Social Network Analysis Lab: Data Capture in Twitter

Puerta de Toledo

Master in Big Data Analytics

Web data analytics and usage

Ion Bueno Ulacia NIA 100364530

Daniel Martín Cruz NIA 100384121

Ion Bueno Ulacia NIA 100364530
Daniel Martín Cruz NIA 100384121

uc3m | Universidad Carlos III de Madrid

Contents

1. Introduction

The 4th of March 2022, the Argentinian producer Bizarrap and the Puerto Rican rapper René Residente released the theme called *BZRP Music Sessions #49*, the 49th chapter of the series of collaborations known as *sessions* that the Argentinian producer is doing with international artists, mainly Hispanic. This last 9-minutes-long session has been the most controversial of all reaching to more than 17 million views in the release day, a total record in the Spanish-speaking community. The reason for this great impact is that, in this song, Residente retakes the *beef* he had with the Colombian artist JBalvin, that started in September 2021 when J Balvin wanted to discredit the Latin Grammy's. Residente didn't agree with this affirmation because he thought that it was disrespectful with a lot of artist that were nominated for the first time to these awards. Their battle was in *standby* until René Residente released this song in which he attacks the Colombian artist non-stop during 9 minutes, accusing him of doing everything for money and not being even capable of writing his own songs.

It is because the great impact of this polarized event that we have decided to analyze how Twitter's users have lived the new chapter of this battle between two very relevant artists of Hispanic music.

2. Declared network

In order to obtain a declared network related to this topic, we have thought of a list of 32 accounts that belong to this Hispanic music world or are very close to it and are somehow related with the issue in question. Some of the being even mentioned in the song.

- Residente: Puerto Rican rapper.
- JBALVIN: Colombian singer
- bizarrap: Argentinian producer.
- rosalia: Spanish singer
- DONOMAR: Puerto Rican singer.
- LatinBillboards: Latin Music awards.
- Billboard: American magazine specialized in music.
- UniversalMusica: Music production company.
- maluma: Colombian singer.
- IbaiLlanos: Spanish influencer.
- Anuel_2bleA: Puerto Rican singer.
- ArcangelPrrra: American/Puerto Rican singer.
- rubenblades: Panamanian singer-songwriter.
- MykeTowers_: Puerto Rican rapper.
- chocquibtown: Colombian Hip-Hop band.
- ozuna: Puerto Rican singer.
- TEGOCALDERON: Puerto Rican singer.
- DJLUIAN: Puerto Rican DJ.
- MariaBecerra22: Argentinian singer.
- Tyga: American rapper.
- Spotify: Multimedia services company.
- losxavales00: Spanish influencers.

- RecordingAcad: Grammy's awards account.
- AmericaTV: Argentinian music TV channel.
- Moluskein: Puerto Rican influencer.
- europa_fm: Spanish radio station.
- karolg: Colombian singer.
- NathyPeluso: Argentinian singer.
- CamiloMusica: Colombian singer.
- FarrukoOfficial: Puerto Rican singer.
- DukiSSJ: Argentinian rapper.
- SebastianYatra: Colombian singer.

As we said, we have generated the declared network with these 32 profiles so we could observe the following relations existing between them. Using Gephi, we have processed this network using the layout method `ForceAtlas 2` with a scaling of 200 and preventing the overlap. We also wanted to detect possible communities present in this network, that's why classified the nodes according to the Modularity Class assigned by Gephi applying a resolution of 1. This is the resulting network:

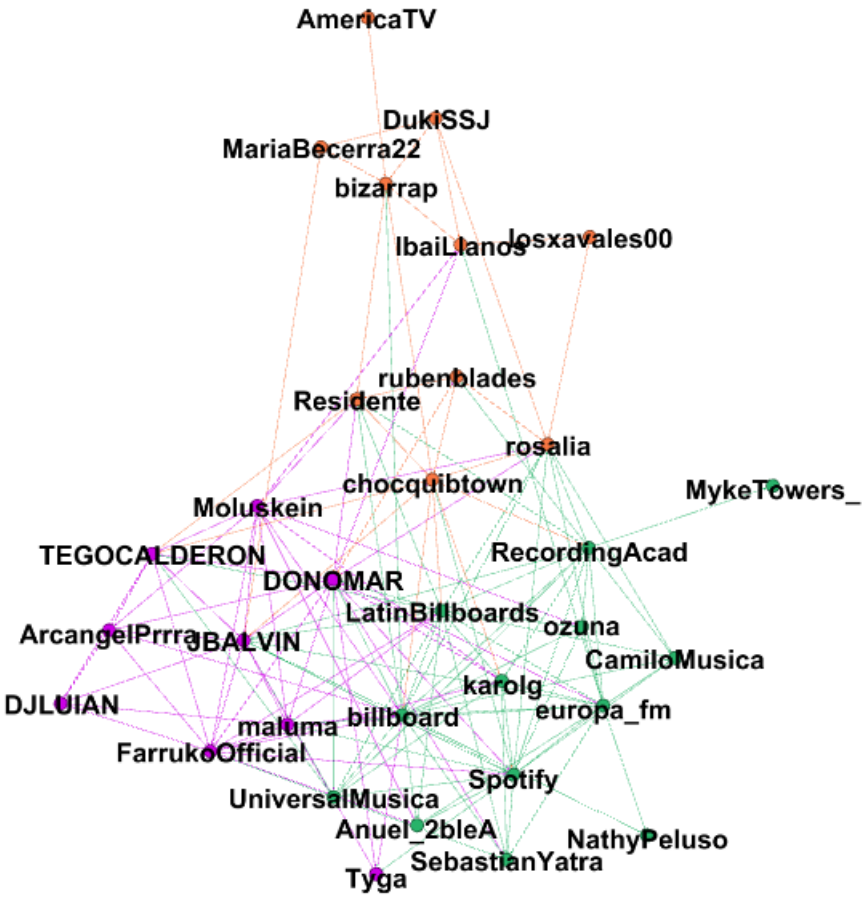


Figure 1: Declared network composed by users

In the proposed clustering that Gephi did for us, we can see three differentiated groups characterized with different colors, we will analyze if this grouping makes actual sense:

- **Orange group**: in this group we can see accounts like Ibai, *losxavales00*, Duki, Rosalia, etc. Mainly artists and influencers with international fame but who started their career not long ago coming from the *underground scene* we could say. There are also artists with a great career behind them like Residente and Ruben Blades but that also share those underground origins.
- **Pink group**: this is the smallest group and is formed by Latin American artists with millions of followers around the world and long and successful careers. This main difference between this group and the previous one is that they are not rising starts, their careers are long and backed up by important music production companies.
- **Green group**: this group is characterized for being made up by big companies like Billboard or Spotify and highly commercial artists like Karol G, Ozuna, Sebastián Yatra, etc. We could say that is the group situated closer to the show business world.

2.1. Centrality metrics

In this subsection, we will analyze different centrality metrics to have an insight of the most influential nodes of the network.

2.1.1. Degree

The degree of a node indicates the number of edges incident to a given user (in and out). These are the ten nodes with the highest degree of this network:

User	Degree
billboard	29
DONOMAR	27
Spotify	24
JBALVIN	22
LatinBillboards	22
FarrukoOfficial	19
RecordingAcad	16
europa_fm	16
chocquibtown	15
Residente	14

Table 1: Top 10 users with the highest degree of the network

In this list we observe that the nodes a higher degree are mainly big companies and some artists that have had important and successful careers like Don Omar or JBalvin.

2.1.2. Closeness

The closeness centrality metric is defined as *the length of the average shortest path between a node and all other nodes in the network*. In other words, it indicates how close you are to influential nodes in the network. These are the 10 nodes with a higher closeness in our declared network:

User	Closeness
SebastianYatra	1
billboard	0,71
DONOMAR	0,67
LatinBillboarda	0,64
Spotify	0,62
europa_fm	0,62
chocquibtown	0,59
UniversalMusica	0,58
Moluskein	0,58
FarrukoOfficial	0,58

Table 2: Top 10 users with the highest closeness of the network

The accounts appearing in this list are more or less the same we saw in the degree list but there are remarkable differences like the appearance of Sebastian Yatra, a very famous artist with collaborations with a lot of different companies and performers.

2.1.3. Betweenness

The last centrality metric that will be studied using this declared network will be betweenness. This metric basically measures how many shortest paths between two nodes that go through a given node. Talking about Twitter users, the betweenness translates to how influential an user is, being able to put in common different people. These are the 10 accounts with a higher betweenness in this network:

User	Betweenness
DONOMAR	166
billboard	97
bizarrap	88
Residente	84
JBalvin	71
Spotify	62
rosalia	48
RecordingAcad	40
FarrukoOfficial	33
chocquibtown	31

Table 3: Top 10 users with the highest betweenness of the network

In this table we see users we have seen in the top part of the previous networks like Don Omar and Billboard. In the next positions, we observe accounts like the ones of Bizarrap, Residente or JBalvin, world-known artists and characterized for having collaborated in different ways with many artists and companies, this is the reason they have such a high betweenness value.

3. Dynamic network

3.1. Data collection

As mentioned before, from the Thursday 4 of March, the topic has been very controversial, being one of the main discussed topics in Twitter and the most listened for YouTube music. For this reason, the easiest way to capture most of the available information was employing a **query**.

Before that, a research was carried out to determine which words were the most used in the tweets to get different opinions. The pattern was simple, because most of the tweets contain one or both names of the artists, employing or not hashtag. In addition, a huge amount of people commented only naming the producer with the hashtag, #BZRP. Of course, due to the sense of humour in this social network, the topic #RIPJBalvin was very common, referring to the hard allegations received by the artist J Balvin. Taking into account these points, the used query was:

```
(residente OR #residente) OR (Residente OR #Residente) OR
((balvin OR #jbalvin) OR (Balvin OR #JBalvin))) OR (JBalvin OR jbalvin) OR
(#BZRPMusicSessions) OR (#RIPJBalvin OR #BZRP)
```

Where all different combinations were imposed. We collected a total of **54.000** tweets to carry on the study. The dynamic network generated is composed by this set, analyzing the relationships through **retweets**. Remember that a retweet is a re-posting of a tweet. Twitter's retweet feature helps people share that tweet with all of their

followers. Someone can retweet the own tweets or tweets from someone else. This feedback is understood as someone agrees to the content published or what this person has said. For this reason is a good metric to study communities and how the different opinions evolve and the corresponding relationships.

3.2. Network analysis

Despite of counting with 54.000 tweets, when generating the dynamic network with *t-hoarder* we only got **36.312** nodes and **40.840** edges connecting them. At first point, the number of links is very small compared with number of nodes. Nevertheless it makes sense, since the normal scenario is someone retweets a post of a person but this person does not return the retweet. In addition, the people who interact with a given tweet, they do not usually have any relationship between them, so it is likely that there are not mutual retweets. The network is presented in figure 2.

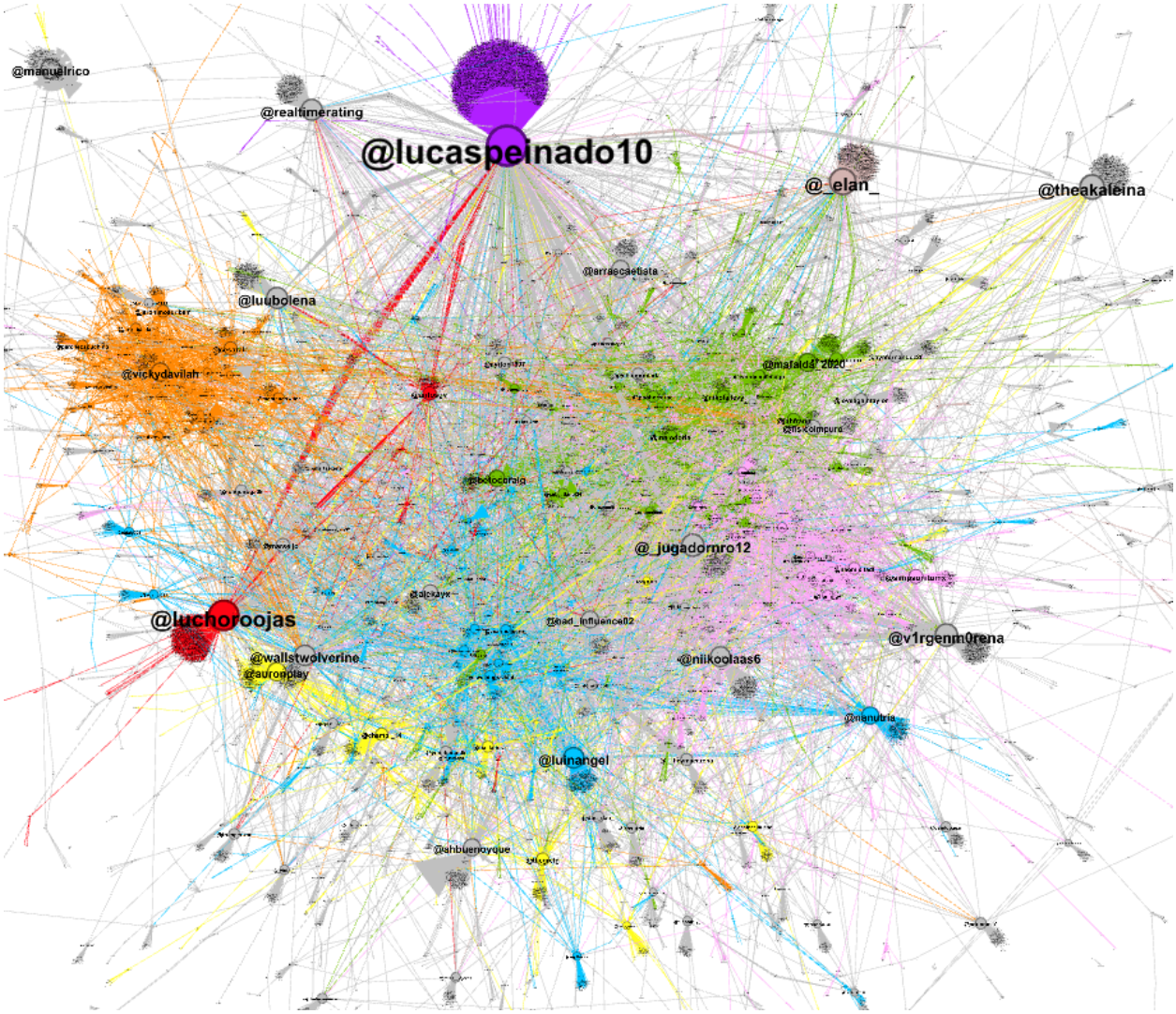


Figure 2: Dynamic network composed by retweets

To obtain this plot several steps have been carried out:

- Space the nodes applying the *ForceAtlas2* algorithm.
- Apply the *Modularity* statistic with resolution of 1,0 to look for possible communities. After that, we applied the corresponding colors to the discovered groups.
- Set the size of each node according to the *In-Degree* attribute. It reflects how much nodes point to a node, so tweets with a large degree means this post has been retweeted for a lot of people. The size was established with the *Spline* option, applying a smooth logarithmic function. This configuration, as well as the minimum and maximum size, were selected because they provide the clearest visualization of the network and the relevant people.
- The same criteria was applied with nodes' labels.

Once the communities are discovered, the best attribute to consider is *In-Degree*, since as explained, it is going to provide the best explanation according to the pattern given by retweets. For example, we checked the betweenness of the nodes, but the graph did not provide any useful information due to the main connection is unidirectional from one node to another.

In the image 2 different communities are shown according to their colors. Some of them are very scattered as the green or the yellow, whereas other as the purple or the red are highly concentrated. Mention the original picture is cut, because around the core there were many dispersed points without any pattern. These correspond with single tweets without a lot of interaction and no interest in the study. So we are going to start analyzing what is happening.

We can clearly see **viral** tweets, that explain why these patterns appear in the plot and the proportion between nodes and edges. These post are very popular in the community due to different reasons; the humor or the controversy caused are the most common ones. Consequently, they start spreading very quickly, what means that a huge amount of people without any interaction between them are going to converge in this node. The clearest examples in this case are the **purple** and **red** communities.

3.2.1. Purple community

As mentioned before, this is a clear example of a viral tweet, a lot of people re-posting it. It is represented as a node with many in-edges from other nodes. The reason it is so popular it is not due to choosing or supporting one of both artist in a clear way, but the tweet makes an humorous reference to one song produced by René Residente. It is about how brave this artist is and recommending to JBalvin to be careful as a joke.

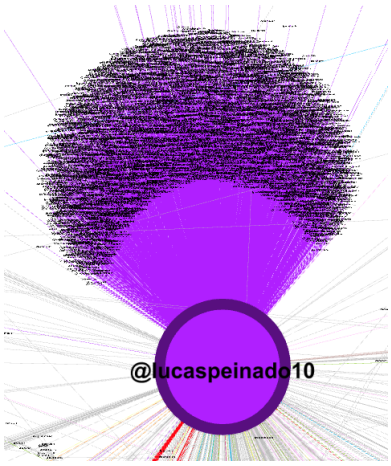


Figure 3: Most viral tweet about a song of René Residente

It received more than 16,000 retweets by the 8th of March, and we can clearly see in the graph how it has been the most influential comment at the moment. In table 4, we can see how the in-degree metric is much larger than the next ones, from **4682** links in the mentioned tweet respect the second one, **1426**.

Label	Id	Incoming edges
@lucaspeinado10	1	4.682
@luchoroojas	2	1.426
@_elan_	3	982
@manuelrico	4	974
@theakaleina	5	829

Table 4: Number of input nodes for the first 5 nodes with higher in-degree

It reflects how this topic has been mainly a motivation for jokes and funny comments in the community.

3.2.2. Red community

Doing a research between the main red nodes, this community groups people who is exposing a critique. But the interesting points is this is not directly related with the problem between both artist, the tweets comment how a lot of people are supporting the song of Residente which last 8 minutes, whereas no one thinks the same when the singer Taylor Swift did something similar.

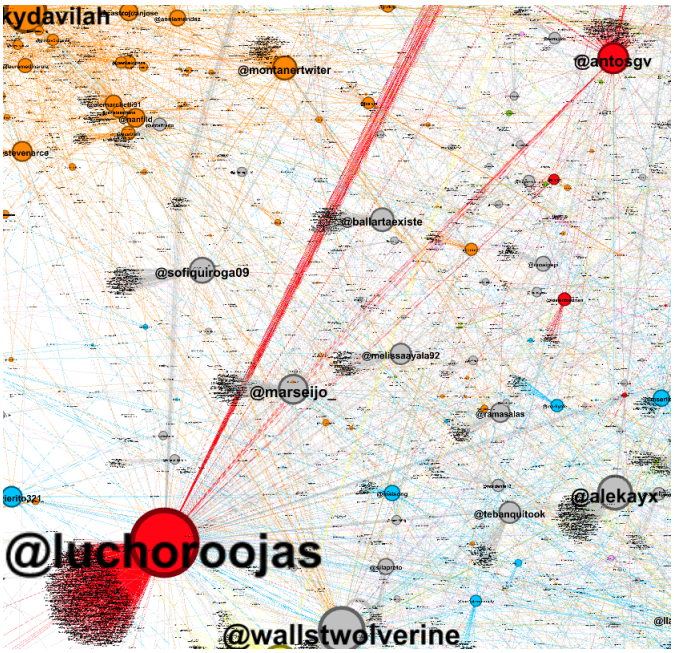


Figure 4: Viral tweets highlighting the different treatment given to Taylor Swift and René Descartes respect to the same situation

As it can be seen, it is community composed mainly by viral tweets, where the biggest node in terms of incoming edges it is the second one in all the network, we can see it in the table 4, with **1426**.

Mention also how a lot of people who retweeted these type of tweets, they also shared the one commented before in the purple section. At the we can see here profiles which expose a negative comment respect other people rather than directly to one of the implied artists and who likes jokes about the discussion.

Apart from these two communities which are in essence popular tweets, most of the collected tweets have positioned respect one artist or the other.

3.2.3. Orange and blue communities

In one hand the **orange** and the **blue** communities criticise the song written by Residente and are in favour of the attacked artist, JBalvin. Nevertheless, there is a factor which is why they are distinguished instead of being grouped together.

The orange corresponds mainly with famous and influential people, most of the profiles are verified and they have a lot of followers. They are journalists, writers or other singers.

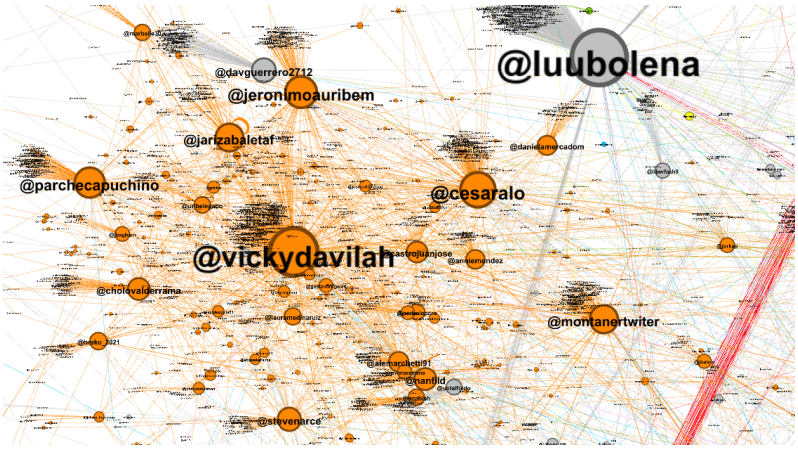


Figure 5: Famous and influential people supporting JBalvin

In general, their opinions are written with education in a formal way. Most of them clearly show their support but they also analyse in detail the situation, highlighting the risks and dangers of this action.

The perspective given by the blue community is completely different. The profiles are not so influential and the number of followers is more reduced.

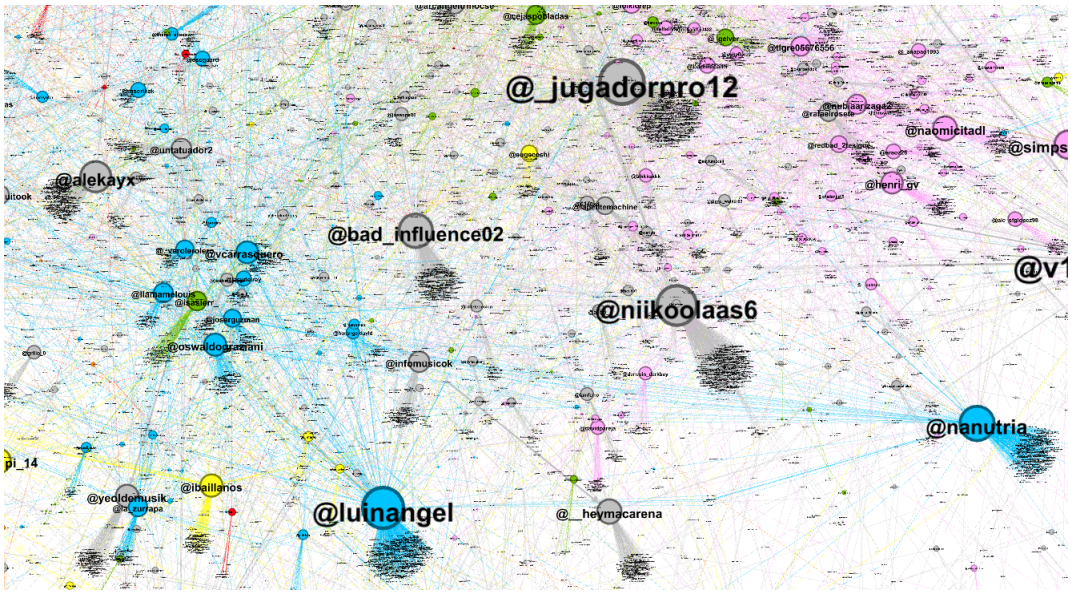


Figure 6: Other people supporting JBalvin

The criticism are more direct and often without any constructive thoughts. In some cases, the tweets are even composed by derogatory comments or insults to Residente. Others are just jokes or funny comments which try to mock the singer.

3.2.4. Green and pink communities

In the other hand we have the **green** and the **pink** communities who support the song. Tweets from both groups are in favour of the criticizes to JBalvin and highlight the great performance from Residente. We can find tweets from jokes trying to be viral to serious comments. The difference here is not given due to the gathered profiles, but for a geographical reason.

The green community encloses as general people from Colombia. As it is normal, there are going to be profiles from other countries, but not most of them.

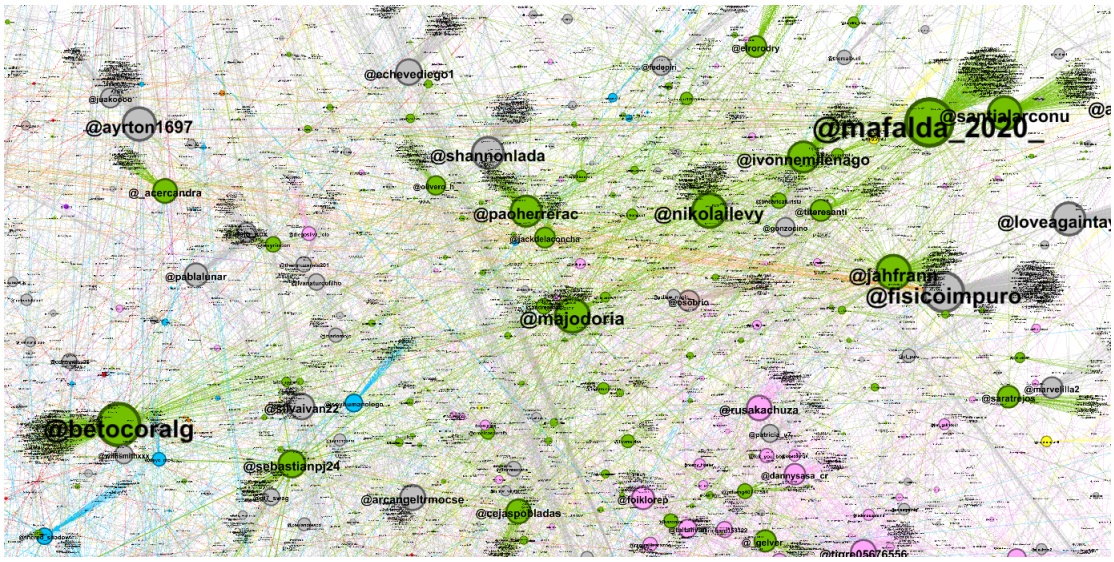


Figure 7: Tweets from Colombia supporting Residente

In the pink group we can see a wide range of countries, as Mexico, Argentina, etc. Mention the South American countries have an important presence in the network, as it is normal because both artists are from there and they are very listened by their populations.

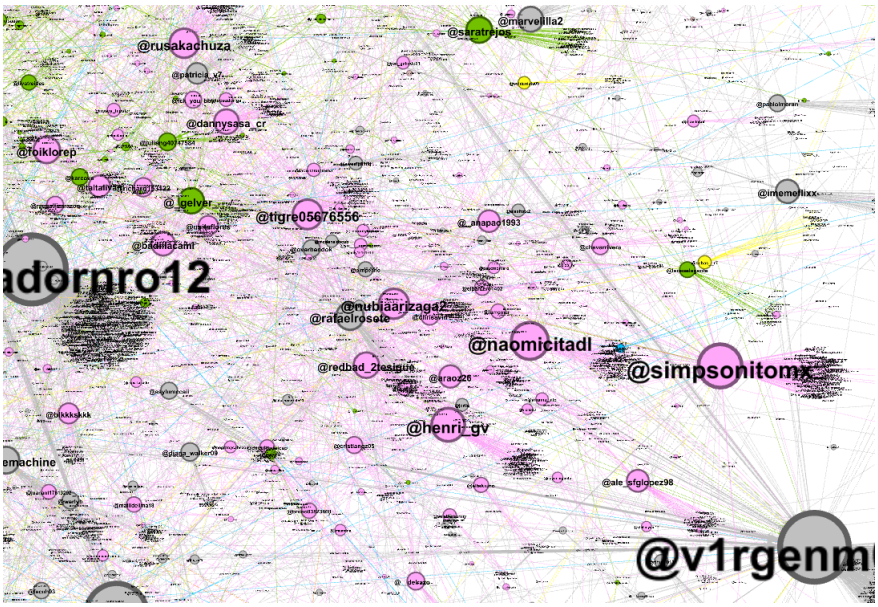


Figure 8: Other tweets supporting Residente

An interesting point to comment is how both communities are closer and more connected between them respect the opposite position represented by the orange and the blue. It is normal because the profiles are similar, whereas in the other case the positions of the people were very different.

3.2.5. Yellow community

Finally, the last community detected is really interested how it is separated from the rest. As we can see, it is mainly composed by only a few of viral tweets which have a lot of impact. If we inspect into the names, we can find how they correspond with very famous streamers or youtubers.

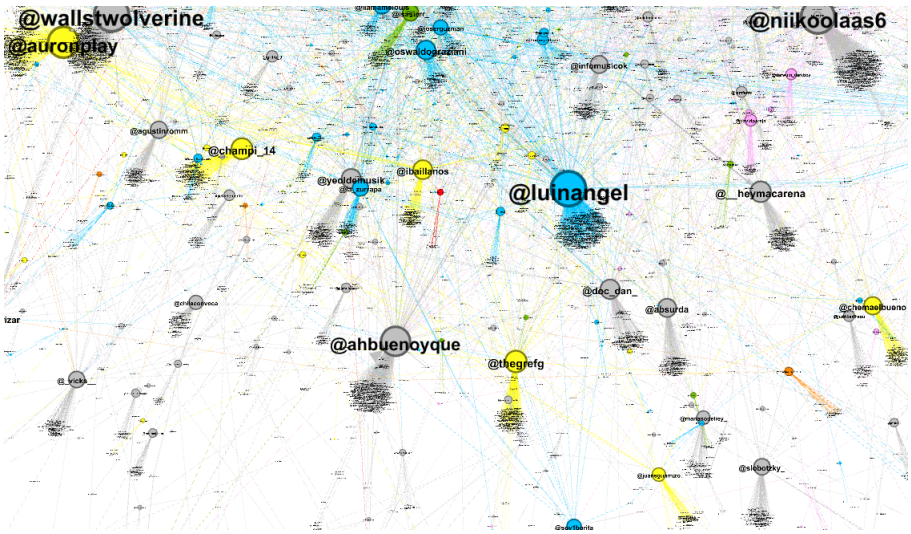


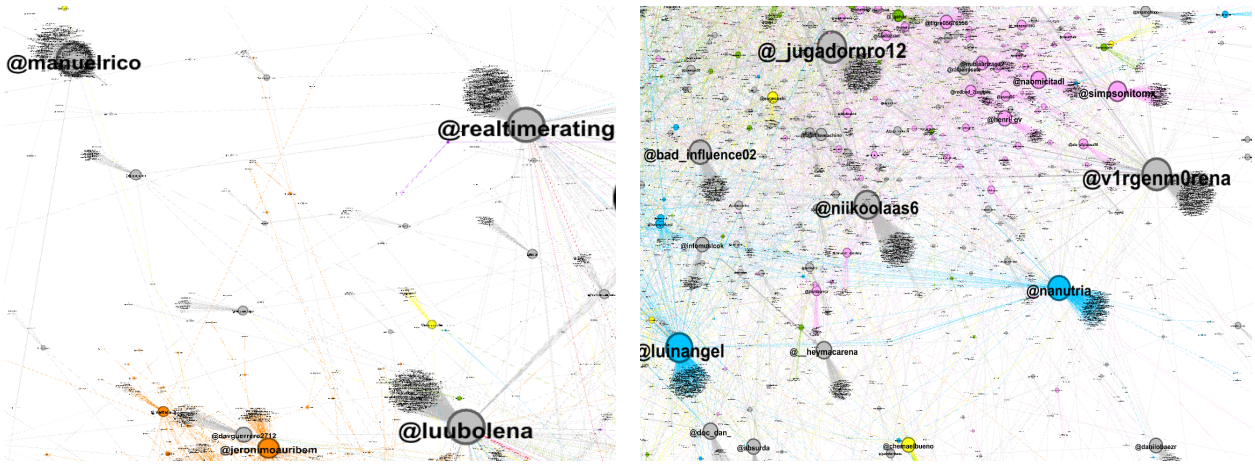
Figure 9: Tweets from streamers

For this reason, this community is mainly gathering the opinions from famous roles in the streaming and multimedia world. Most of them are jokes about how bad looked JBalvin due to the song.

3.2.6. Rest of nodes

The nodes which are not included in any community present a grey color. If we look into the most relevant ones, we can see how they are mainly jokes and tweets written in order to get a high number of interactions.

They are not positioned for any artist, but we can see how there are many with a considerable number of incoming edges. This confirms again the theory that the most relevant and shared tweets have been the ones which provide a touch of humour to the situation, rather than positioning and generating argument. Some examples are plotted in figures 10a and 10b.



(a) Upper left in the network

(b) Below right in the network

Figure 10: Examples of viral tweets not grouped in any community

4. Conclusion

Thanks to social networks, we live in a connected world where we can know realities and opinions from people all over the world in a instant. This simplicity and speed to spread information through the internet makes crucial to have tools to capture all this data and process in the correct way in order to have an insight of what is going on.

In the case of this study, we explored the impact provoked by a song released by an immensely famous rapper together with probably the most influential producer nowadays. It has been thanks to means like *t-hoarder* and *Gephi* that we have been able to capture a huge quantity of information, discover how it is related just by looking at their retweets and process it to discover influential messages, communities, points of view and how supported an opinion is.

A topic like this may seem trivial to be analyzed, but, in a world governed by metrics, it may throw some useful insights for example to companies interested in working with one of the artists implied in order to predict the impact of any kind of advertising campaign or musical collaboration.