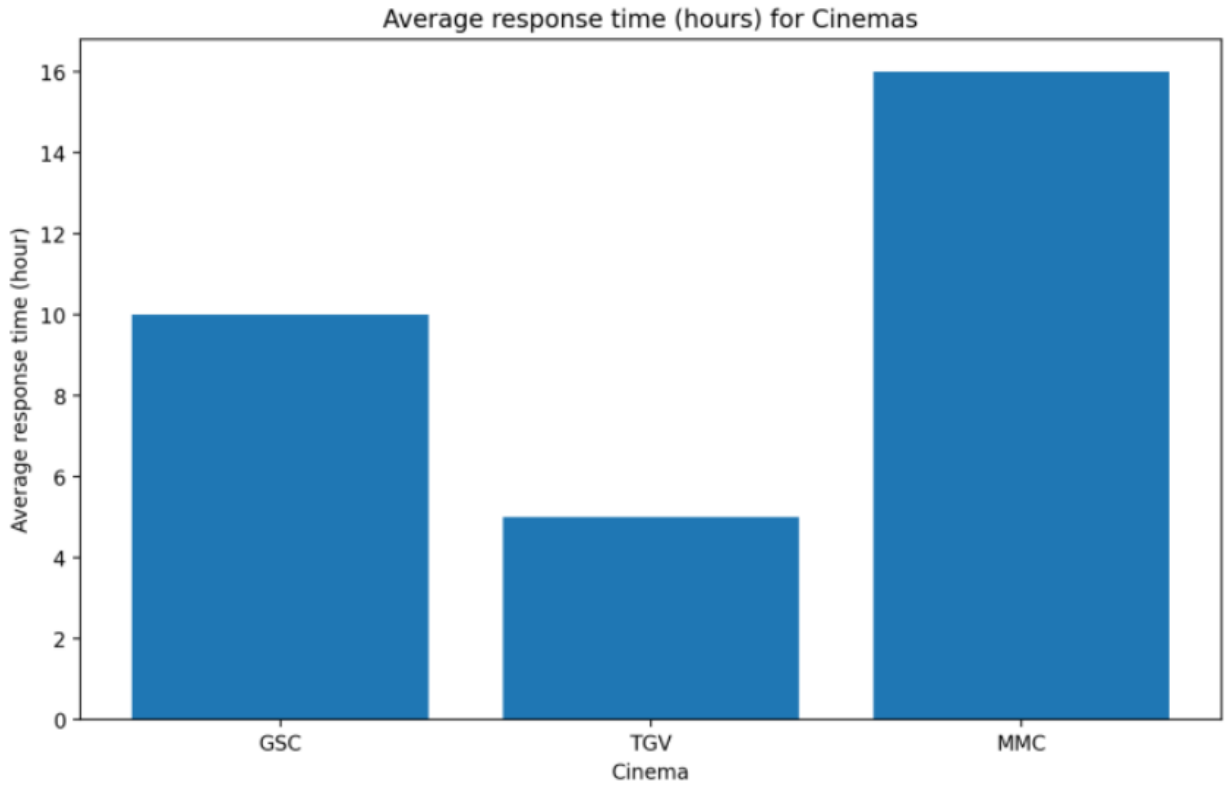
The three line charts above display the followers count of GSC, TGV and MMC for a two weeks period starting from 7/3/2022. It is obvious that both GSC and TGV have an increasing followers growth rate but not MMC as the almost flat line suggested. The followers of GSC increased for 0.22% while for TGV it increased by 1.21%. Lastly for MMC, it only gained 7 more followers in a two weeks period. Overall, TGV has the highest followers growth rate among these three cinemas because it gained almost 1000 followers in two weeks.

Active Rate



This multiple line chart shows the number of tweets that these three cinemas posted within the two weeks period. TGV and GSC are more active compared to MMC as they can tweet at least five times to at most 40 times in one day. While for MMC, it only tweeted 5 times on 8/3/2022 and 19/3/2022. For the other days, it tweeted less than 5 times or did not post any tweets at all.

Responsiveness



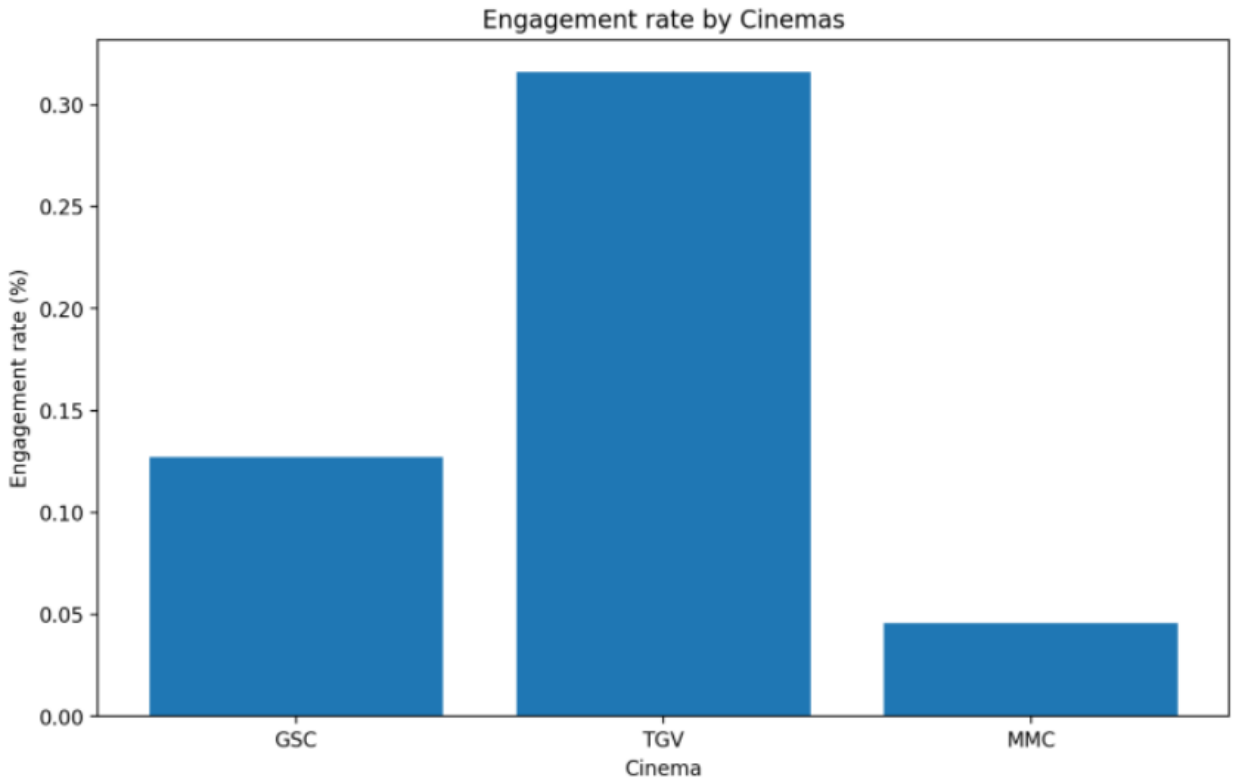
GSC average replied in 10 hours for 169 tweets

TGV average replied in 5 hours for 129 tweets

MMC average replied in 16 hours for 1 tweets

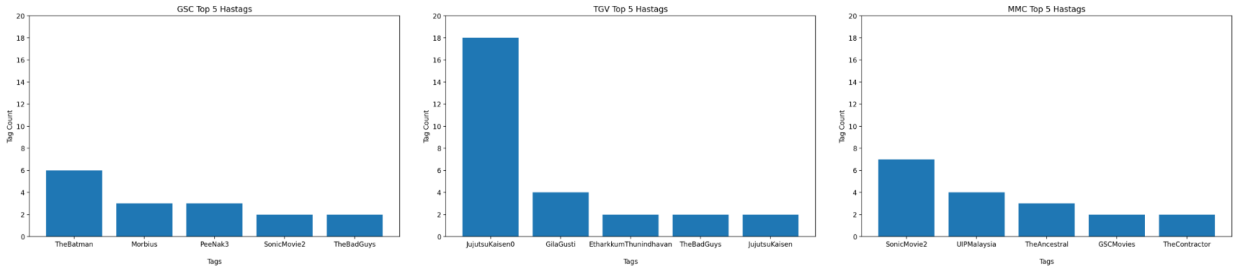
TGV took the least time which is only 5 hours to respond or answer the users when being mentioned. The second fastest responding cinema is GSC which took 10 hours to respond. MMC has the poorest responsiveness towards the questions from the users. It tooks 16 hours to answer only one tweet. Meanwhile for GSC and TGV, they took less than 16 hours to respond to over 100 tweets which is far more responsive than MMC.

Engagement Rate



From the bar chart shown, we can tell that TGV has the highest engagement rate which is 0.3%. GSC has an engagement rate of around 0.12% and MMC has the lowest engagement rate of only 0.04%.

Top 5 Hashtags Used



From the top 5 hashtags used for each of the cinema, we can tell that GSC and TGV used the most hashtags to promote some popular and mainstream movies such as The Batman and Jujutsu Kaisen. TGV used the most hashtags which is 18 hashtags to promote the Jujutsu Kaisen movie. This is a famous anime movie, so by using more hashtags TGV increases the chance of their tweets being exposed, viewed or searched by the users. However, GSC and MMC did not use many hashtags in their tweets. MMC used 7 hashtags for Sonic Movie 2 but it is not a movie that will interest most people.

Result

Based on our opinions, TGV used the social media platform (especially on Twitter) the best. It seems to have the highest followers growth rate (1.21%) and it is quite an active account too. It tweeted from 5 to 32 times per day within the two weeks period. This indicated that TGV is active in using Twitter to post about the movies they will be screening. The followers of TGV can get to know the latest news or details of the movies that they are interested in.

Besides, TGV also responded in a short period of time after it was mentioned by their followers. It took around 5 hours to reply for TGV which is the fastest among the three cinemas with quite a number of tweets.

The highest engagement rate which is 0.3% is achieved by TGV too. According to (Mee, 2021), the author has defined the following engagement rate ranges for Twitter after their analysis of millions of data points:

- An engagement rate between 0% and 0.02% is considered to be low. An influencer with a low engagement rate on Twitter could expect between 0 - 0.2 reactions for every 1000 followers.
- Engagement rates between 0.02% and 0.09% are considered to be good. An influencer with a good engagement rate on Twitter could expect between 0.2 - 0.9 reactions for every 1000 followers.
- An engagement rate between 0.09% and 0.33% is considered to be high, where an influencer would expect 0.9 - 3.3 reactions for every 1000 followers on Twitter.
- Finally, an engagement rate between 0.33% and 1% is considered to be very high, with expected reactions to be between 3.3 - 10 for every 1000 Twitter followers.

By using these engagement rate ranges, it is observed that TGV considered having a high engagement rate as their engagement rate is falling within 0.09% to 0.33%.

Lastly, TGV also utilized the hashtags in their tweets. They used the most hashtags on the mainstream and popular movies because they know this can increase the chance of their tweets being searched and thus increase the exposure.

Part 2

Data Collection

Details

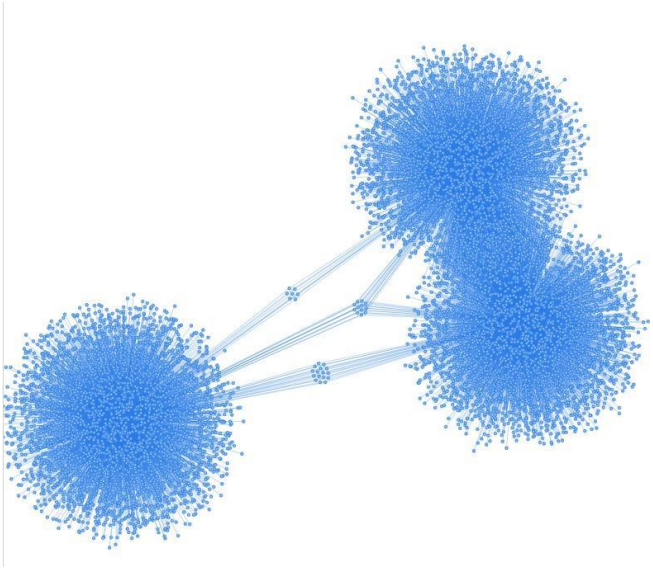
- Data collection date is on 20/3/2022.
- The collected data is the followers details of each cinema. Note that only 3500 follower details are collected because one of the cinemas has a lesser number of followers.
- The data is collected by using the Tweepy library functions with name api.followers.

Scripts

- Collect data about the followers:

```
usernames = [username_gsc,username_tgv,username_mmc]

for username in usernames:
  with open(username+'_followers.json','w') as f:
    for page in tweepy.Cursor(api.followers, screen_name=username,count=100).pages(35):
      for user in page:
        f.write(json.dumps(user._json)+'\n')

  files.download(username+'_followers.json')
```

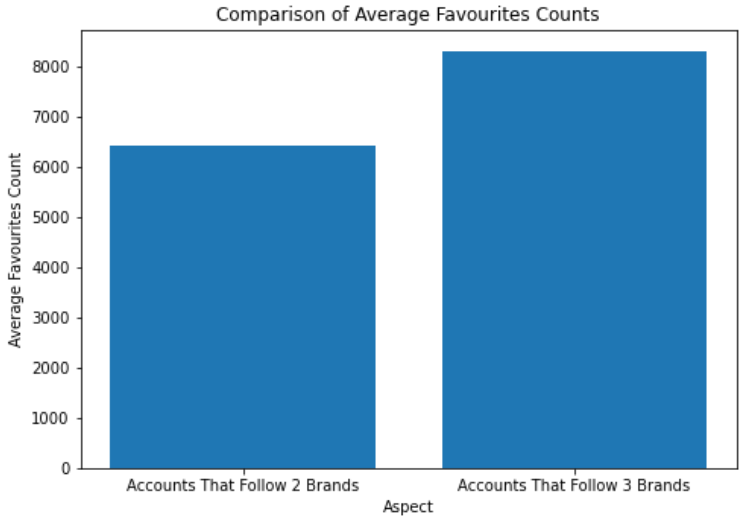
Network



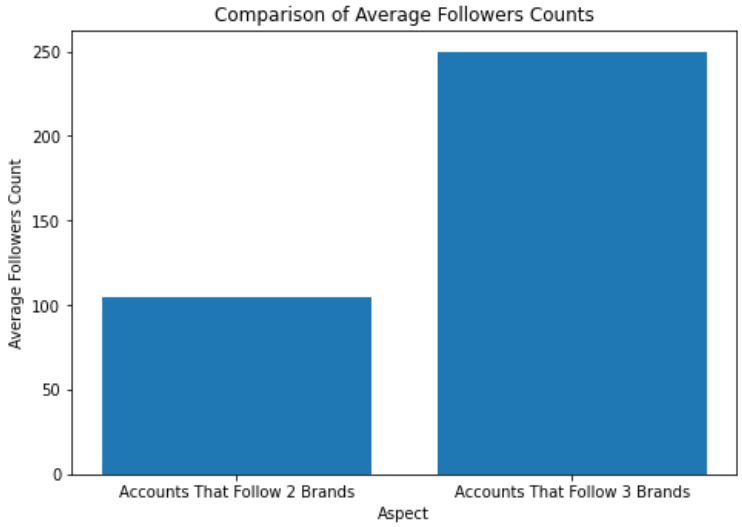
The figure above shows the network generated for the part 2 assignment. Each of the blue points represents one Twitter account. The three large groups of points are the accounts that followed to the three cinemas: GSC, TGV and MMC. As observed, there are some linkages or edges between the three large groups. These are the accounts that followed two or three cinemas.

Brand Analysis and Comparisons

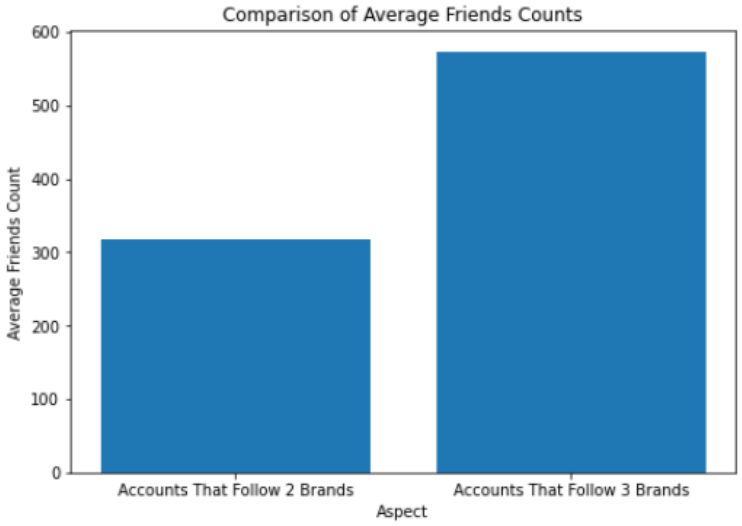
Accounts That Follow More Than One Brand



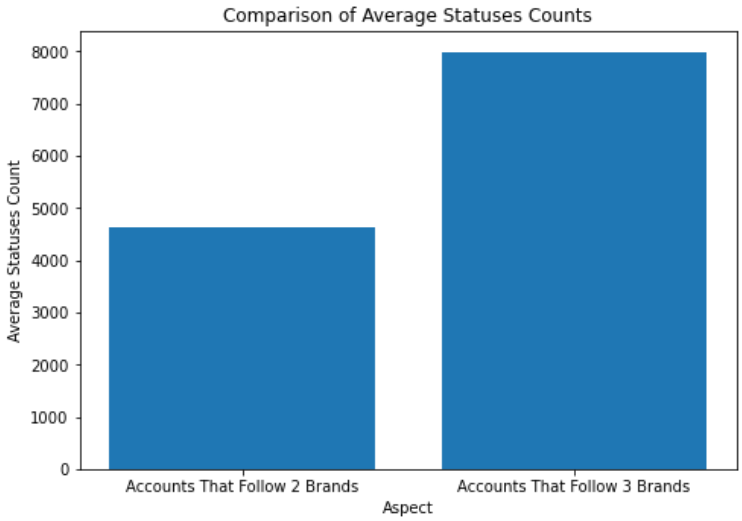
According to the bar graph, we can see that accounts that follow 3 brands have a higher average favorite count than accounts that follow 2 brands. It can be said that accounts with higher favorites count are more active on social media and thus have more interaction with others which lead to high favorite count of the account. They tend to follow more accounts to have more exposure on the latest update on social media. It can also be said that accounts that follow more brands, which is “accounts that follow 3 brands” in this case have a higher chance to be public figure accounts. Public figure accounts, for example, influencers usually hold a huge pool of followers therefore they will have a higher average favorite count.



From the bar graph above, we can observe that the average followers count is higher for accounts that follow 3 brands as compared to the accounts that follow 2 brands. From this observation, we can assume that accounts that follow 3 brands are spending more time on social media, therefore gathering a bigger number of average follower count. Accounts that follow 3 brands are also more likely to follow more brands of other sectors on social media thus winding them up with more average followers counts.

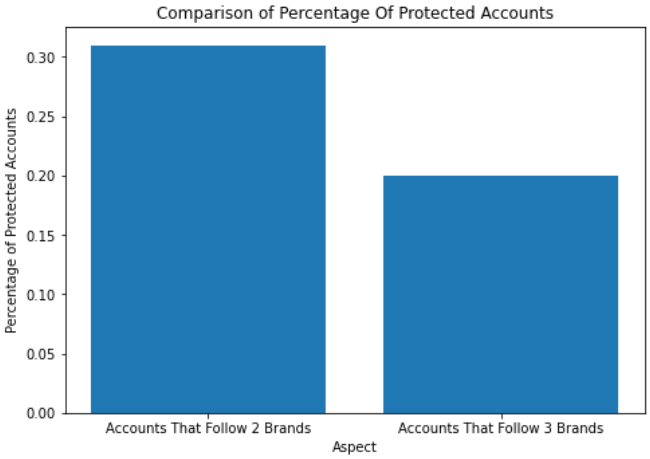


For the bar graph above, it is observed that accounts that follow 3 brands have a higher number of average friends count as compared to accounts that follow 2 brands. As mentioned before, accounts that follow more brands tend to be more active on social media. With this reasoning, it is safe to say that the accounts that follow 2 brands are less likely to be active accounts and therefore resulting in a lower average friends count.



The bar graph above shows that accounts that follow 3 brands have a higher average statuses count than accounts that follow 2 brands. We can assume that accounts that follow 3 brands are more likely to be more active as compared to accounts that follow 2 brands and are more likely to express their opinions and update their status. Also, accounts that follow 3 brands have a higher chance to be corporate brand accounts or public figure accounts that constantly post announcements and advertisements on social media platforms, therefore explaining the higher average count of statuses.

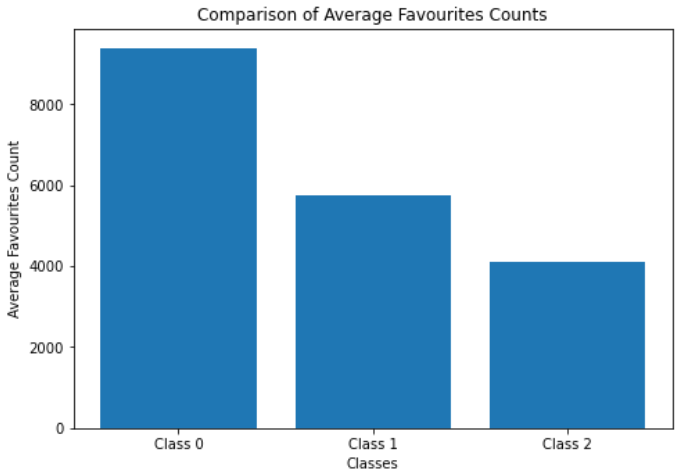
From the bar graph above, we can see that accounts that follow 2 brands have a higher percentage of protected accounts as compared to accounts that follow 3 brands. From the graph we can assume that accounts that follow 3 brands are more likely to be corporate with personal accounts, corporate brand accounts, public figure accounts, etc. Meanwhile, accounts that follow 2 brands are more likely to be pure personal accounts whereby most accounts are protected. Most



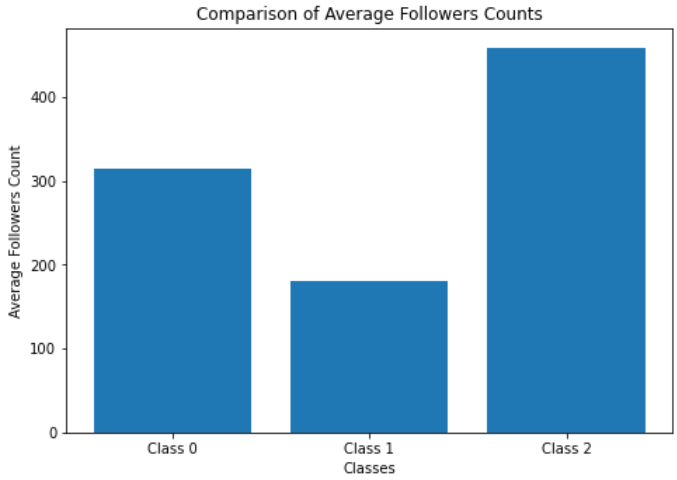
pure personal accounts are more likely to be protected accounts to secure themselves from threats or unwanted attention on social media.

Community Similarities

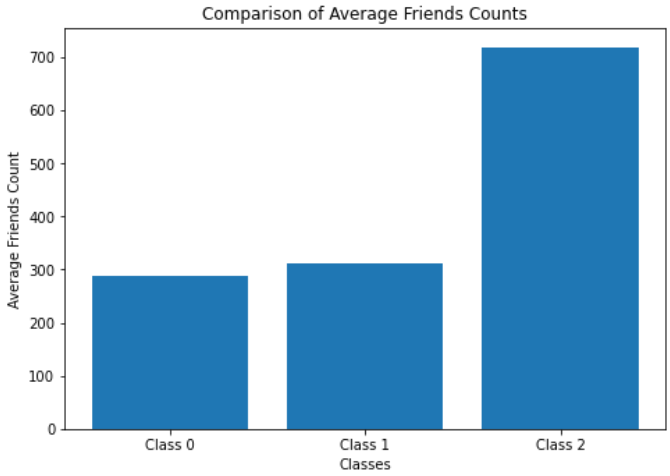
There are a total of three communities sorted out based on the network generated previously. Class 0 has 3106 nodes, Class 1 has 3101 and Class 2 has 3479 nodes. TGV is in Class 0, GSC is in Class 1 while MMC is in Class 2.



From the bar graph above, we can observe that Class 0, Class 1, and Class 2 have at least approximately 4000 thousand average favorites count. With Class 0 having the highest number of approximately 9000, and Class 1 having a number of approximately 5800 and Class 2 having the lowest number of approximately 4000 of average favorites counts, it can be assumed that the average favorite counts does not directly relate to the amount of followers on the brands' social media account. We can assume that since the social media account of GSC is created at an earlier time than TGV and they have gathered a bigger follower count. But at the same time this might result in a bigger amount of inactive followers as social media accounts created at an earlier period may be abandoned and turned inactive. The same goes for MMCineplex which holds the lowest number of both following and followers. The account is inactive and has a low social activity rate therefore, MMCineplex followers tend to be inactive as well. So, it is safe to conclude that the three classes can be said as the classes that have high, medium and low followers count in this case. For example, TGV falls in Class 0 while MMC falls in Class 2.

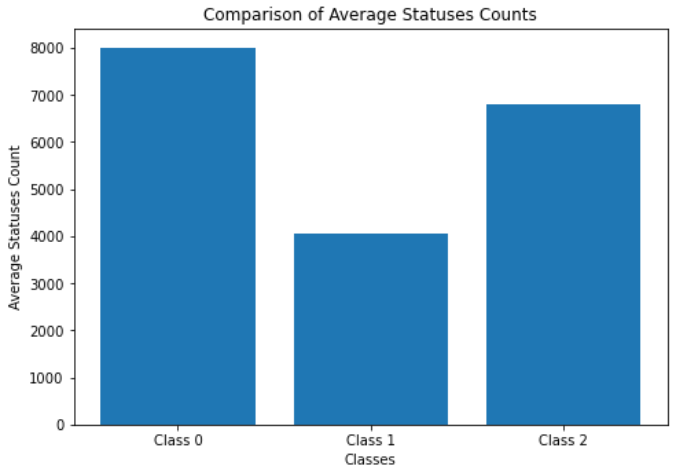


We can observe that Class 2 contains the highest number of average followers with the number of approximately 450, followed by Class 0 with a number of approximately 300, followed by Class 1 with the lowest number of approximately 180. Following our previous findings, MMCineplex is categorized in Class 2 but at the same time MMCineplex contains the lowest overall follower and following numbers and the lowest activity rate. The finding regarding comparison of average followers count is contradicting with the overall analysis of MMCineplex. To conclude, the average followers count is not the similarity within the community.

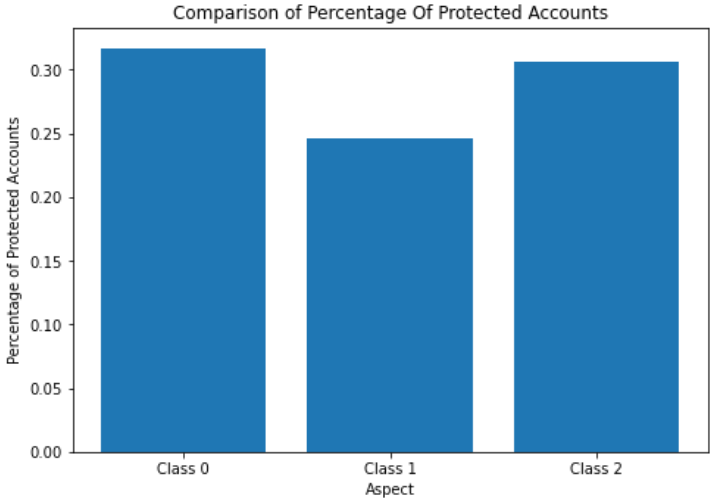


For comparison of average friends counts, Class 2 is significantly higher than both Class 1 and Class 0 with the average friends counts of approximately 700, 300, and 290 respectively. Having MMCineplex categorized in Class 2, it is inaccurate to relate the comparisons to our previous findings. TGV holds the largest number of friends on their social media account followed by GSC and MMCineplex respectively. Since, MMCineplex has the lowest activity rate and the

lowest friends count. To summarize, the average friends count is not the similarity within the community as well.



Moving on to the comparison of average statuses counts. Class 0 which contains TGV has a number of approximately 8000, Class 1 which contains GSC has a number of approximately 4000, and Class 2 which contains MMCineplax has a number of approximately 6800 average statuses counts. Among all 3 brands, TGV has the highest activity rate which satisfies the comparison. However, Class 1 which contains GSC should have a higher number than Class 2 which contains MMCineplex. Therefore, average statuses count is not the similarity within the community as well.



By observing the bar chart above, the comparison of percentage of protected accounts indicated that Class 0 which contains TGV has the highest percentage of approximately 0.32, followed by Class 2 which contains GSC has the percentage of approximately 0.30, and Class 1

which contains MMCineplex has the lowest percentage of approximately 0.24. We can observe that most accounts listed in Class 0 are personal accounts based on our previous findings. But unfortunately, all 3 brands are unprotected accounts thus this comparison is not related to the similarity within the community.

Reference

Mee, G. (2021, May 10). *What is a good engagement rate on Twitter?* scrunch. Retrieved April 12, 2022, from https://scrunch.com/blog/what-is-a-good-engagement-rate-on-twitter