

Projet IC05 – A24

Analyse des Top50 Spotify



Céline Ozkaya, Yowen Pierre et Caio Artus

Sommaire

Sommaire.....	1
Table des figures.....	2
Introduction.....	2
Collecte des données.....	4
1. Playlists Top50.....	4
a) Liste des playlists.....	4
b) Contenu des playistes et caractéristiques musicales.....	5
2. Artistes.....	6
Analyse des attributs associés aux morceaux.....	7
1. Recherche de corrélations.....	7
2. Analyse de la variance.....	8
Exploitation de l'outil Gephi.....	9
1. Lancement des graphes interactifs.....	9
2. Graphe des artistes et des genres.....	10
a) Coloration par continents.....	10
b) Coloration et taille par popularité.....	13
c) Coloration par “danceability”.....	14
3. Graphe des artistes et des pays.....	15
Statistiques sur les artistes.....	19
1. Artistes ayant le plus d'occurrences tout Top50 confondu.....	19
2. Artistes apparaissant dans le plus de Top50 différents.....	20
Conclusion.....	20
Annexes.....	21

Table des figures

- Figure 1 : Position des chansons en fonction de la valeur de différents attributs musicaux
- Figure 2 : BoxPlots de la valence pour chaque continent
- Figure 3 : Graphe complet artistes-genres coloré par continent
- Figure 4 : Légende pour les colorations par continent des artistes
- Figure 5 : Zoom sur le cluster “funk”
- Figure 6 : Zoom sur l’Amapiano dans le graphe des artistes coloré par continents
- Figure 7 : Graphe complet artistes-genres avec mise en évidence du nombre de followers
- Figure 8 : Zoom sur la pop dans le graphe de popularité
- Figure 9 : Zoom sur le rap dans le graphe de popularité
- Figure 10 : Graphe complet artiste-genre coloré par “danceability”
- Figure 11 : Graphe complet artiste-pays
- Figure 12 : Zoom sur le petit cluster du graphe artiste-pays
- Figure 13 : Zoom sur le grand cluster du graphe artiste-pays
- Figure 14 : Zoom sur l’Afrique du Sud dans le graphe artistes-pays
- Figure 15 : Extrait des résultats du top 10 des artistes les plus récurrents dans les Top50
- Figure 16 : Code R pour l’analyse des variances

Introduction

La musique, en tant que reflet des tendances culturelles et sociales, offre une opportunité d'analyse à travers les données disponibles sur des plateformes numériques comme Spotify. Dans le cadre du projet de l'UV IC05, nous avons exploité les données rendues accessibles par l'API Spotify, un outil permettant d'accéder à une multitude de données sur les morceaux, les artistes et les genres, au travers d'une étude sur les Top50 générés par la plateforme. Inspiré du projet "Data Analysis with Spotify API" d'Erick Lopez (2023), qui portait sur le Billboard Hot 100, ce travail s'inscrit dans une perspective à l'échelle mondiale, continentale et nationale. En utilisant Python et ses librairies d'analyse et traitement de données, l'objectif est de mettre en lumière les critères influençant la sélection des morceaux, de cartographier les relations entre artistes, pays et genres présents dans les Top50, et d'analyser les facteurs déterminants de la popularité musicale. Ce projet cherche ainsi à comprendre en quoi les réseaux d'artistes, genres et pays et les caractéristiques musicales attribuées par Spotify reflètent les préférences des utilisateurs de la plateforme ?

Remarque importante : Tous les scripts, documents, fichiers Gephi, etc. cités dans ce rapport sont disponibles sur le répertoire GitLab du projet disponible à ce [lien](#). Pour simplifier l'accès aux différents graphes interactifs, il est recommandé de cloner ce répertoire. Sinon il faut télécharger un fichier pour chaque graphe.

Collecte des données

Les données ont été collectées via des appels à différents endpoints de l'API Spotify. Dans un premier temps, nous avons récupéré les données liées aux playlists Top50. Ensuite, les données des artistes issus des playlists ont été collectées. Finalement, nous avons obtenu la liste des genres musicaux associés à chaque artiste.

Par ailleurs, nous avons fait en sorte de collecter les données tôt dans le semestre afin que le contenu des playlists Top50 ne soit pas impactés par les fêtes de fin d'années. La collecte date ainsi du 11 novembre 2024.

1. Playlists Top50

Le code permettant la collecte des données relatives aux playlists a été implémenté dans les fichiers [recup_donnees.ipynb](#), pour la liste des playlists, et [recup_features.ipynb](#), pour le contenu des playlists et les caractéristiques musicales associées à chaque morceau.

a) Liste des playlists

Pour récupérer l'ensemble des playlists Top50, un script `search_for_top50_playlist()` ([recup_donnees.ipynb](#)), s'appuyant sur le endpoint `/search1` de l'API Spotify, a été développé. Ce script permet de récupérer le premier résultat de la recherche d'une playlist ayant le nom "Top50 {pays}" pour un pays donné. Pour l'utiliser, une liste contenant tous les pays du monde a été générée à l'aide de ChatGPT.

Le script récupère à chaque fois le premier résultat de la requête et vérifie simplement si la playlist contient exactement 50 morceaux, car c'est le standard des playlists officiels. Il était prévu au départ d'effectuer des vérifications plus précises, mais les titres des Top50 étant très variables, il n'était pas possible de prendre en compte tous les cas. Nous avons donc décidé qu'il valait mieux récupérer le plus possible et filtrer à postériori.

A ce stade, nous avons obtenu un grand nombre de playlists dans un fichier `top50_raw_playlists_data.csv`. Cependant, certaines n'avaient pas été créées par l'utilisateur Spotify officiel et d'autres ne correspondaient pas au bon pays. C'est le cas, par exemple, de l'Albanie qui ne présente aucun résultat ou bien de Barbados dont le résultat de la recherche retourne la playlist des Etats-Unis. Il a donc fallu nettoyer les données.

¹ Search for item. (s. d.) Spotify for Developers. Consulté le 14/12/2024. Disponible sur : <https://developer.spotify.com/documentation/web-api/reference/search>

Pour ce faire, le script `clean_top50_data()` a été développé afin de ne conserver que les pays présentant une playlist valide dans un fichier propre nommé `top50_cleaned_playlists_data.csv`.

Pour certains pays comme les Etats-Unis, le premier résultat de la requête sur le endpoint `/search` n'était pas le Top50 officiel, mais une autre playiste très populaire. Dans ce cas le script `clean` élimine ce pays même si en réalité il possède bel et bien un Top50 officiel. Il a donc fallu ajouter à la main les quelques pays manquants.

Au total sur 195 pays nous avons récupéré 68 playlists officielles exploitables pour un total de 3400 morceaux.

b) Contenu des playlists et caractéristiques musicales

L'ensemble des fonctions sont implémentées dans le fichier `recup_features.ipynb`.

Une fois, le fichier des playlists établi, le contenu des playlists a été collecté via une fonction `get_playlist()`, faisant appel à l'endpoint `/get-playlist`². Ensuite, les caractéristiques musicales que la plateforme Spotify associe aux morceaux des playlists ont été récupérées au travers d'une fonction `get_playlist_tracks_features()`, utilisant `/audio-features`³. Finalement, ces fonctions sont appelées dans le script `get_top50_features()`, qui permet de renseigner, pour chaque morceau et dans un fichier `all_features.csv`, les informations suivantes : `title`, `artists_id`, `danceability`, `energy`, `key`, `loudness`, `mode`, `speechiness`, `acousticness`, `instrumentalness`, `liveness`, `valence`, `tempo`, `type`, `id`, `uri`, `track_href`, `analysis_url`, `duration_ms`, `time_signature`, `country`. La liste ci-dessous décrit le sens de chacune de ces caractéristiques :

- **Acousticness** : Évalue la probabilité qu'une piste soit acoustique, variant de 0.0 (pas du tout) à 1.0 (très acoustique). Ex : Une valeur de 0.9 indique un morceau probablement acoustique.
- **Danceability** : Indicateur de la capacité d'un morceau à être dansé, basé sur le tempo, la stabilité du rythme, la force du beat et la régularité. Les valeurs s'étendent de 0.0 (peu dansant) à 1.0 (très dansant).
- **Energy** : Mesure de l'intensité et de l'activité, variant de 0.0 (faible énergie) à 1.0 (forte énergie). Ce facteur prend en compte la rapidité, la densité sonore

² Get Playlist. (s. d.) Spotify for Developers. Consulté le 14/12/2024. Disponible sur : <https://developer.spotify.com/documentation/web-api/reference/get-playlist>

³ Get Several Tracks' Audio Features. (s. d.) Spotify for Developers. Consulté le 14/12/2024. Disponible sur : <https://developer.spotify.com/documentation/web-api/reference/get-several-audio-features>

et le niveau d'activité. Ex : Un morceau de métal est perçu comme étant très énergique.

- **Instrumentalness** : Évalue la probabilité qu'un morceau soit instrumental, de 0.0 à 1.0. Plus la valeur est élevée, plus la probabilité qu'il soit instrumental est forte. Les valeurs au-dessus de 0.5 suggèrent des morceaux sans voix.
- **Liveness** : Probabilité qu'un morceau soit enregistré en public (live). Les valeurs au-dessus de 0.8 indiquent des pistes avec une forte présence de public.
- **Loudness** : Niveau sonore moyen en décibels (dB), moyenne sur l'ensemble du morceau. Varie de -60 à 0db.
- **Speechiness** : Mesure la quantité de mots parlés. Des valeurs proches de 1.0 indiquent principalement des paroles (comme un podcast). Les morceaux musicaux auront des valeurs plus basses. Entre 0.33 et 0.66, on considère qu'il y a une combinaison de mots parlés et de musique (ex : rap). En dessous de 0.33 on considère qu'il s'agit de musiques sans paroles.
- **Valence** : Mesure de la positivité d'un morceau, de 0.0 (triste, colère) à 1.0 (joyeux). Exemples : valeurs hautes pour les morceaux heureux et optimistes.
- **Tempo** : Vitesse estimée en battements par minute (BPM).
- **Key & Mode** :
 - Key : La tonalité principale de la piste, indiquée de 0 à 11, correspondant à C, C#/D b , D, etc. Ex : C=0 (Do majeur), C#/D b = 1 (Do#/ Ré b), etc.
 - Mode : Indique si le morceau est en majeur (1) ou mineur (0).

2. Artistes

Les différentes fonctions permettant de collecter les données liées aux artistes issus des playlist Top50 ont été développées dans un Jupyter Notebook nommé artistes.ipynb.

Tout d'abord, une fonction `extract_artist_country_mapping()` a été implémentée afin de lier chaque artiste aux pays dans lesquels ils apparaissent. Ensuite, `json_to_csv()` a permis de convertir les données JSON fournies par Spotify en un fichier csv contenant les champs `artists`, `followers`, `genres`, `name`, `popularity` et `country`. Ces deux fonctions ont finalement été appelées dans `get_artistes_info()` qui, à partir de la liste des identifiants des artistes issus du fichier `cleaned_top50_playlists_data.csv`, effectue une requête `/artists4` sur l'API et retourne les informations listées ci-dessus. Le fichier `/data/artists.csv` contient ces données pour l'ensemble des artistes recensés dans les playlists. Nous avons également généré un fichier de ce type par continent.

⁴ Get Several Artists. (s. d.) Spotify for Developers. Consulté le 14/12/2024.
Disponible sur : <https://developer.spotify.com/documentation/web-api/reference/get-multiple-artists>

Analyse des attributs associés aux morceaux

1. Recherche de corrélations

La première chose que nous avons voulu faire était de voir s'il y avait une corrélation entre la valeur d'un attribut musical et la position de la musique dans le Top50. Avant de nous lancer dans une régression linéaire nous avons décidé de visualiser les données pour voir si c'était plausible. La fonction `graphique_attribut_position()` dans [graphiques.py](#) réalise cette tâche. Cependant, peu importe l'attribut sélectionné, il ne semble avoir aucun lien linéaire entre les attributs musicaux et la position dans le Top50.

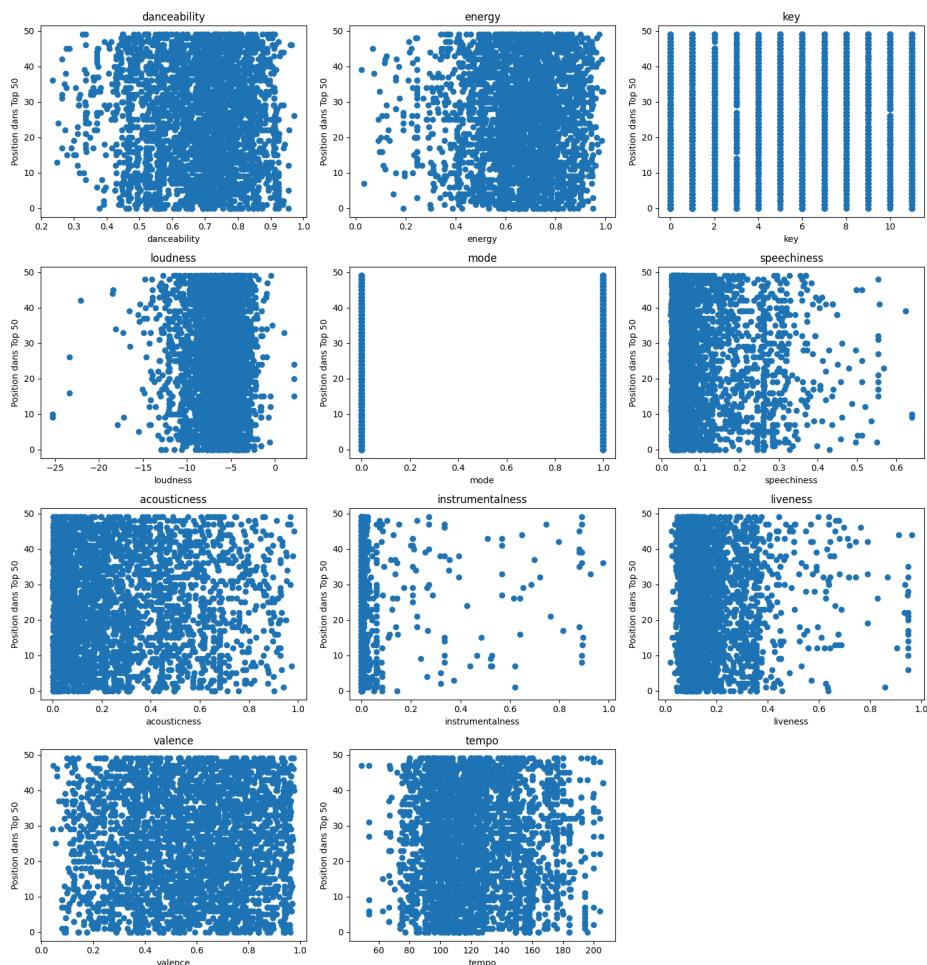


Figure 1 : Position des chansons en fonction de la valeur de différents attributs musicaux

2. Analyse de la variance

Une autre piste était de voir s'il y avait des variations significatives entre les différents pays pour un attribut donné. Pour ce faire, nous avons réalisé des Box Plot pour chaque attribut et pour chaque continent. Cela nous permet de comparer à la fois la médiane de chaque continent mais aussi l'ampleur de la variation autour de cette médiane. Les boxplots que nous avons réalisés sont [disponibles sur le gitlab](#). Pour réaliser ces Box Plot nous avons utilisé la fonction `comparaison_attributs_continents()` également disponible dans [graphiques.py](#).

Un attribut où des différences semblent significatives est la valence. En effet, elle semble être plus élevée pour l'Amérique Centrale et l'Amérique du Sud (fig. 2). Non seulement les médianes sont plus hautes mais la variance autour de la médiane semble limitée indiquant que cette différence de médiane pourrait être significative. Pour rappel la valence est une mesure, comprise entre 0 et 1, de la "joie" d'un morceau.

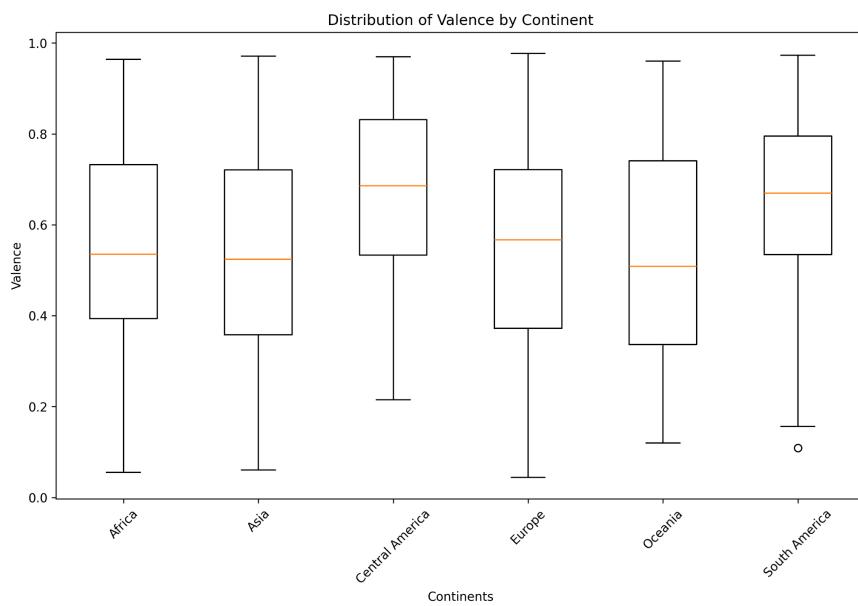


Figure 2 : BoxPlots de la valence pour chaque continent

Afin d'avoir une idée plus précise, il faut réaliser un test statistique pour l'analyse de la variance. En premier lieu, on voudrait savoir si les données suivent une loi normale. On réalise donc un test de Shapiro-Wilk sur chaque jeu de données. Ces tests, réalisés avec la fonction R `shapiro.test()`, ont révélé que toutes les données ne suivent pas des lois normales. Il faut donc réaliser un test de Kruskal-Wallis qui n'exige pas la normalité. Voici le test d'hypothèses que l'on vise à résoudre:

Hypothèse nulle : La valence suit la même distribution pour tous les continents

Hypothèse alternative : Il existe au moins un continent ayant une distribution différente pour la valence

Avec le script R disponible en annexe (fig. 16), nous avons réalisé le test de Kruskal-Wallis à l'aide de la fonction `kruskal.test()`, d'abord avec toutes les données. Ce test renvoie une p-value < 2.2e-16 ce qui nous permet de rejeter l'hypothèse nulle et d'affirmer avec beaucoup de confiance qu'au moins un des continents admet une distribution différente pour la valence.

Pour vérifier qu'il s'agit bien de l'Amérique Centrale et l'Amérique du Sud qui sont différentes, nous avons effectué le test en éliminant les données de ces deux continents. Ce test renvoie une p-value ≈ 0.153 ce qui ne nous permet pas de rejeter l'hypothèse nulle. Il semble donc que la valence suit une distribution différente pour l'Amérique Centrale et l'Amérique du Sud que pour les autres continents.

Une interprétation possible de ces résultats serait de dire que l'on préfère généralement les musiques plus "joyeuses" dans ces continents.

Exploitation de l'outil Gephi

1. Lancement des graphes interactifs

Pour chaque graphe une version interactive est disponible avec le projet. Ils sont dans des répertoires "network" présent dans les répertoires de chaque graphe. Il est préférable de cloner le projet avec `git clone https://gitlab.utc.fr/ozkayace/ic05-analyse-des-top50-spotify` pour les avoir tous en même temps.

Pour exécuter chaque graphe il faut faire tourner un serveur sur votre ordinateur. Pour ce faire il existe deux possibilités:

- a) Avec Pycharm (préférable)

Dans l'IDE Pycharm il suffit d'exécuter le `index.html` présent dans le fichier `network` pour chaque graphe. L'IDE se charge de faire tourner le serveur.

- b) Dans le terminal avec python

Ouvrir un terminal et se placer dans le répertoire `network` du graphe souhaité en utilisant la commande `cd <Chemin du répertoire>`.

Exécuter la commande `python3 -m http.server` ou `python -m SimpleHTTPServer` en fonction de la version de python que vous avez.

Ouvrir le navigateur et aller sur localhost:8000.

2. Graphe des artistes et des genres

Le premier graphe que nous avons réalisé avec l'outil Gephi montrait les liens entre les artistes et les genres musicaux présents dans les Top50. Spotify catégorise très finement les genres des artistes menant à une très grande quantité de genres. Nous avons choisi de n'inclure que les genres partagés par au moins 50 artistes des classements pour une question de lisibilité.

Pour générer les fichiers des nœuds et des liens nécessaires au fonctionnement de Gephi nous avons utilisé le script [gephi_artistes_genres.py](#). Dans le fichier des nœuds nous avons inclus des informations tel que la popularité de chaque artiste, leur nombre de followers et leur continent principal. Le continent principal d'un artiste est déterminé par le pays où il apparaît le plus fréquemment dans le Top50, proportionnellement à ses apparitions dans les Top50 des autres pays.

La position des nœuds a été déterminée par l'algorithme ForceAtlas disponible dans Gephi. Cet algorithme simule un système physique où les nœuds se repoussent comme des particules chargées, tandis que les liens agissent comme des ressorts qui attirent les nœuds connectés. Ainsi, les nœuds fortement connectés sont rapprochés les uns des autres, tandis que les nœuds faiblement connectés sont repoussés vers l'extérieur.

Nous avons, ainsi, pu réaliser différents graphiques à partir de nos fichiers de noeuds et de liens en mettant en évidence différentes données sur les artistes.

a) Coloration par continents

Ce graphe est disponible dans sa version interactive sur Gitlab. Il suffit de télécharger le dossier [network](#) et lancer un serveur. Pour le lancer, suivez [les instructions ci-dessus](#). Cette version permet d'ajouter un filtre sur le continent en utilisant "Select Group" et en s'aidant de la légende (fig. 4).

Dans le graphique ci-dessous (fig. 3), chaque nœud, correspondant à chacun des artistes, prend la couleur correspondant à son continent principal. Les nœuds et labels des styles sont de taille proportionnelle aux nombres de degrés entrants, donc au nombre d'artistes liés aux genres.

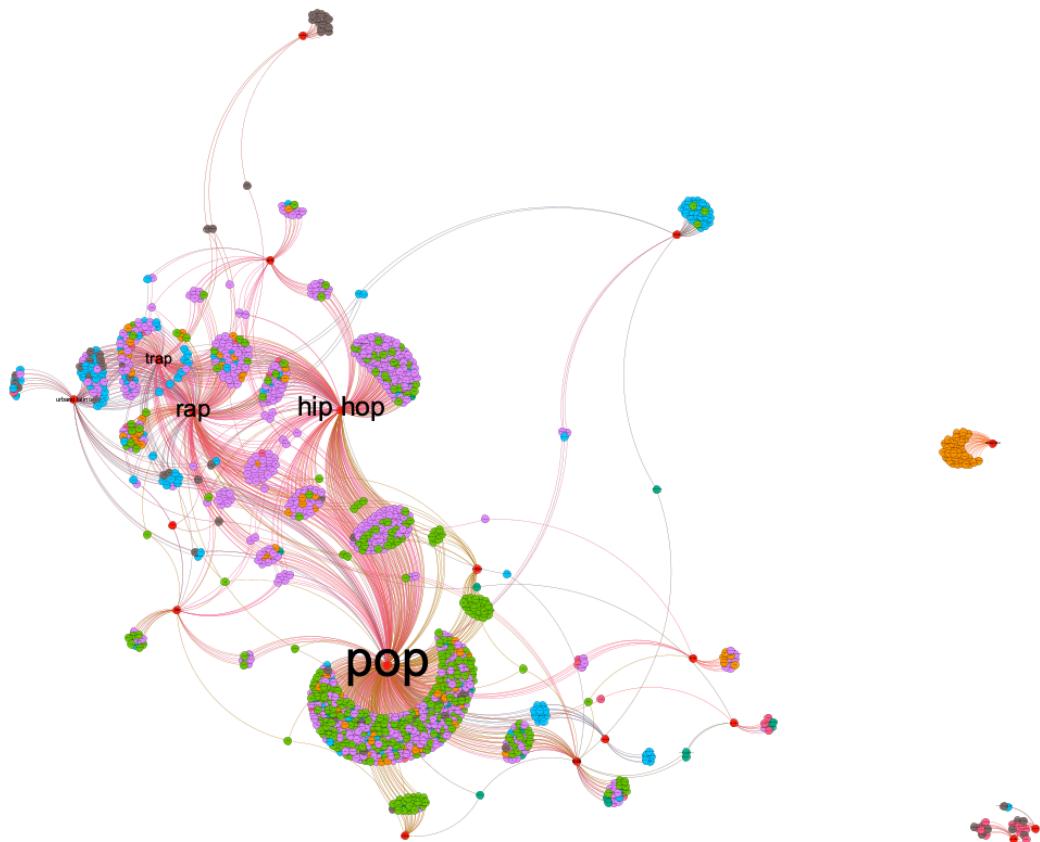


Figure 3 : Graphe complet artistes-genres coloré par continent

[Pink square]	Europe	(42.22%)
[Green square]	Asia	(25.38%)
[Blue square]	South America	(13.27%)
[Orange square]	Africa	(10.26%)
[Grey square]	Central America	(4.35%)
[Red square]	North America	(2.79%)
[Teal square]	Oceania	(1.73%)

Figure 4 : Légende pour les colorations par continent des artistes

Le graphique (fig. 3) montre que le genre le plus présent chez les artistes peuplant les Top50 est la Pop. Ce dernier est partagé par presque tous les continents, en particulier l'Asie (k-pop, v-pop, j-pop), en vert, et l'Europe, en rose.

Par ailleurs, on remarque qu'il existe une préférence pour les artistes faisant du Rap et le Hip Hop en Europe.

De plus, le graphique présente de petits clusters d'artistes de genres spécifiques à des continents ou pays. Par exemple, le "fun" est un style principalement écouté au Brésil mais aussi repris dans l'Europe de l'Est dans des remix appelés "phonk". On voit alors que les artistes autour du funk ont pour continent principal l'Amérique du Sud et l'Asie (présence de nœuds bleus et verts).

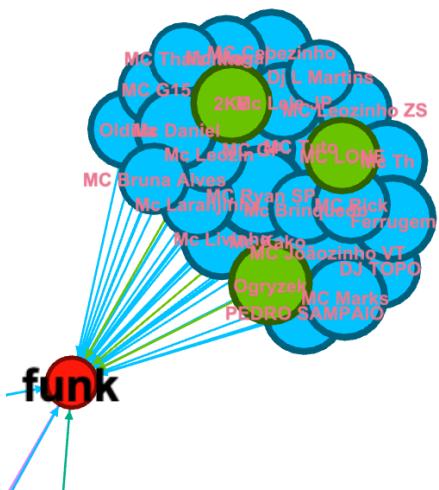


Figure 5 : Zoom sur le cluster "funk"

Un autre élément intéressant à relever est l'existence de genres musicaux isolés, très peu ou pas du tout connectés aux autres. Ces genres se retrouvent naturellement en périphérie du graphe en raison de l'utilisation de l'algorithme ForceAtlas. Ce phénomène est particulièrement visible pour les styles de musique niche qui, bien que très populaires dans certaines régions spécifiques, restent peu liés au reste du réseau musical mondial. L'Amapiano en est un excellent exemple. En effet, ce genre de musique électronique sud-africain, malgré sa popularité locale significative, ne présente pas de connexions avec les autres styles musicaux et n'est présent qu'en Afrique (fig. 6).

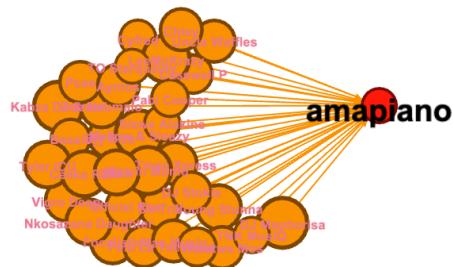


Figure 6 : Zoom sur l'Amapiano dans le graphe des artistes coloré par continents

b) Coloration et taille par popularité

Dans ce graphe, la couleur représente la popularité d'un artiste. Ainsi, plus son nœud est foncé, plus l'artiste a de followers sur la plateforme Spotify. La taille des nœuds d'artistes est aussi proportionnelle au nombre de followers. Ce graphe est disponible en version interactive dans le fichier [network](#). Pour le lancer, téléchargez le fichier et suivez [les instructions ci-dessus](#).

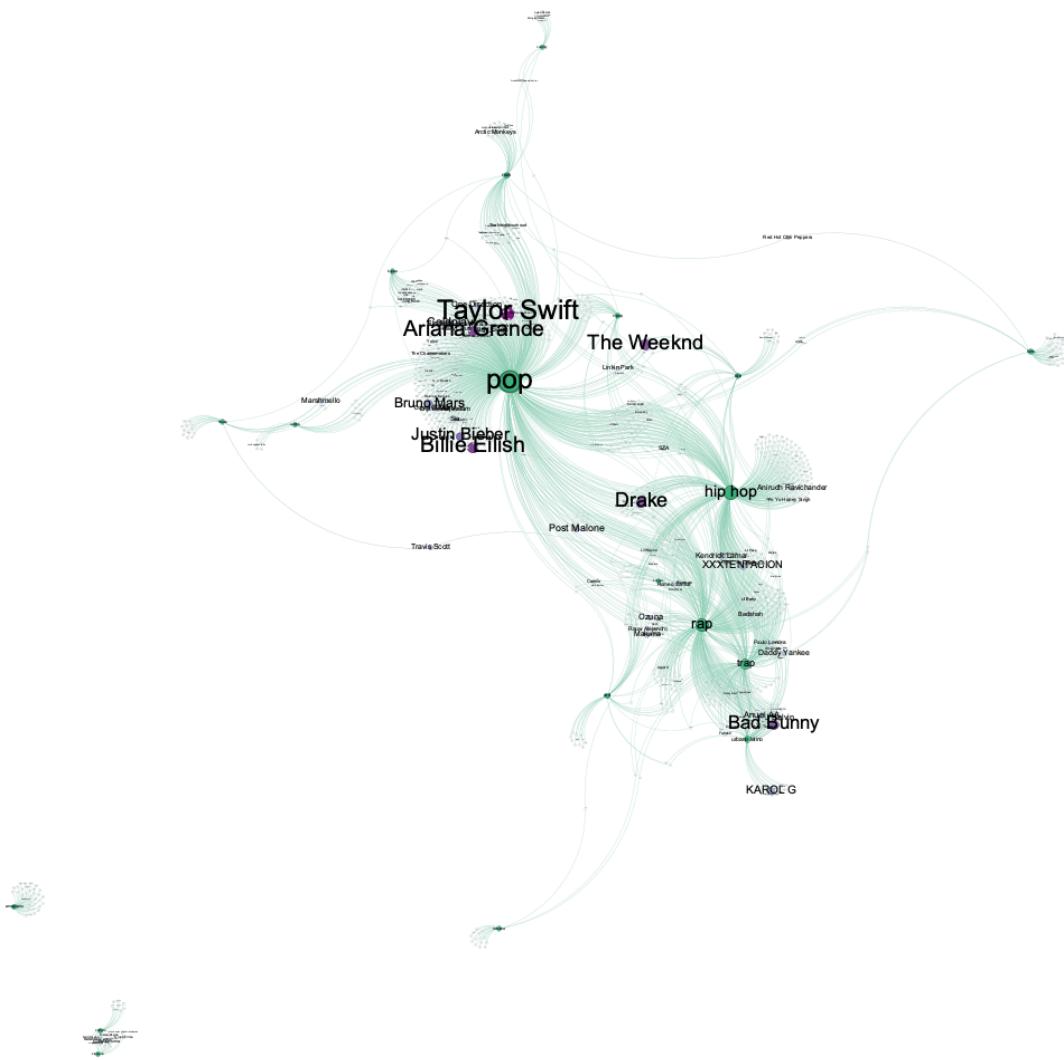


Figure 7 : Graphe complet artistes-genres avec mise en évidence du nombre de followers

Un premier élément frappant est la grande disparité de la popularité des artistes. Il n'y a que quelques artistes très populaires comparé à la majorité. De ce fait, le graphe contient majoritairement de très petits nœuds. On peut également comprendre cela par le fait que ces artistes, de par leur présence dans les Top50, ont tous une popularité et un nombre de followers plus important que la moyenne. Bien que le graphe permette de mettre en évidence les plus gros artistes, il donne l'impression que le reste est moins signifiant malgré leur popularité relativement élevée puisque l'on compare que des artistes très connus.

Il est intéressant de noter que les artistes les plus populaires se concentrent autour du genre Pop (fig. 8), une distribution qui correspond aux attentes et observations relevées sur le précédent graphe. De plus, les genres plus nichés (ex : Amapiano), situés en périphérie du graphe, sont associés à des artistes de moindre popularité étant donné qu'ils touchent un public restreint. Pour finir, un résultat plus inattendu est la relative sous-représentation d'artistes à forte popularité dans les sphères du Rap et du Hip Hop (fig. 9), et ce malgré l'importance culturelle et commerciale de ces genres. En effet, on constate que les artistes principaux du genre Pop comme Taylor Swift ou Justin Bieber ont une popularité bien plus importante que des artistes du genre Rap ou Hip Hop comme Kendrick Lamar.

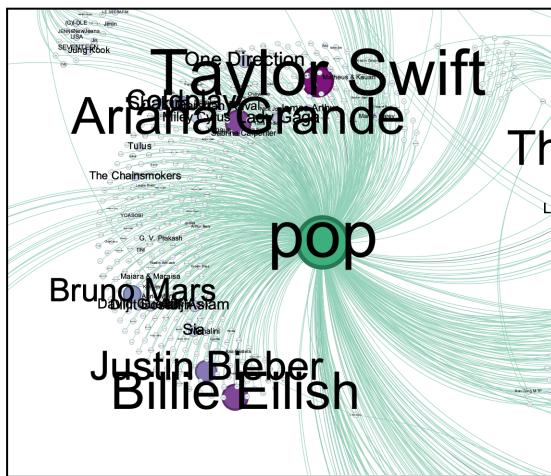


Figure 8 : Zoom sur la pop dans le graphe de popularité

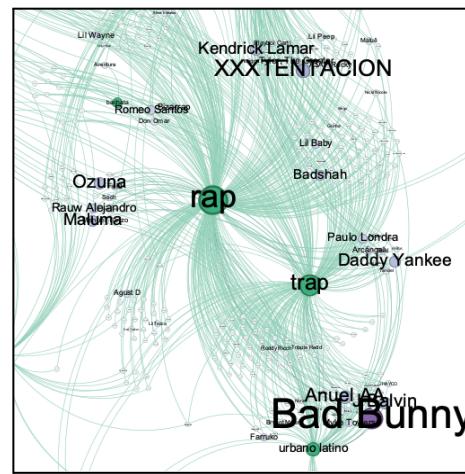


Figure 9 : Zoom sur le rap dans le graphe de popularité

c) Coloration par “danceability”

Dans ce graphe (fig. 10), la couleur des nœuds représentant les artistes reflètent la “danceability” moyenne des morceaux de celui-ci. Cette moyenne est calculée à partir de la “danceability” de chacun des morceaux de l'artiste présents dans les Top50. Plus un nœud est **bleu**, plus sa “danceability” est haute et plus un nœud est **rose** plus elle est faible.

Ce graphe est disponible dans sa version interactive sur Gitlab. Il suffit de télécharger le dossier [network](#) et lancer un serveur. Pour le lancer, suivez [les instructions ci-dessus](#).

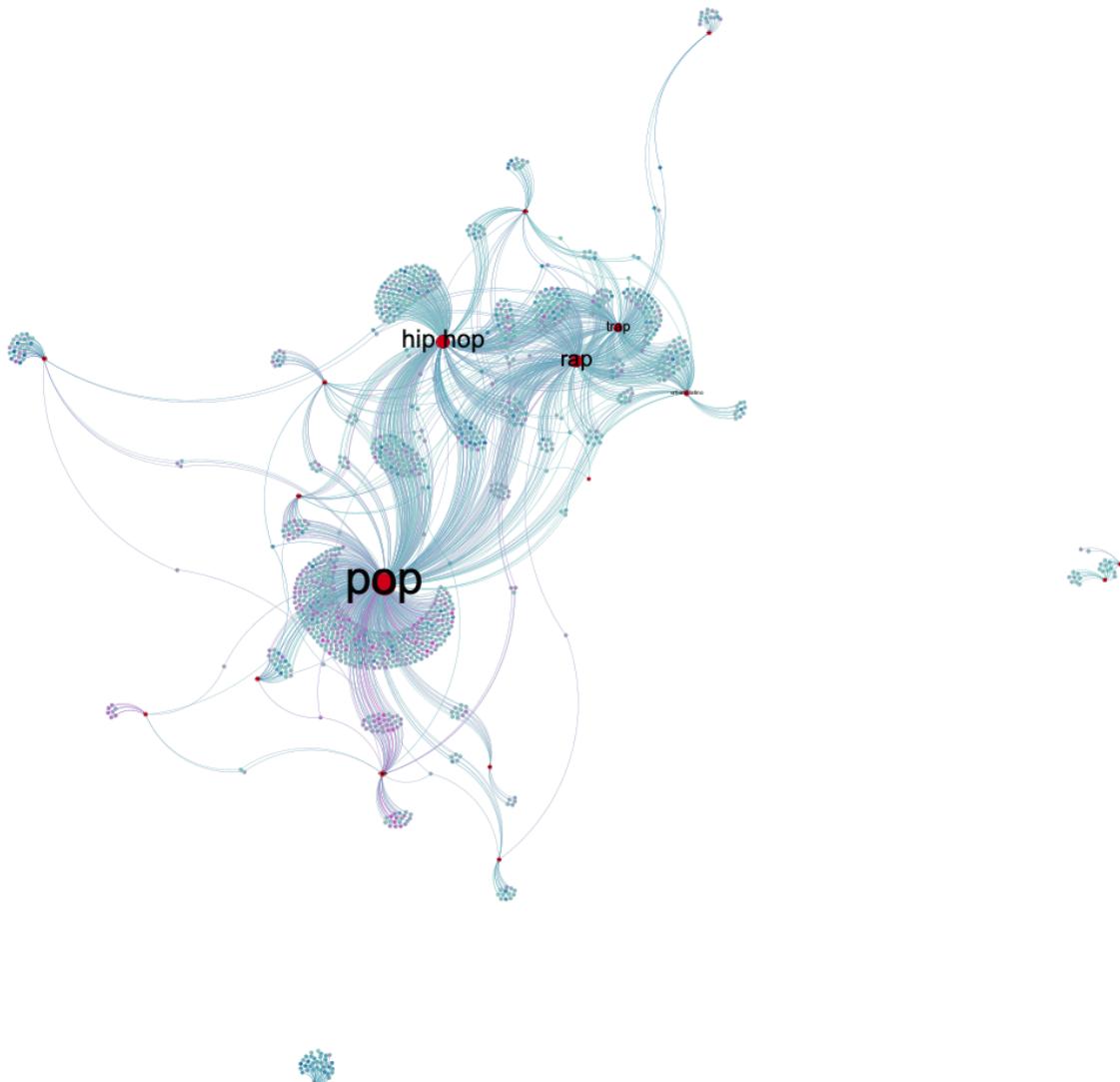


Figure 10 : Graphe complet artiste-genre coloré par “danceability”

On remarque tout d’abord que la “danceability” est plutôt élevée pour tous les genres présents et pour la majorité des artistes peuplant les Top50.

Par ailleurs, bien que la Pop soit un style généralement considéré comme dansant, une partie des artistes qui y sont liés présentent des morceaux ayant une “danceability” plutôt faible.

3. Graphe des artistes et des pays

Nous avons réalisé un graphe (fig. 11) mettant en relation les artistes et les pays dans lesquels ils sont écoutés. Un artiste est lié à un pays lorsque l’un de ses morceaux est présent dans le Top50 de celui-ci. Ici, la taille des nœuds des artistes correspond au degré sortant, donc au nombre de playlists dans lesquelles ils sont présents. De plus, les nœuds sont colorés en fonction du nombre de followers de l’artiste qu’ils représentent. Ainsi, plus les artistes sont populaires, plus leur nœud

est foncé. Dans ce graphe, chaque lien entre un artiste et un pays a un poids proportionnel au nombre de fois où l'artiste apparaît dans le Top50 du pays.

Ce graphe est aussi disponible en format interactif sur le GitLab, il suffit de télécharger le fichier [graph artistes pays/network](#) et lancer un serveur. Pour le lancer, suivez [les instructions ci-dessus](#).

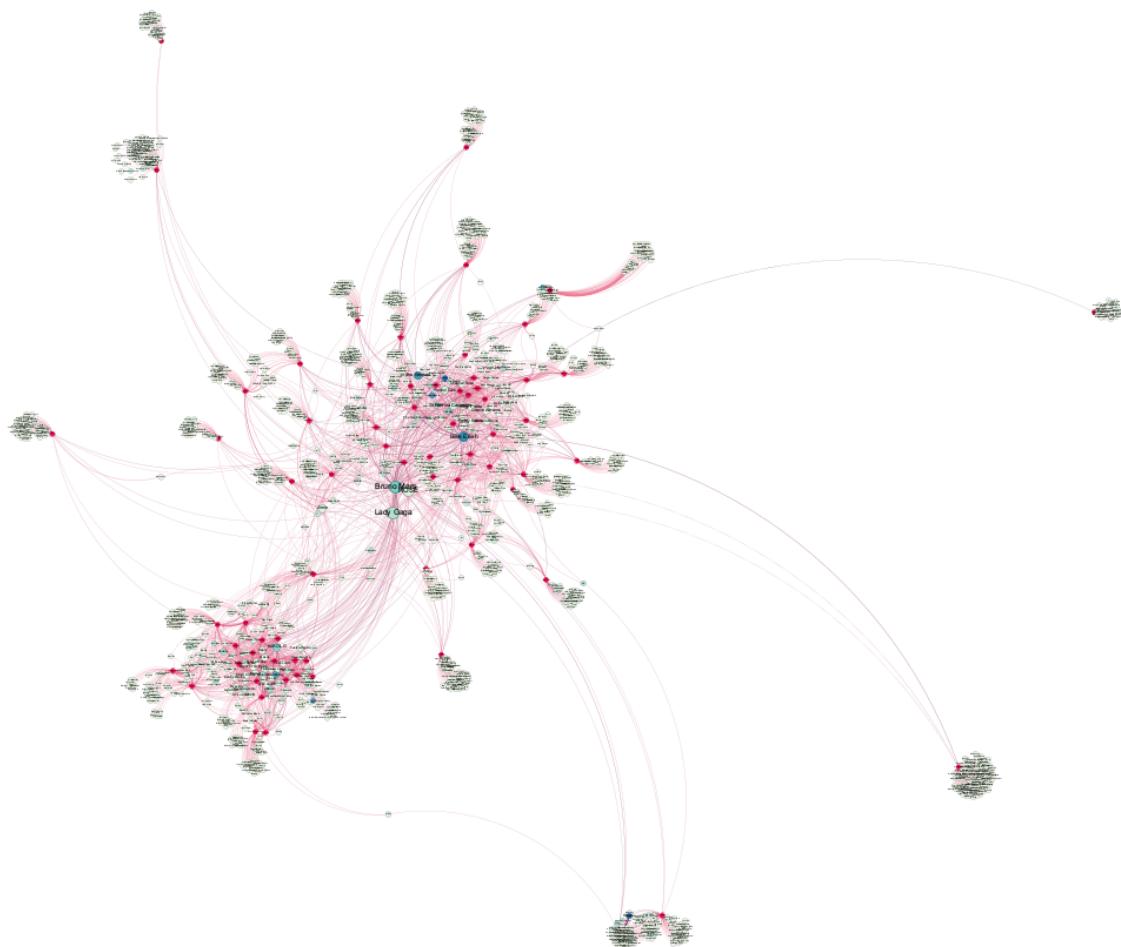


Figure 11 : Graphe complet artiste-pays

Ce graphe nous permet d'observer plusieurs phénomènes intéressants. Tout d'abord, nous remarquons qu'il existe deux clusters principaux. Dans le cluster plus petit (fig. 12), en bas à gauche, et un cluster plus grand (fig. 13), au centre. Le

cluster plus grand contient des pays de plusieurs continents, notamment l'Europe et l'Asie alors que le plus petit cluster contient presque exclusivement des pays d'Amérique du Sud et Centrale ainsi que d'autres pays hispanophones. Ces clusters se forment souvent autour de certains artistes populaires. Par exemple, le grand cluster est autour de Billie Eilish, une chanteuse américaine très populaire, alors que le petit cluster est centré sur des artistes comme KAROL G et Bad Bunny qui sont des grands artistes hispanophones. Certains artistes se trouvent en tension entre les deux clusters, indiquant qu'ils sont particulièrement internationaux étant présents dans des pays des deux clusters. C'est le cas par exemple de Bruno Mars ou Lady Gaga.

Les artistes ayant les plus gros nœuds et donc présents dans le plus de playlists Top50 sont majoritairement Américains. Cette observation révèle le soft power des Etats-Unis dont le rayonnement culturel se répand dans beaucoup de pays.

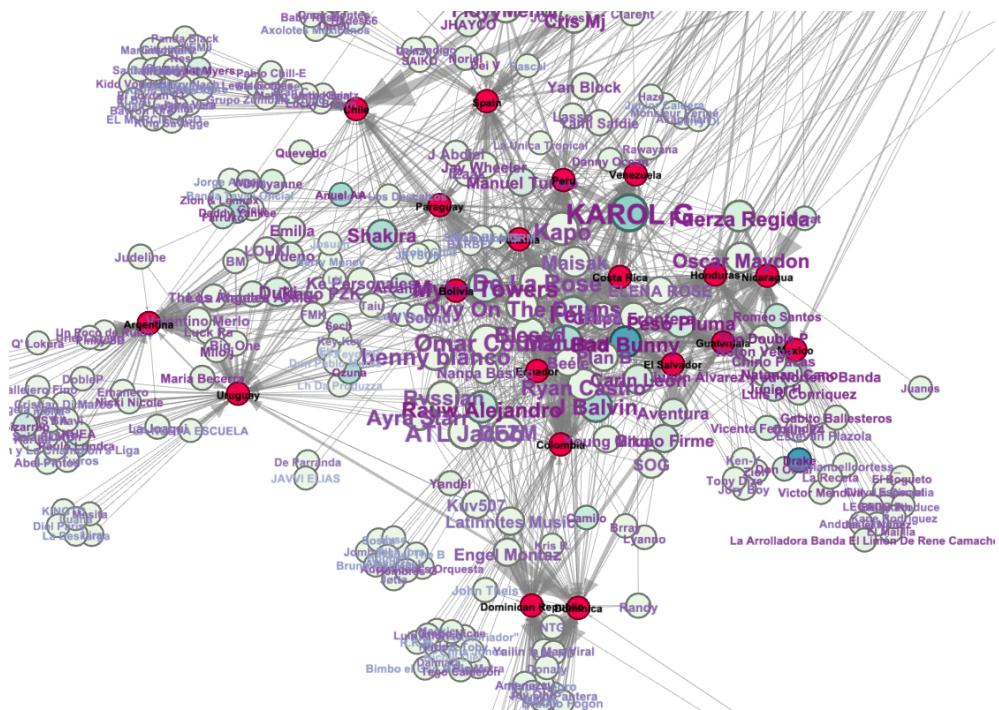


Figure 12 : Zoom sur le petit cluster du graphe artiste-pays

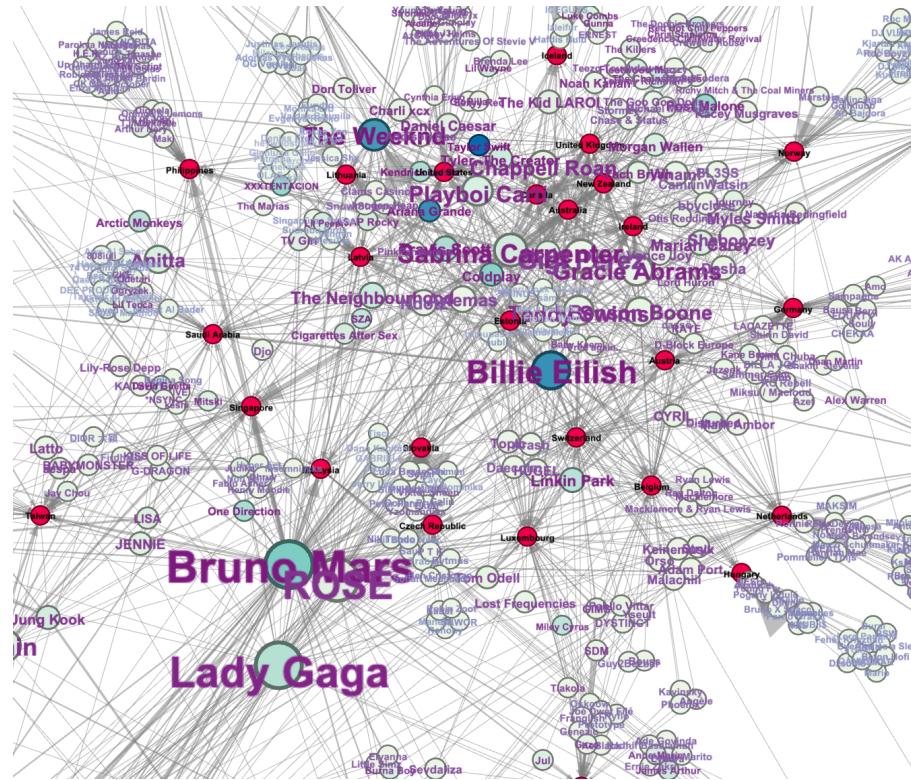


Figure 13 : Zoom sur le grand cluster du graphe artiste-pays

Il est particulièrement intéressant d'observer les pays qui se trouvent en périphérie du réseau, comme le Brésil, l'Égypte et l'Afrique du Sud (fig. 14). Cette position excentrée s'explique par le fait que leurs playlists Top50 partagent très peu d'artistes avec celles des autres pays. Cette singularité est souvent liée à des scènes musicales locales dynamiques, avec des genres propres à leurs cultures : l'Amapiano en Afrique du Sud, le Samba et le Funk Carioca au Brésil, ou encore le Mahraganat en Égypte. Dans ces pays, on trouve des artistes qui, bien qu'extrêmement populaires sur leur marché local, ont une reconnaissance internationale plus limitée. Cette différence de popularité se reflète visuellement dans le graphe par la couleur claire des nœuds autour de ces pays.

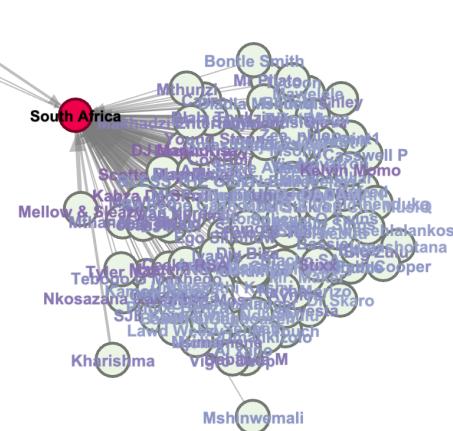


Figure 14 : Zoom sur l'Afrique du Sud dans le graphe artistes-pays

A contrario, les pays se trouvant dans les gros clusters partagent beaucoup d'artistes avec d'autres pays et ont donc un Top50 moins spécifique.

Finalement, on observe une distinction intéressante entre la popularité globale des artistes (représentée par la couleur foncée des nœuds) et leur présence dans les différents Top50 nationaux (représentée par la taille des nœuds). Bien qu'il existe une corrélation générale entre ces deux mesures, on trouve des cas particuliers révélateurs : certains artistes très populaires apparaissent dans relativement peu de Top50 nationaux, tandis que d'autres, malgré une popularité plus modérée, sont présents dans de nombreux classements nationaux. Cela peut être attribué au fait que les Top50 sont des playlists temporaires, très influencés par les modes actuelles, par exemple les phénomènes TikTok.

Statistiques sur les artistes

Pour conclure nos analyses, nous avons décidé de réaliser des statistiques sur les différents artistes présents dans les Top50 afin d'identifier qui peuple le plus de classements. La fonction permettant de trouver ces résultats est implémentée dans la section "Statistiques sur les artistes" du notebook [artistes.ipynb](#). Nous avons choisi de présenter les dix meilleurs artistes de chaque catégorie.

1. Artistes ayant le plus d'occurrences tout Top50 confondu

Les résultats complets avec les occurrences par pays sont consultables sur le notebook. La figure 15 en présente un extrait. Voici les résultats retournés, du meilleur au moins bon :

- | | |
|----------------------|----------------|
| 1. Bruno Mars | 6. Lady Gaga |
| 2. Feid | 7. Myke Towers |
| 3. KAROL G | 8. Maluma |
| 4. Billie Eilish | 9. Aventura |
| 5. Sabrina Carpenter | 10. ROSÉ |

```
Top 10 des artistes les plus fréquents avec les pays dans lesquels ils sont écoutés:
Bruno Mars: {'United Kingdom': 4, 'Ukraine': 6, 'Switzerland': 9, 'Sweden': 4, 'Spain': 3, 'Slovakia': 2, 'Croatia': 1, 'Portugal': 1, 'Norway': 1, 'Greece': 1}
Feid: {'Spain': 12, 'Andorra': 2, 'Mexico': 8, 'Venezuela': 24, 'Uruguay': 6, 'Peru': 23, 'Paraguay': 6, 'Brazil': 5, 'Chile': 3, 'Argentina': 2, 'Colombia': 2, 'Ecuador': 2, 'Dominican Republic': 2, 'Puerto Rico': 1, 'Bolivia': 1, 'Uruguay': 1}
KAROL G: {'Switzerland': 2, 'Spain': 8, 'Netherlands': 3, 'Luxembourg': 2, 'Italy': 2, 'France': 4, 'Germany': 3, 'Belgium': 2, 'Austria': 1, 'Portugal': 1, 'Slovenia': 1, 'Croatia': 1, 'Hungary': 1}
Billie Eilish: {'United Kingdom': 6, 'Ukraine': 8, 'Switzerland': 4, 'Sweden': 8, 'Slovakia': 7, 'Romania': 5, 'Croatia': 3, 'Portugal': 2, 'Hungary': 2, 'Netherlands': 2, 'Greece': 1, 'Slovenia': 1, 'Croatia': 1, 'Hungary': 1}
Sabrina Carpenter: {'United Kingdom': 14, 'Ukraine': 3, 'Switzerland': 3, 'Sweden': 5, 'Slovakia': 2, 'Croatia': 2, 'Portugal': 2, 'Hungary': 2, 'Netherlands': 2, 'Greece': 1, 'Slovenia': 1, 'Croatia': 1, 'Hungary': 1}
Lady Gaga: {'United Kingdom': 2, 'Ukraine': 2, 'Switzerland': 2, 'Sweden': 2, 'Spain': 2, 'Slovakia': 1, 'Croatia': 1, 'Portugal': 1, 'Hungary': 1, 'Netherlands': 1, 'Greece': 1, 'Slovenia': 1, 'Croatia': 1, 'Hungary': 1}
Myke Towers: {'Spain': 13, 'Andorra': 2, 'Venezuela': 2, 'Uruguay': 6, 'Peru': 6, 'Paraguay': 16, 'Ecuador': 5, 'Brazil': 4, 'Chile': 3, 'Argentina': 2, 'Colombia': 2, 'Domingo Republic': 2, 'Uruguay': 2, 'Croatia': 1, 'Hungary': 1}
Maluma: {'Spain': 2, 'Mexico': 5, 'Venezuela': 4, 'Uruguay': 2, 'Peru': 10, 'Paraguay': 5, 'Ecuador': 5, 'Brazil': 4, 'Chile': 3, 'Argentina': 2, 'Colombia': 2, 'Domingo Republic': 2, 'Uruguay': 2, 'Croatia': 1, 'Hungary': 1}
Aventura: {'Ecuador': 5, 'Panama': 2, 'Nicaragua': 42, 'Honduras': 12, 'Guatemala': 6, 'El Salvador': 5, 'Costa Rica': 2, 'Argentina': 2, 'Brazil': 2, 'Chile': 2, 'Uruguay': 1, 'Croatia': 1, 'Hungary': 1}
ROSÉ: {'United Kingdom': 2, 'Ukraine': 2, 'Switzerland': 2, 'Sweden': 2, 'Slovakia': 2, 'Romania': 2, 'Croatia': 1, 'Portugal': 1, 'Hungary': 1, 'Netherlands': 1, 'Greece': 1, 'Slovenia': 1, 'Croatia': 1, 'Hungary': 1}
```

Figure 15 : Extrait des résultats du top 10 des artistes les plus récurrents dans les Top50

2. Artistes apparaissant dans le plus de Top50 différents

Voici les résultats retournés, du meilleur au moins bon :

- | | |
|------------------|----------------------|
| 1. Bruno Mars | 6. Sabrina Carpenter |
| 2. Lady Gaga | 7. Jimin |
| 3. ROSÉ | 8. The Weekend |
| 4. Billie Eilish | 9. KAROL G |
| 5. Gigi Perez | 10. Playboi Carti |

Ces résultats obtenus de façon analytique sont en accord avec les résultats graphiques donnés par Gephi.

Conclusion

L'analyse des Top50 Spotify à travers les différents pays du monde a permis de mettre en évidence des différences culturelles d'un point de vue musical. Les résultats montrent que certains genres musicaux, comme la pop et le rap, sont prédominants dans les Top50 mais des particularités régionales subsistent, telles que le funk brésilien ou l'Amapiano sud-africain.

Nous avons pu montrer, grâce à l'outil Gephi, une domination des artistes américains qui sont très présents dans les Top50 des pays à travers le monde, c'est le cas de Lady Gaga par exemple. Nous avons également pu mettre en évidence la présence d'artistes populaires à l'échelle d'un pays où l'on écoute massivement des genres musicaux locaux et peu populaires dans le reste du monde.

Finalement, ce projet nous a permis d'apprendre à utiliser l'API Spotify pour collecter des données que nous avons analysées à l'aide de l'outil Gephi en vue de conduire une étude sur les classements musicaux de la plateforme Spotify.

Annexes

```

africa_df = read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/Africa.csv")
asia_df =read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/Asia.csv")
central_america_df = read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/Central_America.csv")
europe_df =read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/Europe.csv")
oceania_df = read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/Oceania.csv")
south_america_df = read.csv("/Users/caioartus/Desktop/Uni
Stuff/IC05/Projet/ic05-analyse-des-top50-spotify/data/continent_fe
atures/South_America.csv")

all_values <- c(africa_df$valence,
                 asia_df$valence,
                 central_america_df$valence,
                 europe_df$valence,
                 oceania_df$valence,
                 south_america_df$valence)

groups <- factor(c(rep("Africa", nrow(africa_df)),
                     rep("Asia", nrow(asia_df)),
                     rep("Central_America",
nrow(central_america_df)),
                     rep("Europe", nrow(europe_df)),
                     rep("Oceania", nrow(oceania_df)),
                     rep("South_America", nrow(south_america_df)))))

kruskal.test(all_values ~ groups) # pvalue < 2.2e-16

# refaire le test sans Amerique du sud et centrale

```

```
all_values_filtered <- c(africa_df$valence,
                           asia_df$valence,
                           europe_df$valence,
                           oceania_df$valence
                         )

groups_filtered <- factor(c(rep("Africa", nrow(africa_df)),
                            rep("Asia", nrow(asia_df)),
                            rep("Europe", nrow(europe_df)),
                            rep("Oceania", nrow(oceania_df))
                          ))

kruskal.test(all_values_filtered ~ groups_filtered) # pvalue ≈
0.153
```

Figure 16 : Code R pour l'analyse des variances