

Group members:

Edoardo Bastianello: 2053077

Stefano Binotto: 2052421

Gionata Grotto: 2052418

Finding Significant Statistics And Relationships In Music Networks With Node Features

IMPORTANT NOTE: This PDF exceeds the max limit of six pages because we included several tables containing part of the outputs of our analysis. The code with the entire outputs is included in the Drive folder we linked at the bottom of this report, as well as the PDF with the extended results (top 10 higher scores for each test).

INTRO AND MOTIVATIONS:

In this project, we created and analyzed three different classes of music related networks:

- Artist Collaboration Networks;
- Genre Collaboration Networks;
- Hit Songs Network.

Each network was created from a dataset containing data of 896 music genres, 13880 songs and 3611 artists.

The motivations of this project are the following:

- Figuring out which artists are most relevant, based on year and geographic area;
- Figuring out which music genres are most relevant, based on year and geographic area;
- Figuring out which songs are most relevant, based on music genre, year, geographic area and involved artists;
- Comparing the results between the exact closeness centrality algorithm and the approximated Eppstein-Wang algorithm.

DESCRIPTION OF METHODOLOGY:

For the “Artist Collaboration Network” dataset, the “Genre Collaboration Network” dataset and the “Hit Songs” dataset we read the significant characteristics from several countries and created their respective graphs.

For each graph, we computed the exact methods provided by NetworkX to obtain:

- the degree centrality scores;
- the betweenness centrality scores;
- the closeness centrality scores.

Finally, we confronted the Eppstein-Wang approximated centrality scores with the related NetworkX exact closeness centrality scores on a given graph.

ARTIST COLLABORATION NETWORKS

DESCRIPTION OF THE NETWORKS CREATED FROM THE DATASET:

These are networks of collaborating artists obtained from Spotify Charts for each considered market and year. Each node of the network is an artist. Different nodes/artists are linked when they collaborated on a hit song (i.e. this song appears on a Spotify Top 200 Chart). These networks are subdivided by geographic areas (US, Japan, France, and other five areas) for the year 2019. We also created three different Global networks containing the data of all the countries, one for each of the following years: 2017, 2018, 2019.

The number of nodes for each of these networks is up to 600, while the number of edges is up to 1000.

These networks are undirected, and each edge has a weight corresponding to the number of hit songs in which both artists collaborated.

IMPORTANT NOTE ON THE FOLLOWING RESULTS: in this report we reported only a part of the all the tables and we reported only 4 entries/outputs for most of them; the extended results with 10 entries/outputs for each table are available in the separate PDF file, whose link is at the bottom of this report.

RESULTS FOR EACH COUNTRY (2019):

1)Australia (check the extended results)

2)Brazil

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
Anitta : 0.1085271317829	Anitta : 0.3368435613114	Anitta : 0.2461349308080
Pineapple StormTv: 0.09043927648578	Vitão : 0.08285102615459	Vitão : 0.21326896051498
Salve Malak: 0.077519379	Wesley Safadão: 0.073099	Luísa Sonza: 0.206818481
MC Kevin o Chris : 0.04909	Cardi B : 0.070659323065	Wesley Safadão: 0.206428

3)Canada (check the extended results)

4)Germany

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
Capital Bra : 0.048736462093862815	Capital Bra : 0.09002876778630674	Lil Baby : 0.1639984773123727
Ed Sheeran : 0.04151624548736462	6ix9ine : 0.08374040973321178	Veysel : 0.1611049355954213
Post Malone : 0.036101083032490974	Nicki Minaj : 0.0822980908352324	6ix9ine : 0.1608303249097473
David Guetta : 0.03429602888086643	A Boogie Wit da Hoodie : 0.07809155844648631	Ufo361 : 0.1603747149241673

5)France (check the extended results)

6)United Kingdom (check the extended results)

7)Japan (check the extended results)

8)United States of America (check the extended results)

GLOBAL RESULTS FOR EACH YEAR:

9)Global 2017

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
Drake : 0.051470588235294115	David Guetta : 0.07657473761188938	Major Lazer : 0.1979494816427058
J Balvin : 0.051470588235294115	J Balvin : 0.06953189097114873	Camila Cabello : 0.19679972666704632
Calvin Harris : 0.049019607843137254	Bebe Rexha : 0.06544017697225891	Nicki Minaj : 0.19660939811127548
Travis Scott : 0.04656862745098039	Major Lazer : 0.06491226898625103	Young Thug : 0.19604061489591015
Future : 0.04656862745098039	Camila Cabello : 0.0603324024685393	Travis Scott : 0.19547511312217197
Quavo : 0.04166666666666664	Future : 0.059155379756213745	Future : 0.18964003511852504
Nicki Minaj : 0.03676470588235294	Daddy Yankee : 0.05493914645503251	Drake : 0.1870231073110017
Daddy Yankee : 0.03676470588235294	Justin Bieber : 0.051743817735189726	Ty Dolla \$ign : 0.18347844552983647
Logic : 0.03676470588235294	French Montana : 0.04920889652053614	Bebe Rexha : 0.1824902312810223
Lil Wayne : 0.03431372549019608	Nicki Minaj : 0.04333985331682205	Ed Sheeran : 0.18200010532413505

10)Global 2018

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
Ozuna : 0.05405405405405406	J Balvin : 0.09813731139348462	J Balvin : 0.2167526815306295
Bad Bunny : 0.05197505197505198	Cardi B : 0.0749877416273662	Ty Dolla \$ign : 0.21609084128168096
J Balvin : 0.0498960498960499	Quavo : 0.0626837652570817	Cardi B : 0.2147792125030365
Nicki Minaj : 0.04573804573804574	Nicki Minaj : 0.0610354670904826	Nicki Minaj : 0.21396749968178538
Cardi B : 0.04158004158004158	Khalid : 0.057577849502113525	Drake : 0.2090686869121138
Logic : 0.04158004158004158	Ty Dolla \$ign : 0.05566951823289086	Quavo : 0.20768818934629646
Travis Scott : 0.0395010395010395	David Guetta : 0.05022334424464341	David Guetta : 0.20753592527786074
Swae Lee : 0.035343035343035345	Logic : 0.050077237787347295	Maluma : 0.20632580326457875
Drake : 0.035343035343035345	Drake : 0.04813625668011303	Lil Wayne : 0.2055766173413232
Anuel AA : 0.035343035343035345	Lil Wayne : 0.047923025235578946	French Montana : 0.20483285244500876

11)Global 2019

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
J Balvin : 0.06273764258555133	J Balvin : 0.12351204549812299	J Balvin : 0.2111292482430332
Nicky Jam : 0.051330798479087454	Cardi B : 0.07536423598361569	Cardi B : 0.200805079624792
Farruko : 0.051330798479087454	Khalid : 0.05449910282227925	Tainy : 0.18982113459138655
Anuel AA : 0.043726235741444866	Ed Sheeran : 0.04866219521976986	Diplo : 0.18779560419178337
Ed Sheeran : 0.043726235741444866	Anitta : 0.046742473873335466	Farruko : 0.18623418949328782
Sech : 0.04182509505703422	Diplo : 0.0461859487695752	benny blanco : 0.18567282042171798

Ozuna : 0.04182509505703422	Farruko : 0.03739195133763731	Post Malone : 0.18483708379566374
Post Malone : 0.04182509505703422	Nicki Minaj : 0.0355318634680943	Bad Bunny : 0.18456017430683128
Daddy Yankee : 0.03802281368821293	Post Malone : 0.03499461562782575	Anitta : 0.18442203046378725
Dalex : 0.034220532319391636	Nicky Jam : 0.030605572631017902	Chris Brown : 0.1831879797023195

CONSIDERATIONS ON THE RESULTS

General considerations:

The obtained results are reasonable. In fact, after checking the dataset with the metadata about the artists, we verified that artists that obtained a high score of centrality were very popular artists. However, one can notice that there is not a direct relation between the centrality scores and the popularity. For example, J Balvin, which is the artist that in the global chart of 2019 obtained all the highest centrality scores, in the metadata dataset he has a lower popularity than The Weeknd. In fact, while J Balvin in 2019 was involved in several different collaborations with other artists, the Weeknd was involved in only a few collaborations, and that's why his centrality is really low.

For this reason, the centrality scores are not intended as a direct measurement of the following of an artist, rather they're a measurement of his involvement in different collaborations with other artists in making one or more songs in the top 200 chart in Spotify.

Consideration about different geographic areas:

The results are similar within different geographic areas, but there are still a few differences. In particular, it can be seen that certain artists from a certain country tend to be famous only within their geographical area. For example, the artist Alitta (who is a brazilian singer and dancer) obtained really high scores of the centrality only in the Brazil chart. Another example is Capital Bra, a German rapper that obtained a high score only in Germany.

Consideration about different types of centrality scores:

It can be seen that different centrality values are mostly in agreement with each other. The one with the most different values is the degree centrality. In fact, this centrality is the one that contains the least meaning in its definition, as it is only a measure of the number of connections with other nodes. Therefore, the other two centralities showed more meaningful values relative to what was expected from the charts.

Side note: one can see that the vast majority of the highest ranked artists with respect to their centrality scores are mostly involved in pop songs, as we will analyze further in the next class of networks.

GENRE COLLABORATION NETWORKS

DESCRIPTION OF THE NETWORKS CREATED FROM THE DATASET:

These are networks of music genres obtained from Spotify Charts for each considered market and year. Each music genre is a node. There is an edge between two genres/nodes when artists belonging to these genres collaborated on one or more hit songs. These networks are subdivided by geographic areas (US, Japan, France, and other five areas) for

the year 2019. We also created three different Global networks containing the data of all the countries, one for each of the following years: 2017, 2018, 2019.

The number of nodes for each of these networks is up to 300, while the number of edges is up to 3000.

These graphs are undirected, and each edge has a weight corresponding to the number of hit songs which are a collaboration between artists from these genres.

Note: one can note that, as expected, the genre collaboration networks are denser than the artist collaboration networks. In fact, it's more likely to have a song that crosses two or more genres than a song between two particular artists.

IMPORTANT NOTE ON THE FOLLOWING RESULTS: in this report we reported only a part of the all the tables and we reported only 4 entries/outputs for most of them; the extended results with 10 entries/outputs for each table are available in the separate PDF file, whose link is at the bottom of this report.

RESULTS FOR EACH COUNTRY:

1)Australia (check the extended results)

2)Brazil

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
funk carioca : 0.5460122699386504	post-teen pop : 0.10219558872627484	electropop : 0.43096081267561137
baile pop : 0.5337423312883436	pop rap : 0.09439740128753471	post-teen pop : 0.42979605372243407
pop : 0.5337423312883436	dfw rap : 0.0831132875428906	r&b : 0.4130507529280535
dance pop : 0.43558282208588955	funk carioca : 0.08039093046688273	moombahton : 0.41198067325725546

3)Canada (check the extended results)

4)Germany (check the extended results)

5)France (check the extended results)

6)United Kingdom (check the extended results)

7)Japan

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
pop : 0.7305389221556887	post-teen pop : 0.14862316880409548	dance pop : 0.44518328781530425
dance pop : 0.5808383233532934	rock : 0.12246257028563284	post-teen pop : 0.43531773019058834
post-teen pop : 0.5568862275449102	dance pop : 0.11377223189662937	tropical house : 0.4213128702380761
pop rap : 0.437125748502994	pop : 0.10254885774965761	indie popitism : 0.41795133137979357

8)United States of America (check the extended results)

GLOBAL RESULTS FOR EACH YEAR

9)Global (2017)

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
pop : 0.7003891050583657	pop : 0.10355628148207137	dance pop : 0.4370897878276191
dance pop : 0.5953307392996109	canadian pop : 0.08080980843359929	electropop : 0.4330048365395105

pop rap : 0.5642023346303502	dance pop : 0.07691469644458296	pop : 0.42979144257632307
rap : 0.5252918287937743	electropop : 0.05578090938738278	canadian pop : 0.42741252315246886

10)Global (2018)

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
pop : 0.5816326530612245	pop : 0.07831678112633134	tropical : 0.4088526617800622
dance pop : 0.5374149659863945	dance pop : 0.061724346584012074	post-teen pop : 0.4070328428107148
rap : 0.4931972789115646	rap : 0.0557726323980559	latin pop : 0.40403554647190243
pop rap : 0.4693877551020408	post-teen pop : 0.04938577485624924	indie pop : 0.40403554647190243

11)Global (2019)

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
pop : 0.6148867313915858	post-teen pop : 0.08769799651098627	post-teen pop : 0.39860613911678805
dance pop : 0.5210355987055016	pop : 0.07685295406375021	pop : 0.3963316390362358
pop rap : 0.49190938511326865	dance pop : 0.059290098214068214	uk pop : 0.3940829488856756
post-teen pop : 0.44336569579288027	trap latino : 0.056433316198762615	dance pop : 0.39352475773994516

CONSIDERATIONS ON THE RESULTS

General considerations:

The obtained results are reasonable. In fact, we verified that genres that obtained a high score of centrality are the ones that one would expect, with a predominance of pop, hip pop and rap, followed by a minority of rock (see the extended results file).

We can see that new variants of pop are gaining popularity in the most recent years (like k-pop and dance pop).

Consideration about different geographic areas:

While pop and rap are globally dominant, there are still a few peculiar types of genres related to a specific geographic area. For example, the genre of j-pop is mainly present only in Japan. Another example that highlights this fact even more is the relevance of the afrofuturismo brasileiro genre in Brazil (see extended results file).

Consideration about different types of centrality scores:

The same consideration we made regarding this aspect while analyzing the artist collaboration network can be made in this case, with the closeness and betweenness centralities obtaining more meaningful results.

HIT SONGS NETWORK

DESCRIPTION OF THE NETWORKS CREATED FROM THE DATASET

This is a single network of hit songs obtained from Spotify Charts for all considered markets and years. Each hit song is a node. There is an edge between two songs/nodes when both songs share at least one artist. This network contains all hit songs who made into Weekly Top 200 Spotify Charts in all considered markets (2017-19). The number of nodes for each of this network is 11136, while the number of edges is 211710.

The graph is undirected, and each edge has a weight corresponding to the number of artists that those songs share.

RESULTS

<u>Degree</u>	<u>Betweenness</u>	<u>Closeness</u>
Sturmmaske auf - Gold war gestern RMX : 0.026044005388414906	Réseaux - Remix : 0.030546621314612155	Réseaux - Remix : 0.2445251816057098
Champions : 0.024966322406825326	Afro Trap Pt. 7 (La puissance) - Major Lazer Remix : 0.02323117300891534	Know No Better (feat. Travis Scott, Camila Cabello & Quavo) : 0.24278296894034
Down for Life (feat. PARTYNEXTDOOR, Future, Travis Scott, Rick Ross & Kodak Black) : 0.024158060170633138	INTERNATIONAL GANGSTAS: 0.02322475537348	Maghreb Gang (feat. French Montana, Khaled & HAMZA) - Saucegod Remix : 0.24136273520019494
Woah : 0.02299057027391109	Olha A Explosão - Remix : 0.022364866522034153	Champions : 0.24088736105145192
Jusqu'à minuit (feat. JuL) : 0.02263134261338123	Maghreb Gang (feat. French Montana, Khaled & HAMZA) - Saucegod Remix: 0.02096075000770037	A.W.A : 0.2401441161521662
Et je deviens fou : 0.02263134261338123	Fendi Drip : 0.02094406814650586	Olha A Explosão - Remix : 0.23907118352389023
Tel Me : 0.02263134261338123	Hola Señorita : 0.018026904005721624	Writing on the Wall (feat. Post Malone & Cardi B) : 0.23850507841264515
Iced Out My Arms (feat. Future, Migos, 21 Savage & T.I.) : 0.022182308037718902	A.W.A : 0.016921401462311268	Welcome to the Party (with French Montana & Lil Pump, feat. Zhavia Ward) - from Deadpool 2 : 0.23655281953360266
J'suis loin : 0.022182308037718902	I Like It (feat. Kontra K and AK Ausserkontrolle) : 0.015379917324924861	Out Of Your Mind (feat. Chris Brown) : 0.23635690305666132
Waffen : 0.022092501122586438	VVS : 0.015206056859598931	VVS : 0.23611246334721217

CONSIDERATIONS ON THE RESULTS

From the obtained results, it can be seen that the obtained centrality values do not directly correspond to the popularity of the song intended as the number of times it has been streamed. This is due to the fact that this method does not aim to directly evaluate the popularity of songs, but to evaluate the success of songs based on the notoriety of the artists at the top of the charts with whom the songs are related.

For example, a viral/popular song made by an artist who has only made that song and has never collaborated with other artists will have a centrality of 0. On the other hand, a song among the last of the top 200 may have a high centrality score if the artist who made it tends

to collaborate with a circle of successful artists, even if the song itself is not among the most popular.

Thus, a song's score is highly dependent on its artists' collaboration with other artists in the top 200. As a result, a less popular hit song may get a higher score due to the notoriety gained from collaborations with other artists in the top 200.

In this way we can highlight how influential collaborations with other commercial chart artists helped to make the song a hit.

In particular, the song that obtained the highest score of closeness and betweenness centrality, “Réseaux - Remix”, featured the following artists: Niska, Quavo and Stefflon Don. After analyzing the dataset, it can be seen that the number of hit songs with those artists are a lot. Then, each one of these artists collaborated with other artists, creating a very dense network. In fact, Quavo can be also seen collaborating to the song with the second highest closeness centrality value (“Know No Better (feat. Travis Scott, Camila Cabello & Quavo)”).

To clarify, here is an example. Again, the song with the highest scores, “Réseaux - Remix” is a remixed version of “Réseaux” (which is not in the top 10 scores on the table above). The original Réseaux only featured Niska but has **133 million streams**, while the remix, which not only featured Niska, but also Quavo and Stefflon Don, only has **9.5 million streams**. However, since the remix version includes other notorious artists (who are involved on other hit songs) other than Niska, it has a way higher centrality score.

EPPSTEIN-WANG ALGORITHM FOR APPROXIMATED CLOSENESS CENTRALITY

DESCRIPTION

We implemented the pseudo code of the Eppstein-Wang algorithm to approximate the closeness centrality that we studied during the course. We tested it on the Artist Collaboration Network graph of Australia and confronted the results obtained with 3 different values for the numbers of nodes used to approximate the distances (value of parameter **k**). The dataset we used has one huge connected component and only about ten other small connected components (with at most 5 nodes). For this reason, we decided to apply the Eppstein-Wang approximation only on the big component and then re-normalize the scores using the size of the whole graph.

Exact value	Eppstein-Wang (k=30)	Eppstein-Wang (k=15)	Eppstein-Wang (k=5)
Post Malone : 0.1945162314109539	Post Malone : 0.279645588056803	Post Malone : 0.2933537051184109	Ty Dolla \$ign: 0.383616383616383
Khalid : 0.1924224076067456	Khalid : 0.27451447634933873	Cardi B : 0.2877122877122877	Khalid : 0.3562152133580705
Cardi B : 0.1915974455162558	Cardi B : 0.2695682695682695	Ed Sheeran : 0.277056277056277	Post Malone : 0.332467532467532
Ty Dolla \$ign: 0.1873798916841369	Ed Sheeran : 0.2647971497529019	A Boogie Wit da Hoodie : 0.2770562	The Chainsmokers : 0.311688311688

Ed Sheeran : 0.1865975121781489	A Boogie Wit da Hoodie: 0.26479714	Khalid : 0.277056277056277	Nicki Minaj : 0.3116883116883117
A Boogie Wit da Hoodie: 0.18543611	DJ Khaled : 0.2647971497529019	French Montana : 0.2671614100185	Kanye West : 0.3116883116883117
DJ Khaled : 0.18002056059080226	Ty Dolla \$ign : 0.2601919819311123	Ty Dolla \$ign : 0.2624743677375256	6LACK : 0.3116883116883117
21 Savage : 0.17457071940104166	21 Savage : 0.24935064935064935	DJ Khaled : 0.26247436773752564	Young Thug : 0.29335370511841097
Young Thug : 0.17355380258899675	benny blanco : 0.2472899001824622	Young Thug : 0.25794894760411996	A Boogie Wit da Hoodie : 0.2933537
Camila Cabello : 0.17221620102761	Camila Cabello : 0.24526293378752	Meek Mill : 0.25794894760411996	Halsey : 0.29335370511841097

CONSIDERATIONS ON THE RESULTS

As expected, from the results obtained, it can be seen that the approximated values obtained with the EPPSTEIN-WANG algorithm improve as the value of the parameter k increases.

In fact, as we have seen during the course, for values of k high enough, the approximated value of the closeness centrality converges toward the exact value with high probability.

In particular, it can be seen that the approximated values tend to overestimate the true value of the closeness centrality, but the approximation improves with $k = 30$.

With $k = 5$, instead, the approximated values of the closeness centrality are considerably far from exact values, with some artists who did not get the highest exact values of the closeness centrality appearing in the top 10.

For example, in the exact computation, Cardi B gets the third best score, while in the approximation with $k = 5$ she is not even included in the top 10.

LINKS TO THE CODE AND EXTENDED RESULTS

Link to the Google drive folder:

https://drive.google.com/drive/folders/1MmLsV2DUqzQculx9Z2P7lyslz8LZLhO8?usp=share_link

This folder contains the PDF with the extended version of the results and the Colaboratory notebook with the code.

REFERENCES

Link to the datasets:

- 1) Main dataset we used to create the networks (it contains also some metadata):

<https://opgabriel.github.io/ISMIR2020/>

- 2) Additional dataset with metadata we used to evaluate our results:

<https://marianaossilva.github.io/DSW2019/#tables>

CONTRIBUTIONS OF EACH MEMBER OF THE GROUP

We subdivided the work equally among us. We coded mostly together via Zoom and then we equally divided between us the tests and the analysis.