BLACKBEAR ANALYSIS

4008ICT Big Data Analytics

Damian Garcia
S5054158

Contents

1 Case Study Setting 2
1.1 Artist Information 2
1.2 Purpose of social media analytics 4
2 Data Selection & Exploration 5
2.1 Data Retrieval 5
2.2 Top 5 Users 7
2.3 Top 10 Terms 13
2.4 Retweets/accounts 15
2.5 Spotify Related Artists 16
..... 17
2.6 YouTube views/likes 19
2.7 Term-document Matrix & top 10 Terms 22
2.8 Centrality analysis 24
2.9 Community analysis 27
3.1 K-means 34
3.2 Decision Tree 35
3.3 Sentiment analysis 36
3.4 LDA topic modelling 38
3.8 Social media analytics findings 39
3.9 Suggested actions 39
3.10 Possible refinement 40

1 Case Study Setting

1.1 Artist Information

Matt Tyler Musto, more commonly known as “Blackbear” is an American musician, songwriter and record producer who was born in November in 1990. The Blackbear Spotify account falls under the pop: pop-rap genre with currently over 294 thousand followers.

	name	id	popularity	followers	genres
1	blackbear	2cFrymmkijnjDg9SS92EPM	86	2944018	pop;pop rap

Blackbear currently has 5 individual albums on Spotify; Anonymous, Cybersex, Digital Druglord, Help and Roses as shown in the data retrieved. His first album on Spotify was ‘Deadroses,’ which was released on the 14th of February in 2015 which contained 10 songs. All together Blackbear has published 108 songs on Spotify.

	id	name	album_type	available_markets
1	5DVpGdikHfxFWlELotC3FL	ANONYMOUS	album	
2	3Hft6FQNGdRRQ9dCqyaTYK	ANONYMOUS	album	
3	1nXBX3wP3DQgwpqvVtJl8H	cybersex	album	
4	3uqSZZwlzYhlRMjMHrplBx	cybersex	album	
5	30QoUTsSdpQuZunPNFsLnm	digital druglord	album	
6	4PcOLASOD0VikPG4yB3bxg	Help	album	
7	1TkwzY3l4LqAfrQwBAx45Q	Deadroses	album	

Looking more in depth into the album “Digital Druglord” here displays the durations of each song in milliseconds as well as the track number, for this album we have 10 songs. As we can see from the data his songs range from 154,452ms to 220,588ms averaging at around 193,700ms which is roughly 3 minutes and 10 seconds. The longest song being “moodz” and the shortest being “wish u the

best”.

	id	name	duration_ms	track_number
1	7mmf56XtAneYnuomOuvySJ	hell is where i dreamt of u and woke up alone	168428	1
2	7cZbVLrtLIFDWh5YpRBS9M	moodz (feat. 24hrs)	220588	2
3	1nMYtxDrONcoGnKRvxTwPv	i miss the old u	211127	3
4	3Q3myFA7q4Op95DOpHplaY	do re mi	212027	4
5	23xup5SPK9QGT5Rs5qUzOm	wish u the best	154452	5
6	2LRj8su5PjrGkbp3TZZG8D	juicy sweatsuits (feat. Juicy J)	183943	6
7	7is4HpXFrjZ0ShTk2yv16H	double	180495	7
8	7pF0xW7MH9Tucfpb1Oys0R	if i could i would feel nothing	184500	8
9	13JyykwyYQ3T5QxxL34ukQ	chateau	212926	9
10	1V8c3uV6p7uTsfQof5oRXg	make daddy proud	208235	10

Going even further when looking into both songs individually and their features. Here where we can compare their energy, key, loudness and many other attributes.

Blackbear - moodz:

danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
0.753	0.486	8	-8.864	1	0.144	0.197	2.02e-05	0.115	0.15	101.921	220588

Blackbear – wish u the best:

danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
0.802	0.495	2	-6.154	0	0.188	0.0232	0	0.0933	0.47	114.991	154452

As shown, there are quite a few noticeable differences such as the valence and tempo of the song. The song with less valance which is ‘moodz’, is quite a higher duration in time while ‘wish u the best’ is a shorter song but with a higher valance. Comparing these two pieces of data is comparing the fastest and slowest song of the ‘Digital Druglord’ album and their differences, this could help us achieve trends and decide in the future on what type of song to put out next. Now if we compare these two to the most popular song on the album “do re mi” you can see that there is a big gap in valence between the last two songs and this one. The main differences are valance, speechiness and time. Speechiness detects the presence of spoken words on a track. The data that we are receiving from these three songs can easily be analysed and show how the song ‘do rei mi’ was the most popular. Along with that it also gives a rough overview on what his listeners were looking for in the album and what they liked the most and least, whether it’s time, loudness, valence or tempo.

danceability	energy	key	loudness	mode	speechiness	acousticness	instrumentalness	liveness	valence	tempo	duration_ms
0.745	0.593	8	-6.35	1	0.0526	0.00522	5.25e-06	0.123	0.17	111.002	212027

1.2 Purpose of social media analytics

The purpose of this document is to use social media analytics to improve popularity and give the right listeners what they want from who they want. Improving popularity can be done in many ways, one of the most important ways is to give the listeners what they want. While that might seem broad and difficult this is where data analysis come in. Data analysis shows the most important information relating to the songs and listeners, this can all be further investigated and ultimately can guide an artist to gaining popularity. This analysis contains data extraction using R studio along with the help of Spotify and Twitter API to get the data we need. With this data we can compare songs features and find out what works and what doesn't along with seeing data on similar artists and comparing them to the top 100 Spotify playlist. Twitter data is used to visualise who is interacting with artists the most and how much reach those followers get.

2 Data Selection & Exploration

2.1 Data Retrieval

Select social media platforms (twitter, spotify) and retrieve data. Make sure to choose keywords for data retrieval that are most relevant to your artist\band. However, try not to bee too narrow . As a rough guide you should at retrieve at least 1000 data items (eg. Tweets from twitter), explain what you have done.

When retrieving data for this analysis through Spotify first we extracted a list of all the blackbears and managed to locate the correct blackbear using an artist ID. This gives us information regarding the artist name, id, popularity, followers and genres.

Blackbear's Info

	name	id	popularity	followers	genres
1	blackbear	2cFrymmkijnjDg9SS92EPM	86	2944018	pop;pop rap

Using the same blackbear artist ID there was a query written to extract all the data of all the albums and their details.

Blackbear's Albums

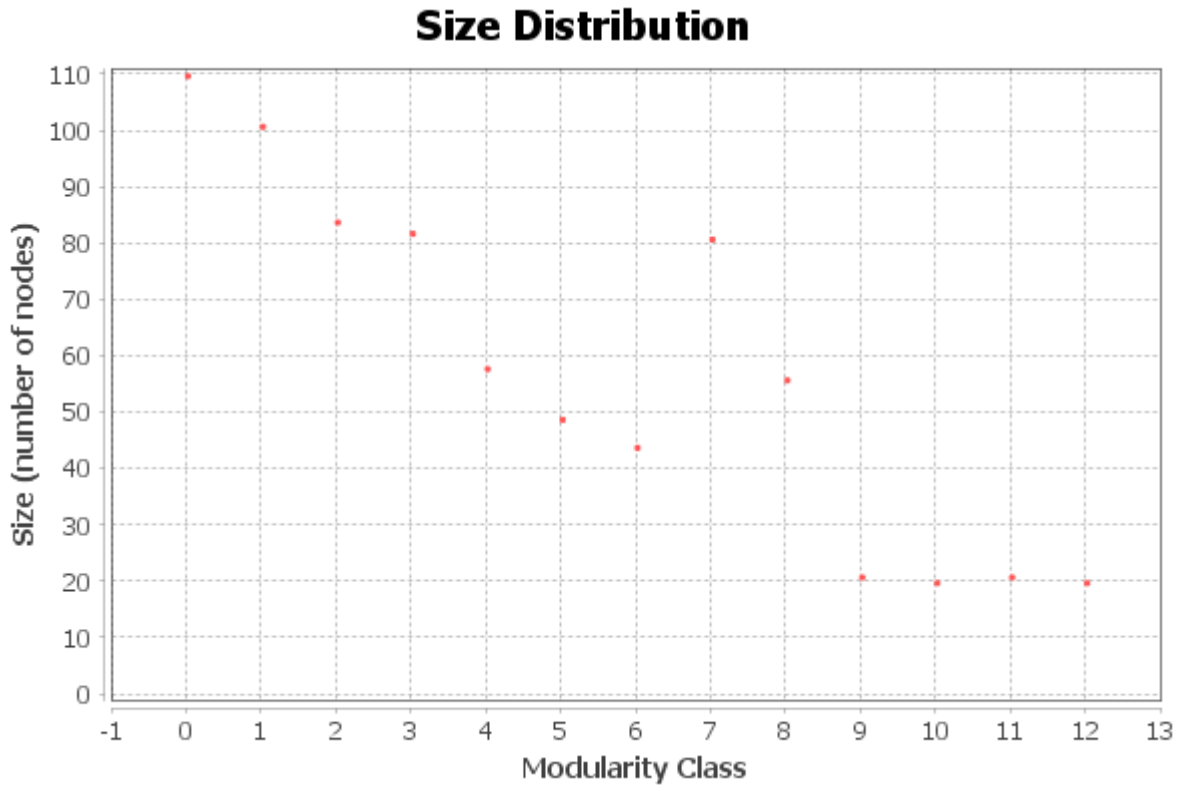
	id	name	album_type	available_markets
1	5DVpGdikHfxFWIELotC3FL	ANONYMOUS	album	
2	3Hft6FQNGdRRQ9dCqyaTYK	ANONYMOUS	album	
3	1nXBX3wP3DQgwpqvVtJI8H	cybersex	album	
4	3uqSZZwIzYhIRMjMHrpIBx	cybersex	album	
5	30QoUTsSdpQuZunPNFsLnm	digital druglord	album	
6	4PcOLASOD0VikPG4yB3bxg	Help	album	
7	1TkwzY3I4LqAfrQwBAx45Q	Deadroses	album	

The album ‘Digital Druglord’ was selected for further analysis and a query was written to show all the album songs name, duration and track number.

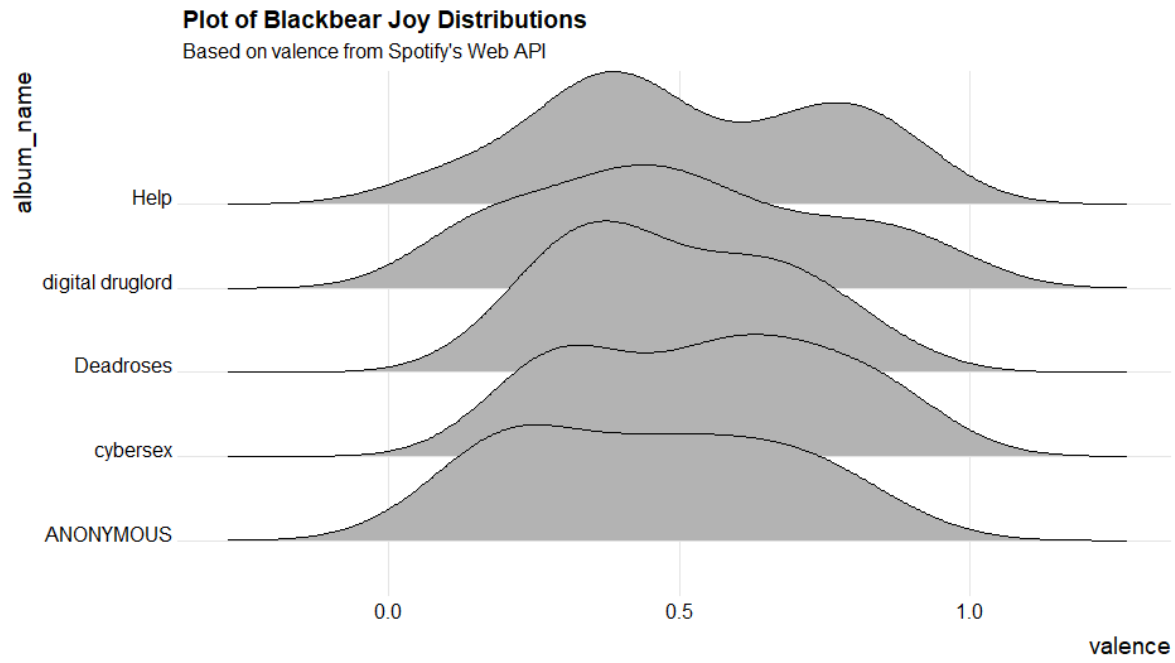
Blackbear’s Digital Druglord Album

	id	name	duration_ms	track_number
1	7mmf56XtAneYnuomOuvySJ	hell is where i dreamt of u and woke up alone	168428	1
2	7cZbVLrtLIFDWh5YpRBS9M	moodz (feat. 24hrs)	220588	2
3	1nMYtxDrONcoGnKRvxTwPv	i miss the old u	211127	3
4	3Q3myFA7q4Op95DOpHplaY	do re mi	212027	4
5	23xup5SPK9QGT5Rs5qUzOm	wish u the best	154452	5
6	2LRj8su5PjrGkbp3TZZG8D	juicy sweatsuits (feat. Juicy J)	183943	6
7	7is4HpXFrjZ0ShTk2yv16H	double	180495	7
8	7pF0xW7MH9Tucfpb1Oys0R	if i could i would feel nothing	184500	8
9	13JyykwyYQ3T5QxxL34ukQ	chateau	212926	9
10	1V8c3uV6p7uTsfQof5oRXg	make daddy proud	208235	10

Using the extracted graph created from blackbears albums data along with data from the top 100 playlist on spotify. This following graph shows the size distributions between the modularity classes. As shown you can see that modularity class 1 is the highest of nodes. Blackbear falls into modularity class 8.



This is a visual representation made by ggplot showing the differences in valence according to what album Blackbear has. As seen Help contains the most valence.

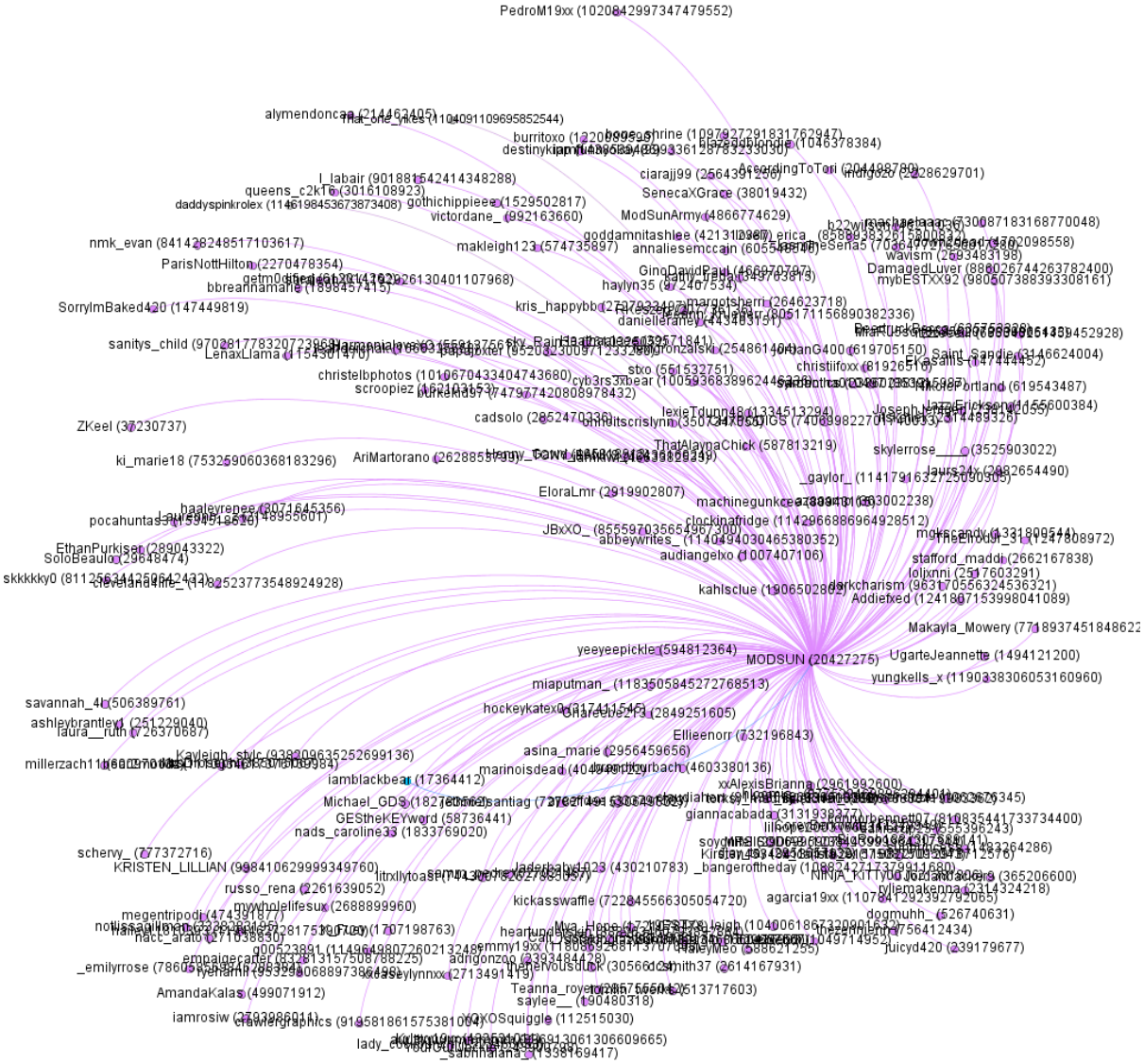


2.2 Top 5 Users

Below is the code and output that was extracted using R studio. It shows the 5 most influential users for Blackbear, one being himself and 4 being other twitter users. Below you can see the graphs where you can see that they are connected to some of the same nodes and share similarities through that.

```
> #finding top 5 more important users according to blackbear search
> pageRank_blackbear_actor <- sort(page.rank(g_twitter_actor_graph)$vector, decreasing=TRUE)
> head(pageRank_blackbear_actor, n=5)
      MODSUN  iamblackbear   mariela_g_g
  0.04574877    0.04191378    0.03437753
Siege_vibes belovedsheis_
  0.03282735    0.01500027
```

MODSUN:


PedroM19xx (1020842997347479552)
MODSUN (20427275)
yeeyeepickle (594812364)
UgarteJeannette (1494121200)
yungkells_x (1190338306053160960)
Makayla_Mowery (771893745184862)
Ellieenorr (732196843)
asina_marie (2956459656)
iamblackbear (17364412)
giannacabada (3131938377)
xxAlexisBrianna (2961992600)
KRISTEN_LILLIAN (998410629999349760)
russo_rena (2261639052)
mywholelifesux (2688899960)
megentripodi (474391877)
dogmuhh_ (526740631)
Juicyd420 (239179677)
AmandaKalas (499071912)
Teanna_royer (2867556042)
saylee__ (1904803l8)
QXOSquiggle (112515030)
iamrosiw (279298041l)
ZKeel (37230737)
ki_marie18 (753259060368183296)
EloraLmr (2919902807)
ThatAlaynaChick (587813219)
skylerrose____ (3525903022)
gaylor (1141791632725090305)
haaleyrenee (3071645356)
JBxXO_ (855597035654967300)
abbeywrites_ (1140494030465380352)
audiangelxo (1007407106)
stafford_maddi (2662167838)
lolixnni (2517603291)
Addiefxed (1241807153998041089)
savannah_4l (506389761)
ashleybrantley1 (251229040)
laura__ruth (726370687)
hockeykatex0 (317411545)
marinoisdead (4049[illegible]6380136)
nads_caroline33 (1833769020)
GEStheKEYword (58736441)
scherv_ (77372716)
ParisNottHilton (2270478354)
nmk_evan (841428248517103617)
makleigh123 (574735897)
annaliesemccain (605548848)
SenecaXGrace (38019432)
ciarajj99 (2564391260)
ModSunArmy (4866774629)
victordane_ (992163660)
gothichippieeex (1529502817)
daddyspinkrolex (1146198453673873408)
queens_c2kT6 (3016108923)
l_labair (901881542414348288)
alymendoncaa (214463405)
That_one_vines (1104091109695852544)
SorrylmBaked420 (147449819)
bbreannamarie (1898457415)
LenaxLlama (154301470)
sanitys_child (970281778320723968)
christelbphotos (1010670433404743680)
scroopiez (162103153)
cadsolo (2852470336)
machaelaaac (730087183168770048)
wavism (2693483198)
DamagedLuver (886026744263782400)
mybESTXX92 (980507388393308161)
haylyn35 (972407534)
GinqDavidPaul (466970797)
kris_happybo (2737933407)
kahlsclue (1906502803)
Kirsten_Lillian (1088242717379911680)
NINJA_KITTY00 (1078412923927920[illegible])
agarcia19xx (1107841292392792065)
tylienakenna (2314324218)
juicyd420 (239179677)
Saint_Sandie (3146624004)
EVAsalls (1474144452)
Rhicher_Portland (619543487)
Addiefxed (1241807153998041089)
Makayla_Mowery (771893745184862)
UgarteJeannette (1494121200)
yungkells_x (1190338306053160960)
Ellieenorr (732196843)
asina_marie (2956459656)
iamblackbear (17364412)
giannacabada (3131938377)
xxAlexisBrianna (2961992600)
KRISTEN_LILLIAN (998410629999349760)
russo_rena (2261639052)
mywholelifesux (2688899960)
megentripodi (474391877)
dogmuhh_ (526740631)
Juicyd420 (239179677)
AmandaKalas (499071912)
Teanna_royer (2867556042)
saylee__ (1904803l8)
QXOSquiggle (112515030)
iamrosiw (279298041l)
ZKeel (37230737)
ki_marie18 (753259060368183296)
EloraLmr (2919902807)
ThatAlaynaChick (587813219)
skylerrose____ (3525903022)
gaylor (1141791632725090305)
haaleyrenee (3071645356)
JBxXO_ (855597035654967300)
abbeywrites_ (1140494030465380352)
audiangelxo (1007407106)
stafford_maddi (2662167838)
lolixnni (2517603291)
Addiefxed (1241807153998041089)
savannah_4l (506389761)
ashleybrantley1 (251229040)
laura__ruth (726370687)
hockeykatex0 (317411545)
marinoisdead (4049[illegible]6380136)
nads_caroline33 (1833769020)
GEStheKEYword (58736441)
scherv_ (77372716)
ParisNottHilton (2270478354)
nmk_evan (841428248517103617)
makleigh123 (574735897)
annaliesemccain (605548848)
SenecaXGrace (38019432)
ciarajj99 (2564391260)
ModSunArmy (4866774629)
victordane_ (992163660)
gothichippieeex (1529502817)
daddyspinkrolex (1146198453673873408)
queens_c2kT6 (3016108923)
l_labair (901881542414348288)
alymendoncaa (214463405)
That_one_vines (1104091109695852544)
SorrylmBaked420 (147449819)
bbreannamarie (1898457415)
LenaxLlama (154301470)
sanitys_child (970281778320723968)
christelbphotos (1010670433404743680)
scroopiez (162103153)
cadsolo (2852470336)
machaelaaac (730087183168770048)
wavism (2693483198)
DamagedLuver (886026744263782400)
mybESTXX92 (980507388393308161)
haylyn35 (972407534)
GinqDavidPaul (466970797)
kris_happybo (2737933407)
kahlsclue (1906502803)
Kirsten_Lillian (1088242717379911680)
NINJA_KITTY00 (1078412923927920[illegible])
agarcia19xx (1107841292392792065)
tylienakenna (2314324218)
juicyd420 (239179677)
Saint_Sandie (3146624004)
EVAsalls (1474144452)
Rhicher_Portland (619543487)

Blackbear:


JenDegtjarewsky (253879577)
alanfanciful (10371293394296
SophieLGilbert (15357521)
mariela_g_g (
battingpractice (118358
grizzlygirl87 (748207868)
save1me1jebus (367331144)
GoofballAnimals (9895784861920
ericdaryl (94418264)
thorbites (81189754)
EileenRouhani (3876384616)
EckertLeckert (465257387)
lheartnoise (15273924)
Nic_Goodridge (23517606)
tcdixon3 (2319
Waaskea (569557355)
Brogan_Holcombe (2365646963)

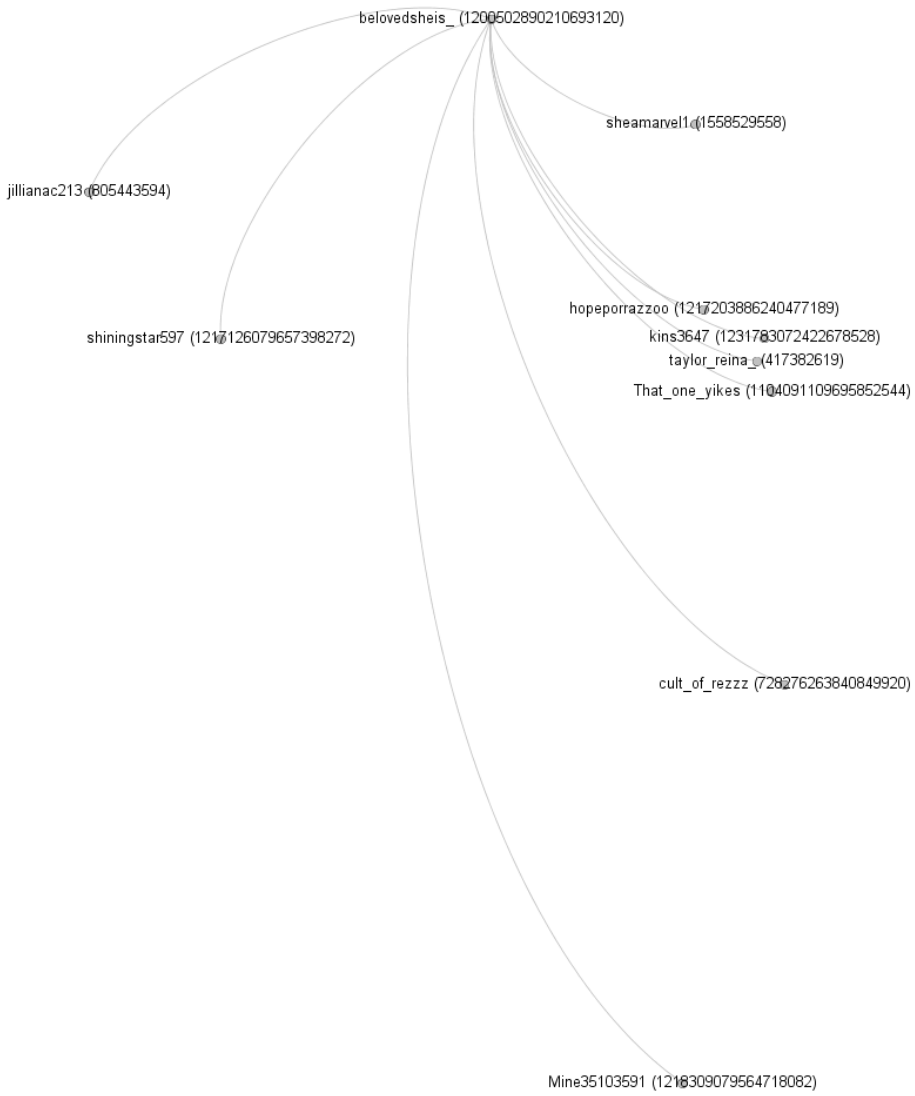
Mariela:


50_Cynt (385210931)
njooniechild (1224372275455000576)
Tanya25019476 (1086880823289708544)
BCR1MusicBot (2164819154)
hot21radio (55649769)
greedypjm (1072578687554473986)
french3xit (1218428059130482689)
MODSUN (20427275)
iamblackbeat (17364412)
incog_neat_HO3 (956035367785893891)
messwasmade (3040835287)
abbyvm17 (840945426)
lewis_hullyXx (616159865)
curlysluke (1122136560928002049)

Siege:


LiquidLo (377754260)
Coronel2_ (1255052347)
lizzzeeth_ (3192654991)
kennacav (3190129886)
hannahisab (1023449948656824320)
jewelsgarciaa (2546598544)
james_austin32 (928616143702564864)
Siege_Vibes (16192998)
jaethompsonn (319902029)
cap10sav (783790442
aali (1090500199116460035)
ThatGuyMike67 (2748927979)
KandiPandiXO (4833496503)
Avartaka (576837058)
adrixnbxzzel (1031865601)
LexinFlexin (237133593)
KlondikeVibes (1549311804)

Beloved:


belovedsheis_ (1200502890210693120)
sheamarvel1 (1558529558)
jillianac213 (805443594)
hopeporrazzoo (1217203886240477189)
shiningstar597 (1217126079657398272)
kins3647 (1231783072422678528)
taylor_reina_ (417382619)
That_one_yikes (1104091109695852544)
cult_of_rezzz (728276263840849920)
Mine35103591 (1218309079564718082)

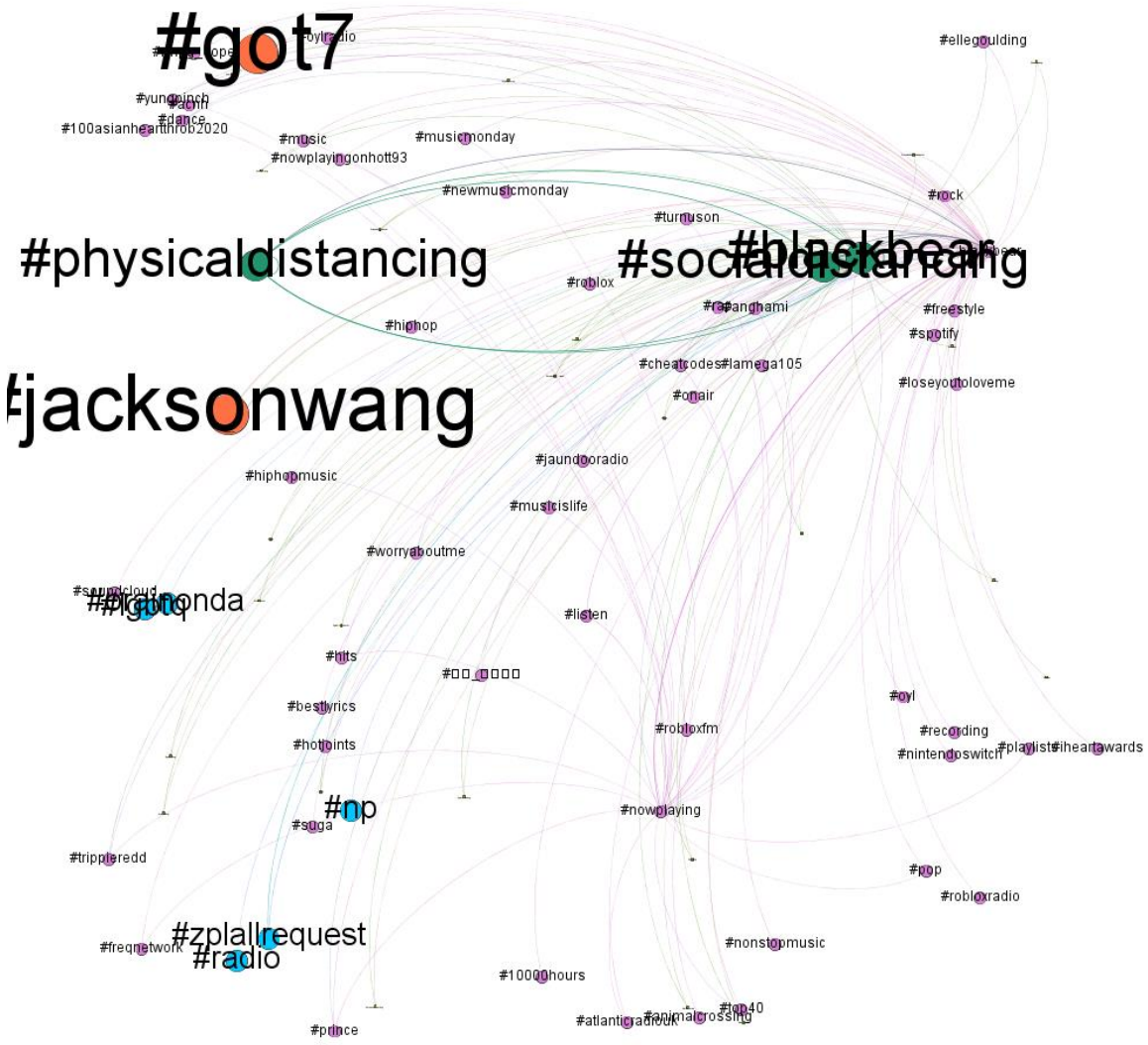
2.3 Top 10 Terms

List the top 10 most important terms that appear together with your keywords related to your artist/band. Explain the results

Bellow is a node representation of the most influential terms related to the hashtag #blakbear. The top 10 most important terms that appear together with blackbear that are music/artist related is shown below in order.

- #nowplayinghott93
- #trippieredd
- #prince
- #rap
- #yungpinch
- #newmusicmonday
- #100asianheartthrob2020
- #oylradio
- #soundcloud
- #ellegoulding

Using these music/artist related hashtags can be very beneficial when gaining popularity. It is always good to look at the trends that relate to an artist to find better ways of gaining popularity through different networks such as getting in contact with some of the online radios and pages that promote you already and going from there. This data creates so many opportunities that you might not have known were there. Through the other artists you can find what type of music they make to get a better understanding of what your audience or a similar audience wants from songs and artists as well as where they go for it.


#got7
#ellegoulding
#yungpinch
#dance
#100asianheartthrob2020
#music
#musicmonday
#nowplayingonhott93
#newmusicmonday
#rock
#turnuson
#physicaldistancing
#socialdistancing
#blackbear
#roblox
#hiphop
#freestyle
#spotify
#cheatcodes#lamega105
#loseyourtoloveme
#onair
#jacksonwang
#jaundooradio
#hiphopmusic
#musicislife
#worryaboutme
#listen
#hits
#bestlyrics
#hotpoints
#np
#suga
#nowplaying
#oyl
#robloxfm
#recording
#nintendoswitch
#playlist#iheartawards
#tripplereddd
#pop
#robloxradio
#zplallrequest
#radio
#freqnetwork
#nonstopmusic
#10000hours
#top40
#animalcrossing
#prince

2.4 Retweets/accounts

For your twitter dataset, calculate how many of your retrieved tweets are retweets. Alternatively, if you filtered out retweets in your query, calculate how many unique user accounts there are in your dataset. What do the results tell you?

Bellow is the data extracted from the last 1800 tweets that contained the word blackbear that were not retweets. There were 135 unique users, this shows just how many people are generating original tweets as well as how much reach they are getting with the number of retweets.

d_at	screen_name	text	source	display_text_width	reply_to_status_id	reply_to_us
3-22 23:19:57	wzpl	The #ZplAllRequest Show has started! 🤩 WOOHOO Hit up ...	Twitter for iPhone	272	NA	NA
3-23 17:31:44	mariela_g_g	If you are still confused, here a handy #blackbear guide on ...	Twitter Web App	121	NA	NA
3-24 06:18:40	Futoagon614	花見の季節‥ポゴナかコロネみたいなもので #ツキノワグ...	Twitter for iPhone	172	NA	NA
3-25 03:22:58	Futoagon614	ʕ•ᴥ•ʔわぁ〜い！ ʕ•ᴥ•ʔ‥‥‥ □□ □□ かぶり (あんまりクマ...	Twitter for iPhone	170	NA	NA
3-23 12:37:39	Futoagon614	#リゼのミニチュア ほぼっぺにも乗るもん！ ʕ•ᴥ•ʔおりろー...	Twitter for iPhone	134	NA	NA
3-21 17:24:47	bleudawn7	There I fixed it @AngryandAsian @Destigmatize @QuareF...	Twitter for Android	166	NA	NA
3-20 11:48:07	Futoagon614	🍀ブログ更新しました🍀 のんほいパークに離出没！？ htt...	Twitter for iPhone	186	NA	NA
3-21 20:11:08	MikeAdamOnAir	US pop radio callouts -- top songs by net positive audience ...	Twitter Web App	249	NA	NA
3-18 03:17:32	Futoagon614	#ホワイトデー から日数が経っても、チョコがまだあるの...	Twitter for iPhone	168	NA	NA
3-19 11:52:31	iwatatube	#八木山動物公園 #ツキノワグマ #BlackBear #YagiyamaZoo...	Twitter Web App	157	NA	NA
3-17 21:01:15	grizzlygirl87	It is time to reveal our socially distant #KnockKnockWhosBe...	Twitter for Android	211	NA	NA
3-23 02:39:37	sincityhighroll	Prince Zuko😈 hates the water and Athena😈 loves the wat...	Twitter for Android	269	NA	NA
3-20 17:10:00	brettbdrums	Got the toms sounding pretty darn good! Decided to test th...	Twitter for iPhone	118	NA	NA
3-20 07:43:00	MSieboldPhotos	This black bear might seem to be a little bit off target in its ...	Twitter for iPhone	277	NA	NA
3-19 00:18:41	989MagicFM	Newish @ 6ish 🥊 w/ @DylanOnTheRadio It's the defendin...	Twitter for iPhone	262	NA	NA

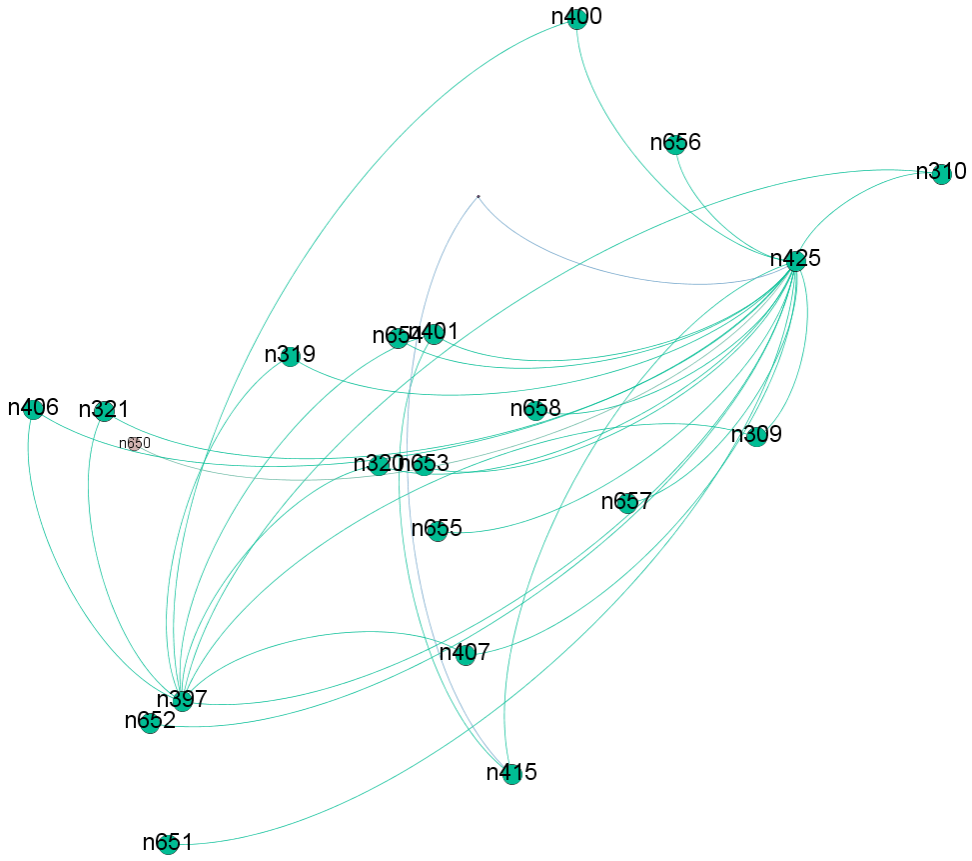
Showing 1 to 19 of 135 entries, 90 total columns

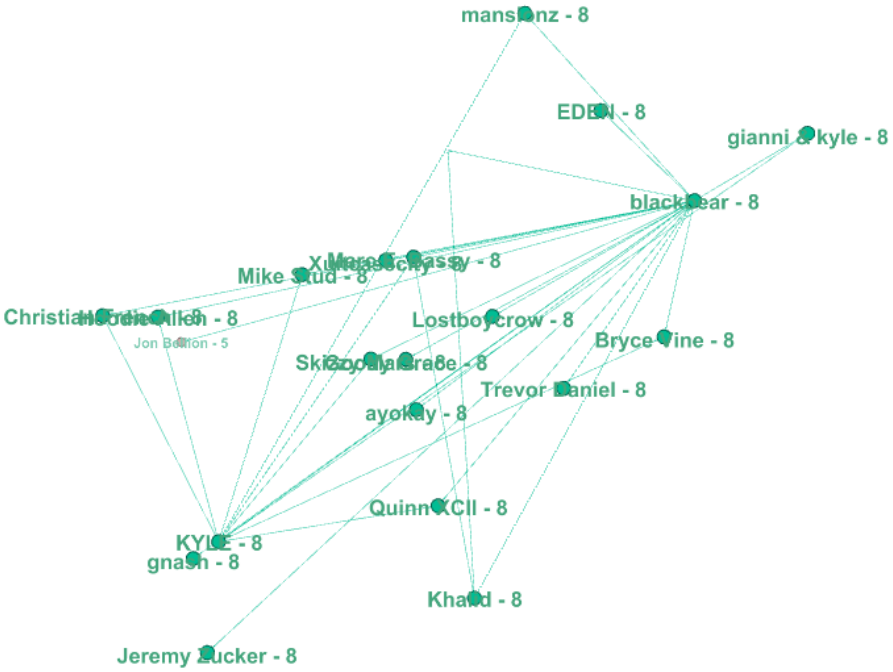
th	reply_to_status_id	reply_to_user_id	reply_to_screen_name	is_quote	is_retweet	favorite_count	retweet_count	quote_count	repl
	NA	NA	NA	FALSE	FALSE	180	35	NA	NA
	NA	NA	NA	FALSE	FALSE	59	22	NA	NA
	NA	NA	NA	FALSE	FALSE	36	14	NA	NA
	NA	NA	NA	FALSE	FALSE	30	11	NA	NA
	NA	NA	NA	FALSE	FALSE	28	8	NA	NA
	NA	NA	NA	FALSE	FALSE	7	8	NA	NA
	NA	NA	NA	FALSE	FALSE	31	7	NA	NA
	NA	NA	NA	FALSE	FALSE	44	5	NA	NA
	NA	NA	NA	FALSE	FALSE	27	3	NA	NA
	NA	NA	NA	FALSE	FALSE	28	3	NA	NA
	NA	NA	NA	TRUE	FALSE	20	2	NA	NA
	NA	NA	NA	FALSE	FALSE	1	1	NA	NA
	NA	NA	NA	FALSE	FALSE	6	1	NA	NA
	NA	NA	NA	FALSE	FALSE	0	1	NA	NA
	NA	NA	NA	FALSE	FALSE	7	1	NA	NA

owing 1 to 19 of 135 entries, 90 total columns

2.5 Spotify Related Artists

Below shows modularity class 8 which is the module that blackbear is in. This shows artists that are related to blackbear under the name of other nodes as well as it provides you a visual representation on which node is worth more than the other and how they can possibly connect.




mansonz - 8
EDBL - 8
gianni & kyle - 8
blackbear - 8
Mike Stud - 8
Lostboycrow - 8
Jon Bellion - 5
Bryce Vine - 8
Trevor Daniel - 8
ayokay - 8
Quinn XCII - 8
KYLE - 8
gnash - 8
Khalid - 8
Jeremy Zucker - 8

MILESTONE 2

2.6 YouTube views/likes

This data has been extracted using the Youtube API by searching up the term “Blackbear” and retrieving the top 5 searches in relation to blackbear listed below. Each video contains its own ID so that we can further investigate the songs individually and compare them.

Video list and IDs

Video No	Name of Video/Song	Video_id
1	Hot girl bummer	yMlKJGKyoCo
2	Higher	L_FRaiw7hg4
3	Do re mi	dfzlUHOFQ5k
4	Me&ur ghost	HQM_T-ijA_I
5	Worry about me	b4qC6UORQOQ

Bellow I have used the get_stats function provided by the youtube API to extract each top 5 songs statistics which include viewCount, likeCount, dislikeCount, favouriteCount, commentCount.

Hot girl bummer

video1	list [6]	List of length 6
id	character [1]	'yMlKJGKyoCo'
viewCount	character [1]	'42432870'
likeCount	character [1]	'966112'
dislikeCount	character [1]	'12518'
favoriteCount	character [1]	'0'
commentCount	character [1]	'18461'

Higher

video2	list [6]	List of length 6
id	character [1]	'L_FRaiw7hg4'
viewCount	character [1]	'113088'
likeCount	character [1]	'17481'
dislikeCount	character [1]	'93'
favoriteCount	character [1]	'0'
commentCount	character [1]	'924'

Do re mi

video3	list [6]	List of length 6
id	character [1]	'dfzIUHOFQ5k'
viewCount	character [1]	'140682557'
likeCount	character [1]	'1689082'
dislikeCount	character [1]	'54500'
favoriteCount	character [1]	'0'
commentCount	character [1]	'53965'

Me & ur ghost

video4	list [6]	List of length 6
id	character [1]	'HQM_T-ijA_I'
viewCount	character [1]	'6872467'
likeCount	character [1]	'169396'
dislikeCount	character [1]	'1679'
favoriteCount	character [1]	'0'
commentCount	character [1]	'2684'

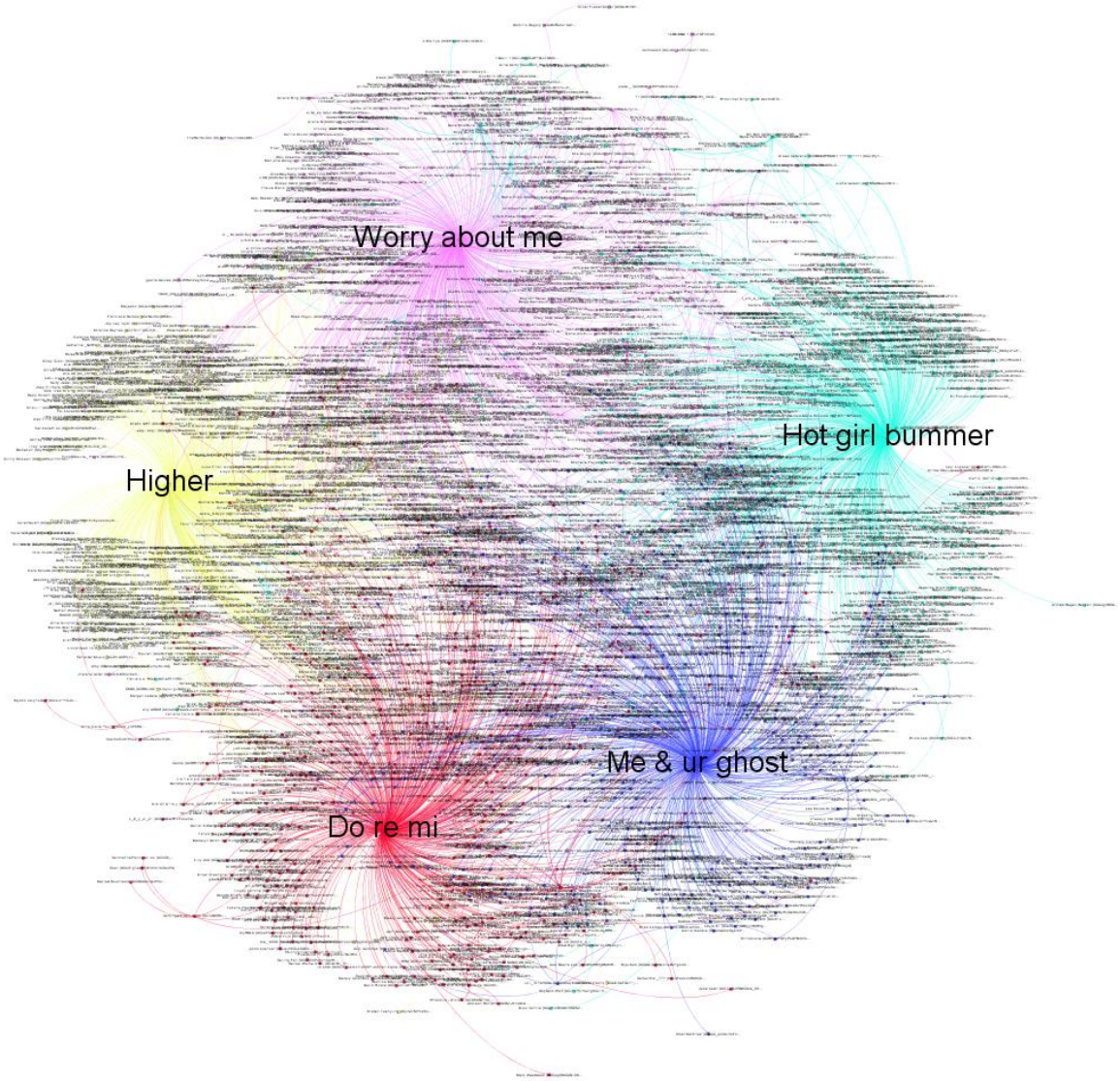
Worry about me

video5	list [6]	List of length 6
id	character [1]	'b4qC6UORQOQ'
viewCount	character [1]	'13364218'
likeCount	character [1]	'135592'
dislikeCount	character [1]	'4059'
favoriteCount	character [1]	'0'
commentCount	character [1]	'3049'

Name	Views	Likes
Hot girl bummer	42432870	966112
Higher	113088	17481
Do re mi	140682557	1689082
Me & ur ghost	6872467	169396
Worry about me	13364218	135592

As shown through the data extracted the video “Do re mi” is the most viewed and most liked. Higher is currently the lowest views and likes but this is due to it being the most recent song.

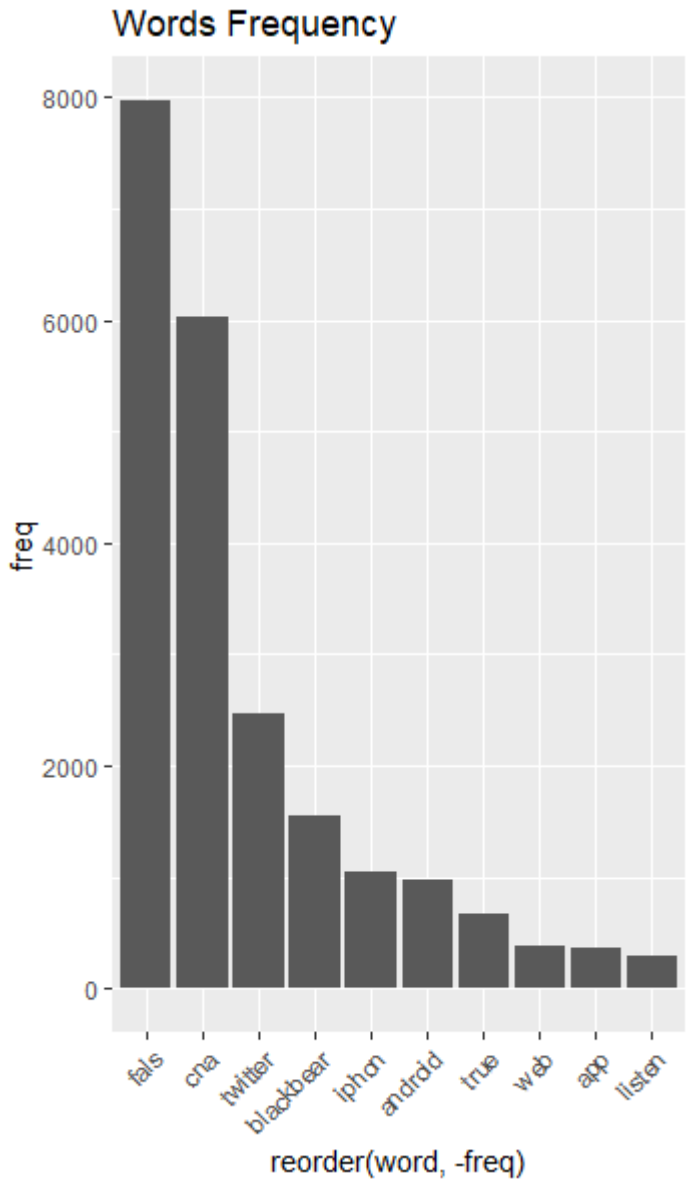
Actor network



2.7 Term-document Matrix & top 10 Terms

Using the search term “blackbear” I have created a TD matrix which shows the terms frequency of the top 10 terms in relation to blackbear. As seen Fals, can and twitter were the most frequent.

Top 10 Terms



Top 10 from the Term-Document Matrix

	word	freq
fals	fals	5985
cna	cna	4514
twitter	twitter	1899
iphon	iphon	816
android	android	797
blackbear	blackbear	579
true	true	547
listen	listen	295
que	que	263
web	web	248

Term-Document Matrix

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
aalanysv	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
abbymilamm	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
abrizaam	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
accidentalbear	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
adamoao	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
adeliina	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
adidasxguccii	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
adoringaveryz	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
adrismil	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
aglagluficumico	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
agusstina	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
ahappiermurd	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
aimeetaylor	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
aintsleep	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
aixacelest	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
alchemillion	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0
alddoo	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
aleccharron	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
alejandroricov	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
alesdo	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0

2.8 Centrality analysis

Top 20 items from subgraph ordered by in-degree and out degree

Out-degree

```
> V(blackbear_bimodal_sub)[order(oud, decreasing=TRUE)[1:20]]
+ 20/309 vertices, named, from 2492893:
 [1] @larshenson1     @lynnpotter11    @twitsareangry   @bansmacking     @freeyamind424
 [6] @smakhanian      @stachanna       @tucsonrosie2001 @radioquo        @fashionabanon
[11] @lavenderlives   @listening4his   @katiesc16350078 @freqnetwork     @rickywaddas
[16] @coffee4danz     @maryana29983592 @yerfavoritegreg @kdrsplaylist    @instrumentals13
```

In-degree

```
> V(blackbear_bimodal_sub)[order(ind, decreasing=TRUE)[1:20]]
+ 20/309 vertices, named, from 2492893:
 [1] @youtube         #blackbear       #nowplaying      @cannabijesus    @cannafrom
 [6] @cgilday         @chiefstonefox   @crazyjane125    @dibywon         @girlsundry
[11] @gracealijah     @hhmubanksia     @hummingbirdpass @joebloww13      @johnsummerssr4
[16] @johntitor33621  @kravitzq        @lbf777          @loriblackbear   @mikecoulson48
```

Degree

This is a measure of the most used terms and most connected users. Below is the differentiation between in and out degree of the top 30 terms and users.

In Degree Based on actor term relationship

```
> sort(degree(blackbear_bimodal_sub, mode="in"), decreasing=TRUE)[1:30]
      #blackbear         @youtube   #blackbearlive      #nowplaying    @iamblackbear
              29               29               19               17               14
   @cannabijesus       @cannafrom         @cgilday   @chiefstonefox    @crazyjane125
              12               12               12               12               12
       @dibywon      @gerrymeany      @girlsundry     @gracealijah     @hhmubanksia
              12               12               12               12               12
@hummingbirdpass      @joebloww13  @johnsummerssr4  @johntitor33621        @kravitzq
              12               12               12               12               12
         @lbf777   @loriblackbear   @mikecoulson48 @mycannabiseeds1   @nemesisnibiru
              12               12               12               12               12
   @protesturban          @ps9714   @randolphtrent @realdonaldtrump @redfeatherheat1
              12               12               12               12               12
```

Out Degree Based on actor term relationship

```
> sort(degree(blackbear_bimodal_sub, mode="out"), decreasing=TRUE)[1:30]
   @lynnpotter11   @twitsareangry     @larshenson1     @bansmacking   @freeyamind424
              50              50               50               50               49
     @smakhanian       @stachanna @tucsonrosie2001   @fashionabanon   @lavenderlives
              49              49               49               40               40
  @listening4his @katiesc16350078        @radioquo     @coffee4danz @maryana29983592
              40              38               14               13               13
@yerfavoritegreg @instrumentals13  @tbeatzofficial     @freqnetwork     @islamiviola
              13              11               11               10               10
   @tammiehedden  @1usmarine4maga @goboulotlaradio  @awesomegroomer     @rickywaddas
              10               8                8                7                7
 @accidentalbear     @brooke_9312  @miraclejacob12    @realtntradio  @edwardsearnest
               6               6                6                6                5
```

Closeness

Closeness centrality is how fast a node can reach others. Bellow is the exported outcomes of the closeness of names and terms in both decreasing and increasing order.

By using this data you can better the reach and find out the closest terms and names and use it to your advantage. Such as marketing to the right market.

```
> sort(closeness(blackbear_bimodal_sub), decreasing=FALSE)[1:30]
             #anghami              @alltimelow               #adventure
         1.050729e-05             1.050729e-05             1.050729e-05
           #blackbear               #havingfun                  #hiking
         1.050729e-05             1.050729e-05             1.050729e-05
            #nature              #photography #shenandoahnationalpark
         1.050729e-05             1.050729e-05             1.050729e-05
           #wildlife               #nowplaying           #animalrescue
         1.050729e-05             1.050729e-05             1.050729e-05
            #animals                     #bear                  #nevada
         1.050729e-05             1.050729e-05             1.050729e-05
       @safehavenrescu1               #anonymous           @iamblackbear
         1.050729e-05             1.050729e-05             1.050729e-05
            @youtube               @g0nervinyl                 @iammp2
         1.050729e-05             1.050729e-05             1.050729e-05
         @svandez1234           #blackbearlive             #bouvidoodle
         1.050729e-05             1.050729e-05             1.050729e-05
          #flandoodle #flandoodlesofinstagram              #fluffydogs
         1.050729e-05             1.050729e-05             1.050729e-05
            #holycow              #ludothebear             @2018tsalagi
         1.050729e-05             1.050729e-05             1.050729e-05
Warning message:
```

```
> sort(closeness(blackbear_bimodal_sub), decreasing=TRUE)[1:30]
  @fashionabanon    @lavenderlives    @listening4his @katiesc16350078 @tucsonrosie2001
     1.277759e-05      1.277759e-05      1.277759e-05      1.267765e-05      1.263280e-05
     @bansmacking    @freeyamind424       @smakhanian        @stachanna    @twitsareangry
     1.258416e-05      1.253557e-05      1.253557e-05      1.253557e-05      1.252787e-05
    @lynnpotter11      @larshenson1         @radioquo      @coffee4danz @maryana29983592
     1.250485e-05      1.248876e-05      1.100255e-05      1.096864e-05      1.096864e-05
@yerfavoritegreg @instrumentals13   @tbeatzofficial      @islamiviola     @tammiehedden
     1.096864e-05      1.089515e-05      1.089515e-05      1.085871e-05      1.085871e-05
     @freqnetwork  @1usmarine4maga @goboulotlaradio  @awesomegroomer      @rickywaddas
     1.085753e-05      1.078656e-05      1.078656e-05      1.075084e-05      1.074934e-05
  @accidentalbear      @brooke_9312  @miraclejacob12    @realtntradio    @warren_smith8
     1.071536e-05      1.071536e-05      1.071536e-05      1.071536e-05      1.071513e-05
```

Betweenness

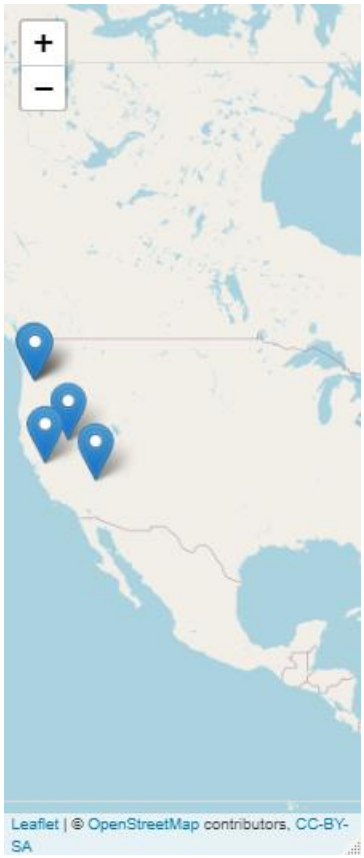
Betweeness centrality is defined by the number of shortest paths going through a vertex. These are the most common terms and names that are on twitter in relation to blackbear. This can be used to better marketing by using the hashtags and names since its almost garenteed due to it being the most related to the topic and able to get the most reach the fastest.

```
> sort(betweenness(blackbear_bimodal_sub, directed = FALSE), decreasing=TRUE)[1:30]
   @iamblackbear       #blackbear         @youtube      #nowplaying  @miraclejacob12
        31404.279        24269.524        19565.525        17999.612        17493.612
    @islamiviola         @csack99    @rostern_mark  @andrea72146420         @radioquo
        10550.360         9685.000         9072.000         8429.360         6431.000
  #blackbearlive     @freqnetwork       @bearglows           #radio    @tammiehedden
         6296.000         4939.112         4842.000         3875.250         2727.000
   @realtntradio     @swiftenrose @tucsonrosie2001  @1usmarine4maga @goboulotlaradio
         2674.000         2365.000         2131.255         2128.000         2128.000
     @smakhanian       @stachanna   @freeyamind424     @bansmacking  @awesomegroomer
         2001.193         2001.193         1981.701         1862.683         1827.000
            #pop             #rap  @accidentalbear     @brooke_9312   @fashionabanon
         1671.105         1653.577         1525.000         1525.000         1481.641
>
```

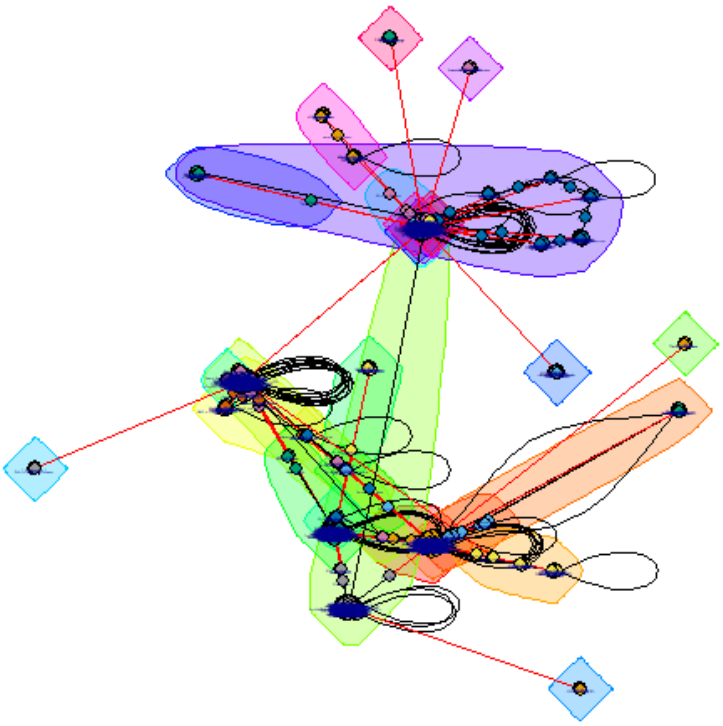
2.9 Community analysis

Infomap

An infomap was made with the term blackbear so that we can pinpoint people in America who have tweeted about blackbear (assuming their location is on). This can be used more in depth with more terms to see where exactly your reach is and where you should be aiming for. Who knows it could be a possibility of the next place an artist can tour with this data.



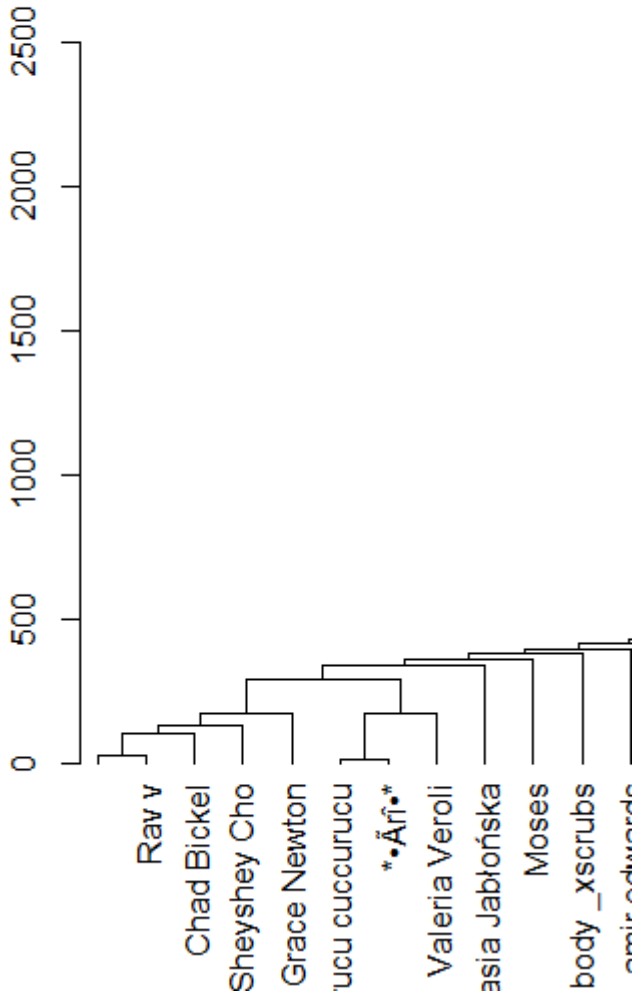
Girvan-Newman



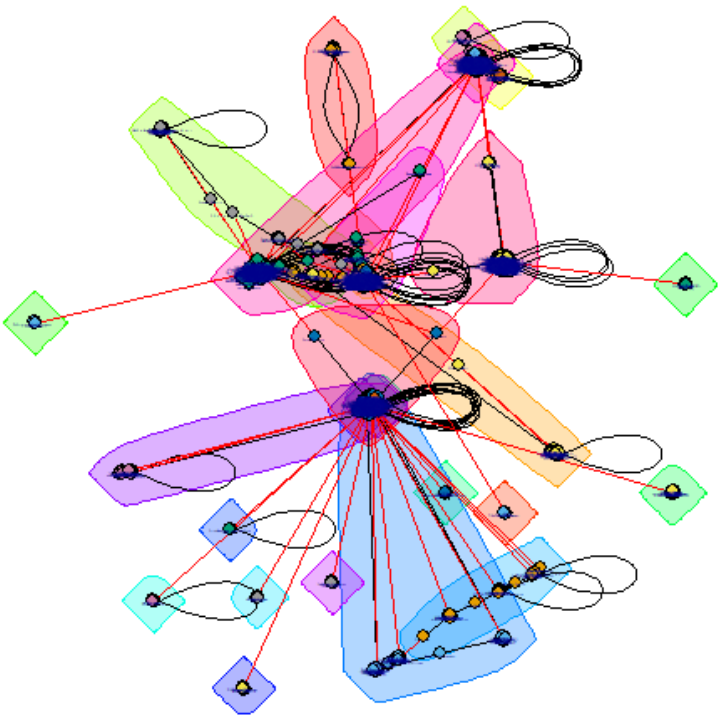
This algorithm detects the communities by progressively removing edges from the original network. The original networks as shown are the youtube videos. The visual representation shows how they link and where the nodes connects from there to other similar nodes.


Hot girl bummer
Me & ur ghost
Worry about me
Do re mi
Higher

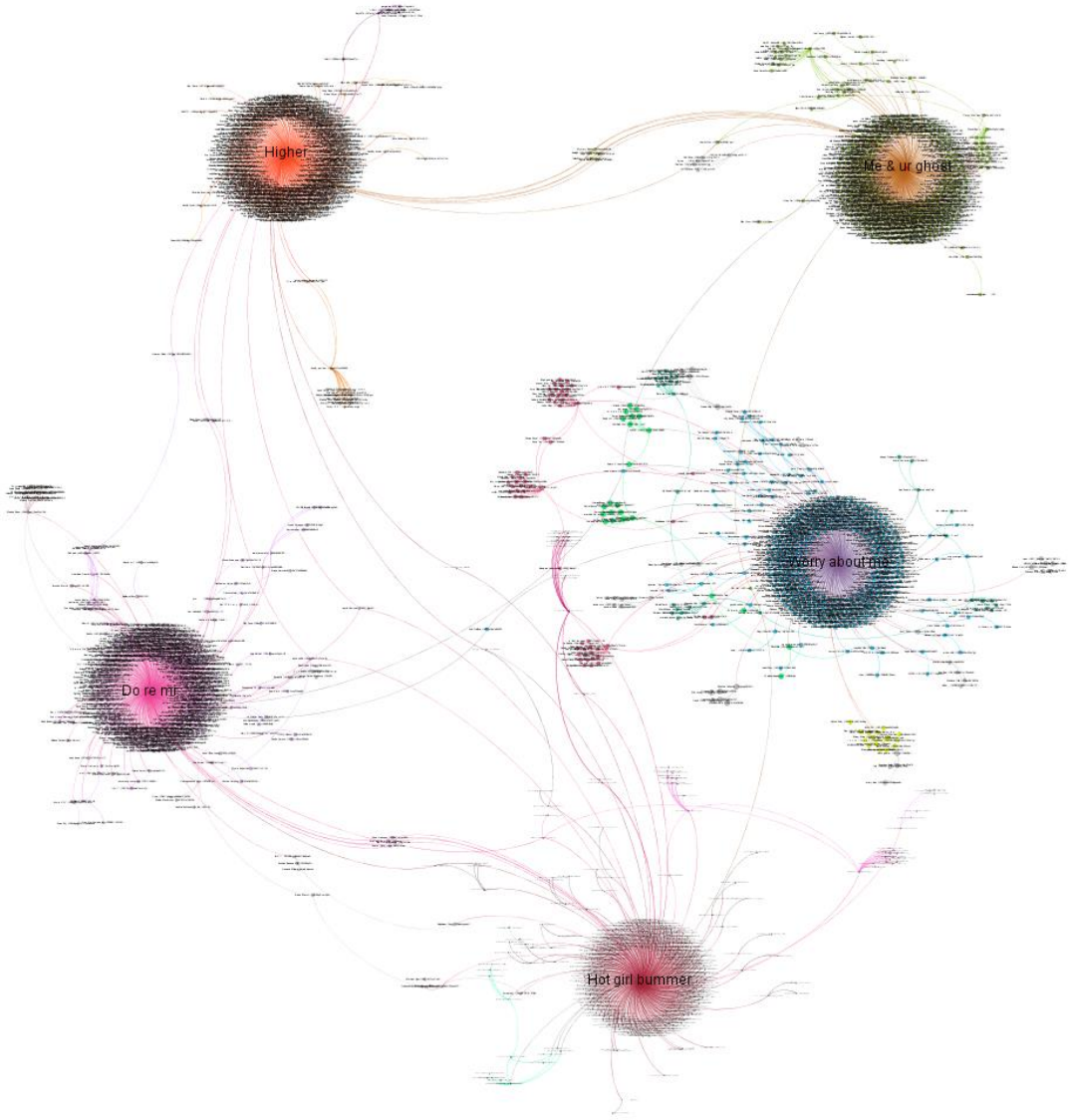
Edge betweenness hierarchy



Louvain analysis



The Louvain method is for community detection and is an algorithm used for detecting communities in the network. This algorithm is hierarchical starting from the single nodes connecting into the main nodes which are the top 5 searched videos. As seen they all connect somehow through similar nodes or just to each other. The biggest nodes connecting to the searched video nodes are the main nodes which are communities that have the most in common using the algorithm.


Higher
Me & ur ghost
worry about me
Do re mi
Hot girl bummer

Milestone 3

3.1 K-means

Using K-means clustering algorithm we can classify a user's friends and followers into clusters. In one of the previous sections we extracted the top 5 influential users for the iamblackear twitter account. The most influential user in relation was MODSUN which is why he has been chosen to further extract information using K-means clustering to create a better idea on what we know about our audience from the most influential user.

Bellow shows a table representation of the centre point of each cluster and their Follower and Friends count. This can be used to help identify how popular some of the clusters are with friends and followers.

```
> user2Means.log$centers
  logFollowersCount logFriendsCount
1         5.6360561        6.033581
2         4.0297787        5.479878
3         7.3120130        7.188308
4         2.4930696        4.669923
5         0.7756755        3.738353
>
```

3.2 Decision Tree

Here we use Blackbear as an artist and find songs that have the same or similar audio features. We create a comparison to the top 100 songs playlist created by spotify. With this data we can find out whether or not the songs are related to blackbears songs audio features. We use decision trees to create classification which is the task of predicting which category a specific data item belongs to, in this case to blackbear. In this example we investigate the audio feature named Valence and predict how often Blackbears songs valence is the same or similar in relation to the top 100 songs. We then use the Classification and Regression Trees (CART) model to compare how accurate our data is.

```
CART

166 samples
 11 predictor
  2 classes: '0', '1'

No pre-processing
Resampling: Bootstrapped (25 reps)
Summary of sample sizes: 166, 166, 166, 166, 166, 166, ...
Resampling results across tuning parameters:

  cp          Accuracy   Kappa
  0.04545455  0.6473928  0.2697512
  0.09090909  0.6053237  0.1778525
  0.10606061  0.5991837  0.1625891
```

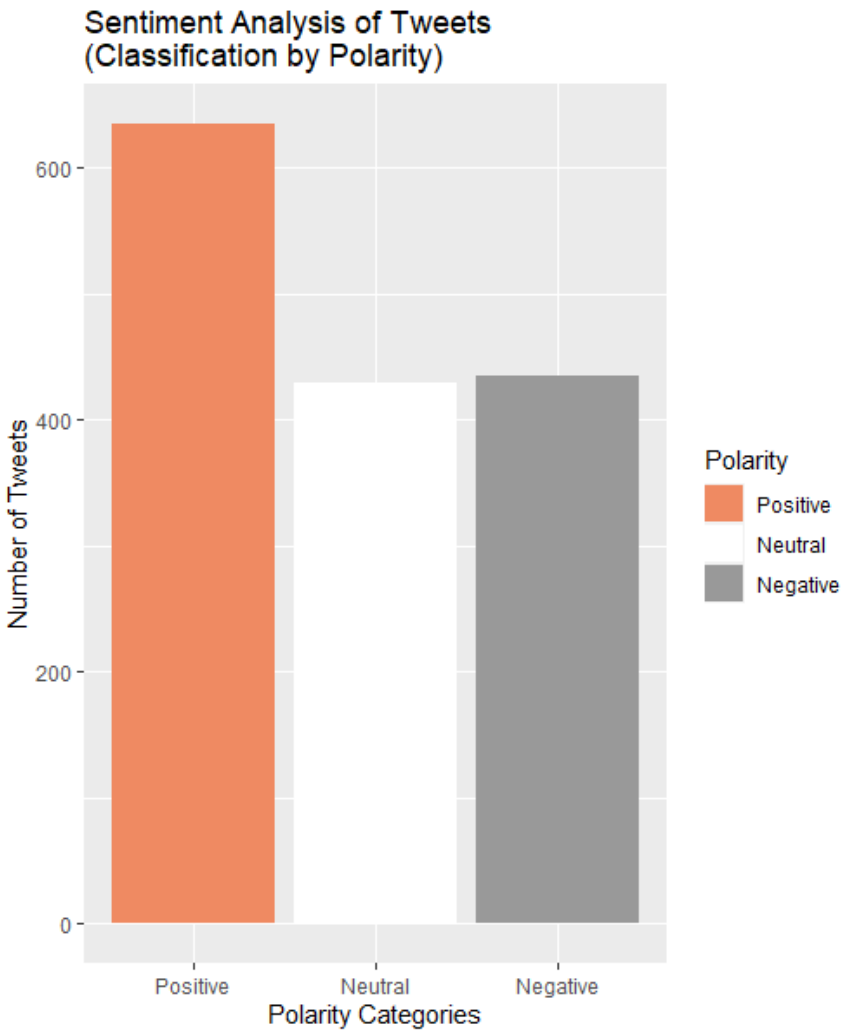
```
Bootstrapped (25 reps) Confusion Matrix

(entries are percentual average cell counts across resamples)

          Reference
Prediction    0    1
         0 40.9 17.1
         1 18.2 23.8

 Accuracy (average) : 0.6475
```

A confusion matrix is then used on the data to measure the performance of the algorithm. The result is that Blackbears audio valence compared to the top 100 songs is averaging 0.6475 which is not bad for the top 100 songs. This same thing can be done to measure other artists Valence average compared to the top 100 and see how they are performing in comparison.

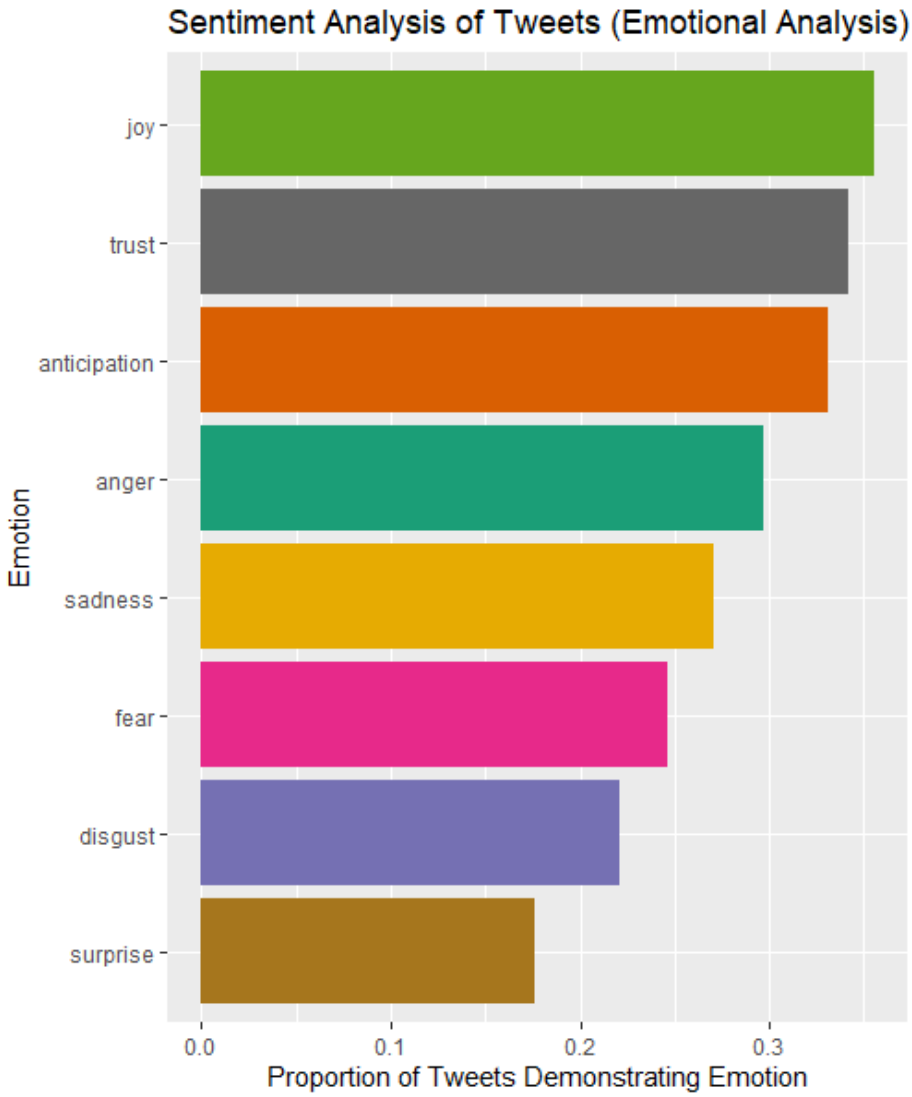
3.3 Sentiment analysis

Sentiment analysis is used to track human response towards a certain event or entity and is measured by how they feel by emotions and polarity.

As you can see there is a high positive polarity that include the word blackbear. This meaning a large number of the tweets we extracted were positive as opposed to neutral and negative which ended up being the same.



Bellow is our emotional classification on the tweets that were related to the term “blackbear”. The data extracted shows the tweets that have been converted to eight types of emotional levels. The bar chart shows the proportion of tweets containing each emotion in relation to the search term. This helps us provide a view on Blackbears public image and see how people feel about him or things related to him. You can even go further to specify events that blackbear has been in and how people are feeling according to the tweets in relation to the event and blackbear. This only scratches the surface on how much information you can retrieve in accordance to how a chosen audience can feel about a subject.



3.4 LDA topic modelling

Topic modelling is when we use a set of text and divide them into different topics which is done by weighing the text words and finding related meanings and common features.

	Topic 1	Topic 2	Topic 3	Topic 4	Topic 5	Topic 6	Topic 7	Topic 8	Topic 9	Topic 10	Topic 11	Topic 12
1	idfc	music	never	listening	song	good	see	much	think	can	yes	heart
2	got	time	heart	need	time	back	can	song	make	life	said	get
3	make	can	people	people	yes	time	people	get	listening	need	never	see
4	back	good	get	can	back	life	get	can	song	make	get	never
5	said	heart	said	back	said	got	said	life	yes	never	song	listening
6	much	make	think	think	got	make	got	need	people	back	back	life
7	need	life	yes	make	life	yes	think	back	heart	yes	make	think
8	never	back	life	yes	music	can	back	make	back	think	can	back
9	song	yes	need	life	make	think	make	yes	can	people	life	make
10	yes	think	much	heart	can	people	yes	think	life	heart	think	yes

The information extracted is related to text Blackbear which created 12 topics and divided all the tweets into their topics. All these terms in the topics are all related to blackbear. As you can see most are music related along with sad emotions which is the target audience.

3.8 Social media analytics findings

What I have gathered from this information is a quite a lot of data in relation to the artist Blackbear that can be very beneficial to finding who your audience is and what you can do to perform and better yourself. Three API's were used in extracting data, these include Twitter, Spotify and Youtube. Using twitter, we created visualisations of modularity classes which shows where blackbear falls under a category of relation and similar twitter users. Joy distributions were also extracted comparing them by album name. Top 5 influential users were extracted in relation to blackbear and visualised along with the top 10 terms that appeared with the term "Blackbear". Highest retweets and accounts that follow them were also extracted using the twitter API. The spotify API was used to extract information about blackbear and show related artists in a visualisation and how they compare. The youtube API was used to extract data like views, likes, video details and actor networks which show the relation between youtube videos and other youtube videos. The top 10 frequent terms in the comments were also extracted. Centrality analysis was conducted to measure the most used terms and most connected users. Infomaps were created to find out where people were tweeting about blackbear. The Girvan Newman algorithm was used to detect the youtube communities and how they link to similar other communities. Louvain analysis was used to detect a community in the network and shows other communities that are most in common in relation to the top 5 searched blackbear videos.

K-means clustering algorithm was used to classify users friends and followers into clusters by using the most influential user MODSUN on twitter. Then a decision tree was used to compare the audio features of blackbear to the audio features of the top 100 songs on spotify. Sentiment analysis found that we could track human response towards the search term blackbear. Visualisation on whether the tweets that relate to blackbear are positive, neutral, or negative. An emotional analysis was done which included using words from the dictionary and comparing it to the data text extracted from twitter to create a bar graph of 8 separate emotion and their proportions were visualised. LDA topic modelling was used to create topics according to twitter data and separating them by weighing the text words and finding related meanings and common features. All this was the data that was extracted and found.

3.9 Suggested actions

Improving popularity is very possible using this data. It provides you with quite a lot of audience information so it would be easy to use marketing when you know the audience along with what they like and how to find their similarities. Twitter data can be used to find out geographically where your audience is and you can focus on them, the people who want to hear you. Youtube data can be used to find similar videos that your audience watch which can help you gauge on what the audience likes to see. Along with that comparing your own video data can be useful to find out what is working best and what is working least. Youtube provides quite a lot of categories of information that can be used as useful information. Spotify data is probably one of the best in improving music because they provide quite a lot of categories to do with music features and they are easily comparable. Music is the main thing that blackbear produces and this information is vital for progression and finding out what music is working.

3.10 Possible refinement

Possible refinement in the data that was extracted is very possible due to me only scratching the surface. There are endless possibilities on how much more you can get from using the API's that can better popularity and possibly help direct towards what the audience wants. Going further extraction of related artist data can be helpful in comparing Blackbear to their data and quite a lot of direction and information can come out of that by just expanding the search properties. This is one of many ways that data extraction and analysis can benefit artists in the music industry generate better music, what the people want with as much detail as possible, every little bit does count in making a difference.