Spotify Network Analysis

Carlos Cañellas, Martina Armas, Rodrigo Camarena, Nicole Vega

Social Network Analysis

October, 2020

INDEX

Objective:

We find quite interesting how Spotify decides which artist recommends us when we want to listen to similar music and we believe that analyzing networks can be helpful for the digital music industry to better understand today's trends and people's preferences. Therefore we decided to see what urban music artists are related to our favorite reggaeton singer Bad Bunny on Martina's Spotify account. We used only one account because Spotify gives related artists according to the account information.

Even though this is an egocentric network that is built from one starting actor (Bad Bunny), it will allow us to observe also its related artists as well as the most relevant artists of similar genres (reggaeton, urban music, trap, latin american, etc), and relations among them to see which one is the most popular and has more followers as well as which one is the most recommended in the network.

Steps Followed:

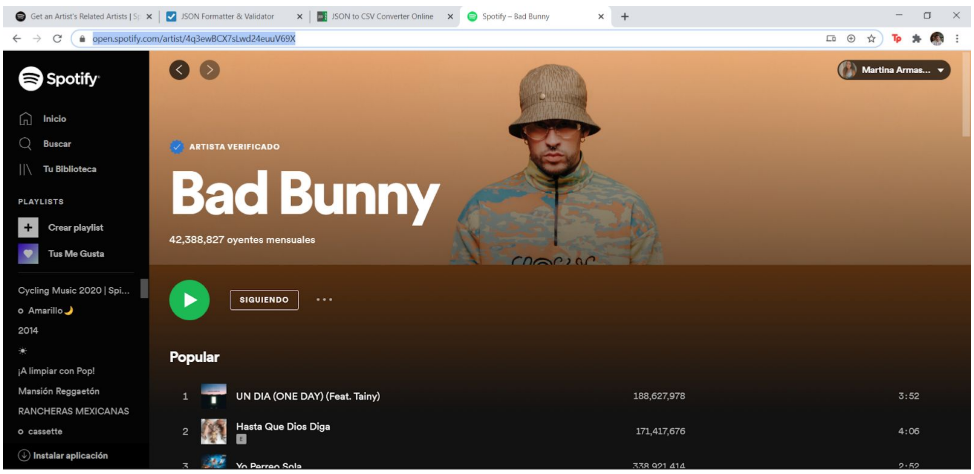
1. Interpreting the phenomenon under investigation as a network

Spotify presents users a list of maximum 20 artists related to the visited artist. In the Spotify Community pages it is explained that the related artists are determined by algorithms that explore what people listen to along the artist music and also music discussions and online trends. Therefore we decided that for this social network analysis we would use artists as nodes that relate to each other through edges to their 20 related artists.

2. Collection, cleaning and refining of the data

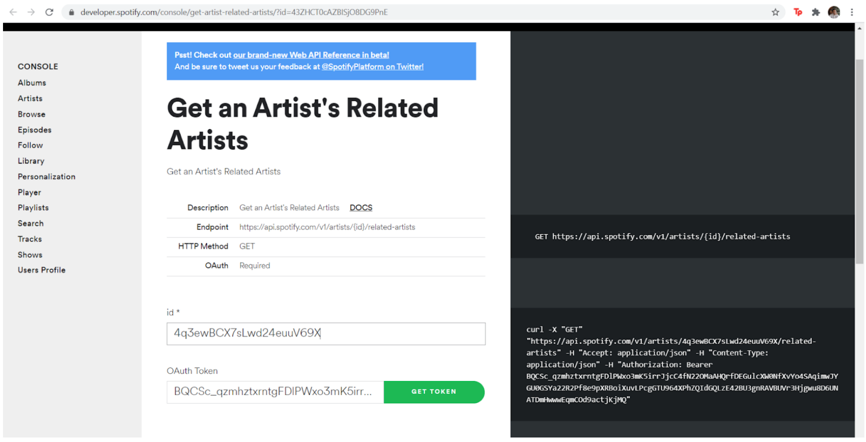
Our main source of information is Spotify. As we collected our data we cleaned it and refined it as it had a lot of information that wasn't very useful for our analysis. Below is our entire process.

1. Find artist ID (https://open.spotify.com/artist/4q3ewBCX7sLwd24euuV69X)



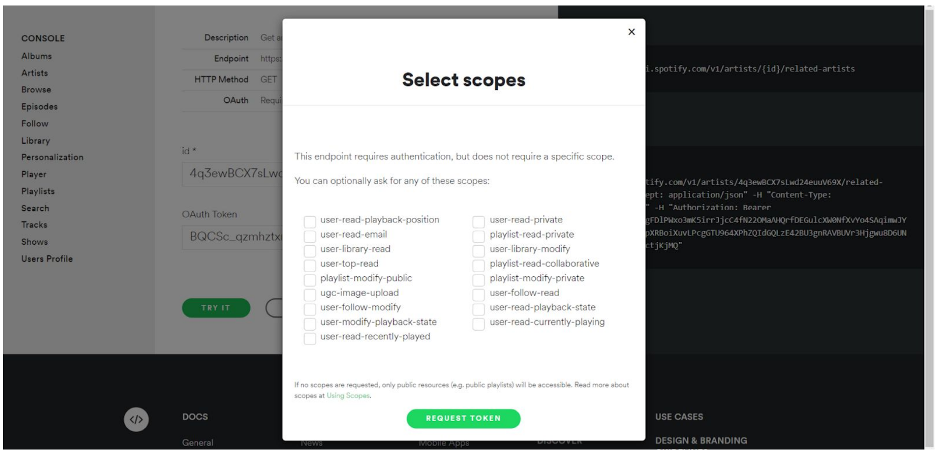
In order to begin our analysis, we first need to choose an artist from which the related artists' database will revolve around. As explained previously, we are not only going to focus on a single artist but rather in all of the results. Nevertheless, as it is a requirement from spotify developers' webpage, we must choose one. In this case, we have chosen Bad Bunny due to his relevant influence on music and how many collaborations he has done during the last years. In order to collect this token ID, we went to spotify web app and searched for the Bad Bunny profile. In the profile link we can see the artists' token ID. In this case, as we can see highlighted in the image, Bad Bunny's token ID is: **4q3ewBCX7sLwd24euuV69X**.

2. Paste ID in Spotify for developers



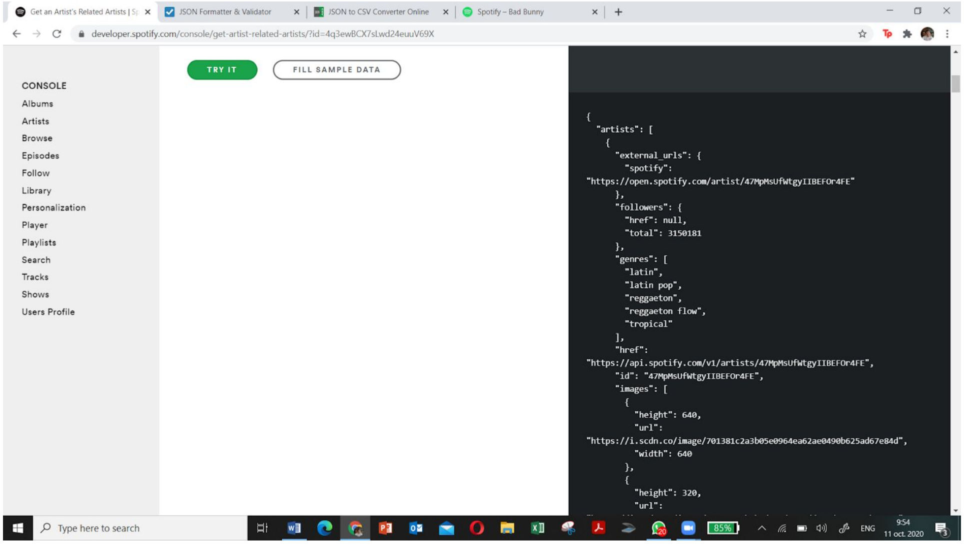
As explained in step 1, once we have Bad Bunny's token ID. We added it to the form in order to get all the artists related to him.

3. Get token (introducing Spotify account)



In this step, before getting all the results. We first need to authenticate we are Spotify's users, by clicking on the REQUEST TOKEN button we will be automatically redirected to the authentication link.

4. Click on “TRY IT”

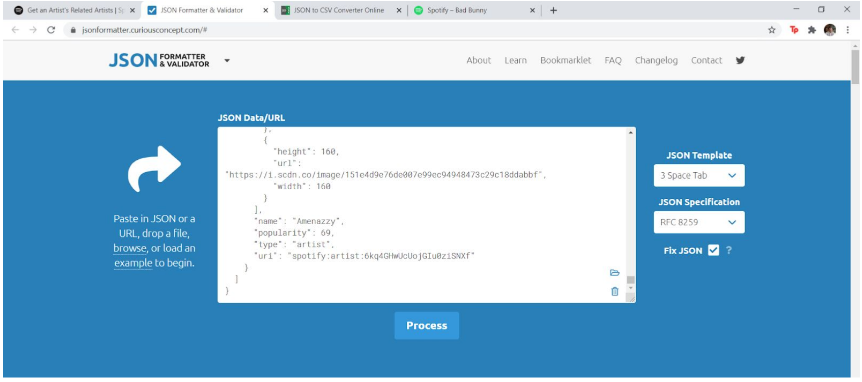


Once introduced to the form both tokens (artists and users' tokens), we click on the FILL SAMPLE DATA button and automatically results will be displayed. As you can see on the right side of the image.

5. Copy the entire JSON file

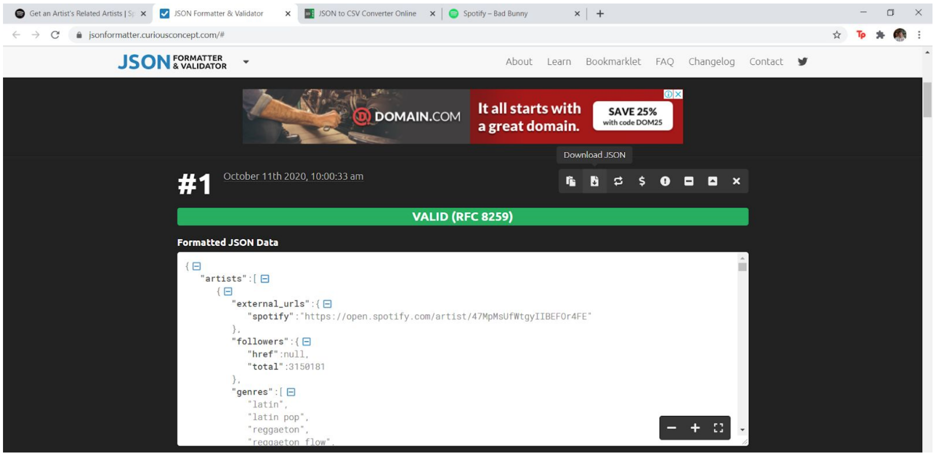
Nevertheless, as we have explained beforehand. It is important to organize and clean our results. In this step we proceed to copy the results.

6. Paste in JSON formatter and validator, click on “Process”



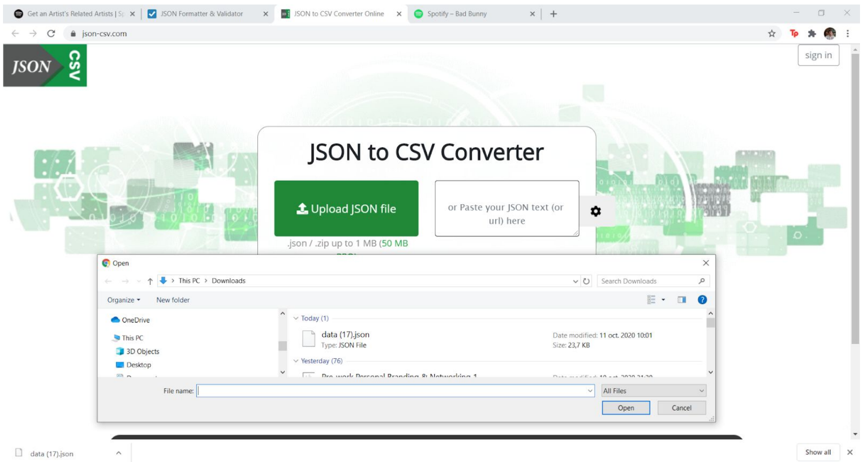
In this step, we paste the results we collected from the Spotify webpage and paste it to an online JSON formatter and validator site.

7. Click on download JSON



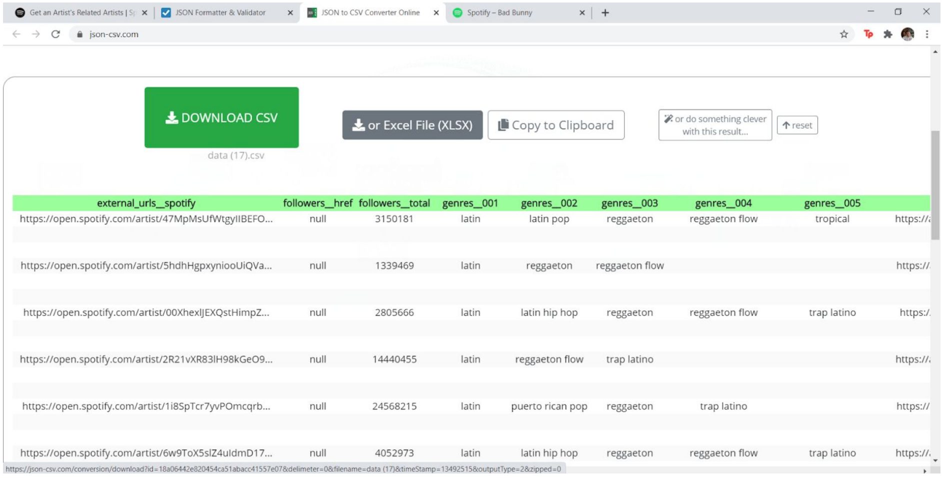
Now, we just download the organized data as a JSON file to our computer.

8. Upload the downloaded file of json-csv.com

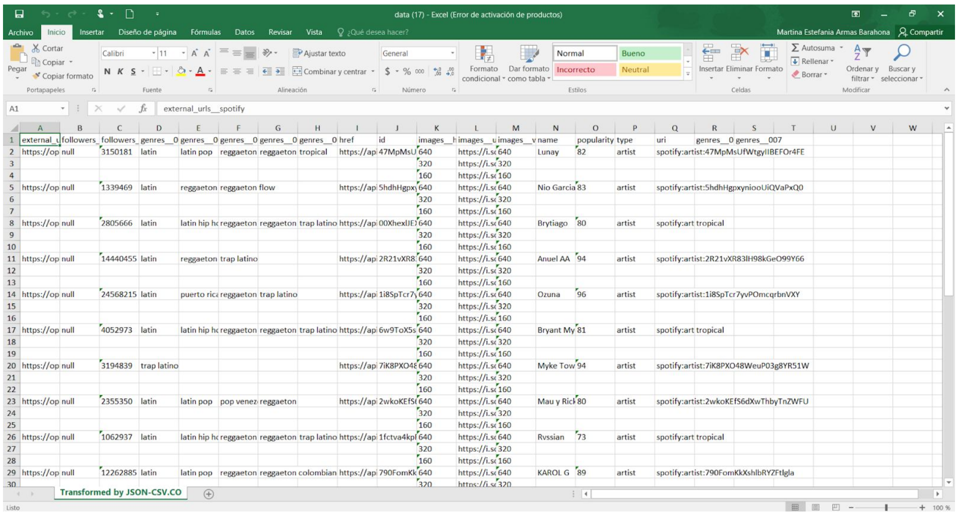


This is a really important step. As we know, gephi does not accept JSON files. That is why we proceeded to convert it to a CSV/EXCEL.

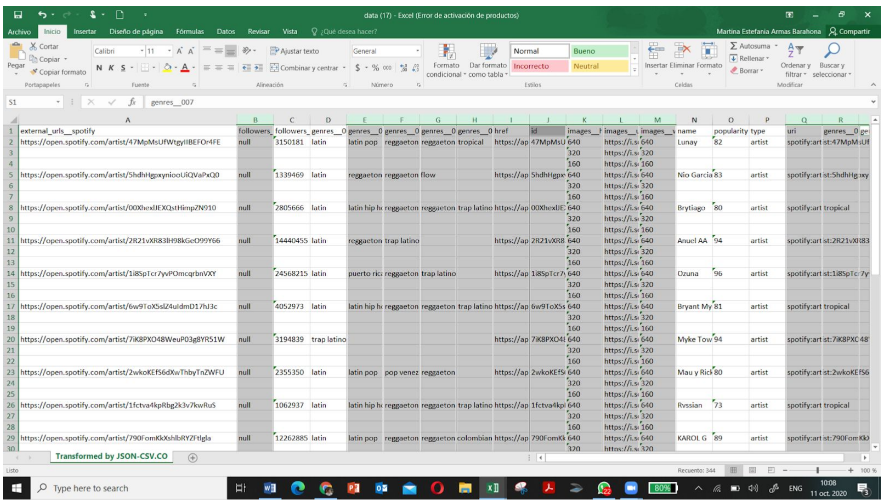
9. Download as excel file



10. Import file to Excel.

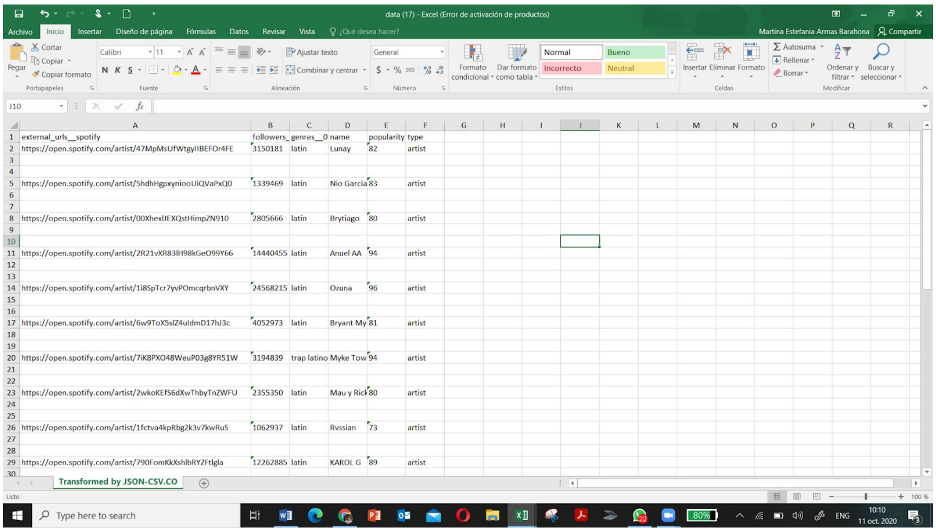


11. Clean data set



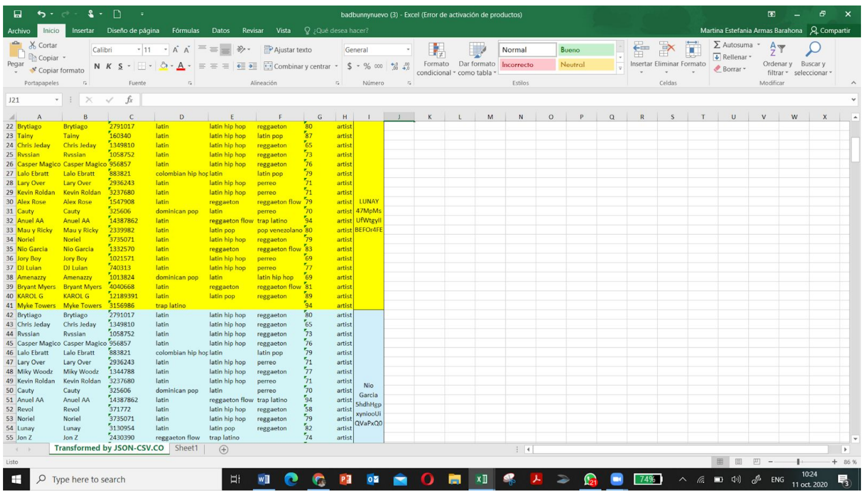
There are plenty of empty rows or with invalid data that must be trimmed out. In this and the next step, is where we are going to be cleaning and organizing the data.

12. Organize the data set

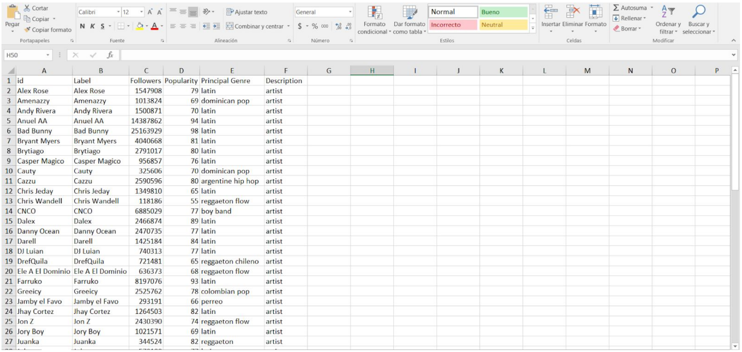


	id	Label	Followers	Popularity	Principal Genre	Description
1	id	Label	Followers	Popularity	Principal Genre	Description
2	Lunay	Lunay	3130954	82	latin	artist
3	Nio Garcia	Nio Garcia	1332570	83	latin	artist
4	Brytiago	Brytiago	2791017	80	latin	artist
5	Anuel AA	Anuel AA	14387862	94	latin	artist
6	Ozuna	Ozuna	24515042	96	latin	artist
7	Bryant Myers	Bryant Myers	4040668	81	latin	artist
8	Myke Towers	Myke Towers	3156986	94	trap latino	artist
9	Mau y Ricky	Mau y Ricky	2339982	80	latin	artist
10	Rvssian	Rvssian	1058752	73	latin	artist
11	KAROL G	KAROL G	12189391	89	latin	artist
12	Miky Woodz	Miky Woodz	1344788	77	latin	artist
13	Rauw Alejandro	Rauw Alejandro	2262499	92	latin	artist
14	Lary Over	Lary Over	2936243	71	latin	artist
15	Casper Magico	Casper Magico	956857	76	latin	artist
16	Jon Z	Jon Z	2430390	74	reggaeton flow	artist
17	Noriel	Noriel	3735071	79	latin	artist
18	Jhay Cortez	Jhay Cortez	1264503	82	latin	artist
19	Alex Rose	Alex Rose	1547908	79	latin	artist
20	Kevin Roldan	Kevin Roldan	3237680	71	latin	artist
21	Amenazzy	Amenazzy	1013824	69	dominican pop	artist

13. Repeat the same process with every artist related to Bad Bunny

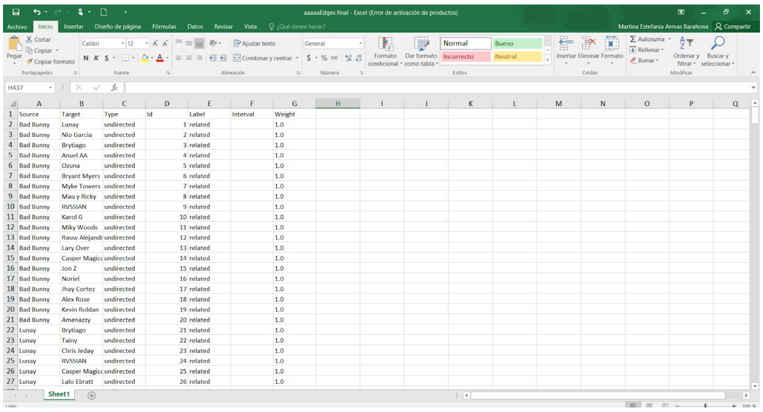


14. Create nodes table with all the names of the artists



id	Label	Followers	Popularity	Principal Genre	Description
Alex Rose	Alex Rose	1547908	79	latin	artist
Amenazzy	Amenazzy	1013824	69	dominican pop	artist
Andy Rivera	Andy Rivera	1500871	70	latin	artist
Anuel AA	Anuel AA	14387862	94	latin	artist
Bad Bunny	Bad Bunny	25163929	98	latin	artist
Bryant Myers	Bryant Myers	4040668	81	latin	artist
Brytiago	Brytiago	2791017	80	latin	artist
Casper Magico	Casper Magico	956857	76	latin	artist
Cauty	Cauty	325606	70	dominican pop	artist
Cazzu	Cazzu	2590596	80	argentine hip hop	artist
Chris Jeday	Chris Jeday	1349810	65	latin	artist
Chris Wandell	Chris Wandell	118186	55	reggaeton flow	artist
CNCO	CNCO	6885029	77	boy band	artist
Dalex	Dalex	2466874	89	latin	artist
Danny Ocean	Danny Ocean	2470735	77	latin	artist
Darell	Darell	1425184	84	latin	artist
DJ Luian	DJ Luian	740313	77	latin	artist
DrefQuila	DrefQuila	721481	65	reggaeton chileno	artist
Ele A El Dominio	Ele A El Dominio	636373	68	reggaeton flow	artist
Farruko	Farruko	8197076	93	latin	artist
Greeicy	Greeicy	2525762	78	colombian pop	artist
Jamby el Favo	Jamby el Favo	293191	66	perreo	artist
Jhay Cortez	Jhay Cortez	1264503	82	latin	artist
Jon Z	Jon Z	2430390	74	reggaeton flow	artist
Jory Boy	Jory Boy	1021571	69	latin	artist
Juanka	Juanka	344524	82	reggaeton	artist

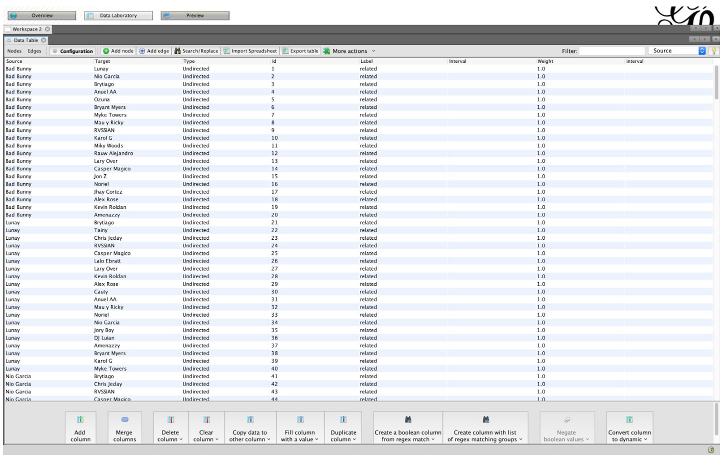
15. Create Edges table



Source	Target	Type	Id	Label	Interval	Weight
Bad Bunny	Lunay	undirected	1	related		1.0
Bad Bunny	Nio Garcia	undirected	2	related		1.0
Bad Bunny	Brytiago	undirected	3	related		1.0
Bad Bunny	Anuel AA	undirected	4	related		1.0
Bad Bunny	Ozuna	undirected	5	related		1.0
Bad Bunny	Bryant Myers	undirected	6	related		1.0
Bad Bunny	Myke Towers	undirected	7	related		1.0
Bad Bunny	Mau y Ricky	undirected	8	related		1.0
Bad Bunny	RVSSIAN	undirected	9	related		1.0
Bad Bunny	Karol G	undirected	10	related		1.0
Bad Bunny	Miky Woods	undirected	11	related		1.0
Bad Bunny	Rauw Alejandr	undirected	12	related		1.0
Bad Bunny	Lary Over	undirected	13	related		1.0
Bad Bunny	Casper Magicc	undirected	14	related		1.0
Bad Bunny	Jon Z	undirected	15	related		1.0
Bad Bunny	Noriel	undirected	16	related		1.0
Bad Bunny	Jhay Cortez	undirected	17	related		1.0
Bad Bunny	Alex Rose	undirected	18	related		1.0
Bad Bunny	Kevin Roldan	undirected	19	related		1.0
Bad Bunny	Amenazzy	undirected	20	related		1.0
Lunay	Brytiago	undirected	21	related		1.0
Lunay	Tainy	undirected	22	related		1.0
Lunay	Chris Jeday	undirected	23	related		1.0
Lunay	RVSSIAN	undirected	24	related		1.0
Lunay	Casper Magicc	undirected	25	related		1.0
Lunay	Lalo Ebratt	undirected	26	related		1.0

Taking the artist searched as the source and the given related artist as the target, all are undirected, we assign an id to every relationship and a weight of 1.0

16. Add nodes and edges spreadsheet to a Gephi workspace and start the analysis.



Source	Target	Type	Id	Label	Interval	Weight
Bad Bunny	Lunay	Undirected	1	related		1.0
Bad Bunny	Nio Garcia	Undirected	2	related		1.0
Bad Bunny	Brytiago	Undirected	3	related		1.0
Bad Bunny	Anuel AA	Undirected	4	related		1.0
Bad Bunny	Ozuna	Undirected	5	related		1.0
Bad Bunny	Bryant Myers	Undirected	6	related		1.0
Bad Bunny	Myke Towers	Undirected	7	related		1.0
Bad Bunny	Mau y Ricky	Undirected	8	related		1.0
Bad Bunny	RVSSIAN	Undirected	9	related		1.0
Bad Bunny	Karol G	Undirected	10	related		1.0
Bad Bunny	Miky Woods	Undirected	11	related		1.0
Bad Bunny	Rauw Alejandro	Undirected	12	related		1.0
Bad Bunny	Lary Over	Undirected	13	related		1.0
Bad Bunny	Casper Magico	Undirected	14	related		1.0
Bad Bunny	Jon Z	Undirected	15	related		1.0
Bad Bunny	Noriel	Undirected	16	related		1.0
Bad Bunny	Jhay Cortez	Undirected	17	related		1.0
Bad Bunny	Alex Rose	Undirected	18	related		1.0
Bad Bunny	Kevin Roldan	Undirected	19	related		1.0
Bad Bunny	Amenazzy	Undirected	20	related		1.0
Lunay	Brytiago	Undirected	21	related		1.0
Lunay	Tainy	Undirected	22	related		1.0
Lunay	Chris Jeday	Undirected	23	related		1.0
Lunay	RVSSIAN	Undirected	24	related		1.0
Lunay	Casper Magico	Undirected	25	related		1.0
Lunay	Lalo Ebratt	Undirected	26	related		1.0
Lunay	Lary Over	Undirected	27	related		1.0
Lunay	Kevin Roldan	Undirected	28	related		1.0
Lunay	Alex Rose	Undirected	29	related		1.0
Lunay	Cauty	Undirected	30	related		1.0
Lunay	Anuel AA	Undirected	31	related		1.0
Lunay	Mau y Ricky	Undirected	32	related		1.0
Lunay	Noriel	Undirected	33	related		1.0
Lunay	Nio Garcia	Undirected	34	related		1.0
Lunay	Jory Boy	Undirected	35	related		1.0
Lunay	DJ Luian	Undirected	36	related		1.0
Lunay	Amenazzy	Undirected	37	related		1.0
Lunay	Bryant Myers	Undirected	38	related		1.0
Lunay	Karol G	Undirected	39	related		1.0
Lunay	Myke Towers	Undirected	40	related		1.0
Nio Garcia	Brytiago	Undirected	41	related		1.0
Nio Garcia	Chris Jeday	Undirected	42	related		1.0
Nio Garcia	RVSSIAN	Undirected	43	related		1.0
Nio Garcia	Casper Magico	Undirected	44	related		1.0

3. Network layout and obtaining SNA measures

The network is built from one starting actor (Bad Bunny), hence it is an egocentric network. Our undirected network has 61 nodes and 420 edges, every node is an artist on Spotify and the edges link them.

We decided to make every relationship undirected because the information given by the API only gives us information on Martina's interactions and data. As interacting more with some artists may change the results, we are assuming that every actor is connected to the others within the network, directly or through others.

After having created the Gephi graph, we decided that in order to see better the relationship between the nodes it would be useful to color the nodes by degree and make the size according to the number of followers. Moreover, we decided to use Force Atlas 2 because of the formula of repulsion and attraction it uses.

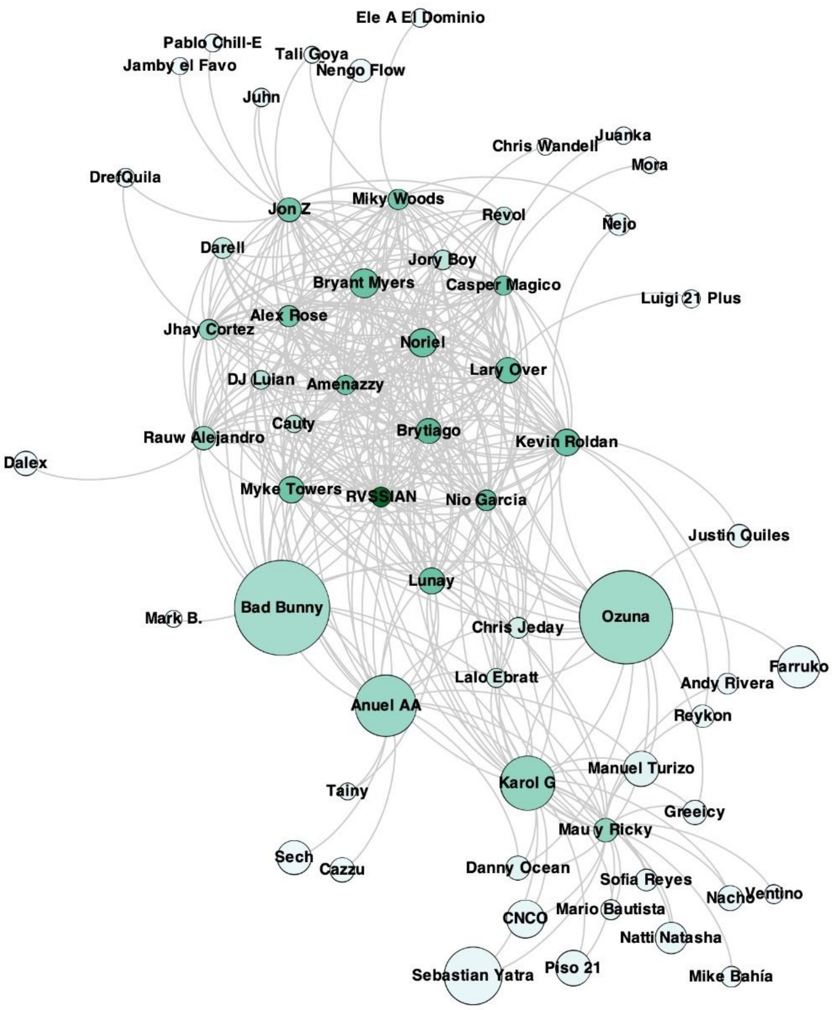
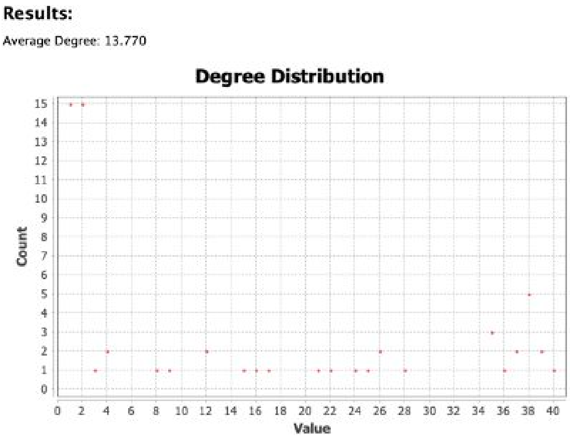
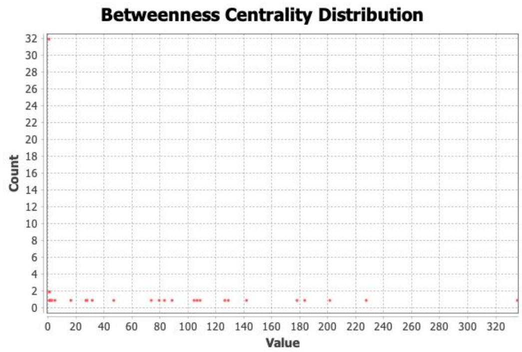


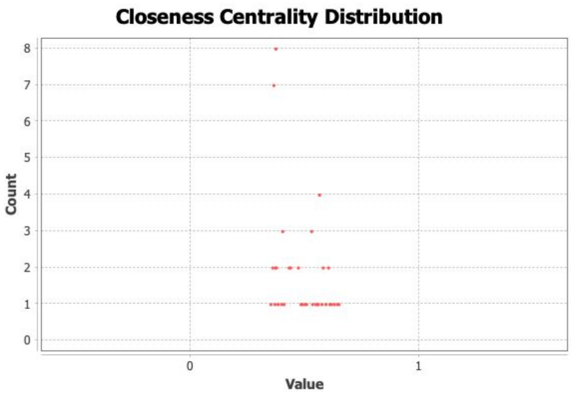
Figure 1: *Network render of Bad Bunny related artists in Spotify. The nodes represent the artists and the edges between them represent they are a related artist . The size of nodes represent the number of followers they have on Spotify and the color, the degree.*

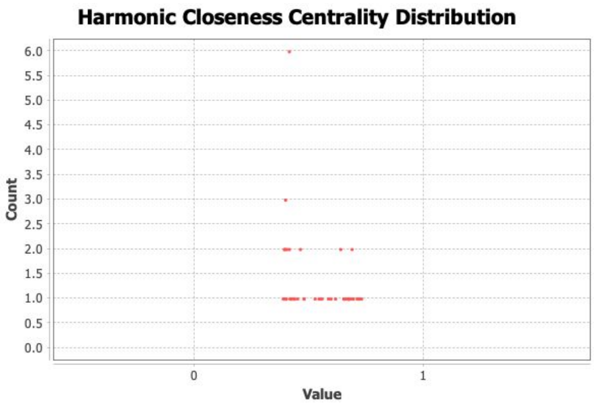
SNA Measures

1. **Network Density**= 0.23.
2. **Average distance**= 2.26
    a. **Diameter:** 4
3. **Average Degree:** 13.77

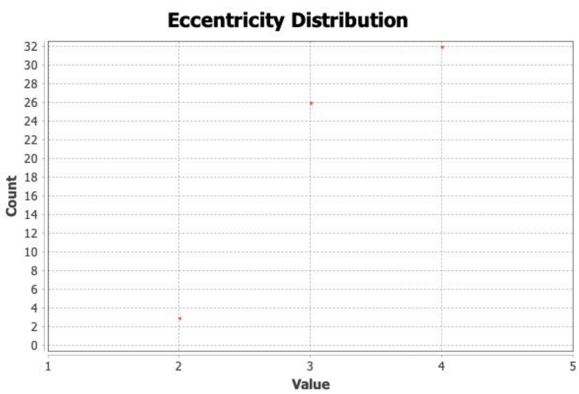


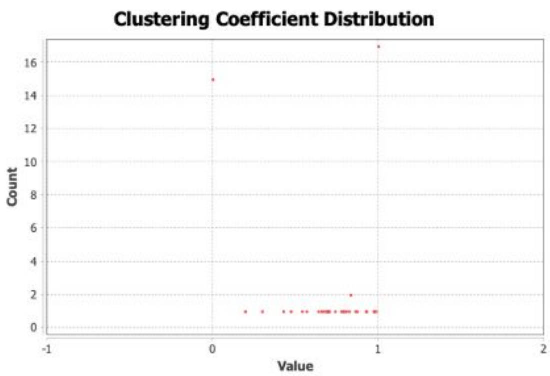




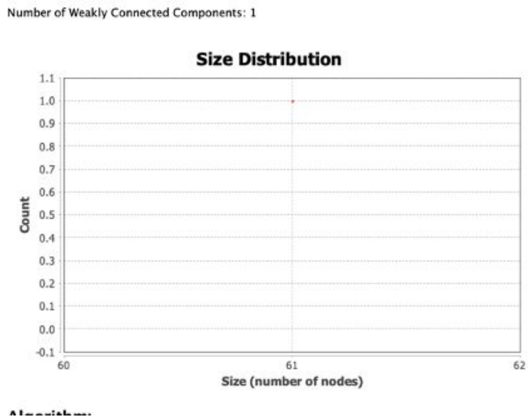


4. **Average Clustering coefficient:** 0,823





5. **Number of components:** 1



4. Analysis and drawing conclusions based on the graph and measures

This graph (figure 1) shows some basic information at first glance. We can observe who the artists in the network are and who they are connected to. Moreover, node's size shows the artists that have the highest number of followers, which artists are more central and which ones are located in the outer edges. Furthermore, we can see a cluster of artists including Bryant Myers, Noriel, Cauty, Amenazzy, Alex Rose, and others in the upper middle of the graph, which means that they share a lot of relationships between themselves. It is clear that Bad Bunny and Ozuna are the artists with most followers from the music genres presented, meaning that they have a big audience and may not need as much reference as others.

We can also observe that there are many artists that are related to just one. We got a network density of 0.23 which means that we could have more relationships within our network. This is an opportunity to keep exploring, obtaining data from more than the main artists we explored. As we are talking about a very similar genre (reggaeton) there is a high probability that the network density would grow.

When building an egocentric network, the connections usually don't go beyond the first order zone, meaning that there is always one step from the ego (Bad Bunny) to any other node. All geodesic distances from the ego to the other nodes would then be 1 by definition. The case with this Spotify's network, however, differs for one, the network expands beyond first-order connection, as it incorporates to some extent artists that are also related to Bad Bunny related artists. Hence, for the average distance in this SNA we got 2.26. The diameter is the length of the longest minimum network path and in this SNA is 4, which indicates that our network isn't very wide.

The nodes have an average of 13.77 connections as the statistic showed. We expected this number to be high considering that we explored data of 21 reggaeton singers related artists but we didn't consider that there could be artists in the outer edges of the graph. We could increase this by analysing more artists seen that have fewer connections as Mike Bahia, Dalex, Farruko and others. The darker the node, the higher the degree. From this graph we can see that "RVSSIAN" is the most recommended artist despite not having as many followers as others. After double checking on the edges table, RVSSIAN was referred by 19 other artists.

The clustering coefficient is 0.823 which means that there is a lot of interaction between the different groups of nodes and that there is a high probability that if we choose two random nodes they will be connected to each other. We can confirm this seeing the amount of clustering between the edges.

The network is made up of one component, meaning that every single node is connected to another directly or through others. We can clearly see that in the render or our network where there are no nodes that are outside or not connected to the network.

After doing this SNA, we realized that this network can be useful for music producers and artists representatives to create new content with its related singers. It could also be a performance indicator to see how much they are being referred and how many followers they have compared to other artists.

References:

Spotify for Developers (no date) Spotify, Retrieved from: https://developer.spotify.com/technologies/metadata-api/search/

Spotify Community (no date) "Related artists" – how do Spotify decide what artists are related to a specific artist?" Spotify. Retrieved from: https://community.spotify.com/t5/Content-Questions/quot-Related-artists-quot-how-do-Spotify-decide-what-artists-are/td-p/1185597.

Marsden, P.V. (2002), "Egocentric and Sociocentric Measures of Network Centrality", Social Networks, vol. 24, no. 4, pp. 407–422, doi:10.1016/S0378-8733(02)00016-3.

Spotify For Artists (nd) "How does Fans Also Like work?" Spotify, online. Retrieved from :https://artists.spotify.com/faq/profile#can-i-create-and-edit-my-artist-playlists-using-spotify-for- artists

Jacomy. M, Venturini, T. Heymann, S et at (June 10, 2014) "ForceAtlas2, a Continuous Graph Layout Algorithm for Handy Network Visualization Designed for the Gephi Software" Retrieved from: https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0098679

BookDown. (nd). “Introduction to Ego-Centric Networks” Retrieved from: https://bookdown.org/chen/snaEd/ego-intro.html